UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE des SCIENCES de la Vie et de la Santé

IGBMC – CNRS UMR 7104 – INSERM U1258

THÈSE présentée par :

Anna-Isavella RERRA

soutenue le : **30 Septembre 2019**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline: Sciences de la Vie et de la Santé

Spécialité: Bioinformatique et biologie des systèmes

Genome-wide analyses of signaling pathways controlled by steroid receptors

THÈSE dirigée par :
  Dr. METZGER Daniel   Directeur de recherches – IGBMC – Université de Strasbourg

RAPPORTEURS :
  Pr. TUCKERMANN Jan   Professeur – Université d'Ulm
  Dr. THOMAS-CHOLLIER Morgane   Maître de conférences – Ecole Normale Supérieure Paris

AUTRES MEMBRES DU JURY :
  Dr. CERALINE Jocelyn   MCU – PH – IGBMC – Université de Strasbourg

Acknowledgments

First and foremost, I would like to thank Pr. Jan Tuckerman, Dr. Morgane Thomas-Chollier and Dr. Jocelyn Ceraline for accepting to evaluate my thesis as well as for taking the time to read my manuscript during the summer period. I hope that your knowledge, feedback, and insights will enrich my thesis work.

I wish to express my gratitude to my supervisor, Dr. Daniel Metzger, for accepting me in his lab and for giving me the possibility to work in his team for four years. He is always available to discuss, to help, to guide and to answer questions. His advice and critical comments are always precious in order to go further with my projects. I thank him for trusting me during these four years.

I sincerely thank the committee of the IGBMC Ph.D. Programme for choosing me among other high-level candidates and for giving me the chance to realize my thesis in an excellent environment, financed by the Labex fellowship for three years. I would like also to express my gratitude to the association "Alsace contre le cancer" for financing the last year of my thesis and giving the means to complete my thesis work.

I would like to thank Valerie, for her kindness, her nice mood and for organizing everything for my thesis defense.

I would like to thank Gilles for the scientific discussions, the collaboration, the advice on the different projects and the nice atmosphere in our box. I would like to thank also Delphine for her good mood, the collaboration in the GR paper, and for her idea to start the C2C12 cells project.

I would like to thank Jean-Marc, the kindest technician, for his humor, his caring, his tartes flambées and of course our discussions. I would like to thank Laetitia and Regis for the good moments we spent in the lab.

I would like to thank new and past members of the lab, especially Justine, Camille, Elise, Kamar and Mohamed, for the time that we spent together and for creating a nice lab atmosphere. I would like also to thank Vanessa and Shilpy with whom I collaborated for the GR paper. I thank Rana with who I shared special moments and of course Margherita, who joined recently the lab, but she was close to me all the hard period of writing.

I would like to say a special thank you to Daniela. She is always there for me, with her nice mood and smile, no matter what, to discuss, to chat, to help, to make me feel better. We traveled together, we shared good and bad moments and I am grateful to her. Definitely, the lab would not be the same without her.

Résumé de thèse

Introduction

Les androgènes (ADs) et les glucocorticoïdes (GCs) sont des hormones stéroïdes exerçant des effets pléiotropes chez les mammifères. Les androgènes contrôlent la prolifération cellulaire, le développement des caractéristiques sexuels et le comportement, ainsi que la masse et la force musculaires, tandis que les glucocorticoïdes contrôlent le rythme circadien, le métabolisme du glucose, des lipides et des protéines, ainsi que les fonctions inflammatoires et immunitaires. Les effets de ces hormones sont relayés par les récepteurs nucléaires, le récepteur des androgènes (AR) et le récepteur des glucocorticoïdes (GR), respectivement.

Comme l'activité de ces récepteurs est également modulée par des ligands synthétiques, ils représentent des cibles pharmacologiques importantes pour de nombreuses maladies, dont le cancer, la sarcopénie, les allergies, et l'asthme. Cependant, même si les GCs synthétiques sont largement utilisés en clinique pour leurs puissantes activités anti-inflammatoires et immunosuppressives, les traitements à long terme sont limités par des effets indésirables, notamment l'atrophie musculaire. De plus, même si les effets anaboliques des androgènes sur les muscles squelettiques sont intéressants pour améliorer la fonction musculaire chez les hommes âgés et chez les patients atteints de diverses maladies (sarcopénie, myopathies et sida), ils stimulent également la prolifération des cellules épithéliales prostatiques et augmentent ainsi le risque de cancer de la prostate. Inversement, les anti-androgènes utilisés comme traitement primaire du cancer de la prostate métastatique provoquent entre autres une atrophie musculaire et diminuent ainsi la qualité de vie des patients.

Le mode d'action classique de GR et AR propose que leurs ligands respectifs induisent la liaison des récepteurs à leurs éléments de réponse à l'ADN (GRE et ARE, respectivement) pour stimuler l'expression du gène cible. Ces éléments sont organisés en répétitions inversées (IR) de motifs de type 5'-AGAACA-3', séparées par trois paires de bases (IR3). Il est important de noter que GR et AR se lient sous forme homodimères à des éléments de liaison IR3 consensus (1, 2). De plus en plus d'évidences convergeant vers une interconnexion entre les voies de signalisation des ADs et des GCs, mais les mécanismes sous-jacents étant inconnus, nous avons identifié les cistromes des AR et GR et leurs gènes cibles dans deux tissus dans lesquels les ADs et les CGs ont des effets opposés, à savoir le muscle squelettique (skm) et la prostate.

Les objectifs de l'étude étaient :

1. déterminer la fonction physiologique du GR dans les muscles squelettiques
2. caractériser les cistromes et transcriptomes des AR et GR dans les muscles squelettiques et la prostate
3. comparer les transcriptomes et les épigénomes des myoblastes, des myotubes et des muscles squelettiques

Résultats

Afin d'identifier les gènes régulés par GR dans les fibres musculaires, le laboratoire a effectué une analyse transcriptomique de muscle gastrocnémien de souris de type sauvage et de souris GR[(i)skm-/-] chez lesquelles GR est sélectivement invalidé dans les myofibres squelettiques au stade adulte. Nous avons identifié environ 1335 gènes exprimés de façon différentielle, dont 677 étaient positivement régulés et 658 étaient négativement régulés. L'analyse des voies de signalisation a révélé que les gènes négativement régulés étaient liés au métabolisme musculaire, et en particulier des gènes codant des enzymes impliquées dans le métabolisme du glycogène. D'autres voies enrichies de gènes négativement régulés ont révélés les termes "facteurs de traduction" et "signalisation de l'insuline". En particulier, nous avons trouvé que les transcrits encodant deux cibles GR connues, Pik3r1 et Ddit4, étaient significativement réduits en l'absence de GR. De plus, les transcrits codant pour les inhibiteurs de traduction Eif2ak1, Eif4ebp1 et Eif4ebp2 étaient moins exprimés chez les souris GR[(i)skm-/-]. En accord avec ces données, les transcrits surexprimés dans les gastrocnémiens déficients en GR faisaient parti de voies associées aux protéines ribosomiques cytoplasmiques et de maturation des ARN messagers, et comprenaient notamment les facteurs anaboliques Akt3, Rps6, et Pik3ca. De plus, les niveaux protéiques de Akt3 étaient augmentés dans le muscle gastrocnémien des souris GR[(i)skm-/-], alors que ceux de Pik3r1 et Ddit4 étaient fortement diminués. Par contre, l'expression de gènes impliqués dans le catabolisme musculaire, y compris le système protéasomique (e.g. Murf, atrogin), était similaire chez les souris de type sauvage et les souris mutantes. De plus, les niveaux protéiques de Foxo1 et Foxo3a, deux régulateurs clés des voies protéolytiques musculaires, n'étaient pas modifiés dans les myofibres déplétées en GR, et même si les niveaux de Foxo1 phosphorylé (forme inactive) étaient diminués, ceux du Foxo3a étaient similaires chez les souris contrôles et GR[(i)skm-/-].

Ainsi, nos résultats démontrent que les niveaux physiologiques de glucocorticoïdes réduisent l'expression de plusieurs facteurs anaboliques et induisent celle des facteurs anti-anaboliques, via GR dans les myofibres, diminuant ainsi la voie anabolique, la taille des fibres musculaires et le poids des muscles, sans stimuler les voies cataboliques.

Pour identifier les gènes cibles de GR, nous avons caractérisé le cistrome du GR dans les tissus de souris par immunoprécipitation de la chromatine suivie par un séquençage massif parallèle (ChIP-Seq) d'ADN (3). A des niveaux physiologiques de GC, nous avons identifié, à l'aide du logiciel MACS2, environ 23000 sites liés par GR (GRBS), localisés dans des régions promotrices (-1 kb ; +100 bp autour des Sites de l'Initiation de la Transcription), les régions intergéniques et les introns. Cependant, l'analyse des motifs *de novo* réalisée avec le logiciel MEME-Suite, n'a identifié aucun GRE dans les régions promotrices, alors que la plupart des sites de liaison intergéniques et introniques étaient des GREs.

Pour caractériser l'environnement génomique des sites de liaison de GR, nous avons effectué une analyse ChIP-seq pour diverses marques d'histones. Nous avons trouvé 21377 pics pour l'histone H3 acétylée à la lysine 27 (H3K27ac, une marque de promoteur et des enhancers actifs), 75523 pour

H3 monométhylée à la lysine 4 (H3K4me1, une marque d'enhancers), 19818 pour H3K4 trimethylée (une marque enrichie aux régions promotrices) et 13053 pics pour le polymerase 2 (Pol2). Les heatmaps produites par seqMINER ont révélé deux groupes de pics de GR, l'un aux promoteurs actifs (11038 pics) définis par la présence de H3K27ac, H3K4me3 et Pol2, et de faibles niveaux de H3K4me1, et un aux enhancers actifs (12158 pics) définis par la présence de H3K27ac, H3K4me1 et Pol2 et des niveaux faibles de H3K4me3.

Étant donné que GR était lié à des enhancers actifs, nous avons identifié des gènes induits par GR dans les myofibres. À cette fin, nous avons croisé les gènes contenant des sites de liaison pour GR, H3K27ac et Pol2 avec les gènes dont l'expression est diminuée chez les souris GR^(i)skm-/-. Nous avons trouvé 375 gènes directement activés par GR. Parmi ces gènes, nous avons trouvé les gènes identifiés par l'analyse transcriptomique, comme les gènes codant pour les facteurs anaboliques, qui sont donc des cibles directes de GR. La caractérisation détaillée de deux d'entre eux, Eif4ebp2 et Pik3r1, a révélé que GR lié au GRE coopère avec Myod1 et le facteur Foxf2 associé à la chromatine au niveau des enhancers, et que GR interagit avec des facteurs liés aux régions promotrices, comme le Nrf1, pour stimuler la transcription de gènes cibles.

Ainsi, à des niveaux physiologiques de GC, GR stimule l'expression des facteurs anti-anaboliques dans les myofibres via des GREs localisés dans des régions enhancer, et diminue l'expression des facteurs anaboliques via des mécanismes GRE-indépendants. Cependant, il ne stimule pas les voies cataboliques classiques. Ainsi, GR limite la synthèse protéique dans les myofibres conduisant à une masse musculaire réduite.

Pour étudier la spécificité tissulaire du GR, nous avons également effectué des analyses ChIP-Seq de GR dans des conditions physiologiques à partir de prostates de souris. Nous avons identifié environ 8000 sites de liaison du GR dans la prostate, principalement situés dans les régions intergéniques et introniques. Près de 3500 gènes contenant des GRBS dans les muscles squelettiques et la prostate ont été identifiés. L'analyse de novo des motifs des gènes communs, à l'aide du MEME-Suite, a révélé des GREs dans les régions intergéniques et introniques, mais pas dans les régions proximales des promoteurs, où GR semble interagir avec divers facteurs selon les tissus. L'analyse des voies de signalisation sur les cistromes partagés, à l'aide du logiciel Webgestalt, a révélé des voies également trouvées dans les cistromes spécifiques du muscle et de la prostate, comme la signalisation MAPK, la signalisation p53, la signalisation de l'apoptose, le PI3K-Akt et le cancer de la prostate, ainsi que des voies tissu-spécifiques comme la signalisation mTOR, le cycle cellulaire la signalisation RAS. Comme les GREs liés par le GR dans les deux tissus ne présentent aucune spécificité de séquence, la sélectivité tissulaire de GR est probablement dictée par la présence sélective de cofacteurs adjacents.

Pour caractériser les sites de liaison à l'ADN de AR dans la prostate de souris, nous avons effectué des analyses ChIP-Seq dans ce tissu. Nous avons identifié environ 3900 sites de liaison d'AR (ARBS) et la plupart était situés dans des régions intergéniques. De plus, près de 2000 gènes contenant à la

fois des ARBS et des GRBS dans la prostate ont été identifiés. L'analyse de novo des motifs des gènes communs a révélé des GREs/AREs dans les régions intergéniques et introniques, mais pas dans les régions proximales des promoteurs, où ils interagissent avec des facteurs distincts selon le récepteur. Nous avons observé que les éléments de réponse des gènes liés par AR et GR dans les régions intergéniques et introniques sont des éléments composés du demi-site 5' d'un ARE canonique et du demi-site 3' d'un GRE canonique, définis par la base de motifs Jaspar. Ainsi, les éléments de réponse liés par les deux récepteurs semblent distincts de ceux liés par uniquement AR ou GR, et la spécificité des liaisons AR et GR dans la prostate est probablement coordonnée par les facteurs adjacents. L'analyse des voies de signalisation des cistromes partagés a révélé des voies également trouvées dans les cistromes spécifiques d'AR et GR, comme le cancer de la prostate, la signalisation p53, la signalisation MAPK et la signalisation Jak-STAT, ainsi que des voies identifiées sélectivement pour un récepteur, comme la signalisation FoxO, la signalisation PPAR, la signalisation mTOR et la glycolyse/gluconéogenèse.

Finalement, nous avons effectué une analyse comparative à l'échelle du génome entre le muscle squelettique murin et les cellules C2C12, à l'aide de données transcriptomiques et cistromiques, en termes d'expression génique, niveaux d'expression, annotations fonctionnelles, familles de facteurs de transcription, motifs de liaison et modifications des histones. Nos analyses montrent que les cellules C2C12 différenciées et le muscle squelettique partagent des caractéristiques communes en termes de traduction, de contraction et de fonction musculaire, de régulation du cytosquelette et de métabolisme. L'analyse de novo a révélé non seulement des motifs communs, mais aussi des motifs spécifiques à chaque étape de la différenciation. Il est important de noter que notre étude met en évidence la sélectivité de l'expression de facteurs de transcription et de familles de facteurs de transcription impliqués dans la myogenèse et dans d'autres processus biologiques. En plus, les analyses bioinformatiques de Pol2 et des marques d'histone ont mis en évidence des caractéristiques spécifiques de chaque étape de la différenciation musculaire. Les cartes de l'état de la chromatine à l'échelle du génome des gènes spécifiques et communs ont révélé plus de différences entre les muscles squelettiques murins et les cellules C2C12 au niveau des enhancers que des régions promotrices, ce qui indique que les enhancers distants assurent une sélectivité dans le processus de différenciation. Ainsi, cette étude fournit une base de données de gènes, voies de signalisation et facteurs de transcription exprimés à differents stades dans les myoblastes, myotubes en culture et muscles squelettiques de souris.

Conclusions

Nos résultats ont permis de caractériser les gènes contrôlés par les androgènes et les glucocorticoïdes dans les muscles squelettiques et la prostate, ainsi que les voies de signalisation.

La disponibilité du site de liaison dépend de l'état de la chromatine, qui est spécifique à chaque type de tissu et de cellule. La liaison d'un récepteur à des motifs spécifiques peut entraîner la transcription dans certains tissus mais pas dans d'autres, et elle est essentielle pour des fonctions spécifiques aux tissus. Nos résultats indiquent que la spécificité des réponses ne repose pas

seulement sur l'expression tissulaire spécifique des récepteurs, mais aussi sur la coopération avec des facteurs de transcription adjacents distincts, plutôt que sur la séquence des éléments de la réponse.

Les similitudes entre les cellules C2C12 différenciées et le muscle squelettique indiquent que les premières peuvent être utilisées dans une certaine mesure comme modèle in vitro qui récapitule la différenciation myogénique. Ainsi, nos résultats fournissent la base d'une compréhension moléculaire de l'activité tissu et/ou promoteur spécifique des androgènes et des glucocorticoïdes, et ouvrent ainsi de nouvelles avenues pour concevoir des criblages pour des analogues induisant sélectivement des gènes, en utilisant des essais cellulaires.

Références

1. Hard, T. et al. Science 249, 157-160, 1990
2. Verrijdt, G., Haelens, A. & Claessens, F. Mol. Gen. and met. 78, 175-185, 2003
3. Joshi, S., Metzger, D. & Davidson, I. JOVE 129, 1-9, 2017

Table of contents

List of tables

List of figures

List of abbreviations

A
AR: Androgen Receptor
ARE(s): Androgen Response Element(s)
ADs: androgens
ARBS: Androgen Receptor Binding Sites
aa: amino acid(s)
ADT: androgen deprivation therapy
AF1/2: activation function 1/2
AP-1: activator protein 1

C
CRPC: castration resistant prostate cancer
ChIP: Chromatin Immunoprecipitation
CTCF: CCCTC-binding factor
Ca2+: calcium
Cre: recombinase Cre
CTD: C-terminal domain

D
Dex: dexamethasone
DHT: Dihydrotestosterone
DBD: DNA-binding domain
DDR: DNA damage response
DR: direct repeat

G
GR: Glucocorticoid Receptor
(n)GRE(s): (negative) Glucocorticoid Response Element(s)
GCs: glucocorticoids
GRBS: Glucocorticoid Responsive Binding Sites
GM: gastrocnemius

H
HDR: homology-directed repair
H3K27ac: histone 3 lycine 27 acetylation
H3K4me1: histone 3 lycine 4 mono-methylation
H3K4me3: histone 3 lycine 4 tri-methylation

I
IR: inverted repeat

L
LBD: ligand-binding domain

M
Myod1: myogenic differentiation 1
Myog: myogenin
Myf6: myogenic factor 6
Myf5: myogenic factor 5
mTOR: mammalian target of rapamycin
MT: myotubes
MB: myoblasts
MGRKO: muscle GR knockdown
MRFs: myogenic regulatory factors

N
NR: Nuclear Receptor
NCoR: Nuclear receptor corepressor
NTD: N-terminal domain
NF-kB: nuclear factor-kappa B

P
PTMs: Post Transcriptional Modifications
PCa: prostate cancer
pro: prostate
Pol2: polymerase 2

S
skm: skeletal muscle
SRC: steroid receptor coactivator

T
TF(s): transcription factor(s)
TSS: transcriptional start site
TTS: transcription termination site

W
WT:wild-type

Introduction

1. Gene transcription and its regulation

All the cells of an organism contain the same DNA information. In order for the cells to differentiate and perform their specific functions, spatio-temporal gene expression needs to be tightly regulated (Hinman and Cary, 2017). Transcriptional regulation is achieved through regulatory networks that exert control through mechanisms such as modulation of chromatin structure, protein availability, transcription initiation, elongation and mRNA maturation and translation (Maston et al., 2006). A key step in transcriptional regulation is the modulation of the initiation phase which involves DNA elements, epigenetic modifications and the recruitment of general and sequence-specific transcription factors (TFs) to their target sites in DNA (Handy et al., 2011). The DNA sequences involved are the promoter, adjacent to the Transcription Start Site (TSS) and other distal elements, such as enhancers and insulators (Spitz and Furlong, 2012) (**Figure 1**).
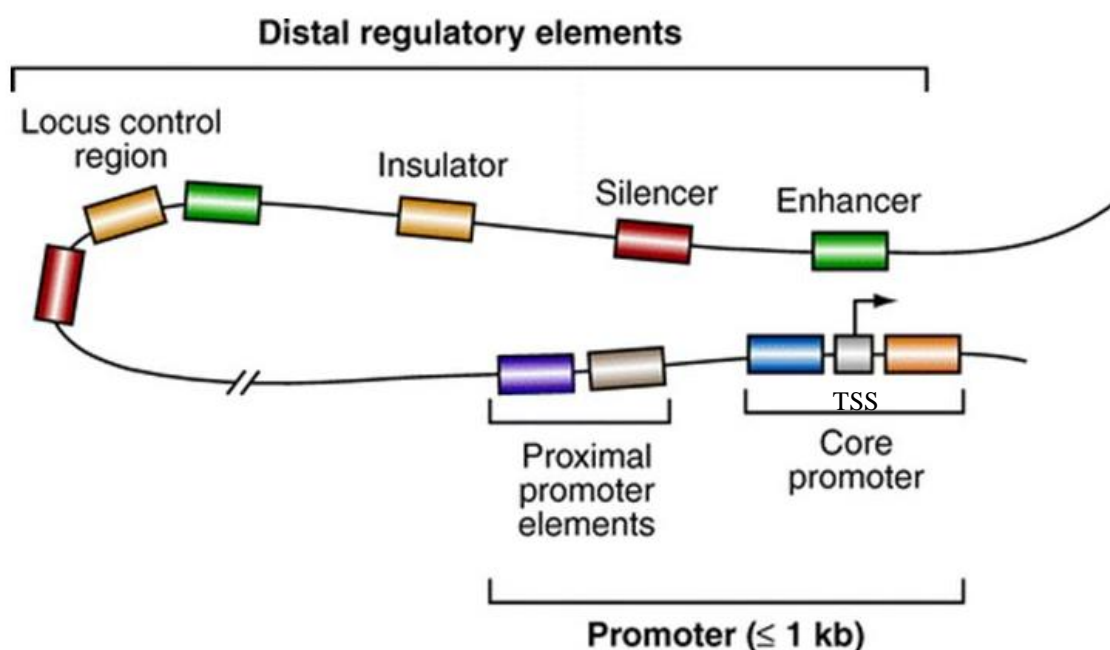
Figure 1. **Scheme of a typical regulatory region of a gene**. The promoter is composed of a core promoter and of proximal promoter elements and typically spans less than 1 kb pairs. Enhancers, silencers, insulators and locus control regions are distal (upstream) regulatory elements and can be located up to 1 Mb pairs from the promoter. The distal elements can contact the core or proximal promoter through a mechanism that involves looping of DNA (adapted from (Maston et al., 2006)).

Promoters and enhancers

The **promoter** (**Figure 1**) is the region where the preinitiation complex (PIC) is assembled before the initiation of transcription (Gupta et al., 2016). The PIC is composed by RNA Pol II, general TFs (TFIIA/B/D/E/F/H) and the Mediator (Luse, 2014). The Mediator is a large protein complex which interacts with sequence specific TFs, transmitting its signals to RNA Pol II, modulating its function and leading to the release from the PIC (Jeronimo and Robert, 2017). The Mediator is involved in transcriptional regulation by changing the chromatin organization and modulating enhancer-promoter looping, transcriptional initiation and elongation (Allen and Taatjes, 2015).

Regulation by **enhancers** (**Figure 1**) is generally believed to occur by binding of specific transcription factors to the enhancer region, which causes the attraction of co-activators such as CBP and p300 (Merika et al., 1998). Since these factors, apart from being co-activators also covalently modify the surrounding histones, it was suggested that enhancers were marked by specific histone modifications (Smith and Shilatifard, 2010). This hypothesis did turn out to be true and since then many chromatin marks have been specifically associated with enhancers (Calo and Wysocka, 2013) (see "Histone modifications" part).

Enhancers usually contain binding sites for several TFs (Mullen et al., 2011; Zeitlinger et al., 2003). Genes can have many enhancers, each of them being active at different developmental stages and/or at different cell types and tissues, or they can act synergistically to drive gene expression (Maston et al., 2006) (Shlyueva et al., 2014). Enhancers contain motif sequences usually of 6 to 12 bp to which TFs bind (Kadonaga, 2004; Shlyueva et al., 2014). However, most of TF binding events in the genome do not lead to expression of neighboring genes (for higher eukaryotes only 20% of binding is functional) (O'Connor and Bailey, 2014). There are many reasons why, such as transcription factor redundancy; random binding events in areas of open chromatin; the need for additional cofactors and functions of the binding (Spitz and Furlong, 2012).

The emergence of genome-wide methods for identifying active enhancers as well as genome-wide methods for analyzing long-range chromatin interactions (such as ChiA-pet (Li et al., 2014) and Capture Hi-C (Mifsud et al., 2015) as well as many studies on single genes have resulted in an update of the model for how enhancers and promoters interact. In this model enhancers and promoters in close physical proximity each other cooperate to increase the local concentration of factors needed for transcription with the help of their local chromatin modification (Andersson et al., 2014) (**Figure 1**). Importantly the active transcription of enhancers is part of this mechanism, since the early termination of enhancer RNAs, a class of relatively short non-coding RNA molecules (50-2000 nucleotides) transcribed from the DNA sequence of enhancer regions, will leave the general transcription machinery in close proximity to the gene promoter (Andersson et al., 2014). This model is based mostly on analysis of steady state data; the temporal details of how transcriptional induction occurs is still unknown.
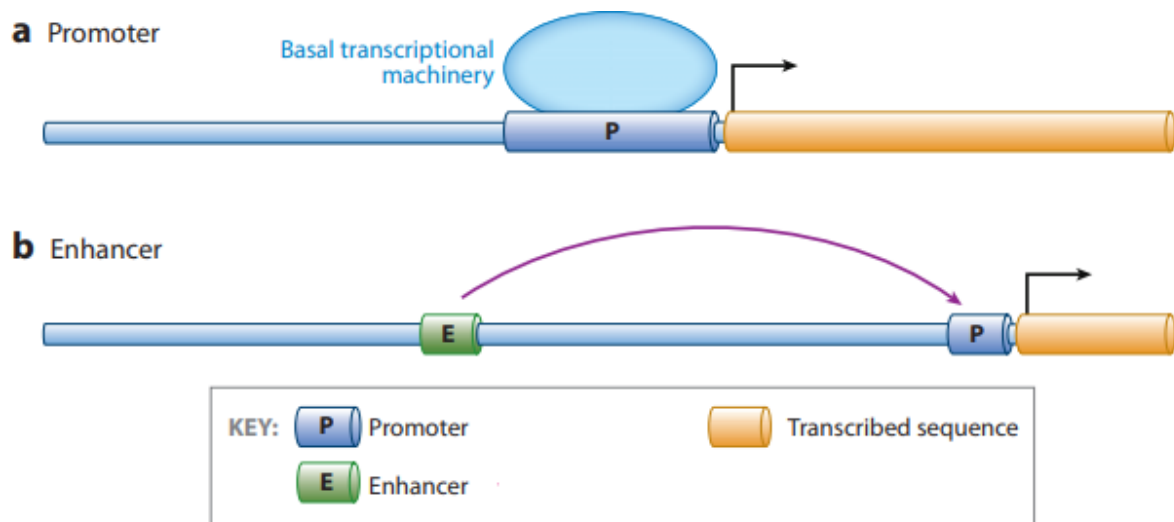
Figure 2. Known regulatory elements. (a) Promoter sequence is binding the general transcriptional machinery to mediate the basal transcriptional control of a transcribed sequence. (b) Enhancer sequence mediate positive effects of transcription through the interaction with the promoter sequence (adapted from (Noonan and McCallion, 2010)).

Enhancer-promoter looping

Enhancers and promoters can also physically interact through chromatin looping that was shown in the repression of bacterial genes and in the mouse beta-globin locus (Ptashne, 1986; Tolhuis et al., 2002). Many studies have identified PIC proteins, such as Mediator subunits and TFs, at enhancers, suggesting that enhancers and promoters are very close during transcriptional initiation (Levine et al., 2014; Lin et al., 2013a) (**Figure 3**). It is not completely understood the enhancer-promoter loop structure and how it is formed, but there is evidence that the two regions are linked by Mediator-cohesin protein complexes, which form rings around DNA, and CTCF, that binds together strands of DNA and anchors the boundaries of chromatin domains (Kagey et al., 2010; Meng and Bartholomew, 2018; Phillips-Cremins and Corces, 2013) (**Figure 3**).

Most of the promoters interact with multiple enhancers and half of the enhancers interact and activate multiple promoters simultaneously, in the same cell, rather than only one at a time (Fukaya et al., 2016; Javierre et al., 2016; Thurman et al., 2012).

Studies suggest that enhancer-promoter loops are maintained and when the enhancer interacts with the promoter, RNA Pol II is recruited. However, it remains paused and later, at the time of gene activation, TFs and/or other enhancers may be recruited (the pausing of the RNA Pol II is released) and the transcription is active (Ghavi-Helm et al., 2014). Other studies suggest that there are changes in looping and enhancer states during development (Heintzman et al., 2009; Simonis et al., 2006; Tolhuis et al., 2002). Moreover, studies support that the enhancer-promoter dynamic looping is highly specific and varies depending on the cell-type and on the gene's activation state (Javierre et al., 2016). Dynamic enhancer-promoter loops allow the reliable transfer of regulatory information over distance (Meng and Bartholomew, 2018). Depending on the gene locus, the number of loops and their dynamics can vary greatly (Meng and Bartholomew, 2018).
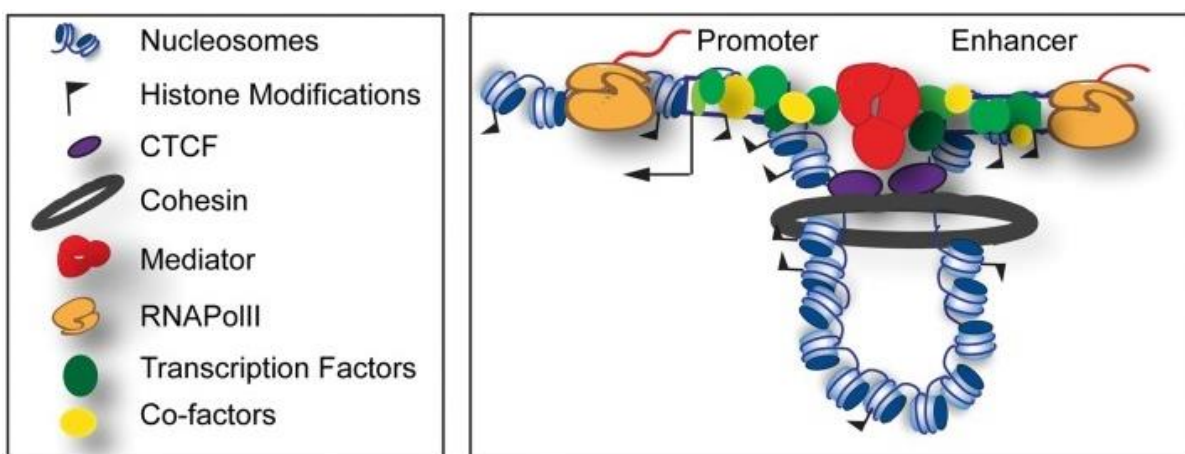
Figure 3. **Enhancer-promoter looping mechanism.** This mechanism is mediated by cohesion, CTCF and the Mediator complex that brings the enhancer close to its target promoter. The Transcription Start Site (TSS) is annotated with an arrow. The TFs and the co-factors bind the enhancer and are brought close to the basal transcription machinery at the promoter (adapted from (Mora et al., 2016)).

Histone modifications

DNA is packaged in the cell in the form of chromatin, which is a DNA-protein complex (Heslop-Harrison and Schwarzacher, 2013). Its basic unit is the nucleosome consisting of 147 bp stretch of DNA wrapped around an octamer of histones – two H2A, two H2B, two H3 and two H4 (Kouzarides, 2007; Widom, 1997).

Histones have N-terminal tails that can be modified, leading to alterations in the nucleosome structure and in DNA accessibility (Marino-Ramirez et al., 2005). There are different types of chemical modifications, namely, acetylation (ac), methylation (me), ubiquitylation (ub) and sumoylation (su) of lysines (K), methylation of arginines (R), phosphorylation (ph) of serines (S) and threonines (T), ADP ribosylation (ar) of glutamic acid (E), deamination (conversion of arginine to citruline) and proline isomerization (**Figure 4**). Methylation of lysines can be mono-, di- or trimethylation (me1, me2, me3) (Kouzarides, 2007; Zhao and Garcia, 2015) (**Figure 4**) (**Figure 5**).
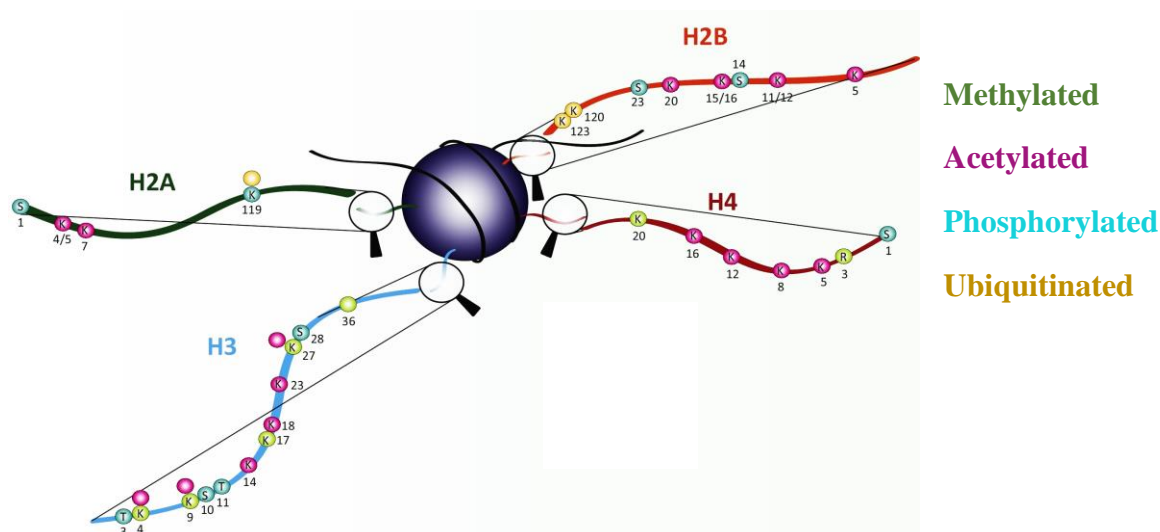
Figure 4. Schematic of post-translational modifications of the histone tails. The location of each modification and the amino acid modified at each position are shown (K=lysine, R=arginine, S=serine, T=threonine). Different colors for the different modifications are used (green=methylation, pink=acetylation, turquoise=phosphorylation, yellow=ubiquitination) (Lawrence et al., 2016).

H3K4me1 (histone 3, lysine 4 monomethylated) was shown to mark active promoters as well as enhancers (**Figure 4**, **Figure 5**). Thus enhancers compared to promoter regions can be identified by a higher ratio of the monomethylated vs trimethylated modification (Heintzman et al., 2007). It was the first histone linked to distal regulatory regions through genomic studies (Heintzman et al., 2007). Moreover, it is highly enriched in TSSs of active genes (Koch et al., 2007). H3K4me1 enhancers can either be active, inactive or poised (Zentner et al., 2011)

H3K4me3 (histone 3, lysine 4 trimethylated) was shown to mark active promoters and the level of this histone modification at a gene's promoter broadly correlates with the transcriptional activity of the gene (Koch et al., 2007; Santos-Rosa et al., 2002) (**Figure 4**, **Figure 5**). H3K4me3 levels at enhancers are low (Sharifi-Zarchi et al., 2017).

H3K9me3 (histone 3, lysine 9 trimethylated) is enriched in the repeat-rich regions of constitutive heterochromatin (Nakayama et al., 2001) where the accessibility is decreased (Wei et al., 2018) (**Figure 4**, **Figure 5**). H3K9me3 plays a role in embryonic stem cells at the beginning of organogenesis during lineage commitment and in lineage fidelity maintenance (Nicetto et al., 2019).

H3K9ac (histone 3, lysine 9 acetylated) is found in actively transcribed promoters (**Figure 4**, **Figure 5**). It is also proposed that H3K9ac can promote progression through the transcription cycle mediating a switch from transcription initiation to elongation (Gates et al., 2017).

H3K27ac (histone 3, lysine 27 acetylated) was shown to mark the active transcription and as a result it was defined as an active enhancer and promoter mark (Spicuglia and Vanhille, 2012) (**Figure 4**, **Figure 5**). H3K27ac is found at proximal and distal regions of TSS (Dao et al., 2017) and it is often used to separate active from poised enhancers with the help of H3K4me1 active and poised enhancers (Creyghton et al., 2010). The enrichment of H3K27ac at active enhancers is probably linked to the recruitment of transcriptional cofactors with histone acetyltransferase (HAT) activity such as p300 and CBP (Camp-response element-binding protein) (Pasini et al., 2010). As a result, H3K27ac association with enhancers shows high levels of cell-type specificity.

H3K27me (histone 3, lysine 27 methylated) (Figure 4, Figure 5) is a modification usually associated with gene repression and has established roles in regulating the expression of genes involved in lineage commitment and differentiation (Wiles and Selker, 2017). Perturbations in the distribution or levels of H3K27me occur due to deregulation at all levels of the process, either by mutation in the histone itself, or changes in the activity of the writers, erasers or readers of this mark. H3K27ac shares a location with H3K27me3 and they interact in an antagonistic manner.

H3K36me (histone 3, lysine 36 methylated) is a common epigenetic mark involved in epigenetic regulation (Suzuki et al., 2017) (**Figure 4**, **Figure 5**). The modifications of H3K36 play roles in processes like DNA replication, transcription, recombination and repair of DNA damage (Lee et al., 2010) and its misregulation is linked to many human diseases (Zhang et al., 2017). H3K36 methylation is linked to transcribed regions of active genes. H3K36me3 exhibits a more 3' end distribution.
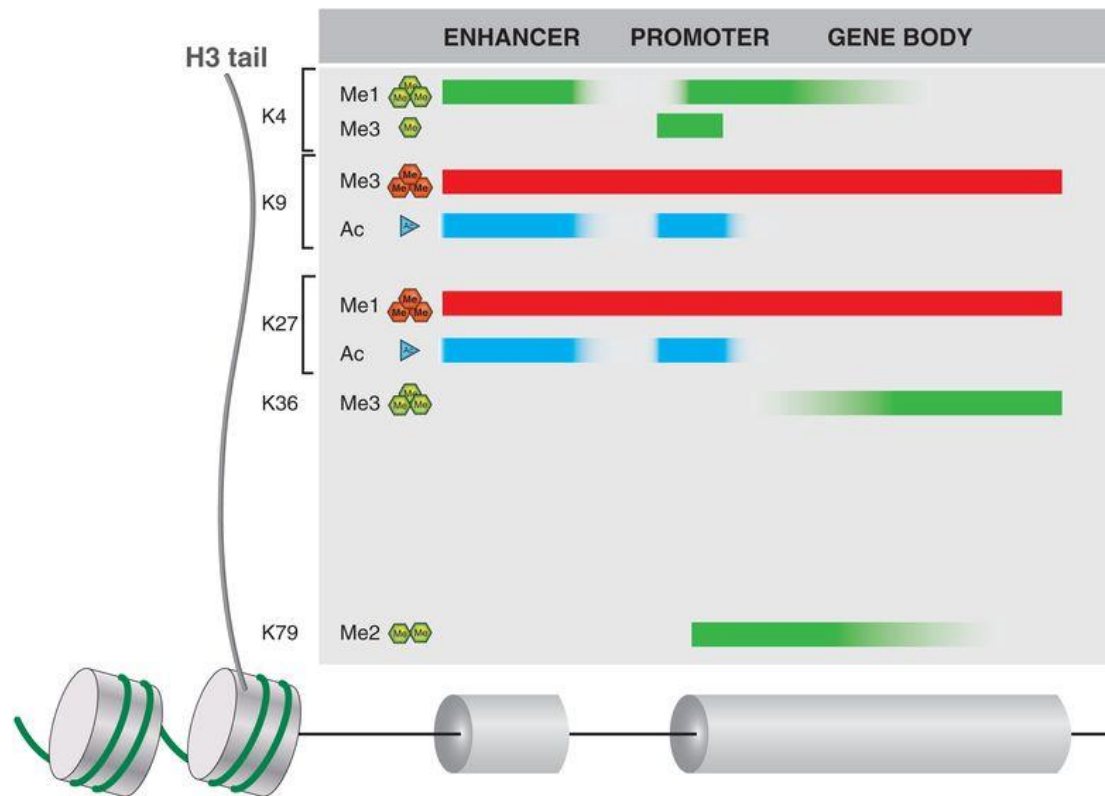
Figure 5. Histone H3 tail lysine residues are subject to post-translation modifications (PTMs). The typical distribution of these PTMs is indicated along the length of gene loci as shaded blocks. Green (methylation) and cyan (acetylation) indicate histone marks associated with active genes, whereas red indicated silent genes (adapted from (Audia and Campbell, 2016)).

2. Nuclear receptors

Nuclear receptors (NRs) are one of the most abundant class of transcriptional regulators in animals (Sasse and Gerber, 2015). They are ligand-activated transcription factors (TFs) which regulate the expression of target genes by binding to specific cis-acting sequences (Beato, 1989; Evans, 1988; Green and Chambon, 1988). When binding to the promoter or enhancer regions of the target genes, the receptor will affect transcription by recruiting specific co-regulators and components of the transcription initiation complex or RNA polymerase II (Acevedo and Kraus, 2004).

The diversity of the NRs has been organized in a phylogeny-based nomenclature (Nuclear Receptors Nomenclature Committee, 1999) of the form **NRxyz**, where x is sub-family, y is the group and z the gene (e.g. NR3C1). The superfamily (**Table 1**) includes receptors for hydrophobic molecules, such as steroid hormones (e.g. estrogens, glucocorticoids, progesterone, mineralocorticoids, androgens, vitamin D3, ecdysone), oxysterols and bile acids, retinoic acids (all-trans and 9-cis isoforms), thyroid hormones, fatty acids, leukotrienes and prostaglandins (Escriva et al., 2000; Laudet and Gronemeyer, 2002). This family also contains genes encoding receptors for unknown ligands or no ligands, described as 'orphan' receptors (Evans, 1988; Moore, 1990).

The superfamily of NRs is specific to animals, and performs many functions, from embryonic development and homeostasis of various physiological functions, to the control of metabolism (Laudet and Gronemeyer, 2002). 48 NRs genes have been identified in the human genome (Robinson-Rechavi et al., 2001; Xiao et al., 2013), 49 genes in the mouse (Robinson-Rechavi and Laudet, 2003), 21 genes in the fly *Drosophila melanogaster* (Adams et al., 2000) and more than 270 genes in nematode worm *Caenorhabditis elegans* (Robinson-Rechavi et al., 2005; Sluder et al., 1999). The zebrafish contains a total of 73 NRs genes, and orthologues of almost all human NRs are present (Schaaf, 2017).

Table 1. Homology classification of NRs

Subfamily	Receptor	Ligand	Abbreviation - Symbol
I	Thyroid hormone receptor	Thyroid hormone	TRα – NR1A1 TRβ – NR1A2
	Retinoic acid receptor	All-trans-retinoic acid	RARα – NR1B1 RARβ – NR1B2 RARγ – NR1B3
	Peroxisome-proliferator-activated receptor	Fatty acids, prostaglandins	PPARα –NR1C1 PPARβ/δ – NR1C2 PPARγ –NR1C3
	Reverse-ErbA	heme	Rev-ErbAα – NR1D1 Rev-ErbAβ – NR1D2
	RAR-related orphan receptor	cholesterol	RORα – NR1F1 RORβ – NR1F2 RORγ – NR1F3
	Liver X receptor	oxysterols	LXRα – NR1H1 LXRβ – NR1H2
	Vitamin D receptor	1a,25(OH)2D3	VDR – NR1I1
II	Hepatocyte nuclear factor-4	Fatty acids	HNF4α –NR2A1 HNF4γ – NR2A2
	Retinoic X receptor	retinoids	RXRα – NR2B1 RXRβ – NR2B2 RXRγ – NR2B3
	Testicular receptor	unknown	TR2 – NR2C1 TR4 – NR2C2
	TLX/PNR	unknown	TLX – NR2E1 PNR – NR2E3
	COUP/EAR	unknown	COUP-TFI – NR2F1 COUP-TFII – NR2F2 EAR-2 – NR2F6
III	Estrogen receptors	estrogens	ERα – NR3A1 ERβ – NR3A2
	Estrogen related receptors	unknown	ERRα – NR3B1 ERRβ – NR3B2 ERRγ – NR3B3
	Glucocorticoid receptor	Glucocorticoids	GR – NR3C1
	Mineralocorticoid receptor	Mineralocorticoids, glucocorticoids	MR – NR3C2
	Progesterone receptor	Progesterone	PR – NR3C3
	Androgen receptor	Androgens	AR – NR3C4
IV	Nerve-growth-factor-Induced B-like	unknown	NGFI-B – NR4A1 NURR1 – NR4A2 NOR1 – NR4A3
V	Steroidogenic factor-like	Oxysterols	SF-1 – NR5A1 LRH-1 – NR5A2
VI	Germ cell nuclear factor-like	unknown	GCNF – NR6A1
0	Dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1	unknown	DAX1 – NR0B1
	Small heterodimer partner	unknown	SHP – NR0B2

Genomic mechanisms of hormone action

The regulation of target gene activity by hormones via their protein receptors is known as genomic mechanism of hormone action (Puzianowska-Kuznicka et al., 2013). These mechanisms engage transcription and translation, and their biological effects are executed by newly synthesized proteins. The first effects of engagement of these mechanisms might be detected 30–60 minutes after the initiation and the maximal effects are observed after several hours (Puzianowska-Kuznicka et al., 2013).

DNA binding by NRs

The composition of the DNA response element determines which NR can bind to it. Response elements are typically composed of two hexameric sequence organized as a **direct**, **inverted** or **everted** repeat (Helsen et al., 2012). Each hexameric sequence or half-site is recognized by a receptor (Roemer et al., 2006). The half sites are generally separated from each other by a spacer with variable length (Lu et al., 2017). The response elements that consist of only one hexameric sequence are less common and are recognized by a NR in a monomeric binding mode (e.g. ROR) (Chen and Young, 2010).

The composition and the recognition by the correct NR is dependent on orientation and sequence of the hexamer and on the spacer length (Pawlak et al., 2012). Steroid receptors recognize the 5'-AGAACA-3'-like motifs, while non-steroid receptors and the ER bind to the 5'-AGGTCA-3'-like motifs (Helsen and Claessens, 2014). The specific DNA binding properties of each receptor will enable or disable binding to a certain response element. Briefly, the NRs can be subdivided into three groups based on their DNA binding characteristics: receptors that **homodimerize**, **heterodimerize** with one of the RXRs or bind as a **monomer** (Pawlak et al., 2012) (*Table 2*).

Table 2. NR family classification and DNA binding

Subfamilies	NRs	Consensus RE	Dimerization	Configuration
Steroid receptors	AR, PR	5'-AGAACA-3'	Homodimer	IR3, DR3 (selective AREs)
	GR, MR	5'-AGAACA-3'	Homodimer	IR3
	ER	5'-AGGTCA-3'	Homodimer	IR3
Non-steroid receptors Heterodimer with RXR	RAR	5'-AGGTCA-3'	Homodimer	IR0
			Heterodimer	DR1, DR2, DR5
	VDR	5'-AGGTCA-3'	Homodimer	DR3
			Heterodimer	DR3
	PPAR	5'-AGGTCA-3'	Heterodimer	DR1
	TR	5'-AGGTCA-3'	Monomer	Half-site
			Homodimer	DR4, IP6, P0
			Heterodimer	DR4
	RXR	5'-AGGTCA-3'	Homodimer	DR1
			Heterodimer	DR1, 2, 3, 4, 5
Orphan receptors	Nur77	5'-AAA AGGTCA-3'	Monomer	Extended half-site
	SF1, ERR2	5'-TCA AGGTCA-3'	Monomer	Extended half-site
	RORα,β,γ	5'-TCA AGGTCA-3'	Homodimer	Extended half-site

The first group of **homodimeric** receptors consists of the steroid receptors: the estrogen receptor (ER), the glucocorticoid receptor (GR), the mineralocorticoid receptor (MR), the progesterone receptor (PR) and the androgen receptor (AR). They can homodimerize on an inverted repeat of 5'-AGAACA-3'-like motifs with a 3-nucleotide spacer (IR3) (5'-AGGTCA-3' for the ER) (Pawlak et al., 2012) (*Table 2*) (**Figure 6**). The AR and the PR can also bind to 3-nucleotide spaced direct repeats of a similar hexamer (DR3), probably through two alternative dimerization interfaces (Denayer et al., 2010; Kerkhofs et al., 2012).

The second group of receptors comprises the receptors that **heterodimerize** with RXR, although some of them also homodimerize (Evans and Mangelsdorf, 2014). They recognize direct repeats of 5'-AGGTCA-3'-like motifs with receptor-specific spacer lengths (*Table 2*) (**Figure 6**) (Rastinejad et al., 1995).

Monomeric DNA binders, such as Nur77 (Meinke and Sigler, 1999), SF1 (steroidogenic factor 1) (Little et al., 2006) and ERR2 (estrogen receptor-related receptor 2) (Gearhart et al., 2003) are known to extend the DBD-DNA interface outside the major groove of the DNA. Additional contacts are formed between the C-terminal extension of the orphan receptor and the minor groove of the DNA upstream of the hexameric consensus sequence (*Table 2*) (**Figure 6**).

Receptors can also be recruited to DNA indirectly (tethering), via other sequence-specific TFs such as AP-1 and CREB1 (see in chapter 3, "Genomic-effects of GR") (Heldring et al., 2011; Sahu et al., 2011).
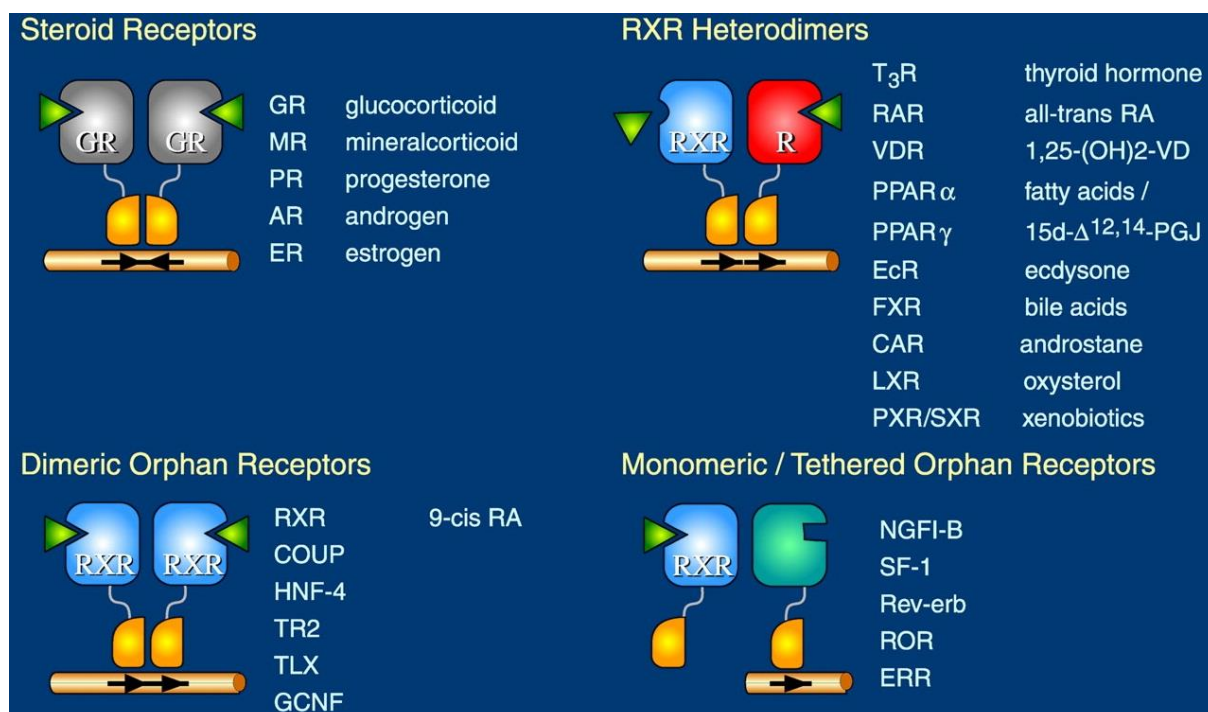
Figure 6. Nuclear receptors share common function domains. Nuclear receptors' classification according to ligand binding, DNA binding and dimerization properties (adapted from (Mangelsdorf et al., 1995)).

Structural organization of NRs

NRs share a common structural organization, which notably includes a conserved DNA binding domain (DBD) and a moderately conserved ligand-binding domain (LBD). The amino-terminal domains of NRs are highly variable in length and in sequence. Structural studies indicate they are flexible and intrinsically disordered (Khan et al., 2011; Kumar and Litwack, 2009; Kumar and Thompson, 2012). The hinge regions which connect the DNA- with the ligand-binding domains are the least conserved among the members of the NR family and their structures are poorly understood (Helsen and Claessens, 2014).

The **N-terminal domain** (A/B domain, NTD) is highly variable, and contains one constitutively active transactivation region (AF-1) and several autonomous transactivation domains (AD) (e.g. in H3 of VDR) (**Figure 7**). NTDs are variable in length, from less than 50 to more than 500 amino acids (aa), and their 3D structure is not known (Robinson-Rechavi et al., 2003). The steroid receptors AR, GR, MR and PR have large NTDs, ranging from 400 to 600 aa with MR having a very large NTD of 602 aa (Yang and Fuller, 2012). The NTDs of the non-steroid receptors are much shorter; for example the NTD of the VDR is only 24 aa long (Campbell et al., 2010). The AF-1 acts as multiple signal input and output domain integrating signals from different pathways, sometimes in cooperation with the signals that modulate the receptors activity via the LBD, the hinge and the DBD. However, the NTD-mediated mechanisms are mostly receptor-specific. While in the absence of binding partners the NTD is believed to be intrinsically disordered, interaction with their binding partners might induce appropriate folding of the activation functions (Helsen and Claessens, 2014).

The most conserved region is the **DNA-binding domain** (DBD, C domain), which notably contains the P-box, a short motif responsible for DNA-binding specificity on sequences typically containing the AGGTCA motif (**Figure 7**). The 3D structure of the DBD has been resolved for a number of NRs and contains two highly conserved zinc fingers – C-X2-C-X13-C-X2-C and CX5-C-X9-C-X2-C– the four cysteines of each finger chelating one Zn2+ ion (Helsen et al., 2012; Robinson-Rechavi et al., 2003). The α-helix in the first zinc finger module enables the sequence–specific interactions with the DNA. The second zinc finger module allows the receptor-DBDs to hetero- or homodimerize. The DBD dimerization has to be compatible with the format of the hormone response element (Kumar and McEwan, 2012; Robinson-Rechavi et al., 2003). The variations in spacer length and hexamer orientations have to be accommodated by the 2nd zinc finger and the C-terminal extension (Richmond and Davey, 2003).

Between the DNA- and ligand-binding domains, a less conserved region (D domain) that behaves as a flexible **hinge** contains a nuclear localization signal (NLS), which may overlap on the C domain (Robinson-Rechavi et al., 2003) (**Figure 7**). It also contains sites for post-translational modifications like phosphorylation, acetylation, methylation and sumoylation (Anbalagan et al., 2012; Clinckemalie et al., 2012).

The largest domain is the moderately conserved **ligand-binding domain** (LBD, E domain), with a secondary structure of 12 α-helixes (Robinson-Rechavi et al., 2003) (**Figure 7**). The 3D structure for several unliganded (apo) or liganded (holo) NRs has been determined (Moras and Gronemeyer,

1998), allowing a much better understanding of the mechanisms involved in ligand binding. The E domain is responsible for many functions, mostly ligand induced, notably the AF-2 transactivation function, a strong dimerization interface, another NLS, and often a repression function (Robinson-Rechavi et al., 2003).

The 12 a-helices of the LBD structures are arranged around a common central hydrophobic pocket, with helices 3, 7 and 10 providing amino acid residues that shape the ligand binding pocket (LBP) (Li et al., 2003; Nagy and Schwabe, 2004; Wurtz et al., 1996). The size of these ligand pockets can range in volume from zero (filled with the receptor's own hydrophobic side chains) to larger than 1500 Å^3 (Li et al., 2003). The C-terminal-most helical segment, helix 12 (H12), is the major architectural feature associated with AF-2 function, and can undergo dramatic shifts in position in response to the molecule in the pocket (de Lera et al., 2007; Nagy and Schwabe, 2004). Other α-helices in the LBDs also shift in positions in subtle ways that can impact receptor's activation (Huang et al., 2010).

The LBDs of the non-steroid receptors and of the ER also contain a surface required for receptor dimerization (Rastinejad et al., 2013). This interface is formed by helix 10, helix 9 and the loop between helix 7 and 8 (Bourguet et al., 1995). Dimerization via the NR-LBDs is known to occur in solution and to facilitate dimerization via the DBDs (Perlmann et al., 1996). While such dimerization via the LBD is well-known for VDR, THR, PPAR, RAR and RXR, there is no clear evidence that it occurs for GR, MR, PR and AR. The presence of a β-strand C-terminally of helix 12 in the AR, GR, MR and PR LBD structures could be a possible explanation for the absence of dimerization, since it covers helix 9, 10–11 and thereby prevents LBD dimerization as known for the non-steroid receptors (Helsen and Claessens, 2014; Schoch et al., 2010) (**Figure 8**).

NRs contain a final domain in the C-terminus of the E domain, the **F domain**, whose sequence is extremely variable and whose structure and function are unclear (Pawlak et al., 2012) (Robinson-Rechavi et al., 2003) (**Figure 7**).
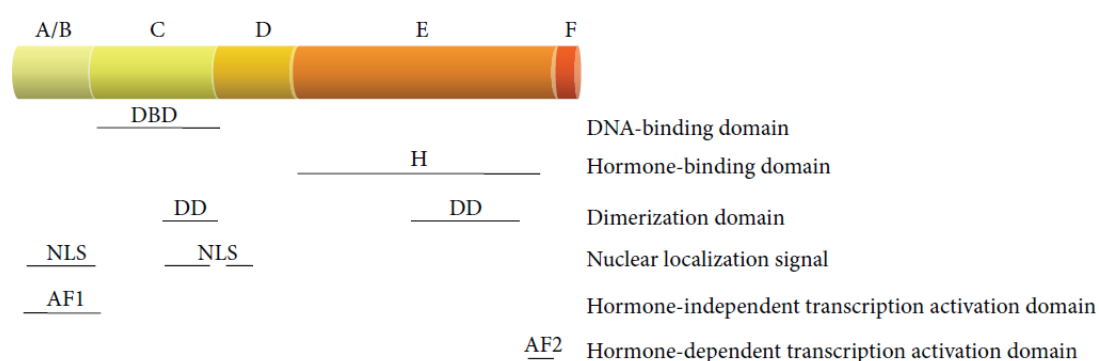
Figure 7. **Schematic diagram of NR structure**. N-terminal domain (A/B), DNA-binding domain (C), hinge (D), hormone-binding domain (E), C-terminal domain (F) (Puzianowska-Kuznicka et al., 2013).
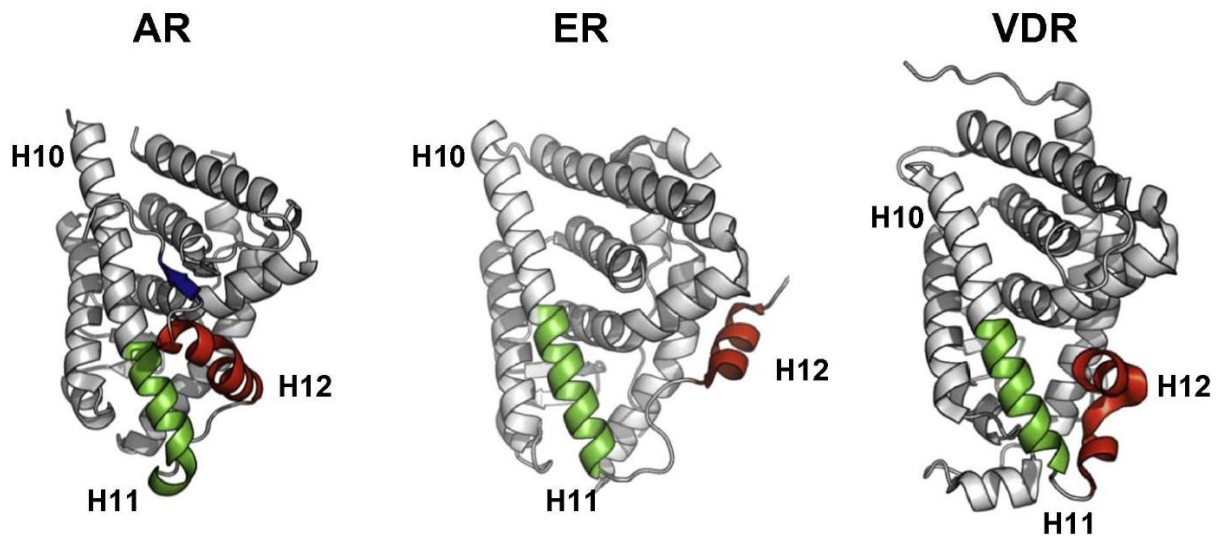
Figure 8. The LBD structure of AR, ER and VDR. PDB IDs: 2AMA, 2QXS and 3A78. Helix 12 (red) closes off the ligand-binding pocket. The β-sheet carboxyterminal of helix 12 in the AR (blue) is not present in ER and non-steroid receptors. Helix 10 and helix 11 (green) are the most important structural elements for homodimerization of ER and for homo- and heterodimerization of VDR. In AR, the C-terminal β-sheet is shielding this interface potentially preventing LBD dimerization (Helsen and Claessens, 2014).

Mode of action of NRs and regulation of transcription

The NRs can be grouped into 4 subtypes based on their mode of action:

Type I receptors, such as the AR, the ER and the PR, are bound in the cytoplasm by chaperone proteins (Echeverria and Picard, 2010) (**Figure 9**). Ligand binding frees the receptor from the chaperones, allowing it to homo-dimerize and exposes the nuclear localization sequence (NLS) thereby stimulating nuclear translocation (Sever and Glass, 2013) (**Figure 9**). Once in the nucleus, the ligand– receptor complex associates with transcriptional coactivators that facilitate binding to and activation of target genes (Bulynko and O'Malley, 2011; Glass and Rosenfeld, 2000) (**Figure 9**). Genome-wide location analysis indicates that most NR binding sites in the genome are located in enhancer elements that are far away from the transcriptional start site, as first documented for the estrogen receptor (Carroll et al., 2006).

Type II receptors, such as the THR and the RAR, in contrast, reside in the nucleus bound to their specific DNA response elements even in the absence of ligand (Maruvada et al., 2003) (**Figure 9**). They form heterodimers with the RXR and in the absence of ligand exert active repressive functions through interactions with NCoR and SMRT corepressor complexes (Chen and Evans, 1995; Horlein et al., 1995) that are associated with histone deacetylases (HDACs) (Watson et al., 2012b) (**Figure 9**). Binding of ligand to the LBD leads to dissociation of corepressors and their replacement with coactivator complexes (Sever and Glass, 2013). Coactivator complexes typically contain proteins with enzymatic functions, including histone acetyltransferases, that help open up chromatin and facilitate activation of target genes (Glass and Rosenfeld, 2000).

Type III receptors (principally NR subfamily 2) (**Figure 9**) function similarly to type I receptors except that the organization of the hormone response elements (HRE) differs (it is a direct repeat rather

than inverted repeat) and **type IV receptors** bind as monomers or dimers to half-site HREs (Mangelsdorf et al., 1995). Examples of type IV receptors are found in most of the NR subfamilies.
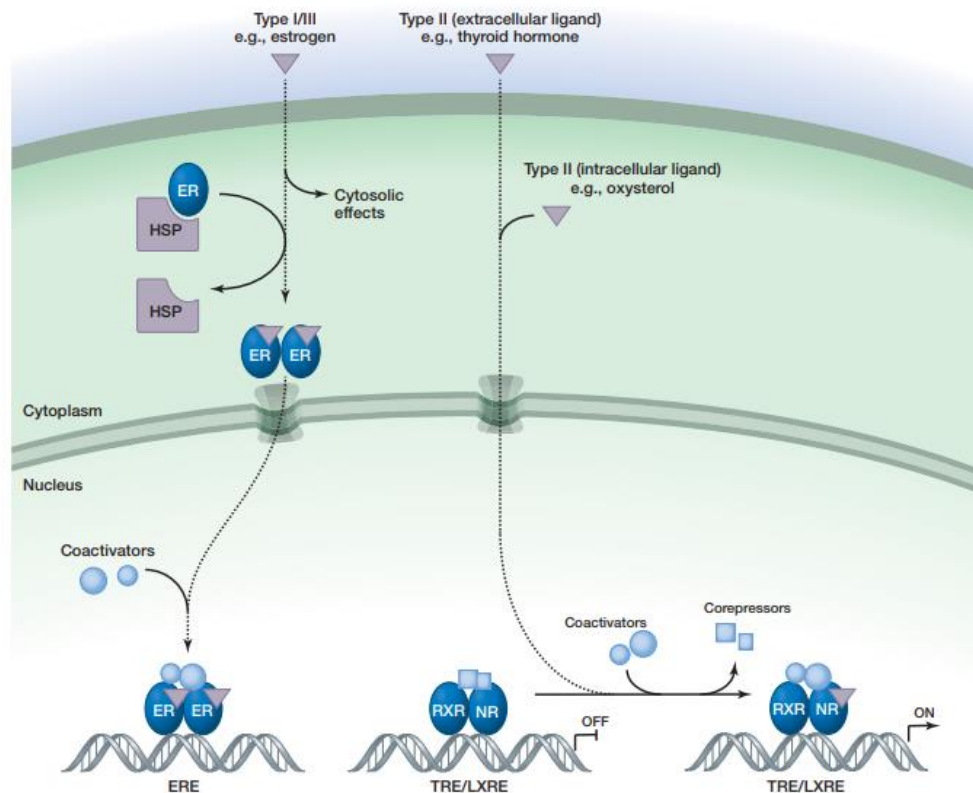
Figure 9. NR signaling (Sever and Glass 2013). The receptors can be grouped into four subtypes based on their mode of action. Type I receptors, such as the androgen, the estrogen and the progesterone receptors, type II receptors, such as the thyroid hormone and the retinoid acid receptors, type III receptors and type IV receptors.

NRs and transcription

NRs mediate a variety of effects on gene transcription. The most common modes of regulation are **ligand-dependent transactivation**, **ligand-independent repression** and **ligand-dependent repression** and **transrepression** of transcription. Much of this regulation is mediated by interactions of NRs with proteins called co-regulators, which include coactivators and corepressors (Lonard and O'Malley, 2012).

Ligand-dependent transactivation

Ligand-dependent activation is the most well understood function of NRs and their ligands (**Figure 10**). The ligand-bound receptor stimulates transcription of the bound target gene and the DBD brings the receptor domains that mediate transcriptional activation to a specific gene (Pawlak et al., 2012) (**Figure 10**). Transcriptional activation itself is mediated primarily by the LBD, which can

function as an independent unit even when it is transferred to a DNA-binding protein that is not related to NRs (Aagaard et al., 2011).

The ligand-bound NRs communicate stimulatory signals to General Transcription Factors (GTFs) on the gene to which they are bound. Ligands specifically recruit a subset of the co-regulators to the NRs LBD (Millard et al., 2013). Positively acting co-regulators, called **coactivators**, specifically recognize the ligand-bound conformation of the LBD and bind to the NRs only when an activating (agonist) hormone or ligand is bound (Bulynko and O'Malley, 2011) (*Table 3*).

The most important determinant of coactivator binding is the position of **H12**, which changes dramatically when activating ligands bind receptors. Along with H3, H4 and H5, H12 forms a hydrophobic cleft that is bound by short polypeptide regions of the coactivator molecules (Feng et al., 1998). These polypeptides, called NR boxes have characteristic sequences of **LxxLL**, in which L is leucine and xx can be any two aa (Heery et al., 1997). Coactivators increase the rate of gene transcription. This is accomplished by enzymatic functions, including histone acetyltransferase (HAT) activity (Berger, 2007).

Table 3. NRs co-regulators

Coactivators	Corepressors
Chromatic remodeling	NCoR (nuclear receptor corepressor)
SWI/SNF complex	SMRT (silencing mediator for RAR and THR)
Histone acetyltransferase	
p160 family (SRCs)	
p300/CBP	
PCAF (p300/CBP-associated factor)	
Mediator	

Ligand-independent repression

Some NRs (e.g. RAR, TR) are bound to DNA in the absence of their cognate ligand (Meyer et al., 2014) (**Figure 10**). DNA-bound receptor actively represses transcription of the target gene (Pawlak et al., 2012) and by reducing the expression of the target gene, this repressive function of the receptor amplifies the magnitude of the subsequent activation by hormone or ligand (Hu and Lazar, 2000) (**Figure 10**).

Unliganded NRs recruit negatively acting co-regulators, called **corepressors**, to the target gene (Hsia et al., 2010). The two major corepressors, NR corepressor (NCoR) and silencing mediator for RAR and TR (SMRT, also known as NCoR2), are large (≈ 270 kDa) proteins (*Table 3*) (Privalsky, 2004). NCoR and SMRT specifically recognize the unliganded conformation of NRs and use an amphipathic helical sequence similar to the NR box of coactivators to bind to a hydrophobic pocket in the receptor (Hsia et al., 2010).

Ligand-dependent repression and transrepression

Some gene targets of hormones are turned off in the presence of the ligand (Puzianowska-Kuznicka et al., 2013) (**Figure 10**). The mechanism of negative regulation is not fully understood and it consists of several mechanisms. One mechanism involves NR binding to DNA-binding sites that confers ligand-dependent negative transcriptional functions on the target gene (i.e., negative response elements) (Pawlak et al., 2012; Surjit et al., 2011) (**Figure 10**). Moreover, NR transrepression results in inhibition of signal-dependent transcriptional activation by other transcription factors associated at target sites (i.e. NF-kB and AP-1-dependent genes) (Pascual and Glass, 2006) (**Figure 10**).
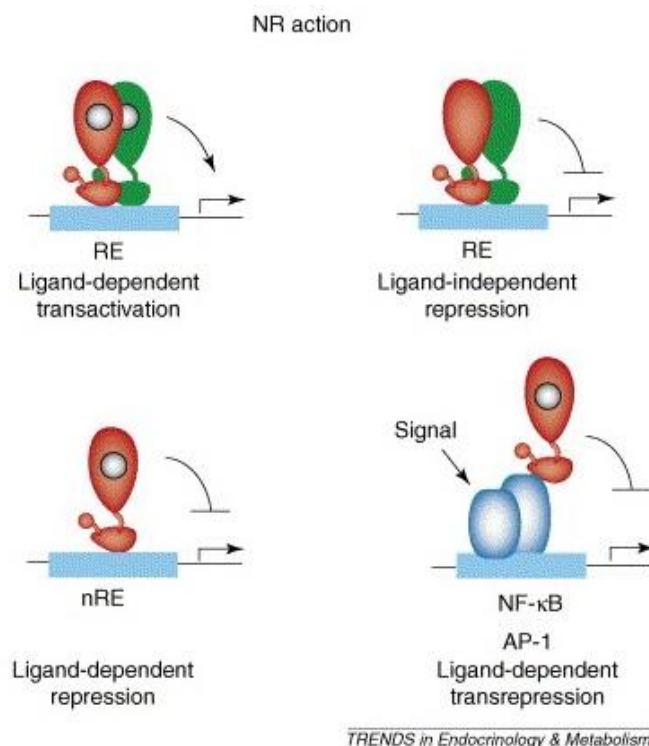
Figure 10. **General mechanisms of NR action** (adapted from (Pascual and Glass, 2006)). In the upper left illustration, ligand-dependent transactivation is highlighted as a sequence-specific DNA-binding event occurring as a consequence of heterodimer or homodimer binding to an NR response element (RE). In the upper right illustration, ligand-independent repression is mediated by some unliganded NR heterodimers, such as TR–RXR. In the lower left illustration, negative nuclear receptor response elements (nREs) are described for the GR, conferring ligand-dependent negative transcriptional functions on the target gene. In the lower right illustration, NR transrepression encompasses several mechanisms of NR action that result in inhibition of signal-dependent transcriptional activation by other transcription factors associated at target sites (e.g. NF-κB and AP-1-dependent genes).

Post-translational modifications (PTMs)

The functions of NRs can be modulated by post-translational modifications, like phosphorylation, SUMOylation, accetylation etc. (Berrabah et al., 2011).

Phosphorylation is one of the best-characterized modification (Liu et al., 2016). It is defined as the covalent addition of phosphate groups to specific amino acids, with the most common in eukaryotic cells being serine, threonine and tyrosine. It is catalyzed by kinases, whereas the removal of phosphate groups is performed by phosphatases (Ardito et al., 2017). Phosphorylation of NRs can alter protein-protein interaction, protein conformation and binding to the receptor to DNA, thus affecting their transcriptional activity (Lalevee et al., 2010). Phosphorylation can activate some NRs independently of ligand binding and function as the major mechanism regulating activities of orphan receptors (Berrabah et al., 2011).

Another modification is **ubiquitinylation**, an energy-dependent process in which an ubiquitin is transferred from an ubiquitin-activating enzyme (E1) to an ubiquitin-conjugating enzyme (E2) and finally to the target protein by a ligase enzyme (E3) (Garside et al., 2006).

SUMOylation is the covalent binding of members of the small ubiquitin-like modifier (SUMO) family to proteins (Wilkinson and Henley, 2010). In mammals, the SUMO family consists of 3 members: SUMO-1, SUMO-2 and SUMO-3 (Flotho and Melchior, 2013). SUMOylation is reversible and uses a specific set of enzymes for processing and attachment, such as the E1 SUMO-activating enzyme subunits 1/2 or members of the E3 ligases protein inhibitor of activated signal transducer and activator of transcription (PIAS) family and removal, known as SUMO peptides (Gareau and Lima, 2010). SUMOylation typically reduces the activation function of NRs and/or promotes repressor activity (Treuter and Venteclef, 2011) (Hua et al., 2016a).

Acetylation of lysine residues was initially identified in histones for their critical role in the control of gene expression (Verdone et al., 2005). Enzymes that add or remove acetyl groups are named histone acetyltransferases and histone deacetylases (HDACs), respectively (Seto and Yoshida, 2014). Approximately 85% of all eukaryotic non-histone proteins are acetylated (Glozak et al., 2005). NRs are acetylated at a phylogenetically conserved motif and more than a dozen NRs have been shown to function as substrates for acetyltransferases with diverse functional consequences (Wang et al., 2011).

Cell type specificity of NRs

The NRs have the ability to regulate specific genes in different cell types (Sever and Glass, 2013). Several studies indicate that tissue-specific responses are a consequence of binding of NRs to enhancer elements that are selected in a cell-specific manner (Pascual and Glass, 2006). Cell-specific enhancer selection is conferred by the key lineage-determining factors for each cell type, which interact in a collaborative manner to generate open regions of chromatin that provide access points for signal-dependent TFs (Bulynko and O'Malley, 2011; Heinz et al., 2010).

The non-genomic mechanism of hormone action

Fast biological effects of hormones, just minutes or even seconds after hormone administration, have been described (Puzianowska-Kuznicka et al., 2013). The rapidity of biological response and its independence from transcription and translation suggested that the genomic mechanism of hormone action is not involved; therefore, this mechanism is called **non-genomic** or **extra-genomic** (Losel et al., 2003). The non-genomic mechanisms of hormone action are multiple, variable, and only partially known (Puzianowska-Kuznicka et al., 2013), and will not be described in details here. Small part of NRs also act outside of the nucleus, in non-genomic mechanisms, which are mediated by processes other than a direct binding of the receptor to DNA (Ordonez-Moran and Munoz, 2009) (Unsworth et al., 2018).

General comments for NRs

Given the fact that many processes are controlled by NRs, their dysregulation can lead to several diseases, like cancer, diabetes and others (Dasgupta et al., 2014). However, they bind small molecules and represent therapeutic targets for which selective agonists and antagonists can be engineered (Burris et al., 2012). As NRs regulate many genes in various tissues, synthetic ligands with beneficial therapeutic effects also exert unwanted side effects, limiting their clinical use (Sever and Glass, 2013). Thus, it is important to better understand the mechanisms underlying their actions in specific cell types in order to allow a selective modulation of their activities.

3. Glucocorticoids and Glucocorticoid Receptor

Glucocorticoids (GCs) play an important role in various biological processes, like metabolism, homeostatic functions, development, inflammatory reactions and stress responses (Patel et al., 2014). The synthesis and release of natural GCs (cortisol in humans and corticosterone in rodents, cholesterol-derived hormones) is subject to a circadian and ultradian rhythm, controlled by the **hypothalamus-pituitary-adrenal axis** (HPA-axis) (**Figure 11**), with the lowest levels reached late night and early morning (Kadmiel and Cidlowski, 2013). Imbalance in GC levels such as chronic elevation or deficiency can result in pathological conditions known as Cushing's disease and Addison's disease, respectively (Kadmiel and Cidlowski, 2013).

Synthetic GCs, such as prednisone/prednisolone, dexamethasone (Dex) and budesonide are drugs that mimic natural GCs (Bindreither et al., 2014). However, synthetic GCs differ from natural ones by their potency, metabolic clearance (He et al., 2014) and by the fact that they do not bind to corticosteroid-binding globulin and are thereby not susceptible to their regulation of available levels (Kadmiel and Cidlowski, 2013). Synthetic GCs are being used for treatment of chronic inflammatory diseases like asthma, skin infections, ocular infections, as well as for immunosuppression in patients undergoing organ transplantation (Yasir and Sonthalia, 2019). In addition to their **anti-inflammatory** properties, corticosteroids have been exploited for their **anti-proliferative** and **antiangiogenic** actions for the treatment of cancers (Vilasco et al., 2011). The iatrogenic effects of the GCs vary from dermatological, ophthalmological, cardiovascular, gastrointestinal problems, to effects on bone and muscle, and metabolic and immune system defects (Yasir and Sonthalia, 2019).

Natural and synthetic GCs transduce their actions by binding to the Glucocorticoid receptor (GR) (Oakley and Cidlowski, 2013). GR is the product of a single gene, **NR3C1**, located on chromosome (chr) 5q31-32 in humans, that undergoes alternative processing to give multiple, functionally distinct subtypes of GR (Kino, 2000) (**Figure 12**). The human **NR3C1** gene contains 9 exons with the protein-coding region from the exon 2 to the exon 9 (Kino, 2000) (**Figure 12**). Exon 1 forms the 5'-untranslated region (Turner and Muller, 2005).

Alternative splicing of GR generates **hGRa** and **hGRb** isoforms, which differ after aa 727 (Oakley and Cidlowski, 2011) (**Figure 12**). The hGRa isoform binds to GCs, translocates to the nucleus, and recruits coregulators to exert transcriptional effects (Lu and Cidlowski, 2006). In contrast, the hGRb isoform resides constitutively in the nucleus and acts as a natural dominant negative inhibitor of the hGRa isoform (Lu and Cidlowski, 2006). The hGRb isoform can directly regulate genes that are not regulated by the hGRa isoform (Lu and Cidlowski, 2006). GRb isoforms are also present in mice and zebrafish, but are generated by an alternative splicing mechanism that is distinct from the GRb in humans (Otto et al., 1997) (Schaaf et al., 2008). The GRa isoform also undergoes alternative translation initiation in exon 2, generating eight additional isoforms of GR with truncated N-terminal (GRa-A, GRa-B, GRa-C1, GRa-C2, GRa-C3, GRa-D1, GRa-D2, and GRaD3) (Lu and Cidlowski, 2006) (**Figure 12**). GRb may also generate eight b isoforms similar to hGRa (Kino et al., 2009).
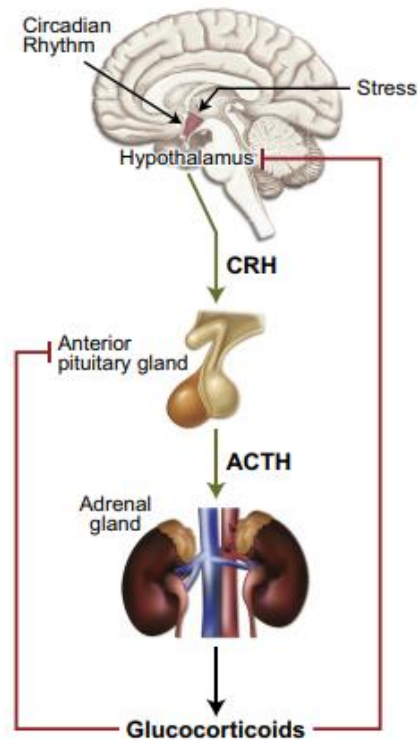
Figure 11. **Regulation of GC hormone secretion by the hypothalamic-pituitary-adrenal (HPA) axis**. CRH, Corticotropin-releasing hormone, ACTC, adrenocorticotrophic hormone (Oakley and Cidlowski, 2013)

The GR has the typical structure of a NR, containing a NTD, a central DBD, a LBD and a flexible hinge region (**Figure 12**). Diversity in GR signaling comes from the actions of different glucocorticoid response elements (**GRE**) and multiple receptor isoforms generated by alternative splicing and alternative translation initiation (Oakley and Cidlowski, 2011) (**Figure 12**). Furthermore, multiple post-translational modifications (PTMs) including phosphorylation (P), acetylation (A), ubiquitination (U), and SUMOylation (S) with small ubiquitin-related modifier proteins can alter the function of this TF (Anbalagan et al., 2012) (Table 4) (**Figure 12**). In the absence of GCs, GR is in the cytoplasm bound to chaperone proteins such, as heat shock protein 90 (hsp90) (Galigniana et al., 1998) (**Figure 13**).
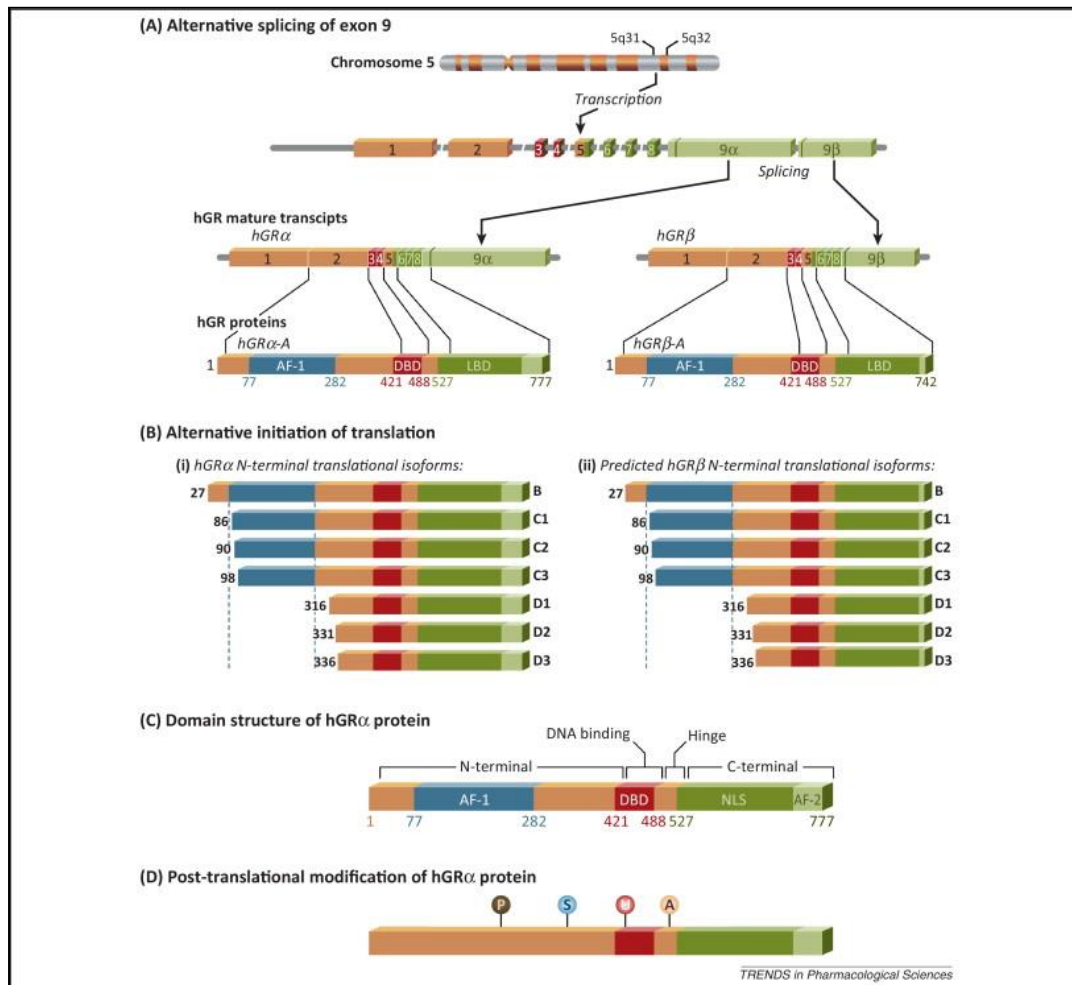
Figure 12. Genomic location and organization of the human GR (Kadmiel and Cidlowski, 2013). Human GR is located on chromosome 5q31–32. (A) GR undergoes alternative processing to yield multiple functionally distinct subtypes of GR. GR contains nine exons, with the protein-coding region formed by exons 2–9. Exon 1 forms the 5' untranslated region. Alternative splicing of GR generates the hGRα and hGRβ isoforms, which differ in their C termini. (B) The GRα isoform undergoes alternative translation initiation in exon 2, generating eight additional isoforms of GR with truncated N termini. GRβ is predicted to also generate eight β isoforms similar to hGRα. (C) The NTD has a strong transcription activation function (AF1) that allows for the recruitment of coregulators and transcription machinery. Glucocorticoids bind the hydrophobic pocket of the LBD, causing the second activation function (AF2), located in the LBD itself, to interact with coregulators. The DBD/hinge region junction and the LBD each contain a nuclear localization signal (NLS) that allows translocation to the nucleus. (D) GR undergoes multiple post-translational modifications including phosphorylation (P), SUMOylation with small ubiquitin-related modifier proteins (S), ubiquitination (U), and acetylation (A).

Table 4. Mechanisms of GR-mediated regulation

		Simple GREs
Genomic effects	Direct	Negative GREs
		Composite GREs
	Indirect	Tethered GREs
Non-genomic effects	Specific	Cytoplasmic GR
		Membrane-bound GR
	Non-specific	Not GR mediated

Genomic effects of the GR

Upon ligand-binding, GR translocates in the nucleus, it binds cofactors and exerts its genomic actions (transcriptional activation or repression) by direct binding to GREs found in the enhancer regions of GC target genes (Oakley and Cidlowski, 2013) (Ramamoorthy and Cidlowski, 2013a) (**Figure 13**). Some examples of genes upregulated by activated GR are the gene encoding GC-induced leucine zipper (GILZ) (Wang et al., 2004), serum/glucocorticoid-regulated kinase 1 (SGK1) (Itani et al., 2002), tristetraproline (Smoak and Cidlowski, 2006) and mitogen-activated protein kinase phosphatase-1 (MKP-1) (Barnes, 2011). Examples of genes negatively regulated by GR are b-arrestin 2 (Oakley et al., 2012), osteocalcin (Barnes, 2011) and the GR gene NR3C1 itself (Ramamoorthy and Cidlowski, 2013b) (Surjit et al., 2011).

The consensus GRE sequence 5'-**GGAACAnnnTGTTCT**-3' is an imperfect palindrome that is comprised of two 6-bp half sites (Oakley and Cidlowski, 2013). The GR binds this element as a homodimer, with each half site occupied by one receptor subunit (Oakley and Cidlowski, 2013). The 3-nucleotide spacing between the 2 half sites is required for the GR to dimerize on this element (Oakley and Cidlowski, 2013). The GRE has been shown to mediate the GC-dependent induction of many genes and is often referred to as an **activating** or **positive** GRE (Oakley and Cidlowski, 2013) (**Figure 13**). Coactivators and chromatin remodeling complexes are also recruited by GR to mediate transactivation (Wallberg et al., 2000).

However, genome-wide studies have revealed that GR occupancy of the canonical GREs can also lead to the **repression** of target genes (Uhlenhaut et al., 2013). A **negative** GC-responsive element (nGRE) that mediates the GC-dependent repression of specific genes has also been described (Surjit et al., 2011). The consensus nGRE sequence **CTCC(n)$_{0-2}$GGAGA** is palindromic but differs from the classic GRE in sequence, in having a variable spacer that ranges from 0 to 2 nucleotides, and in being occupied by 2 GR monomers (that do not homodimerize) (Hudson et al., 2013) (**Figure 13**). Hudson et al. have discovered through structural studies that binding of GR to nGREs prevents dimerization of the GR, whereas the converse is true when the GR binds to activating GREs (Hudson et al., 2013). Upon GR binding, corepressors (NCoR and SMRT) are recruited, which further recruit histone deacetylases (HDACs) to exert gene repression (Kadmiel and Cidlowski, 2013; Scheschowitsch et al., 2017).

Tethering is another way by which GR indirectly regulates gene expression (Xavier et al., 2016). In this case, GR is bound to other TFs and not directly to the DNA (Scheschowitsch et al., 2017) (**Figure 13**). For example, suppression of inflammation in diseases such as asthma occurs by GR tethering with pro-inflammatory TFs such as activator protein-1 (AP-1), nuclear factor kappa B (NF-κB), and signal transducer and activator of transcription 3 (STAT3) (Kassel and Herrlich, 2007; Langlais et al., 2012). Until recently, the mechanism of GR transrepression was thought to be primarily mediated by tethering. However, the finding by Surjit et al. (Surjit et al., 2011) has improved our understanding of the direct role of the GR in transrepression.
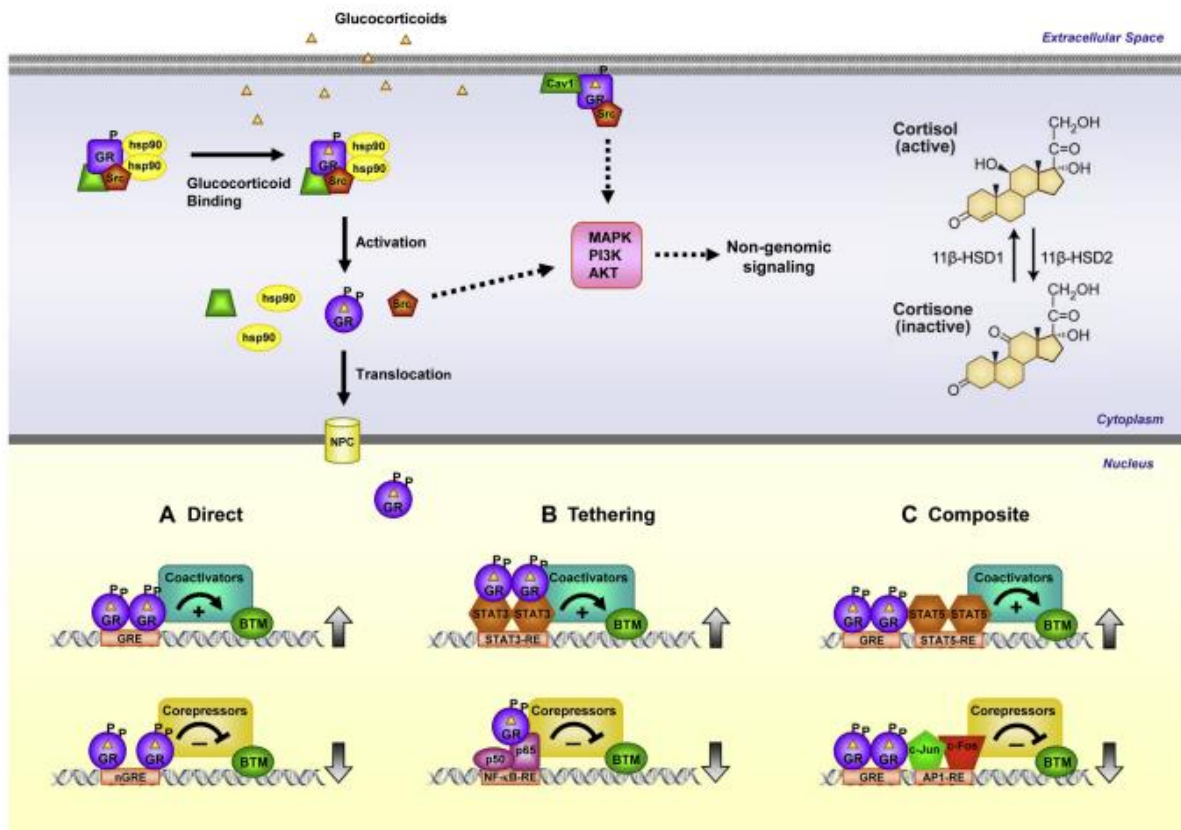
Figure 13. GR signaling pathways (Oakley and Cidlowski, 2013). Glucocorticoid-activated GR regulates gene expression in 3 primary ways: binding directly to DNA (A), tethering itself to other DNA-bound transcription factors (B), or binding directly to DNA and interacting with neighboring DNA-bound transcription factors (C). GR can also signal in a non-genomic manner through alterations in the activity of various kinases.

For STAT3, Langlais et al. demonstrated that GR tethering to DNA-bound Stat3 results in transrepression, whereas Stat3 tethering to GR results in synergirm (Langlais et al., 2012). Moreover, it was shown recently that the GR binds directly to AP-1 recognition motifs to repress inflammatory genes and the tethering is no more required (Weikum et al., 2017). GR also modulates gene expression by binding to **composite** GREs, wherein the target gene contains binding sites for GREs as well as other TFs like AP-1 and signal transducer and activator of transcription 5 (STAT5) (Oakley and Cidlowski, 2013) (**Figure 13**).

Non-genomic effects of the GR

The GCs can also exert their actions in a more rapid (within minutes), non-genomic signaling mechanism that does not require nuclear GR-mediated transcription or translation (Scheschowitsch et al., 2017). These actions are thought to be mediated by the activation of signal transduction pathways such as the mitogen-activated protein kinase (MAPK) pathway, by the membrane-bound GR or the cytoplasmic GR (Ayroldi et al., 2012) (Busillo and Cidlowski, 2013). Additionally, rapid effects, not specific to GR, also occur as a result of physiochemical interactions of GCs with the cell membrane (Song and Buttgereit, 2006). These rapid actions of the GR have been reported in various systems, including the cardiovascular, immune and neuroendocrine (Alangari, 2010; Mitre-Aguilar et al., 2015). These non-genomic GCs effects provide the basis for new drug development with better therapeutic index (Panettieri et al., 2019).

The musculoskeletal system

Skeletal muscle (skm) is a striated muscle tissue that serves critical functions in the organism such as movement and metabolism (Shadrin et al., 2016).

Myocytes or muscle cells are specialized cells with many nucleus due to the fusion of precursor cells known as **myoblasts** (Yin et al., 2013). They appear as striated cells due to myofibrils, an arrangement of intracellular structures (Bray et al., 2008) (**Figure 14**). Individual myofibrils are surrounded by a basal lamina beneath which a population of muscle progenitor cells, called **satellite cells**, is located (Yin et al., 2013) (**Figure 14**). Satellite cells are quiescent in adult muscles, but can be activated upon injury to regenerate muscles (Yablonka-Reuveni, 2011)**. Myofibrils** are composed of thick and thin filaments and each filament contains different proteins according to the function they serve during muscle contraction (Hooper et al., 2008). These filaments are arranged in a structure called **sarcomere**, the smallest unit of contraction (Pollard and Weihing, 1974). The thin filaments are mostly actin protein and compose the Z-line of the sarcomere (Luther, 2009). The thick filament is composed of the motor protein myosin and forms the M-line of the sarcomere (Pollard and Weihing, 1974). The thousands of sarcomeres in muscle are shifting together to contract the tissue to move.
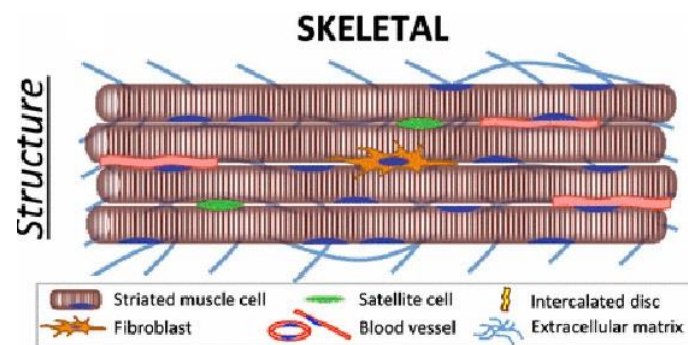
Figure 14. Adult skeletal muscle. It contains uniformly aligned, long multinucleated myofibers, blood vessels and satellite cells with few fibroblasts (adapted from (Shadrin et al., 2016)).

Myogenesis is the process of muscle development. The main events of this process are the specification of cell lineage, proliferation, migration and differentiation (**Figure 15**).

The precursor cells during mouse muscle development express the paired homeobox transcription factors **Pax3 (paired box 3)** and **Pax7 (paired box 7)**. Pax3/7 proteins play a role in tissue specification in several contexts and are not expressed in a muscle-specific manner. The progenitor cells are maintained during further development and are a source to generate all trunk and limbs skeletal muscles and associated satellite cells (Gros et al., 2005; Kassar-Duchossoy et al., 2005; Relaix et al., 2006). Pax7 is expressed in the central area of the dermomyotome and Pax3 is expressed in epaxial and hypaxial lips and more strongly in the latter (Hammond et al., 2007; Relaix et al., 2004). In Pax3-mutant mice all limb muscles are absent while some trunk muscles are still formed (Relaix et al., 2004). Pax7 is not essential during development, but is more important in the postnatal muscle where is expressed in quiescent satellite cells (Oustanina et al., 2004) (**Figure 15**). Pax7-mutant mice, to compromise muscle homeostasis and regeneration, lack satellite cells that progressively die after birth (Oustanina et al., 2004; Seale et al., 2000).

Skeletal muscle differentiation relies on the myogenic regulatory factors (MRFs) (**Figure 15**). The family of MRFs is composed of the determination factors **Myf5**, **MyoD**, **Mrf4** and the differentiation factor **Myogenin** (Myog) (Pownall and Emerson, 1992; Sassoon, 1993). These factors belong to the family of basic-helix-loop-helix (bHLH) transcription factors and bind to ubiquitous bHLH E-proteins and form heterodimers that recognize the E-box consensus sequence (CANNTG) on promoters and enhancers of muscle-related genes promoting their expression (Massari and Murre, 2000). MRFs are acting redundantly, but in the absence of all the three determination factors no skeletal muscle is formed (Rudnicki et al., 1992). Myog has no major effect in the specification of the early myogenic lineage, but when it is absent, myoblast differentiation and myofibre formation are impaired (Hasty et al., 1993; Nabeshima et al., 1993).
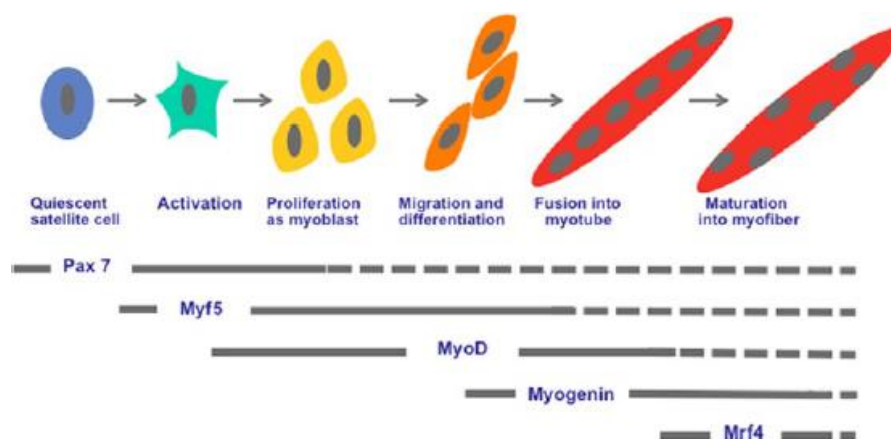
Figure 15. The myogenice regulatory factors pathway during myogenesis. Satellite cells are quiescent and express Pax7. Upon muscle damage, they are activated and express Myf5 to proliferate as myoblast, then expressing MyoD. MyoD is a key MRF which regulates myoblast differentiation during myogenesis. The MyoD-positive cells exit the cell cycle and express myogenin to initiate the differentiation and fusion into mytotubes. The myotubes have central nuclei. The expression of Mrf4 allows maturation of myotubes in myofibers that have peripheric nuclei (Zanou and Gailly, 2013).

The role of GCs in the musculoskeletal system

Skeletal muscle serves as major body store of aa, and GCs induce the catabolism of this tissue, increasing the plasma levels of free aa (Wise et al., 1973). The catabolic effects of GCs are well known for years (Seene and Viru, 1982). Either as drugs used to treat medical conditions or as endocrine hormones released in response to many stress situations, GCs can induce muscle atrophy. The resulting weakness of peripheral and respiratory muscles have major clinical implications, such as loss of quality of life, fatigue, compromised lung function, and poor immune response (Schakman et al., 2008a) (Yasir and Sonthalia, 2019). It is important to distinguish muscle atrophy from inflammatory myopathy. The former involves a reduction in the size of muscle fibers without disruption of the cell membrane, while in the latter there is a marked immune cell infiltration into muscle and loss of membrane integrity (Braun and Marks, 2015).

Role of GCs in muscle atrophy and wasting

The increase in circulating GCs levels is associated with pathological conditions characterized by muscle atrophy (sepsis, cachexia, starvation, metabolic acidosis, etc.) (Lecker et al., 1999), suggesting that these hormones could trigger muscle atrophy observed in these situations. In the case of sepsis, cachexia and starvation, adrenalectomy or treatment with a GR antagonist (RU-486) attenuates muscle atrophy, indicating that GCs are partially responsible for this muscle loss (Schakman et al., 2008a). In addition to GC excess, other factors such as poor nutrition and cytokines may contribute to muscle atrophy observed in these wasting conditions (Hasselgren, 1999).

Characterization of the GC-induced muscle atrophy

Muscle atrophy is characterized by a decrease in the size of the muscle fibers as well as muscle dysfunction characterized by reduced force and weakness (Bonaldo and Sandri, 2013; Shin et al., 2000). GCs have been shown to cause atrophy of fast-twitch or type II muscle fibers with less or no impact on type I fibers (Dekhuijzen et al., 1995). Therefore, fast-twitch glycolytic muscles (i.e., tibialis anterior) are more susceptible than oxidative muscles (i.e., soleus) to GC-induced muscle atrophy (Wang and Pessin, 2013).

Mechanisms of GC-induced muscle atrophy

In muscle, GCs decrease the rate of protein synthesis and increase the rate of protein breakdown contributing to atrophy (Goldberg et al., 1980; Tomas et al., 1979) (Lofberg et al., 2002). The severity and the mechanism for the catabolic effect of GCs may differ with age, as it was shown that GC-induced muscle atrophy results mainly from increased protein breakdown in adult rats, but mostly from depressed protein synthesis in the aged animals (Dardevet et al., 1998).

- Anti-anabolic action of GCs

The inhibitory effect on protein synthesis results from different mechanisms. First, GCs inhibit the transport of aa into the muscle which could limit the protein synthesis (Kostyo and Redmond, 1966). Secondly, GCs inhibit the stimulatory action of insulin, insulin-like growth factor-I (IGF-I) and aa, on the phosphorylation of Eif4E-binding protein 1 (4E-BP1) and the ribosomal protein S6 kinase

1 (S6K1) two factors that play a key role in the protein synthesis machinery by controlling the initiation step of mRNA translation (Shah et al., 2000a, b). Finally, there is also evidence that GCs cause muscle atrophy by inhibiting myogenesis through the downregulation of myogenin, a TF mandatory for differentiation of satellite cells into muscle fibers (te Pas et al., 2000).

- Catabolic action of GCs

The effect of GCs on muscle proteolysis results from the activation of the major cellular proteolytic systems, the ubiquitin-proteasome system (UPS), the lysosomal system (cathepsins) and the calcium-dependent system (calpains) (Hasselgren, 1999). The protein degradation caused by GCs affects mainly the myofibillar proteins, as demonstrated by the increased excretion of 3-methylhistidine (Zamir et al., 1991). To activate protein degradation, GCs stimulate the expression of several components of the UPS either involved in the conjugation to ubiquitin of the protein to be degraded [ubiquitin; 14 kDa (E2), a conjugating enzyme; atrogin-1 and MuRF-1, two muscle-specific (E3) ubiquitin ligases; (Bodine et al., 2001)]] or directly responsible for the protein degradation by the proteasome (several subunits of the 20S proteasome (Mitch and Goldberg, 1996). It was shown that GCs stimulate not only the UPS-dependent proteolysis, but also the calcium-dependent and lysosomal protein breakdown (Hasselgren, 1999).

Signaling pathways involved in GC-induced muscle atrophy

- mTOR

The inhibition of protein synthesis by GCs mainly results from the inhibition of mTOR, the kinase responsible for the phosphorylation of 4E-BP1 and SGK1 (Yoon, 2017). Repression of mTOR signaling results in a reduction in the initiation phase of mRNA translation with downregulation of protein synthesis (Showkat et al., 2014). The repression of mTOR signaling in response to GCs is the result of enhanced transcription of REDD1 (Wang et al., 2006). REDD1 repression of mTOR function leads to decreased phosphorylation of both 4E-BP1 and SGK1. Evidence suggests that mTOR signaling could also be inhibited directly by FOXO (Southgate et al., 2007) (Mori et al., 2014).

- FOXO (**Figure 16**)

Muscle catabolism caused by GCs is thought to be mediated by the FOXO TFs (Sukari et al., 2016). Indeed, exposure of myotubes to GCs increases the FOXO gene expression, particularly FOXO1 (forkhead box O1) and FOXO3 (forkhead box O3) (Imae et al., 2003). FOXO overexpression, in vitro as well in vivo, causes muscle cell atrophy (Sandri et al., 2004) together with activation of several genes characteristic of muscle cell atrophy or atrogenes such as atrogin-1/MAFbx, MuRF-1, autophagy-related genes (Mammucari et al., 2007), myostatin (Allen and Unterman, 2007) and cathepsin L (Sandri et al., 2004). Moreover, overexpression of a dominant negative form of FOXO-3a prevents muscle cell atrophy together with atrogin-1 induction caused by GCs in vitro (Sandri et al., 2004). As FOXO overexpression, but not of atrogin-1, is sufficient to cause muscle atrophy, FOXO TFs activate a variety of genes, in addition to atrogin-1, to induce atrophy.

- p300 – C/EBPbeta (**Figure 16**)

GC-induced muscle proteolysis is at least in part regulated by p300–histone acetyl transferase activity (p300). Indeed, p300 protein levels and activity are increased, in a time- and dose

dependent manner, in Dex-treated myotubes (Yang et al., 2005). Finally, treatment of myotubes with p300 small interfering RNA prevents the Dex-induced increase in protein degradation, whereas overexpression of wild-type p300 potentiates the effect of Dex on protein degradation (Yang et al., 2007).

- MyoD & MyoG (**Figure 16**)

The TF MyoD regulates muscle differentiation and development (Wilson and Rotwein, 2006). It is also required for regeneration and self-renewal of muscle satellite cells (Megeney et al., 1996). The transcriptional activities of MyoD are negatively regulated by a family of inhibitors of DNA-binding (Id) proteins among which Id1 (inhibitor of DNA binding 1, HLH protein) is the most important factor with regard to MyoD binding (Jen et al., 1992). Muscle wasting is characterized by decreased levels of MyoD, reflecting ubiquitin-proteasome-dependent degradation of the TF (Lingbeck et al., 2003). TNF (tumor necrosis factor) can reduce the MyoD protein abundance in muscle cells secondary to NF-kB activation and these effects of TNF may play a role in muscle wasting and cachexia (Guttridge et al., 2000). GCs can also stimulate the degradation of MyoD and this may be a mechanism of GC-induced muscle wasting (Sun et al., 2008).

MyoG is an additional myogenic TF that is involved in muscle differentiation (Sassoon et al., 1989). Unlike MyoD, it is not known if MyoG levels change during muscle atrophy (Macpherson et al., 2011). A study suggests, however, that GC-induced muscle wasting may in fact be characterized by reduced MyoG levels (Jogo et al., 2009) and that MyoG is degraded by the ubiquitin-proteasome pathway after exposure of muscle cells to Dex.

- Hyperacetylation (**Figure 16**)

Acetylation of cistromes and other cellular proteins is regulated by histone acetyl transferases (HAT) and by histone deacetylases (HDAC) (Hasselgren et al., 2010). Hyperacetylation in GC-induced muscle wasting may induce muscle proteolysis (Yang et al., 2007) and prevention of hyperacetylation may be a therapeutic strategy to reduce the loss of muscle mass in catabolic patients and in individuals treated with corticosteroids (Alamdari et al., 2013). Moreover, in clinic, small molecules were developed that can reduce acetylation by activating HDACs (Milne et al., 2007).
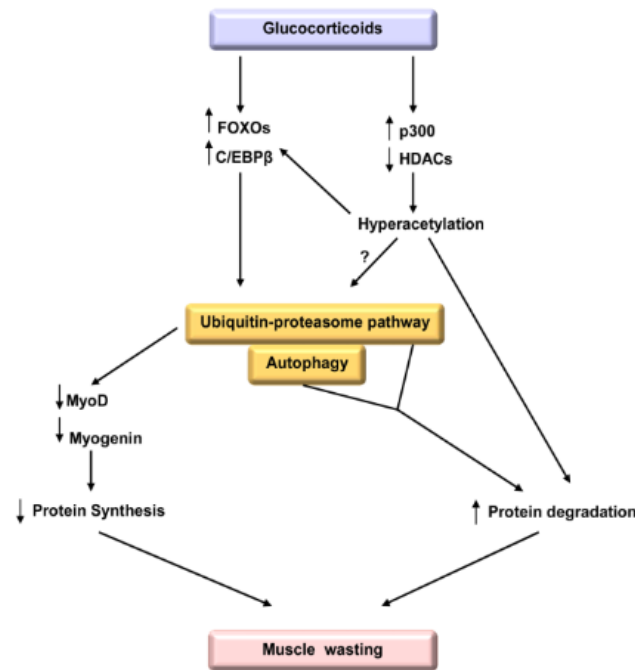
Figure 16. Potential mechanisms involved in GC-induced muscle wasting (Hasselgren et al., 2010). Studies suggest that the expression and activity of FOXO transcription factors and C/EBPβ are upregulated by glucocorticoids and that hyperacetylation caused by increased p300/HAT and decreased HDAC expression and activity may contribute to transcription factor activation. Proteasome-dependent degradation may contribute to reduced expression and activity of the "anabolic transcription factors" MyoD and myogenin, further accentuating the loss of muscle mass. Although there is evidence that glucocorticoid-induced hyperacetylation stimulates muscle protein degradation, the role of hyperacetylation in the regulation of the ubiquitin-proteasome pathway and of autophagic/lysosomal protein degradation is not known at present (as indicated by the question mark).

Role of local growth factors in GC-induced muscle atrophy

- Insulin-like growth factor 1 (IGF-I) (**Figure 17**)

GCs can cause muscle atrophy by altering the production of growth factors that control locally the muscle mass development (Schakman et al., 2008a). GCs inhibit the production by the muscle of IGF-I (Gayan-Ramirez et al., 1999), a growth factor that stimulates the muscle mass by increasing protein synthesis and myogenesis, while decreasing proteolysis and apoptosis (Frost and Lang, 2003). For these reasons, decreased muscle IGF-I has been thought to play a key role in GC-induced muscle atrophy (Nystrom et al., 2009). This hypothesis has been confirmed both in vitro and in vivo. First, by activating the PI3K/Akt/Mtor pathway and blocking nuclear translocation of the TF FOXO, IGF-I downregulates the proteolytic systems (lysosomal, proteasomal, and calpain dependent) and the expression of atrogenes such as atrogin-1, MuRF-1, cathepsin L induced by GCs (Latres et al., 2005). Secondly, IGF-I suppresses muscle cell atrophy induced by GCs in vitro (Sacheck et al., 2004). Thirdly, systemic administration (Tomas et al., 1992) or local overexpression of IGF-I into muscle prevents GC-induced muscle atrophy (Schakman et al., 2005). Taken together, IGF-I has a dominant role, not allowing GCs to turn off catabolism. In addition, decreased muscle IGF-I plays a role in the

atrophy caused by GCs (Schakman et al., 2008b). Therefore, restoration of IGF-I may provide a strategy to reverse the catabolic effects of GC excess (Song et al., 2013).

- Myostatin (Mstn) **(Figure 17)**

GCs also stimulate the production of Mstn by the muscle (Ma et al., 2003), a growth factor that inhibits the muscle mass development by downregulating the proliferation and differentiation of satellite cells (Thomas et al., 2000) and protein synthesis (Taylor et al., 2001). In vitro evidence indicates that Mstn also causes muscle cell atrophy by reversing the IGF-I/PI3K/Akt hypertrophy pathway (Elkina et al., 2011). Through inhibition of Akt phosphorylation, Mstn increases the levels of active FOXO, allowing increased expression of atrogenes (McFarlane et al., 2006). Furthermore, targeted disruption of Mstn gene expression in mice leads to dramatic increase in muscle mass due to fiber hyperplasia and/or hypertrophy (Grobet et al., 2003). Finally, transgenic mice that express Mstn selectively in muscle have muscle atrophy (Reisz-Porszasz et al., 2003).
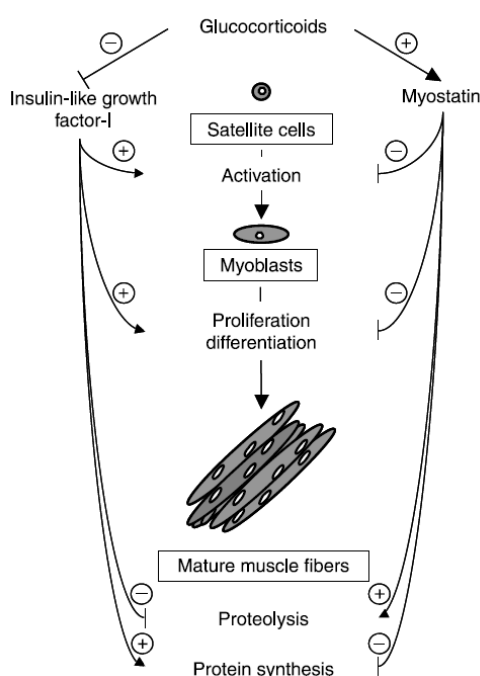
Figure 17. **Local growth factors production plays a crucial role in GC-induced muscle atrophy** (Schakman et al., 2008a). Glucocorticoids can cause muscle atrophy by altering the muscle production of IGF-I and myostatin, two growth factors exhibiting opposite effects on muscle mass development. Decrease in IGF-I together with increase in myostatin both induced by glucocorticoids inhibit satellite cells activation as well as myoblast proliferation and differentiation. In mature muscle fibers, these growth factor changes cause downregulation of protein synthesis and stimulation of protein degradation.

Prevention and treatment of GC-induced muscle atrophy

- Androgens (ADs)

Administration of ADs, such as testosterone or nandrolone, a minimally aromatizable analog, prevents decreased muscle mass and strength induced by GCs in animals (Van Balkom et al., 1998) and humans (Crawford et al., 2003). Although the molecular mechanisms by which testosterone attenuates the effects of GCs are not fully uncovered, testosterone, like many other anabolic stimuli, appears to stimulate muscle IGF-I expression (Wu et al., 2007).

- Dissociated GR agonists

Other potential treatments that have not yet been reported in the context of GC-induced muscle wasting are two novel classes of agents, i.e., dissociated GR agonists and 11β -hydroxysteroid dehydrogenase type 1 (11β -HSD1) inhibitors (Hasselgren et al., 2010). Dissociated GR agonists are designed to induce GC-regulated transrepression pathways while minimizing transactivation activity, the latter being responsible for metabolic side-effects (probably including muscle wasting) of GCs (Rosen and Miner, 2005) (Schacke et al., 2007). Moreover, dissociated GR ligands are useful to prevent loss of muscle mass in conditions characterized by GC-regulated muscle wasting and in patients being treated with GCs.

4. Androgens and Androgen Receptor

Androgens (ADs) are male sex hormones required for development of the male reproductive system and secondary sexual characteristics (Chang et al., 1995). Testosterone is synthesized primarily by the Leydig cells in the testes, under the regulation of luteinizing hormone (LH) produced by the anterior pituitary gland (Ramaswamy and Weinbauer, 2014). Testosterone can be converted by 5α reductase into its more biologically active form, dihydrotestosterone (DHT), and to oestradiol by aromatase (Davey and Grossmann, 2016). Testosterone and DHT mediate their actions via the Androgen Receptor (AR), a ligand-dependent nuclear TF, through a high affinity binding (Chang et al., 1995; Grino et al., 1990) (**Figure 18**).

AR (NR3C4, nuclear receptor subfamily 3, group C, gene 4) is a member of the steroid hormone NR family, located on the X chromosome (**Figure 19**). It is expressed in diverse tissues, such as bone, muscle, prostate, adipose tissue and the reproductive, cardiovascular, immune, neural and haemopoietic systems (Rana et al., 2014). The protein coding region has 2757 nucleotides and spans eight exons (**Figure 19**) (Tan et al., 2015). The AR gene encodes a 110-kDa protein consisting of 919 aa (**Figure 19**) (Gelmann, 2002).

AR, like the other members of the family, comprises three main functional domains: the N-terminal domain (NTD), the DNA binding domain (DBD) and the ligand binding domain (LBD) (**Figure 19**) (MacLean et al., 1997). Given the highly conserved nature of the DBD among the steroid hormone NR family, it has been shown that binding of selective androgen response elements (AREs) allow specific activation of the AR (Shaffer et al., 2004). The LBD mediates the interaction between the AR, heat shock and chaperone proteins, while also interacting with the N-terminus of the AR to stabilize bound ADs (Heinlein and Chang, 2002). The AF-1 (residues 142–485) in the NTD is constitutively active (McEwan, 2004), whereas the AF-2 is ligand dependent (He et al., 1999).
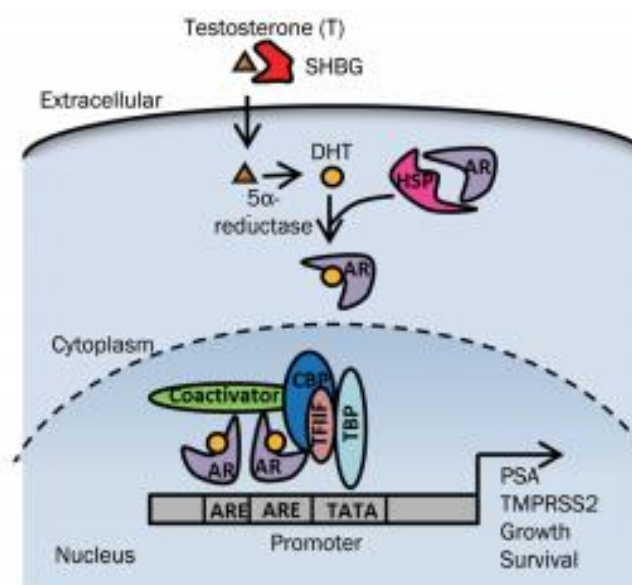
Figure 18. Androgens and AR action in prostate cells
(Tan et al., 2015).
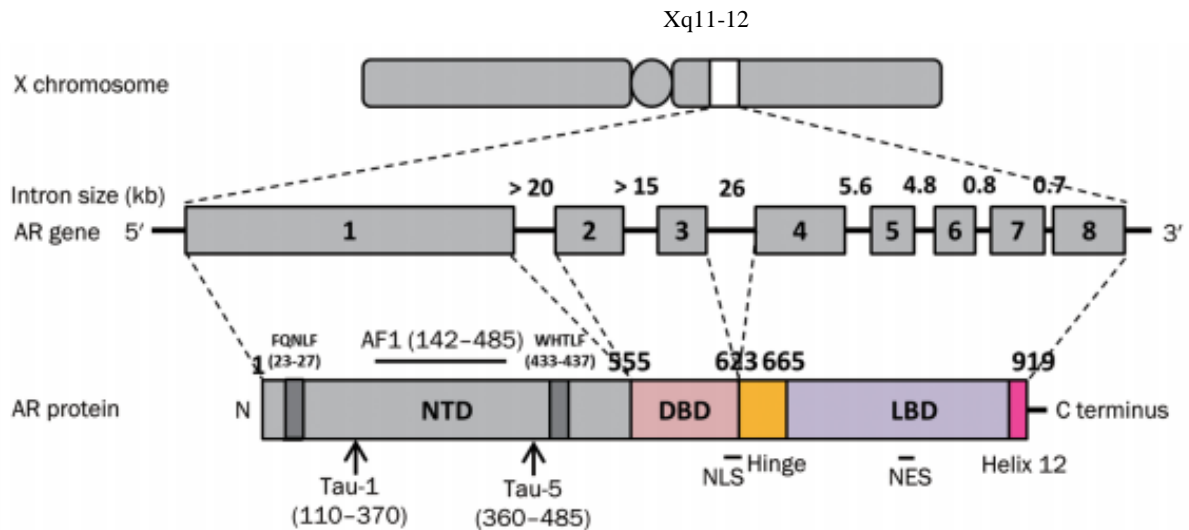
Figure 19. **The functional domain structure of the AR protein** (Tan et al., 2015). The androgen receptor gene is mapped to the long arm of the X-chromosome (locus: Xq11-q12). It contains 8 exons and introns of varying length and codes for a 919 aa consisting of several functional domains. Exon 1 codes for the NTD, exons 2 and 3 encode the DBD, and exons 4 to 8 encode both the hinge and LBD.

DNA binding-dependent actions of the AR

The binding-dependent actions of the AR are also referred to as "genomic", "classical" or "canonical" AR signaling (Davey and Grossmann, 2016) (**Figure 20**).

In the absence of ligand, AR is cytoplasmic, associated with heat shock and other chaperone proteins (Galigniana et al., 2010) (**Figure 18**). ADs bind to AR, resulting in a conformational change, AR is dissociated from the chaperone proteins and the NLS is exposed (Davey and Grossmann, 2016; Srinivas-Shankar and Wu, 2006) (**Figure 18**). The AD/AR complex is translocated to the nucleus where it dimerizes and binds to AREs within promoters of classical target genes to modulate gene transcription (Eder et al., 2001) (**Figure 18**). The transcriptional activity of the AD-bound AR is modulated by specific proteins known as co-regulators (Heemers and Tindall, 2007) (**Figure 18**). Co-regulators bind to the activated AR in a ligand-dependent manner to either enhance (co-activator) or repress (co-repressor) its ability to transactivate the target genes through chromatin remodeling and histone modifications, as well as being involved in the recruitment of the basal transcriptional machinery (Bevan and Parker, 1999) (Shang et al., 2002; van de Wijngaart et al., 2012) (**Figure 18**).

Non-DNA binding-dependent actions of the AR

The DNA binding independent actions of the AR are known as "non-genomic", "non-classical" or "non-canonical" AR signaling (Davey and Grossmann, 2016) (**Figure 20**).

The AD/AR complex can also signal through non-DNA binding-dependent pathways (Estrada et al., 2003). Activation of 2nd messenger pathways including ERK, Akt and MAPK has been identified in a

number of cell lines (Kang et al., 2004). These effects occur within seconds to minutes of ADs treatment and they are too rapid to have been initiated via the DBD actions of the AR to regulate the transcription and translation of target genes (Davey and Grossmann, 2016). Indirect gene transrepression can also occur, by the AR binding and sequestering TFs such as AP-1 that are normally required to upregulate target gene expression in the absence of the AR binding to DNA (Davey and Grossmann, 2016).

The physiological significance of the non-DNA binding-dependent actions of the AR is not yet fully understood and it has been proposed that they serve as a brake to the normal androgen action in target tissues (Davey and Grossmann, 2016).
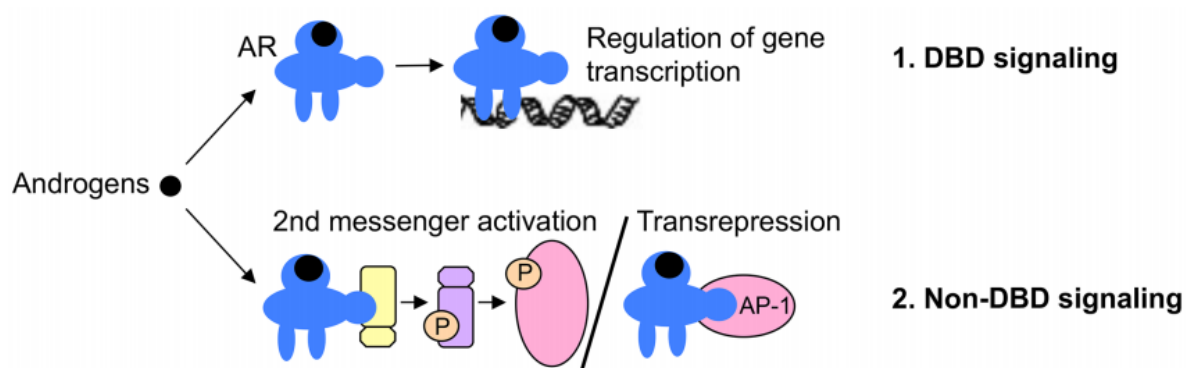
Figure 20. **Mechanisms of ligand-dependent AR action** (Davey and Grossmann, 2016). (1) DNA binding-dependent (DBD) and (2) nonDNA binding (DBD)-dependent. (AP-1 – activator protein 1).

Ligand-independent actions of the AR

There is evidence suggesting that the AR can act in a ligand-independent manner (Weigel and Zhang, 1998). Ligand independent activation of the AR by a number of different growth factors has been demonstrated, via phosphorylation of the AR or following interaction with co-activators (Ueda et al., 2002). One such pathway identified is IL-6, the circulating levels of which are commonly elevated in patients with metastatic prostate cancer (Drachenberg et al., 1999). IL-6 upregulates AR activity in a ligand-independent manner via the protein kinase A (PKA), protein kinase C (PKC) and MAPK pathways, and as such has important clinical implications for prostate cancer patients with low ADs levels as a result of androgen deprivation therapy (Hobisch et al., 1998).

Ligand-independent AR activation is one mechanism through which prostate cancer develops hormone resistance (Hu et al., 2009a). Similar to GR, it has been shown in a prostate cancer cell line that ligand-independent AR regulates a distinct group of target genes compared with ligand-bound AR (Lin et al., 2009). Ligand-independent actions of the AR have also been identified in the C2C12 cell line, where IGF-I stimulates phosphorylation, nuclear localization and DNA binding activity of the AR and upregulation of the expression of known AR target genes via the MAPK pathway (Kim and Lee, 2009). However it is still unclear whether ligand-independent AR pathways are limited to prostate cancer or play a role in normal physiology (Davey and Grossmann, 2016).

AR and prostate

Prostate growth depends on the presence of ADs. Functional AR and its activation by DHT is critical for complete prostate development as men lacking a functional 5α-reductase gene have only a small partial prostate or the lack of prostate (Heinlein and Chang, 2004; Koochekpour, 2010).

Prostate cancer (PCa) is also dependent on the actions of ADs and functional AR expression, and tumors will regress temporarily with castration (Eisermann et al., 2013). AR is expressed in both AD-dependent (ADD) and –independent (ADI) PCa and is sustained throughout progression of the disease to hormone refractory PCa (Knudsen and Penning, 2010; Yuan and Balk, 2009). PCa therapy is based on blocking androgen activity. Androgen ablation therapy in turn causes atrophy of the prostate epithelium (Eisermann et al., 2013). Treatment of metastatic PCa involves androgen deprivation therapy (ADT) through blocking production of ADs by castration and/or by using anti-ADs such as bicalutamide or enzalutamide (MDV3100) (Lin et al., 2013b). When ADs ablation therapies fail, advanced PCa ultimately progresses to a late stage that is refractory to current therapies, also known as castrate resistant prostate cancer (CRPC). This recurrence results from a reactivation of AR activity (Eisermann et al., 2013). AR signaling pathways play critical role in both ADD and CRPC (Hoang et al., 2017).

Decreased AR protein expression levels can reduce both primary localized PCa and CRPC growth (Eisermann et al., 2013). ADT is initially successful in most patients (~80%) resulting in tumor regression and AR suppression. However, these therapies fail at the end and the cancer progresses to a stage where it is unresponsive to blockage of ADs and growth becomes ADI (Eisermann et al., 2013). Overexpression or amplification of the AR in CRPC seems to be induced by hormone suppression (Waltering et al., 2012). Many mechanisms have been proposed to play a role in this reactivation of AR following ADT including: deregulation (causing overexpression of AR), mutation of AR (gain of function), alternative splicing (causing AR to be constitutively active), co-activator gain of function or loss of co-repressor function, and intracrine AD synthesis [reviewed in (Knudsen and Penning, 2010)].

5. Interplay of androgen and glucocorticoid receptors

Increasing evidence indicates that androgen and glucocorticoid signaling pathways are highly interconnected. Indeed, ADs have been shown to downregulate GR expression in various cell lines (Arora et al., 2013; Davies and Rushmere, 1990; Isikbay et al., 2014; Xie et al., 2015). In addition, GCs downregulate ADs synthesis through a feedback inhibitory mechanism of the hypothalamic/pituitary axis (Hardy et al., 2005; Ing et al., 2014). Moreover, recent genome-wide analyses revealed that the two half-sites of natural GR and AR binding sequences are generally imperfect inverted repeats, with the sequence of the first half-site being more conserved than that of the spacer and the second half-site. In addition, GR and AR share 1/3 of their cistromes depending on the cell type, and they regulate similarly distinct sets of target genes (Arora et al., 2013; Sahu et al., 2013; Zhang et al., 2018). In addition, GCs can also bind to AR with mutations in the LBD (L701H and T877A) driving AR signaling and tumor proliferation (Zhao et al., 2000).

As the activity of AR and GR can be modulated by synthetic ligands, they represent important drug targets for a number of diseases, including cancer, sarcopenia, allergies and asthma (Brill et al., 2002; Claessens et al., 2008; Kadmiel and Cidlowski, 2013). Even though synthetic GCs with potent anti-inflammatory and immunosuppressive activities are largely used in the clinic, long-term treatments are limited by adverse effects, including diabetes, osteoporosis, muscle wasting and myopathies. Moreover, anabolic effects of ADs on skeletal muscles are of interest to improve muscle function in elderly men and in patients with myopathies and AIDS. Nevertheless, as ADs also induce prostatic epithelial cell proliferation, they increase the risk of prostate cancer. Conversely, anti-androgens that are used as the primary treatment of metastatic prostate cancer induce muscle atrophy, and thus impair the quality of life of patients and increase the risk of fractures (Bhasin et al., 2006; Nguyen et al., 2015).

These clinical observations indicate that signaling pathways controlled by GCs and ADs are interconnected, and underline the need for AR and GR ligands with increased selective activities.

Control of gene transcription by AR and GR involves highly complex and poorly characterized molecular mechanisms, although there has been enormous interest in elucidating their structure and function. The classical mode of action of GR and AR proposes that cognate ligands promote receptor binding to their response elements (GREs and AREs) to induce target gene expression. These elements are organized as inverted repeats (IR) of 5'-AGAACA-3' like motifs, separated by three base pairs (IR3) (Meijsing et al., 2009; Watson et al., 2013). However, there also exists an AR-selective ARE (5'-AGAACAnnnAGAACA-3') (Claessens et al., 2008; Claessens et al., 2001; Shaffer et al., 2004; Verrijdt et al., 2003). Importantly, GR and AR can bind as homodimers to the consensus IR3 binding elements (*Figure 21*) (Hard et al., 1990). However, the dimerization behavior of AR and GR is not yet well characterized (Billas and Moras, 2013). In addition, like other transcription factors, GR and AR modulate gene expression by recruiting co-regulatory proteins (Dasgupta et al., 2014; Malovannaya et al., 2011).
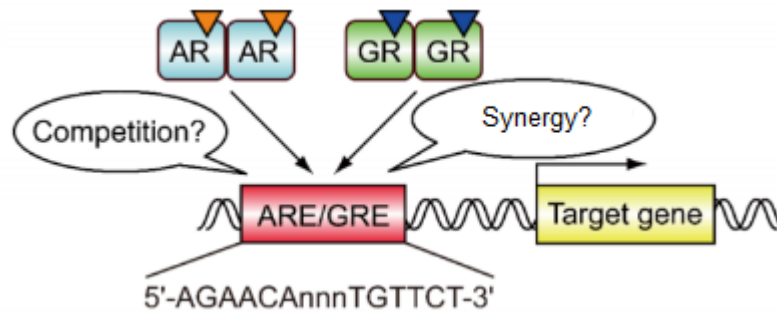
Figure 21. Function of AR and GR as ligand-dependent transcription factors (adapted from (Harada N. et al. 2015)). Androgens and glucocorticoids have competitive and compensatory effects in several physiological and pathophysiological processes.

Androgens and glucocorticoids in skeletal muscle

Skeletal muscle is a tissue known to express both receptors in myofibers and to respond, in both males and females, to ADs and GCs in a rather opposite way, as they have anabolic and catabolic effects, respectively (Qin et al., 2010; Van Balkom et al., 1998; Wu et al., 2010; Zhao et al., 2008).

ADs have anabolic actions on muscle and bones (Yin et al., 2003). Testosterone, the main AD in muscle (Bhasin et al., 2003), increases muscle size and strength both in young (Bhasin et al., 1996) and older men (Bhasin et al., 2005). The testosterone-induced increase in muscle mass is partly due to muscle fiber hypertrophy, reflected by an increase in myonuclear number and cross-sectional area of both type I and type II muscle fibers (Sinha-Hikim et al., 2002). The responsiveness of muscle to ADs could potentially be exploited clinically in the treatment of various chronic diseases that are accompanied by muscle wasting, such as cancer, cachexia, AIDS, chronic obstructive pulmonary disease, chronic renal disease (MacLean and Handelsman, 2009).

The protein hypothesis states that testosterone administration induces an increase in muscle protein synthesis (Ferrando et al., 1998; Urban et al., 1995) and an improved recycling of intracellular amino acids (Ferrando et al., 1998; Sheffield-Moore et al., 1999) (**Figure 22**). The proposed effects of ADs on muscle protein degradation, however, are less clear: short-term treatment does not appear to change the breakdown rate (Ferrando et al., 1998; Sheffield-Moore et al., 1999), whereas treatment for several months decreases muscle protein breakdown (Ferrando et al., 2003; Ferrando et al., 2002) (**Figure 22**). Testosterone induced muscle hypertrophy may thus be explained by changes in muscle protein metabolism (Atherton and Smith, 2012). However, ADs also mediate changes in body composition characterized by an increase in lean body mass accompanied by a concomitant decrease in fat mass (Wittert et al., 2003), which are difficult to explain only by muscle protein synthesis and/or breakdown (**Figure 22**). It is questionable therefore how ADs may induce differential anabolic actions such as changes in body composition as well as muscle hypertrophy (Chambon et al., 2010; Dubois et al., 2012).

Skeletal muscle atrophy occurs in response to conditions such as sepsis, cachexia and glucocorticoid treatment (Fanzani et al., 2012). GCs induce muscle atrophy, especially in fast-twitch fibers (Braun

and Marks, 2015; Schakman et al., 2013). They also increase the expression of atrophy-related genes, the atrogenes such as atrogin-1, MuRF1, FOXO1 and decrease the expression of IFG-1 (Braun and Marks, 2015; Schakman et al., 2013). ADs moderately increase muscle mass and strength in hypogonadal man (Borst, 2004) by decreasing the expression of atrogin-1 and MuRF1 and by increasing the expression of IGF-1 in muscle (Ye et al., 2014). It has been also proposed that testosterone protects from Dex-induced muscle atrophy by increasing PGC-1α levels, thereby inhibiting the expression and/or activity of FOXO1 and FOXO3, two key regulator of the transcription of genes that promote muscle atrophy (e.g. MAFBx = atrogin) (Qin et al., 2010; Zhao et al., 2008), and/or by inhibition of Dex-induced expression of the mammalian target of rapamycin (mTOR) inhibitor REDD1 (Wu et al., 2010).

Moreover, ADs and GCs decrease the expression of GR and AR, respectively (Inder et al., 2010; Ye et al., 2014). Although it is not clear whether these steroid directly affect each other's receptor in muscle cells, GR-signaling at least partly interferes with AR expression, and vice versa, in muscle. The regulation of atrogin-1 and IGF-1 levels by ADs and GCs are also observed in C2C12 myoblasts (Jones et al., 2010; Zhao et al., 2008). Therefore, the expressions of these two genes are considered to be directly modulated by GCs and ADs in muscle cells. IGF-1 has a common ARE/GRE and AR and GR may competitively regulate the expression of IGF-1 in muscle cells (Wu et al., 2007).
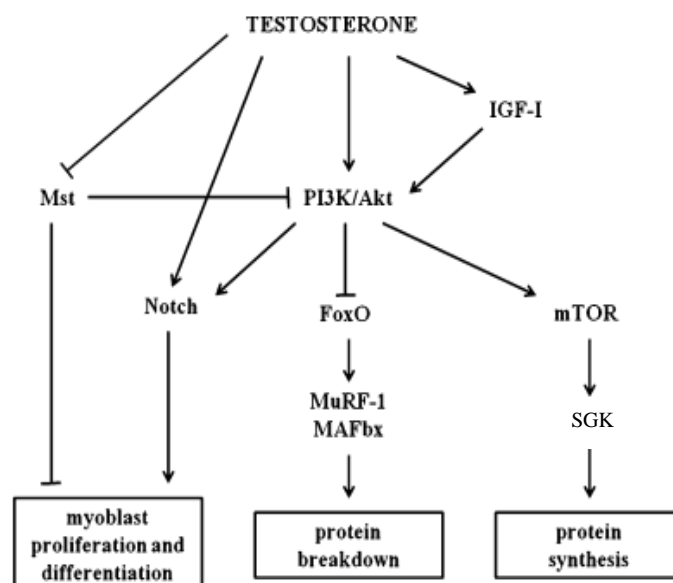
Figure 22 . **Crosstalk between ADs and other signaling pathways in muscle** (Dubois et al., 2012). Testosterone activates PI3K/Akt signaling, either directly or through IGF-I stimulation. Activation of Akt leads to phosphorylation and activation of downstream molecules including mTOR and SGK, resulting in an increase in protein synthesis. Moreover, Akt activation leads to phosphorylation and inhibition of FoxO transcription factors, which are required for upregulation of the ubiquitin ligases MuRF-1 and MAFbx, resulting in a decrease in protein degradation. Testosterone also inhibits expression and activity of Mst, which represses protein synthesis and stimulates muscle atrophy though inhibition of PI3K/Akt signaling and also negatively regulates myoblast proliferation and differentiation. Finally, testosterone increases Notch signaling, which is also a downstream effector of Akt and is essential for satellite cell proliferation and myogenic progression

Androgens and glucocorticoids in prostate

AR is a regulator of cell proliferation in prostate and primary prostate cancer (PCa) (Heinlein and Chang, 2004). Androgen deprivation therapy (ADT) such as surgical or chemical castration (i.e., LH-RH analog and antiandrogen) is a standard therapy for treatment of prostate cancer. However, prostate cancer often recurs as castration-resistance prostate cancer (CRPC) with poor prognosis (Chen et al., 2008).

GCs can be used to relieve pain, to suppress inflammation and ADs and are frequently prescribed to PCa patients undergoing ADT, chemo- and radiotherapy, as they repress the secretion of adrenocorticotrophic hormone (ACTH), thus resulting in reduced expression of adrenal ADs and consequently in a decline in prostate-specific antigen (PSA) and circulating tumor cells (Ndibe et al., 2015). Moreover, their inhibitory role on prostate cancer cell proliferation as well as angiogenesis had been documented in preclinical models (Yano et al., 2006; Yemelyanov et al., 2007). GCs are also currently used in the treatment of metastatic-resistant PCa in combination with docetaxel, cabazitaxel and abiraterone, and as they exert anti-inflammatory effects they can suppress severe therapy related adverse effects (Puhr et al., 2018). However, their independent impact on survival is unclear and unfavorable effects such as osteoporosis and immunosuppression complicate long-term use (Puhr et al., 2018).

Moreover, studies proposed that GR conferred resistance to anti-androgens through bypassing AR-signaling blockade in LNCaP/AR xenograft models (Arora et al., 2013). It was reported that many genes, including PSA are commonly regulated by both receptors and as result it was proposed that GR might have similar functional role as AR in continuously driving AR-targeted gene expressions in tumors undergoing ADT (Arora et al., 2013). In addition, clonal selection of LNCaP xenografts after long-term enzalutamide treatment showed a gain of GR expression, further supporting that GR may compensate the inactivated AR signaling in CRPC tumors (Wang et al., 2005). Moreover, the capacity of GR to drive aggressive phenotypes of CRPC was supported by the observation that rapid tumor progression was correlated with higher GR expression in LNCaP xenografts and human metastatic tumors (Wang et al., 2005). This resistance of GCs to anti-androgens was also demonstrated by the work of Isikbay M. et al. (Isikbay et al., 2014).

The development of CRPC is enhanced by administration of DEX, a common GC agent used in clinic, whereas a GR antagonist or GR silencing reduces the proliferation of CRPC cells without affecting AR expression, indicating that GR compensates for the loss of AR function (Arora et al., 2013). The increased GR activates a similar, but distinguishable, set of target genes, suggesting that CRPC development is not due to complete compensation by GR (Wang et al., 2013). SGK1 is known to be regulated by both AR and GR and is a key protein for the compensation of AR action by GR (Bolton et al., 2007; Itani et al., 2002). This idea is supported by the finding that GR is repressed by AR signaling in hormone-sensitive prostate cancer cells (Isikbay et al., 2014; Xie et al., 2015). Another study has also demonstrated that SGK1 over-expression confers resistance to castration in vivo (Isikbay et al., 2014). These results show that GR partially compensates for the loss of AR function and steadily leads to the development of CRPC.

Xie N. and his team showed that there is an inverse correlation between AR activity and GR protein expression during PCa progression (Xie et al., 2015). Pathological scoring of GR expression in PCa tissues showed increases in GR protein levels under ADT treatment. However, GR levels dropped to pre-ADT levels when the tumors progressed into the CRPC stage. In brief their data demonstrated that GR expression is suppressed by AR signaling dependent on the presence of a negative ARE (nARE).

Puhr M. et al. investigated the role of inhibiting GR for improved anti-androgen therapy (Puhr et al., 2018) and revealed a negative relationship between GR and AR by screening of different human prostate cancer cell lines. This is consistent with reports that suggest that GR is negatively regulated by AR (Chen et al., 1997; Xie et al., 2015). Furthermore, Puhr M. et al support that GR expression is associated with reduced progression-free survival and propose a dual AR/GR blockade to overcome resistance to anti-androgen therapy (**Figure 23**) (Hirayama and Sadar, 2018).
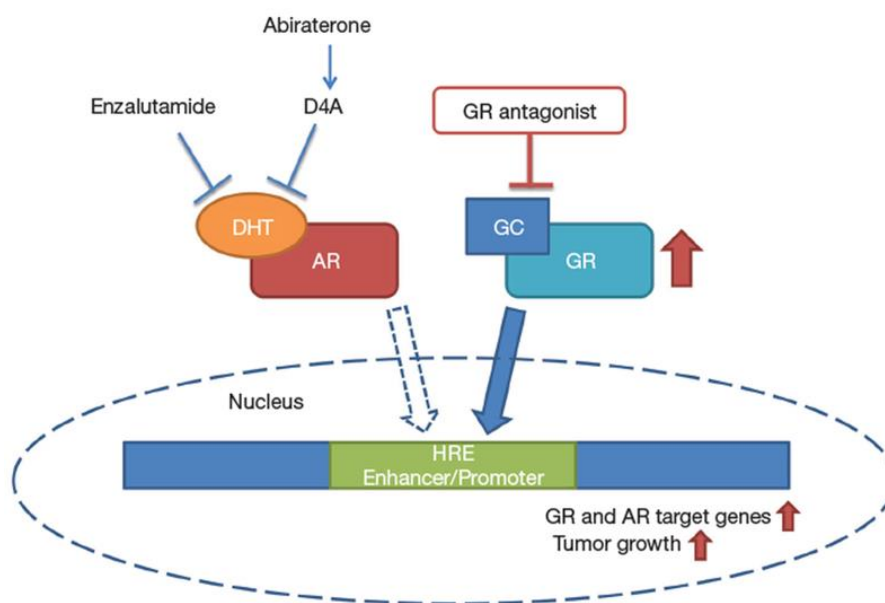
Figure 23. **Upregulation of GR leads to anti-androgen resistance through a potential bypassing pathway**. Blockade of AR by enzalutamide or abiraterone can lead to elevated GR expression which activates AR and GR target genes and tumor growth. GR-driven resistance to anti-androgens might be overcome by combination with GR antagonist (Hirayama and Sadar, 2018).

Therefore, all these observations strongly support an interplay between AR and GR in gene regulation!

6. Mouse models and cell lines to study glucocorticoid and androgen signaling in skeletal muscle

Mouse models of GR invalidation

A powerful tool for studying protein functions in vivo is the use of genetically engineered mice (GEM).

The mice with a targeted invalidation of the GR gene (GR$^{-/-}$) die in the hours following their birth, due to respiratory insufficiency resulting from a severe defect in lung development (Cole et al., 1995). At the axis HPA, GR$^{-/-}$ mice have 20-fold higher levels of ACTH and 2 to 3 times higher levels of circulating corticosterone than control mice (Gjerstad et al., 2018). Similarly, the expression of CRH in the hypothalamus of GR$^{-/-}$ mice is approximately five times less than that of wild-type mice (Kretz et al., 1999). These effects are consistent with a loss of the inhibition of the negative feedback of the HPA axis, and thus confirm the role of GR in this process (Laryea et al., 2015).

Four groups developed mice no longer expressing GR in muscle (MGRKO) in different models of muscle atrophy.

I. Hu et al. created an MGRKO model using Cre-recombinase under the control of the muscle-specific creatine kinase promoter (MCK-Cre) to excise exon 2 of the GR (Hu et al., 2009b). They demonstrated that MGRKO mice are protected from diabetes-induced muscle atrophy and fasting by preventing the decrease of IRS-1 and PI3K activity.

II. The Braun team used the same strategy as Hu et al. (Hu et al., 2009). They showed that muscle atrophy induced by inflammation or cachexia (Braun et al., 2013) or by chemotherapeutic cytotoxic agents (Braun et al., 2014) is greatly reduced in these mice. These results suggest that inflammatory cytokines, instead of acting directly on the muscle, induce the expression of GR in the muscle.

III. Watson et al. created a MGRKO model also using the MCK-Cre to excise the exon 3 of the GR (Watson et al., 2012a). They showed that the GR is essential in setting up atrophy in muscle induced by excess of GCs, but it is only partly required in muscle atrophy induced by fasting and is not involved in mechanisms leading to muscle atrophy induced by denervation.

IV. Recently, the Shimizu team has created a model MGRKO using the Cre-recombinase under the control of the skeletal muscle actin 1 (ACTA1) promoter to excise exon 3 of the GR. They provided evidence that there is a signaling axis between muscle, liver and fat via fibroblast growth factor factor 21 (FGF21) (Shimizu et al., 2015).

Generation of GR$^{(i)skm-/-}$ mice in which GR is selectively ablated in skm myofibers at adulthood generated by our team

GR$^{L2/L2}$ mice bear GR L2 alleles, in which exons 3 and 4 encoding the DNA binding domain are flanked with 2 LoxP sites. Cre-mediated recombination between the two LoxP sites induces a frame shift in the GR sequence, and thus results in a GR-null allele. To selectively ablate GR in skm myofibers of adult mice, GR$^{L2/L2}$ mice are intercrossed with HSA-CreER^{T2} mice that express the

CreER^{T2} recombinase selectively in skm myofibers (Schuler et al., 2005). GR$^{L2/L2}$ mice and HSA-CreER^{T2}/GR$^{L2/L2}$ mice are intraperitoneally injected with Tamoxifen (Tam) (1 mg/mouse/day) at adulthood to generate control (GR$^{L2/L2}$) mice and GR$^{(i)skm-/-}$ mutant mice, respectively.

Mouse models of AR invalidation

Several mouse models in which AR was invalidated in the germ line were generated (Notini et al., 2005; Sato et al., 2003; Yeh et al., 2002). These mice have an external appearance of female, without prostate, seminal vesicles and with greatly atrophied testicles. These animals also have metabolic problems and become obese. MacLean et al. studied the muscle functions of mice in which AR is invalidated (MacLean et al., 2008) and showed that the absence of AR leads to an absence of LA muscle and a decrease in muscle mass (Chambon et al., 2010). Moreover, in these KO mice, the force generated is lower in the fast muscles and this decrease is related to the decrease in muscle mass. But these studies do not allow the characterization of the role of ADs in muscle, since AR is invalidated in all cells of the body.

This is why the host laboratory has developed a mouse model in which AR is selectively invalidated in myofibers of mouse skm (AR$^{skm-/y}$), through conditional targeted somatic mutagenesis. To this end, HSA-Cre mice expressing the Tam-dependent Cre^{T2} recombinase under the control of the human skeletal actin (HSA) promoter elements (Schuler et al., 2005) were intercrossed with mice bearing LoxP-flanked AR alleles to obtain HSA-Cre/AR$^{L2/y}$. Tam administration to male HSA-Cre/AR$^{L2/y}$ mice induces AR ablation selectively in skeletal myofibers, thus generating AR$^{skm-/y}$ mice (Chambon et al., 2010).

AR is needed to structure sarcomeres to generate optimal muscle strength, by regulating autophagy via AR in myofibers. Moreover, during a mechanical overload, AR in the myofibers is essential for the growth muscle (Chambon et al., 2010; Ferry et al., 2014).

C2C12 cells

The C2C12 cells are an immortalized mouse myoblast cell line (Muses et al., 2011). The C2C12 cell line is a subclone of myoblasts established from normal adult C3H mouse leg muscle (Blau et al., 1985) that were originally obtained from a C2 cell line by Yaffe and Saxel at the Weizmann Institute of Science in Israel in 1977 (Yaffe and Saxel, 1977). Wild-type C2C12 cells have a radial branching morphology consisting of long fibers extending in many directions. The cells proliferate in high-serum conditions, and differentiate and fuse in low-serum conditions (Cheng et al., 2014). Moreover, C2C12 cells demonstrate rapid development and maturation into functional skeletal muscle cells or cardiac muscle cells, having the ability to contract and generate force (McMahon et al., 1994). They are also convenient for studying the cell cycle, as they have high division rate (Mamchaoui et al., 2011) and a very useful tool to study aspects of myogenesis, metabolism and muscle biology.

7. Next generation sequencing technologies and genome-wide analysis of cistromic and transcriptomic data

Methods and applications

Next-generation sequencing (NGS) is a broad term referring to multiple sequencing technologies developed since 2005 (van Dijk et al., 2014) which can address many questions in biology. Compared to Sanger sequencing, the NGS technologies simplify library preparation by using vector cloning of the DNA fragments, significantly improve the sequencing throughput by simultaneously monitoring millions of reactions, and highly automate the determination of nucleotides using imaging or semiconductor technologies, instead of electrophoresis. Using NGS, ambitious genomic sequencing projects that target many organisms and large scale studies of sequence variation have become feasible (Stratton, 2008).

Major providers in the next-generation sequencing market are *454 pyrosequencing* by Roche ((Margulies et al., 2005), now discontinued), *Illumina/Solexa (Bentley, 2006)*, *SOLiD* by Life Technologies ((Valouev et al., 2008), previously Applied Biosystems), *Ion Torrent* Personal Genome Machine (PGM) ((Rothberg et al., 2011), now Life Technologies), *Single Molecule Real-Time Sequencing* (SMRT) by Pacific Biosciences (PacBio) (Eid et al., 2009) and more recently *Nanopore sequencing (Feng et al., 2015)*.

A common denominator of the above sequencing techniques is the reliance on DNA polymerase (Niedringhaus et al., 2011). This enzyme is utilized to synthesize deoxynucleotides (dNTPs), which may or may not be labeled, against single-stranded DNA templates. Signals released by base synthesis (hydrogen ions or fluorescent radiation) are "read" by the sequencers and converted into nucleotide sequences (or reads), a common strategy known as sequencing-by-synthesis (SBS) (Buermans and den Dunnen, 2014).

Of course, new technologies come with challenges. For many next generation sequencers, the advantage of deeper and cheaper coverage comes at the cost of shorter reads with higher error rates compared to the Sanger sequencing (Buermans and den Dunnen, 2014). Each sequencing platform has different error profiles. Although some technologies such as the 454 produce reads with an average length of 400 bp, most of these high throughput next generation sequencing systems produce short reads, ranging from 25 bp to 150 bp (Liu et al., 2012). Synthetic **long-read sequencing technology** is a highly accurate, end-to-end solution that can be used to generate synthetic long reads for de novo assembly and genome finishing applications, to sequence traditionally challenging genomes, such as those containing stretches of highly repetitive elements and to perform whole human genome phasing to identify co-inherited alleles, haplotype information, and phase de novo mutations (Midha et al., 2019). **Short read technologies** have been widely used to initiate new applications and sometimes, to replace the existing ones. Some of these applications include genome sequencing, re-sequencing, metagenomics, whole transcriptome analysis, genome methylation analysis, chromatin Immunoprecipitation for TF binding sites detection, microRNA discovery and others (Ansorge, 2009; Marguerat et al., 2008).

Since Illumina sequencers can generate the highest throughput of NGS reads, they have become the most dominant platform in this field (Pareek et al., 2011). The reason behind such wide adoption of Illumina's systems is the large volume of information obtained from a typical sequencing run (e.g. sequencing depth), which, at a good ratio with the cost, compensates for the lower accuracy compared to other competitors (Mardis, 2013). One of the main problems of Illumina reads is the read length (Nakamura et al., 2011). In the library preparation step, the DNA or RNA molecules are chopped into smaller fragments. Each fragment can be sequenced from one end up to 150 bp only. The first form of Illumina reads is *single-end* (**Figure 24**). That is, only one end of the fragment can be sequenced. The major problem of single end reads is the ambiguity when reads are mapped to multiple loci. A simple improvement to the single-end library preparation is to sequence both ends of fragments (scanning both the forward and reverse template strand). The *paired-end* sequencing incorporates the fragment length information, which can significantly improve the mapping and assembly accuracy (**Figure 24**). The typical fragment length of paired-end sequencing is 200-500 bp. In terms of genomic assembly, this fragment length is still too short when scaffolding contigs. A newer library preparation method can produce paired reads (referred to as mate pairs) separated by longer distance (2 kb ~ 8 kb) (Wetzel et al., 2011). In this method, longer fragments are circularized and both ends are sequenced (Van Nieuwerburgh et al., 2012). Mate pairs reads are very useful for the de novo genome.

Figure 24. Single-end read and pair-end read.

Quality control of NGS

To deal with the high error rate of next-generation sequencing in general, both experimental and bioinformatics approaches for quality control are available.

On the experimental side, library preparation protocols have been designed to enhance the number of replicates within samples, thus limiting sequencing errors (Schurch et al., 2016). These experimental approaches reduce error rate by multiple order of magnitudes and allow rare mutations to be identified, but require unconventional library preparations and reduce the effective throughput.

On the bioinformatics side, tools dedicated to remove sequencing errors from high-throughput sequencing data exist. A large class of methods use read filtering and read trimming to discard noisy reads or read segments (Del Fabbro et al., 2013; Guo et al., 2014), based on criteria like sequence quality, alignment quality, and variant calling quality. However, all such methods potentially suffer from the loss of sequencing coverage due to data removal (Trivedi et al., 2014).

Gene expression profiling

Studying changes in gene expression between experimental conditions, different tissues and developmental stages provide insights into gene function. Therefore, methods measuring gene expression of all genes present in a sample are required to obtain an unbiased read-out of the entire transcriptome.

Expression profiling using RNA sequencing (RNA-Seq)

In the past years, RNA-seq has become the method to study the transcriptome composition (Mortazavi et al., 2008; Wang et al., 2009). Compared to microarrays, which constituted the first technology for the high throughput comparison of expression levels across conditions, RNA-seq offers the possibility to study gene expression patterns in a much bigger dynamic range and enables a much broader set of analyses without special experimental designs (Malone and Oliver, 2011). For example, besides standard differential gene expression analysis, popular applications of RNA-seq are the identification of novel transcribed regions, including fusion genes, the deconvolution of allele specific expression, the estimation of transcript expression levels and the study of differential splicing across conditions (Mortazavi et al., 2008). Due to decreasing sequencing costs, RNA-seq is becoming more and more accessible and has almost replaced gene expression analysis using microarrays.

RNA-seq experimental workflow

A typical RNA-seq experiment workflow consists of several steps (**Figure 25**) (Cullum et al., 2011). First, ribosomal RNA (rRNA) are removed from the sample, since it is the predominant RNA species in a cell. Next RNA molecules get randomly fragmented either before or after they are reverse transcribed into double-stranded cDNA. Often a size selection step is performed following 5' and 3' adapter ligation and PCR amplification. The cDNA library can then be sequenced using for example the Illumina Hiseq 4000.

The steps of RNA-seq library preparation are:

  i. Purification of mRNA
 ii. RNA fragmentation
iii. cDNA synthesis
 iv. Adapter ligation and PCR amplification
  v. Size selection
 vi. Sequencing by synthesis
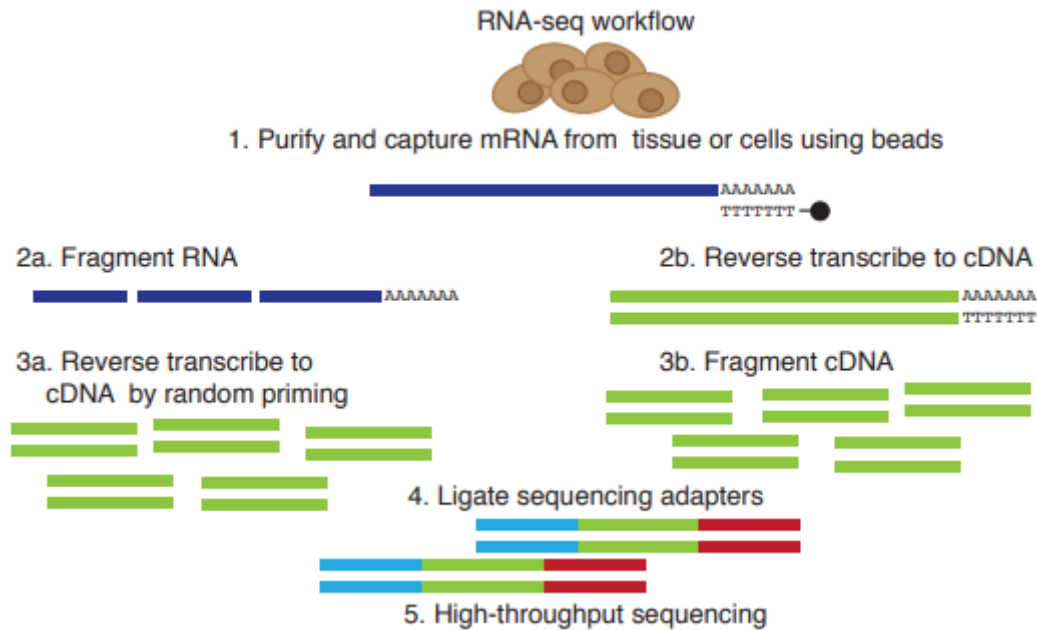
Figure 25. **RNA-seq experiment workflow**. The initial step of RNA-Seq is the purification of mRNA. The RNA is either fragmented prior to reverse transcription (workflow a) or after reverse transcription (workflow b). The double stranded DNA fragments are then ligated to sequencing adapters for the subsequent sequencing on a next generation sequencer (Cullum et al., 2011).

After the experimental procedure, the bioinformatics analysis needs to be performed composed of several important steps.

Read mapping

The next step in the RNA-seq analysis pipeline consists of identifying, for each read, the genomic region from which it has originated (Oshlack et al., 2010). In RNA-seq, this task is equivalent to discovering the loci that are expressed in a given sample. Two different strategies exist to perform this task: on one hand, reads can be aligned to the reference genome or transcriptome, provided that such information is available for the species of interest; on the other hand, they can be directly assembled into contigs (e.g. contiguously expressed regions) with the aim of reconstructing the set of expressed transcripts. The first strategy is a much simpler approach, and it is typically the one when working with model organisms.

Independently of the strategy used, read mapping was typically the most time consuming step of the analysis workflow, and the available tools made use of heuristic parameters such as the maximum number of allowed mismatches per read in order to speed up this task. Such processing can lead to information loss given a decrease of quality at the 3' end of the read, which emerges as a common profile when working with Illumina platforms due to the increased difficulty in interpreting the fluorescent signal as sequencing cycles accumulate (Minoche et al., 2011). In this case the quality control done previously can help. Similarly, reads with overall low quality can be also removed, in order to speed up the subsequent mapping process.

Nowadays, there are ultrafast and memory efficient aligners such as *STAR, Tophat and HISAT.*

Alignment to genome or transcriptome

If the reference genome or transcriptome is available, then the reads can be aligned to those sequences. The main advantage of using a transcriptome is that the alignment is simplified, as there are no intronic sequences; but this limits the number of downstream analysis that can be performed (e.g. alignment to the transcriptome is not compatible with the identification of novel expressed regions nor the study of intronic expression levels). Thus, a good compromise is the use of hybrid approaches, like *TopHat2* (Kim et al., 2013), a read mapping tool specially intended for RNA-seq data, since it enables alignment of the reads to the genome while taking into consideration the existence of splice junctions. It is based on *Bowtie* (Langmead and Salzberg, 2012), an independent algorithm for the alignment of short reads, and its main strength is the ability to detect exon-exon junctions without the need for any a priori knowledge on the annotation.

De novo **assembly**

When there is no reference genome, the strategy to follow is de novo assembly. Additionally, it can be used in situations where the genome composition of a given sample is expected to differ largely from that of the reference assembly (e.g. cancer samples). The goal is to assemble the reads into sets of expressed regions (e.g. contigs), by relying on their overlap. Nonetheless, the short read length make the task even more difficult, and even though the use of paired-end data can simplify this process, lowly expressed regions are often difficult to assemble. The most popular software for this task is *Trinity* (Grabherr et al., 2011).

Quantification of gene expression

Read alignments are used to quantify gene expression levels. The gene raw read count is defined as the number of reads that map to exons of known genes. These raw counts, however, cannot be used to compare gene expression levels between genes within the same sequencing run or the same gene between different sequencing runs. A long gene will have a higher read count compared to a short gene expressed at the same level. Equally, a gene will have a higher read count if the sequencing run resulted in x million reads instead of y million reads. One tool to perform quantification of the read counts is *htseq-count* (Anders et al., 2015).

However, there are some challenges that need to be considered. First, in order not to over-estimate expression levels, reads that map to multiple locations in the genome, and which arise from repetitive or duplicated loci, need to be considered. In this situation, htseq-count adopts the most conservative approach and discards them, but other alternative strategies have been proposed in order to attempt to keep the information from such multi-mapping reads. Second, special attention is required in the case of overlapping features. Htseq-count offers several execution modes to deal with these features, even though in some cases reads remain ambiguously assigned. Finally, despite not being intended for de novo quantification, htseq-count also gives the user some flexibility on how strictly the provided feature coordinates should be taken into account.

Normalization of gene expression

It is necessary to normalize the raw counts both for gene **length** and total **read number**. One popular way of normalizing raw counts for length and total number of reads are *RPKM* values, which stands for reads per kilobase of exon model per million mapped reads ([Mortazavi et al., 2008](#)). *FPKM* (fragments per kilobase of exon model per million mapped fragments) values are the RPKM equivalent for paired-end data where the two reads coming from one fragment are counted as one ([Trapnell et al., 2010](#)). Several slight adaptations to this normalization have been proposed, e.g., only using uniquely mappable regions of genes for length normalization ([Lee et al., 2011](#)).

It is also important, to ensure that the expression levels are comparable across libraries (different samples and different biological conditions). In their paper, Robinson and Oshlack argue that the RPKM model of standardizing the data between samples by scaling the number of reads in a library to a common value across all sequenced libraries in an experiment may not be appropriate for normalization between libraries of different biological conditions ([Robinson and Oshlack, 2010](#)). The total number of reads that can be sequenced in a sequencing lane is limited and counts from very highly expressed genes do not leave space for counts from lowly expressed genes.

A better assumption, which has been widely used with microarrays, is that the RNA output of a core set of genes G is similar between samples. A number of methods, including the ones implemented in the *DESeq* ([Anders and Huber, 2010](#)) and *edgeR* ([Robinson et al., 2010](#)) Bioconductor packages, use this assumption and find a scaling factor for one sample relative to the other accordingly. The way, in which these methods find this scaling factor can be quite different.

DESeq and the more recent *Deseq2* ([Love et al., 2014](#)) start by calculating a geometric mean for each gene in order to capture the variability of the observed measurements across all the libraries (similar to obtaining a reference sample). Then, these values are used to normalize the initial counts, and finally, the library-specific normalization factors are obtained from the median of the calculated ratios.

Differential gene/transcript expression analysis

The most common use of the RNA-Seq is the assessment of differences in expression levels across conditions ([Oshlack et al., 2010](#)). When the counts have been obtained, such analysis can be performed both at the gene and transcript level and one of the most popular tools to achieve this is *DESeq2 ([Love et al., 2014](#))*.

In general terms, DESeq2 relies on the use of Generalized Linear Models (GLMs) of the Negative Binomial (NB) family in order to address the significance of the detected changes in expression levels. The implemented analysis workflow first consists of normalizing the observed counts in order to enable their comparison across libraries, as covered in the previous section. Next, for each gene, an estimate on the amount of variability that can be expected on the measurements from biological replicates is calculated, and finally, the differential expression test is performed. The biological variance of a gene, that is the natural variance of expression levels of a gene within the

same conditions, has to be estimated to identify differentially expressed genes. The usage of biological replicates is also crucial for the estimation of biological variance (Schurch et al., 2016).

Originally, differential expression between conditions was tested using a Poisson model for read counts. This model provides a good fit for technical replicates (Marioni et al., 2008). However, samples from biological replicates show higher variance than predicted by the Poisson model (Anders and Huber, 2010; Robinson et al., 2010). As a result, the NB distribution has been widely adopted to account for such over-dispersion. However, because of the low number of replicates typically available in RNA-seq experiments, such variation cannot be directly calculated, and needs to be estimated from the data instead. Following the assumption that genes with similar expression levels have similar sample-to-sample variance, DESeq2 obtains gene-specific variance estimates by taking into account not only the observed dispersion for each given gene, but also that of all other genes. This is achieved by fitting a regression curve to the data (e.g. average normalized counts vs. observed dispersion), which is subsequently used to modify the observed dispersion values. Finally, by further decomposing the mean into a function of independent variables (i.e. the covariates), it is possible to take all known sources of variation into account.

DESeq2 works on count data and do not consider ambiguously mapped reads, gene structure and cannot identify isoform switching, where genes are expressed at the same level in two or more conditions, but where the major isoform is different. For these cases, other approaches and probabilistic methods have been developed, but they will not be presented here.

Protein-DNA interactions

The identification of chromatin modifications and protein-DNA interactions is essential to characterize transcriptional regulation (Geertz and Maerkl, 2010). Mapping TF binding sites, chromatin modifications or components of the core transcriptional machinery, provides insights into the gene regulatory networks that control transcription. Chromatin immunoprecipitation (ChIP) followed by deep sequencing enables the genome-wide detection of protein-DNA binding events (Furey, 2012; Park, 2009; Pepke et al., 2009).

Chromatin Immunoprecipitation followed by sequencing (ChIP-seq)

Chip-Seq allows the identification of DNA-binding sites at genomic scale in an unbiased fashion, as it does not depend on what is represented on the array. Additionally, it suffers from less noise because signal from cross-hybridization is removed. An early study made by Johnson et al. (Johnson et al., 2007) shows the increased sensitivity and specificity of ChIP-seq genome-wide mapping of TF binding sites as well as the identification of non-canonical binding sites. Similarly, the first ChIP-seq studies of histone modifications suggested new functions for modifications and the importance of combinatorial modification patterns (Barski et al., 2007). ChIP-seq data also require extensive computational analysis.

ChIP-Seq experimental workflow

Chromatin Immunoprecipitation (ChIP) is a commonly used technique to detect protein-DNA interactions. The experimental workflow consists of distinct steps (**Figure 26**) (Park, 2009). Formaldehyde treatment of cells allows cross-links to form between the protein and DNA, thus stabilizing protein-DNA complexes. The complexes are then extracted from lysed cells and sonicated to form 200-600 bp DNA segments. The complexes are immunoprecipitated using an antibody against the protein of interest (e.g. a TF or histones) and the DNA is purified. Finally, the crosslinks are reversed and the released DNA is assayed by PCR in order to quantify the DNA immunoprecipitated segments. After the amplification step, the next generation sequencing is following (Park, 2009). Only regions that show a statistically significant enrichment of signal in the treatment experiment compared to the control experiment are regarded as peaks (piles of short reads).
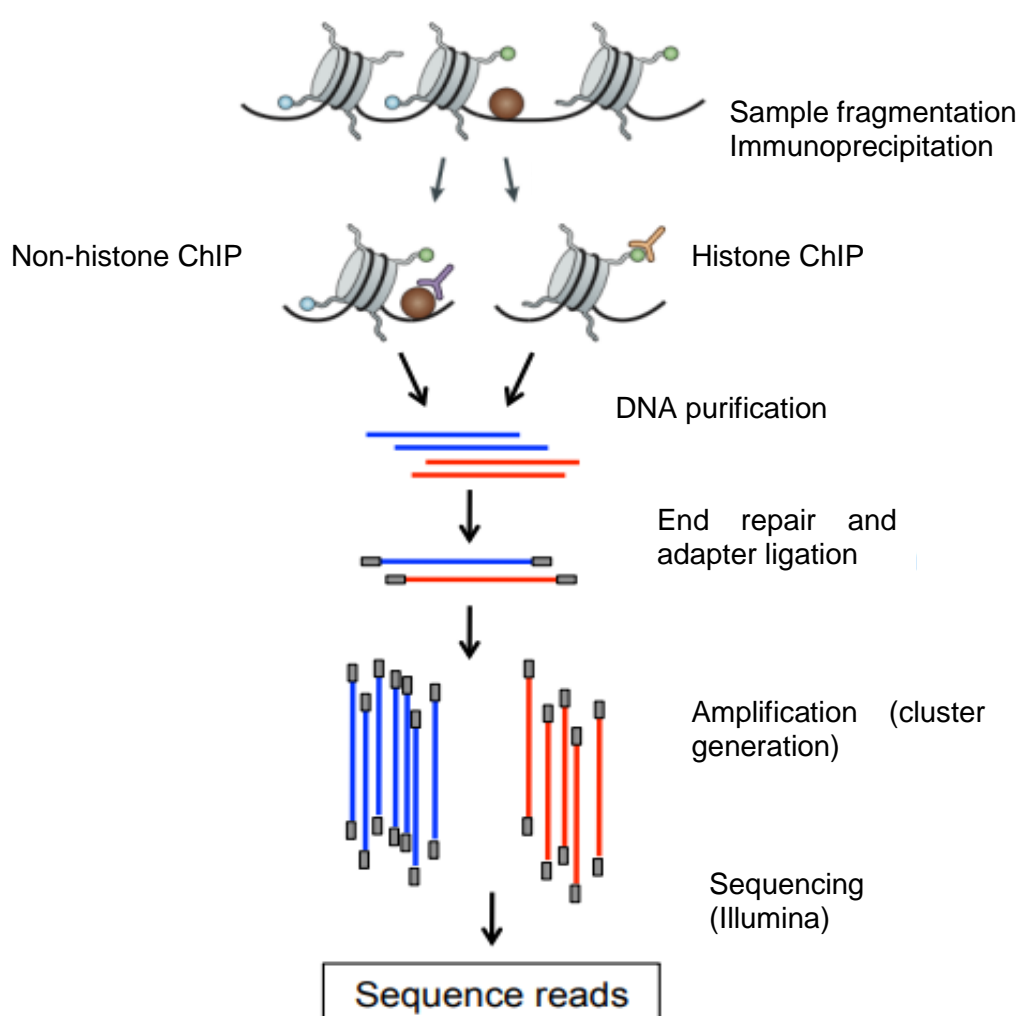
Figure 26. **ChIP-seq experimental workflow** (Park, 2009).

Mapping, peak detection & annotation

For the alignment of the reads to the reference genome traditional DNA aligners can be used like the *Bowtie2* (Langmead and Salzberg, 2012).

After alignment of the reads, the regions that are significantly enriched in the ChIP sample compared to the control sample are identified. Several "peak callers" are available, for example *MACS* (Zhang et al., 2008) and the more recent *MACS2*.

The simplest strategy for peak detection is to calculate the number of reads within a window and the enrichment relative to the number of reads in the control (**Figure 27**). MACS2 models the shift between reads mapped to different strands and uses a Poisson distribution with varying value to characterize the background model (Feng et al., 2012) (**Figure 28**). The False Discovery Rate (FDR) is calculated as the ratio between the numbers of peaks called in the control sample to the ChIP sample.

More advanced methods make use of the directionality of the reads. As the fragments are sequenced from the 5' end, the positions of the aligned reads should form two distributions, one on each strand, with a consistent distance between the two peaks of the distributions. A combined profile is then calculated by shifting each distribution towards the center or by extending all reads to an estimated fragment size and adding the fragments together (**Figure 28**). This approach is mostly applicable to sharp peaks as for example TF binding sites. ChIP-seq analysis of histone modifications typically results in much broader peaks, which was an additional challenge to peak detectors. Several specialized methods are available for the detection of broader peak domains (Xu et al., 2008; Zang et al., 2009). There is also a type of signal that can be localized in broader regions of binding extending up to a few kilobases, for example Pol II and some histone modifications in coding regions.
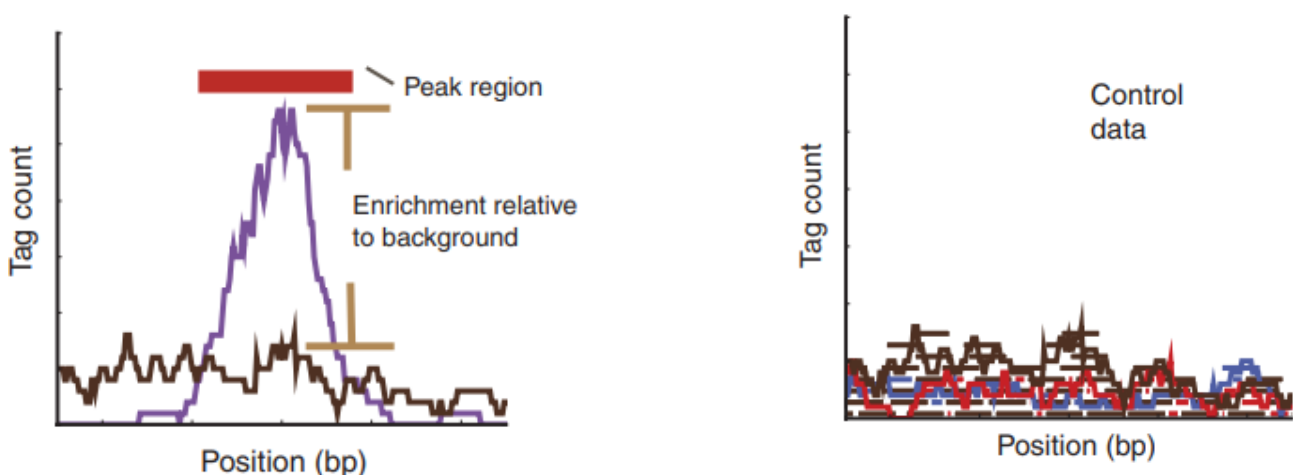
Figure 27. Regions of enrichment of ChIP-seq reads relative to control (left) and the control data (right) (Pepke et al., 2009).
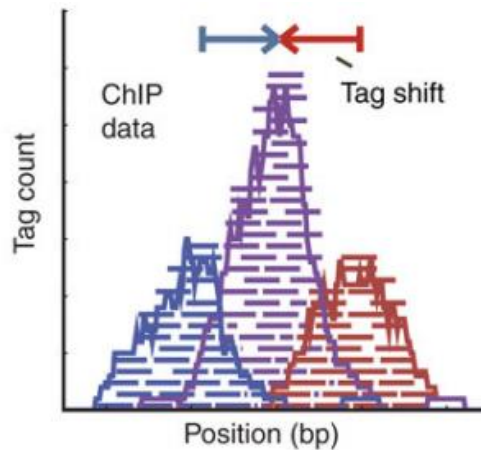
Figure 28. **Identification of peaks in ChIP signal** by shifting the distribution of each strand towards to the center (blue for positive strands, red for negative strands, purple for the combined distribution) (Pepke et al., 2009).

After the detection of the peaks, they have to be annotated. This can be performed by using the *Homer* software (Heinz et al., 2010). The process of annotating peaks/regions is divided into two primary parts. The first determines the distance to the nearest TSS and assigns the peak to that gene. **Homer** determines the closest Transcription Start Site (TSS) by reporting the distance (negative values mean upstream of the TSS, positive values mean downstream). The second determines the genomic annotation of the region occupied by the center of the peak/region in terms of known genomic features (e.g. exon, intron, etc.). However, close proximity does not always indicate that a binding site has a functional role related to the proximal gene.

Downstream analysis

Following peak calling, the downstream analysis depends on the biological process under investigation. For TF binding, a common follow-up analysis is the detection of enriched sequence motifs. These motifs can indicate sequence-specific binding of TFs. To identify such motifs, the sequences of top-scoring peaks is used by motif-finding algorithms such as ***MEME-ChIP*** (Machanick and Bailey, 2011) which belongs to the *MEME-Suite* (Bailey et al., 2009). To get more accurate results it can be advantageous to select only the region surrounding the peak summits for motif discovery (e.g. +/- 100bp around the summit). Once a motif has been identified, methods like **TOMTOM** can be used to find similar known motifs (Gupta et al., 2007).

Another tool for downstream analysis is ***Homer***, which uses a differential motif discovery algorithm. It takes two sets of sequences and tries to identify the regulatory elements that are specifically enriched in on set relative to the other. Moreover, ***RSAT*** (Thomas-Chollier et al., 2012) is a computational pipeline that discovers motifs in peak sequences, compares them with databases, exports putative binding sites for visualization in the UCSC genome browser and generates an extensive report.

All these algorithms can treat thousands of peaks in some minutes, they are memory efficient and they offer user-friendly web interfaces (except Homer), without any restriction on sequence size or number of peaks.

There are also open-access databases of manually curated, non-redundant sets of TF binding profiles like the **Jaspar database** (Khan et al., 2018). The profiles are in a format of position frequency matrices (PFM) and TF flexible models (TFFM) for TFs from different species in six taxonomic groups. The profiles derive from published and experimentally defined TF binding sites for eukaryotes and can be used for scanning genomic sequences.

Functional annotation and biological interpretation of the results

The functional enrichment analysis has a key role in the biological interpretation of high-throughput gene-expression data and it is the last and the most important step in a gene expression study. It requires a fundamental understanding of the biological question and possibly what to expect. Many software programs are available to perform this step of the analysis and can be done using gene-set enrichment methods that implement statistics to analyze differentially expressed genes and link them to particular biological functions or pathways or terms (Tomczak et al., 2018).

WebGestalt (WEB-based GEne SeT AnaLysis Toolkit) (Zhang et al., 2005a) (Wang et al., 2013; Wang et al., 2017) is a web-based integrated data mining system to explore large sets of genes. It is composed of four modules: gene set management, information retrieval, organization/visualization, and statistics. After the last update (14/01/2019) it supports 12 organisms, 354 gene identifiers and 321.251 functional categories from public databases and computational analyses.

DAVID (Database for annotation, visualization and integrated discovery) (Huang da et al., 2009a, b) is a free online bioinformatics resource which aim to provide functional interpretation of large lists of genes derived from genomic studies. The DAVID Bioinformatics Resources consists of the DAVID Knowledgebase and five integrated, web-based functional annotation tool suites: the DAVID Gene Functional Classification Tool, the DAVID Functional Annotation Tool, the DAVID Gene ID Conversion Tool, the DAVID Gene Name Viewer and the DAVID NIAID Pathogen Genome Browser.

Reactome (Fabregat et al., 2018; Joshi-Tope et al., 2005; Vastrik et al., 2007) is a free, open-source, curated pathway database, which provides intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge. It is hosted and mostly curated by European Bioinformatics Institute (EMBL-EBI). It has a web portal at http://www.reactome.org/ , and the current version is v67 released in December, 2018.

KEGG (Kanehisa and Goto, 2000) (Wixon and Kell, 2000) was one of the first pathway databases. Actually it is a collection of databases of genomes, biological pathways, diseases, drugs and chemical substances used in order to decipher different genomes. The databases of KEGG are

categorized into systems, genomic, chemical and health information. In July 2011 KEGG introduced a subscription model for FTP download due to a significant cutback of government funding.

Gene Ontology (GO) (Ashburner et al., 2000) project provides 3 ontologies, which are used for the systematic description of gene products: *biological function*, *cellular component*, and *molecular funct*ion. Each ontology forms a rooted directed acyclic graph in which each node is associated with a GO identifier (or GO term). A gene annotated with any given GO term is also annotated with all ancestral GO terms, allowing for descriptions of the gene product at varying levels of specialization. Generic and species-specific versions of the Gene Ontology are continuously updated based on experimental or electronically derived evidence.

PANTHER (protein analysis through evolutionary relationships) ***Classification System*** (Mi and Thomas, 2009) is a curated biological database of gene/protein families and their related subfamilies and it is used to identify and classify proteins and their genes. Proteins have been classified according to family and subfamily, molecular function, biological process and pathway. The Gene List Analysis tool gives the possibility to analyze gene lists and expression data from high-throughput experiments. It is possible to map lists to multiple annotation data sources and biological pathways and to visualize them using a variety of graphs.

GREAT (Genomic Regions Enrichment of Annotations Tool) predicts functions of cis-regulatory regions (McLean et al., 2010). Whereas previous methods took into account only binding proximal to genes, GREAT is able to properly incorporate distal binding sites and control for false positives using a binomial test over the input genomic regions. GREAT incorporates annotations from 20 ontologies and is available as a web application.

Data visualization

Data visualization is an important part of genomic data analysis. It includes the visualization of information acquired from sequences, genomes, alignments, gene expression and networks. According to the type of data and the purposes, different types of visualization tools can be used.

- Genomes

UCSC genome browser (Kent et al., 2002) is an on-line, interactive genome browser that offers access to genome sequence data from a variety of species and organisms. It support the uploading of personal data but also the downloading of data files.

Integrative Genomics Viewer (IGV) browser (Robinson et al., 2011) is a downloadable and interactive genome viewer of public data as well as local data. It is offering high level performance of data visualization and exploration with a variety of tools for manipulation.

- Protein-Protein Interaction Networks

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) (Snel et al., 2000) is a biological database of known and predicted protein-protein interactions from experimental data, computational prediction methods and public text collections. The known or predicted interactions

can be visualized as interactive networks. It provides also functional annotation (GO terms, pathways), functional partners and clustering of the lists of genes/proteins.

- Gene expression

Cluster 3.0 (de Hoon et al., 2004) provides k-means clustering, hierarchical clustering and self-organizing maps that can be visualized after using the Java TreeView.

Java TreeView (Saldanha, 2004**)** is an interactive application for the visualization of microarray and RNA-Seq data that have been processed before with the Cluster 3.0. Gene expression data is organized into rows and columns, where the rows correspond to genes, and the columns correspond to experiments. Thus, the value in row m, column n, is a measure of the expression of gene m in experiment n. The value used is commonly the log2 of the ratio of the experimental sample to the control. These values are rendered in a two color scale, where one color represents higher expression and the other indicates lower expression in the given experiment. If the data have been clustered before using the Cluster 3.0, this clustering can be displayed.

- Genome-wide data

seqMINER (Ye et al., 2011) allows quantitative and qualitative comparisons between reference set of genomic positions and multiple ChIP-Seq datasets. SeqMINER proposes two complementary methods to analyze the signal enrichment status in multiple other tracks:

i) a qualitative method that computes a density array over a defined window around the reference coordinate.

ii) a quantitative method that computes enrichment value over a defined window around the reference coordinate.

According to the signal enrichment status the reference coordinates are organized in different clusters. The elements of the clusters can be extracted, annotated and visualized as mean profile or merged profile.

8. References

Aagaard, M.M., Siersbaek, R., and Mandrup, S. (2011). Molecular basis for gene-specific transactivation by nuclear receptors. Biochim Biophys Acta *1812*, 824-835.

Acevedo, M.L., and Kraus, W.L. (2004). Transcriptional activation by nuclear receptors. Essays Biochem *40*, 73-88.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.*, et al.* (2000). The genome sequence of Drosophila melanogaster. Science *287*, 2185-2195.

Alamdari, N., Aversa, Z., Castillero, E., and Hasselgren, P.O. (2013). Acetylation and deacetylation--novel factors in muscle wasting. Metabolism *62*, 1-11.

Alangari, A.A. (2010). Genomic and non-genomic actions of glucocorticoids in asthma. Ann Thorac Med *5*, 133-139.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol *16*, 155-166.

Allen, D.L., and Unterman, T.G. (2007). Regulation of myostatin expression and myoblast differentiation by FoxO and SMAD transcription factors. Am J Physiol Cell Physiol *292*, C188-199.

Anbalagan, M., Huderson, B., Murphy, L., and Rowan, B.G. (2012). Post-translational modifications of nuclear receptors and human disease. Nucl Recept Signal *10*, e001.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T.*, et al.* (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-461.

Ansorge, W.J. (2009). Next-generation DNA sequencing techniques. N Biotechnol *25*, 195-203.

Ardito, F., Giuliani, M., Perrone, D., Troiano, G., and Lo Muzio, L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). Int J Mol Med *40*, 271-280.

Arora, V.K., Schenkein, E., Murali, R., Subudhi, S.K., Wongvipat, J., Balbas, M.D., Shah, N., Cai, L., Efstathiou, E., Logothetis, C.*, et al.* (2013). Glucocorticoid receptor confers resistance to antiandrogens by bypassing androgen receptor blockade. Cell *155*, 1309-1322.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T*., et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Atherton, P.J., and Smith, K. (2012). Muscle protein synthesis in response to nutrition and exercise. J Physiol *590*, 1049-1057.

Audia, J.E., and Campbell, R.M. (2016). Histone Modifications and Cancer. Cold Spring Harb Perspect Biol *8*, a019521.

Ayroldi, E., Cannarile, L., Migliorati, G., Nocentini, G., Delfino, D.V., and Riccardi, C. (2012). Mechanisms of the anti-inflammatory effects of glucocorticoids: genomic and nongenomic interference with MAPK signaling pathways. FASEB J *26*, 4805-4820.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res *37*, W202-208.

Barnes, P.J. (2011). Glucocorticosteroids: current and future directions. Br J Pharmacol *163*, 29-43.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Beato, M. (1989). Gene regulation by steroid hormones. Cell *56*, 335-344.

Bentley, D.R. (2006). Whole-genome re-sequencing. Curr Opin Genet Dev *16*, 545-552.

Berger, S.L. (2007). The complex language of chromatin regulation during transcription. Nature *447*, 407-412.

Berrabah, W., Aumercier, P., Lefebvre, P., and Staels, B. (2011). Control of nuclear receptor activities in metabolism by post-translational modifications. FEBS Lett *585*, 1640-1650.

Bevan, C., and Parker, M. (1999). The role of coactivators in steroid hormone action. Exp Cell Res *253*, 349-356.

Bhasin, S., Calof, O.M., Storer, T.W., Lee, M.L., Mazer, N.A., Jasuja, R., Montori, V.M., Gao, W., and Dalton, J.T. (2006). Drug insight: Testosterone and selective androgen receptor modulators as anabolic therapies for chronic illness and aging. Nat Clin Pract Endocrinol Metab *2*, 146-159.

Bhasin, S., Storer, T.W., Berman, N., Callegari, C., Clevenger, B., Phillips, J., Bunnell, T.J., Tricker, R., Shirazi, A., and Casaburi, R. (1996). The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. N Engl J Med *335*, 1-7.

Bhasin, S., Woodhouse, L., Casaburi, R., Singh, A.B., Mac, R.P., Lee, M., Yarasheski, K.E., Sinha-Hikim, I., Dzekov, C., Dzekov, J*., et al.* (2005). Older men are as responsive as young men to the anabolic effects of graded doses of testosterone on the skeletal muscle. J Clin Endocrinol Metab *90*, 678-688.

Bhasin, S., Woodhouse, L., and Storer, T.W. (2003). Androgen effects on body composition. Growth Horm IGF Res *13 Suppl A*, S63-71.

Billas, I., and Moras, D. (2013). Allosteric controls of nuclear receptor function in the regulation of transcription. J Mol Biol *425*, 2317-2329.

Bindreither, D., Ecker, S., Gschirr, B., Kofler, A., Kofler, R., and Rainer, J. (2014). The synthetic glucocorticoids prednisolone and dexamethasone regulate the same genes in acute lymphoblastic leukemia cells. BMC Genomics *15*, 662.

Blau, H.M., Pavlath, G.K., Hardeman, E.C., Chiu, C.P., Silberstein, L., Webster, S.G., Miller, S.C., and Webster, C. (1985). Plasticity of the differentiated state. Science *230*, 758-766.

Bodine, S.C., Latres, E., Baumhueter, S., Lai, V.K., Nunez, L., Clarke, B.A., Poueymirou, W.T., Panaro, F.J., Na, E., Dharmarajan, K*., et al.* (2001). Identification of ubiquitin ligases required for skeletal muscle atrophy. Science *294*, 1704-1708.

Bolton, E.C., So, A.Y., Chaivorapol, C., Haqq, C.M., Li, H., and Yamamoto, K.R. (2007). Cell- and gene-specific regulation of primary target genes by the androgen receptor. Genes Dev *21*, 2005-2017.

Bonaldo, P., and Sandri, M. (2013). Cellular and molecular mechanisms of muscle atrophy. Dis Model Mech *6*, 25-39.

Borst, S.E. (2004). Interventions for sarcopenia and muscle weakness in older people. Age Ageing *33*, 548-555.

Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H., and Moras, D. (1995). Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha. Nature *375*, 377-382.

Braun, T.P., Grossberg, A.J., Krasnow, S.M., Levasseur, P.R., Szumowski, M., Zhu, X.X., Maxson, J.E., Knoll, J.G., Barnes, A.P., and Marks, D.L. (2013). Cancer- and endotoxin-induced cachexia require intact glucocorticoid signaling in skeletal muscle. FASEB J *27*, 3572-3582.

Braun, T.P., and Marks, D.L. (2015). The regulation of muscle mass by endogenous glucocorticoids. Front Physiol *6*, 12.

Braun, T.P., Szumowski, M., Levasseur, P.R., Grossberg, A.J., Zhu, X., Agarwal, A., and Marks, D.L. (2014). Muscle atrophy in response to cytotoxic chemotherapy is dependent on intact glucocorticoid signaling in skeletal muscle. PLoS One *9*, e106489.

Bray, M.A., Sheehy, S.P., and Parker, K.K. (2008). Sarcomere alignment is regulated by myocyte shape. Cell Motil Cytoskeleton *65*, 641-651.

Brill, K.T., Weltman, A.L., Gentili, A., Patrie, J.T., Fryburg, D.A., Hanks, J.B., Urban, R.J., and Veldhuis, J.D. (2002). Single and combined effects of growth hormone and testosterone administration on measures of body composition, physical performance, mood, sexual function, bone turnover, and muscle gene expression in healthy older men. J Clin Endocrinol Metab *87*, 5649-5657.

Buermans, H.P., and den Dunnen, J.T. (2014). Next generation sequencing technology: Advances and applications. Biochim Biophys Acta *1842*, 1932-1941.

Bulynko, Y.A., and O'Malley, B.W. (2011). Nuclear receptor coactivators: structural and functional biochemistry. Biochemistry *50*, 313-328.

Burris, T.P., Busby, S.A., and Griffin, P.R. (2012). Targeting orphan nuclear receptors for treatment of metabolic diseases and autoimmunity. Chem Biol *19*, 51-59.

Busillo, J.M., and Cidlowski, J.A. (2013). The five Rs of glucocorticoid action during inflammation: ready, reinforce, repress, resolve, and restore. Trends Endocrinol Metab *24*, 109-119.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell *49*, 825-837.

Campbell, F.C., Xu, H., El-Tanani, M., Crowe, P., and Bingham, V. (2010). The yin and yang of vitamin D receptor (VDR) signaling in neoplastic progression: operational networks and tissue-specific growth control. Biochem Pharmacol *79*, 1-9.

Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoute, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F.*, et al.* (2006). Genome-wide analysis of estrogen receptor binding sites. Nat Genet *38*, 1289-1297.

Chambon, C., Duteil, D., Vignaud, A., Ferry, A., Messaddeq, N., Malivindi, R., Kato, S., Chambon, P., and Metzger, D. (2010). Myocytic androgen receptor controls the strength but not the mass of limb muscles. Proc Natl Acad Sci U S A *107*, 14327-14332.

Chang, C., Saltzman, A., Yeh, S., Young, W., Keller, E., Lee, H.J., Wang, C., and Mizokami, A. (1995). Androgen receptor: an overview. Crit Rev Eukaryot Gene Expr *5*, 97-125.

Chen, J.D., and Evans, R.M. (1995). A transcriptional co-repressor that interacts with nuclear hormone receptors. Nature *377*, 454-457.

Chen, S., Wang, J., Yu, G., Liu, W., and Pearce, D. (1997). Androgen and glucocorticoid receptor heterodimer formation. A possible mechanism for mutual inhibition of transcriptional activity. J Biol Chem *272*, 14087-14092.

Chen, Y., Sawyers, C.L., and Scher, H.I. (2008). Targeting the androgen receptor pathway in prostate cancer. Curr Opin Pharmacol *8*, 440-448.

Chen, Y., and Young, M.A. (2010). Structure of a thyroid hormone receptor DNA-binding domain homodimer bound to an inverted palindrome DNA response element. Mol Endocrinol *24*, 1650-1664.

Cheng, C.S., El-Abd, Y., Bui, K., Hyun, Y.E., Hughes, R.H., Kraus, W.E., and Truskey, G.A. (2014). Conditions that promote primary human skeletal myoblast culture and muscle differentiation in vitro. Am J Physiol Cell Physiol *306*, C385-395.

Claessens, F., Denayer, S., Van Tilborgh, N., Kerkhofs, S., Helsen, C., and Haelens, A. (2008). Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. Nucl Recept Signal *6*, e008.

Claessens, F., Verrijdt, G., Schoenmakers, E., Haelens, A., Peeters, B., Verhoeven, G., and Rombauts, W. (2001). Selective DNA binding by the androgen receptor as a mechanism for hormone-specific gene regulation. J Steroid Biochem Mol Biol *76*, 23-30.

Clinckemalie, L., Vanderschueren, D., Boonen, S., and Claessens, F. (2012). The hinge region in androgen receptor control. Mol Cell Endocrinol *358*, 1-8.

Cole, T.J., Blendy, J.A., Monaghan, A.P., Krieglstein, K., Schmid, W., Aguzzi, A., Fantuzzi, G., Hummler, E., Unsicker, K., and Schutz, G. (1995). Targeted disruption of the glucocorticoid receptor gene blocks adrenergic chromaffin cell development and severely retards lung maturation. Genes Dev *9*, 1608-1621.

Crawford, B.A., Liu, P.Y., Kean, M.T., Bleasel, J.F., and Handelsman, D.J. (2003). Randomized placebo-controlled trial of androgen effects on muscle and bone in men requiring long-term systemic glucocorticoid treatment. J Clin Endocrinol Metab *88*, 3167-3176.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936.

Cullum, R., Alder, O., and Hoodless, P.A. (2011). The next generation: using new sequencing technologies to analyse gene regulation. Respirology *16*, 210-222.

Dao, L.T.M., Galindo-Albarran, A.O., Castro-Mondragon, J.A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., Charbonnier, G., Griffon, A., Vanhille, L., Stephen, T.*, et al.* (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. Nat Genet *49*, 1073-1081.

Dardevet, D., Sornet, C., Savary, I., Debras, E., Patureau-Mirand, P., and Grizard, J. (1998). Glucocorticoid effects on insulin- and IGF-I-regulated muscle protein metabolism during aging. J Endocrinol *156*, 83-89.

Dasgupta, S., Lonard, D.M., and O'Malley, B.W. (2014). Nuclear receptor coactivators: master regulators of human health and disease. Annu Rev Med *65*, 279-292.

Davey, R.A., and Grossmann, M. (2016). Androgen Receptor Structure, Function and Biology: From Bench to Bedside. Clin Biochem Rev *37*, 3-15.

Davies, P., and Rushmere, N.K. (1990). Association of glucocorticoid receptors with prostate nuclear sites for androgen receptors and with androgen response elements. J Mol Endocrinol *5*, 117-127.

de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453-1454.

de Lera, A.R., Bourguet, W., Altucci, L., and Gronemeyer, H. (2007). Design of selective nuclear receptor modulators: RAR and RXR as a case study. Nat Rev Drug Discov *6*, 811-820.

Dekhuijzen, P.N., Gayan-Ramirez, G., Bisschop, A., De Bock, V., Dom, R., and Decramer, M. (1995). Corticosteroid treatment and nutritional deprivation cause a different pattern of atrophy in rat diaphragm. J Appl Physiol (1985) *78*, 629-637.

Del Fabbro, C., Scalabrin, S., Morgante, M., and Giorgi, F.M. (2013). An extensive evaluation of read trimming effects on Illumina NGS data analysis. PLoS One *8*, e85024.

Denayer, S., Helsen, C., Thorrez, L., Haelens, A., and Claessens, F. (2010). The rules of DNA recognition by the androgen receptor. Mol Endocrinol *24*, 898-913.

Drachenberg, D.E., Elgamal, A.A., Rowbotham, R., Peterson, M., and Murphy, G.P. (1999). Circulating levels of interleukin-6 in patients with hormone refractory prostate cancer. Prostate *41*, 127-133.

Dubois, V., Laurent, M., Boonen, S., Vanderschueren, D., and Claessens, F. (2012). Androgens and skeletal muscle: cellular and molecular action mechanisms underlying the anabolic actions. Cell Mol Life Sci *69*, 1651-1667.

Echeverria, P.C., and Picard, D. (2010). Molecular chaperones, essential partners of steroid hormone receptors for activity and mobility. Biochim Biophys Acta *1803*, 641-649.

Eder, I.E., Culig, Z., Putz, T., Nessler-Menardi, C., Bartsch, G., and Klocker, H. (2001). Molecular biology of the androgen receptor: from molecular understanding to the clinic. Eur Urol *40*, 241-251.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B.*, et al.* (2009). Real-time DNA sequencing from single polymerase molecules. Science *323*, 133-138.

Eisermann, K., Wang, D., Jing, Y., Pascal, L.E., and Wang, Z. (2013). Androgen receptor gene mutation, rearrangement, polymorphism. Transl Androl Urol *2*, 137-147.

Elkina, Y., von Haehling, S., Anker, S.D., and Springer, J. (2011). The role of myostatin in muscle wasting: an overview. J Cachexia Sarcopenia Muscle *2*, 143-151.

Escriva, H., Delaunay, F., and Laudet, V. (2000). Ligand binding and nuclear receptor evolution. Bioessays *22*, 717-727.

Estrada, M., Espinosa, A., Muller, M., and Jaimovich, E. (2003). Testosterone stimulates intracellular calcium release and mitogen-activated protein kinases via a G protein-coupled receptor in skeletal muscle cells. Endocrinology *144*, 3586-3597.

Evans, R.M. (1988). The steroid and thyroid hormone receptor superfamily. Science *240*, 889-895.

Evans, R.M., and Mangelsdorf, D.J. (2014). Nuclear Receptors, RXR, and the Big Bang. Cell *157*, 255-266.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B.*, et al.* (2018). The Reactome Pathway Knowledgebase. Nucleic Acids Res *46*, D649-D655.

Fanzani, A., Conraads, V.M., Penna, F., and Martinet, W. (2012). Molecular and cellular mechanisms of skeletal muscle atrophy: an update. J Cachexia Sarcopenia Muscle *3*, 163-179.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat Protoc *7*, 1728-1740.

Feng, W., Ribeiro, R.C., Wagner, R.L., Nguyen, H., Apriletti, J.W., Fletterick, R.J., Baxter, J.D., Kushner, P.J., and West, B.L. (1998). Hormone-dependent coactivator binding to a hydrophobic cleft on nuclear receptors. Science *280*, 1747-1749.

Feng, Y., Zhang, Y., Ying, C., Wang, D., and Du, C. (2015). Nanopore-based fourth-generation DNA sequencing technology. Genomics Proteomics Bioinformatics *13*, 4-16.

Ferrando, A.A., Sheffield-Moore, M., Paddon-Jones, D., Wolfe, R.R., and Urban, R.J. (2003). Differential anabolic effects of testosterone and amino acid feeding in older men. J Clin Endocrinol Metab *88*, 358-362.

Ferrando, A.A., Sheffield-Moore, M., Yeckel, C.W., Gilkison, C., Jiang, J., Achacosa, A., Lieberman, S.A., Tipton, K., Wolfe, R.R., and Urban, R.J. (2002). Testosterone administration to older men improves muscle function: molecular and physiological mechanisms. Am J Physiol Endocrinol Metab *282*, E601-607.

Ferrando, A.A., Tipton, K.D., Doyle, D., Phillips, S.M., Cortiella, J., and Wolfe, R.R. (1998). Testosterone injection stimulates net protein synthesis but not tissue amino acid transport. Am J Physiol *275*, E864-871.

Ferry, A., Schuh, M., Parlakian, A., Mgrditchian, T., Valnaud, N., Joanne, P., Butler-Browne, G., Agbulut, O., and Metzger, D. (2014). Myofiber androgen receptor promotes maximal mechanical overload-induced muscle hypertrophy and fiber type transition in male mice. Endocrinology *155*, 4739-4748.

Flotho, A., and Melchior, F. (2013). Sumoylation: a regulatory protein modification in health and disease. Annu Rev Biochem *82*, 357-385.

Frost, R.A., and Lang, C.H. (2003). Regulation of insulin-like growth factor-I in skeletal muscle and muscle cells. Minerva Endocrinol *28*, 53-73.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358-368.

Furey, T.S. (2012). ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet *13*, 840-852.

Galigniana, M.D., Echeverria, P.C., Erlejman, A.G., and Piwien-Pilipuk, G. (2010). Role of molecular chaperones and TPR-domain proteins in the cytoplasmic transport of steroid receptors and their passage through the nuclear pore. Nucleus *1*, 299-308.

Galigniana, M.D., Scruggs, J.L., Herrington, J., Welsh, M.J., Carter-Su, C., Housley, P.R., and Pratt, W.B. (1998). Heat shock protein 90-dependent (geldanamycin-inhibited) movement of the glucocorticoid receptor through the cytoplasm to the nucleus requires intact cytoskeleton. Mol Endocrinol *12*, 1903-1913.

Gareau, J.R., and Lima, C.D. (2010). The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. Nat Rev Mol Cell Biol *11*, 861-871.

Garside, H., Waters, C., Berry, A., Rice, L., Ardley, H.C., White, A., Robinson, P.A., and Ray, D. (2006). UbcH7 interacts with the glucocorticoid receptor and mediates receptor autoregulation. J Endocrinol *190*, 621-629.

Gates, L.A., Shi, J., Rohira, A.D., Feng, Q., Zhu, B., Bedford, M.T., Sagum, C.A., Jung, S.Y., Qin, J., Tsai, M.J*., et al.* (2017). Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. J Biol Chem *292*, 14456-14472.

Gayan-Ramirez, G., Vanderhoydonc, F., Verhoeven, G., and Decramer, M. (1999). Acute treatment with corticosteroids decreases IGF-1 and IGF-2 expression in the rat diaphragm and gastrocnemius. Am J Respir Crit Care Med *159*, 283-289.

Gearhart, M.D., Holmbeck, S.M., Evans, R.M., Dyson, H.J., and Wright, P.E. (2003). Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition. J Mol Biol *327*, 819-832.

Geertz, M., and Maerkl, S.J. (2010). Experimental strategies for studying transcription factor-DNA binding specificities. Brief Funct Genomics *9*, 362-373.

Gelmann, E.P. (2002). Molecular biology of the androgen receptor. J Clin Oncol *20*, 3001-3015.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512*, 96-100.

Gjerstad, J.K., Lightman, S.L., and Spiga, F. (2018). Role of glucocorticoid negative feedback in the regulation of HPA axis pulsatility. Stress *21*, 403-416.

Glass, C.K., and Rosenfeld, M.G. (2000). The coregulator exchange in transcriptional functions of nuclear receptors. Genes Dev *14*, 121-141.

Glozak, M.A., Sengupta, N., Zhang, X., and Seto, E. (2005). Acetylation and deacetylation of non-histone proteins. Gene *363*, 15-23.

Goldberg, A.L., Tischler, M., DeMartino, G., and Griffin, G. (1980). Hormonal regulation of protein degradation and synthesis in skeletal muscle. Fed Proc *39*, 31-36.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q*., et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol *29*, 644-652.

Green, S., and Chambon, P. (1988). Nuclear receptors enhance our understanding of transcription regulation. Trends Genet *4*, 309-314.

Grino, P.B., Griffin, J.E., and Wilson, J.D. (1990). Testosterone at high concentrations interacts with the human androgen receptor similarly to dihydrotestosterone. Endocrinology *126*, 1165-1172.

Grobet, L., Pirottin, D., Farnir, F., Poncelet, D., Royo, L.J., Brouwers, B., Christians, E., Desmecht, D., Coignoul, F., Kahn, R*., et al.* (2003). Modulating skeletal muscle mass by postnatal, muscle-specific inactivation of the myostatin gene. Genesis *35*, 227-238.

Gros, J., Manceau, M., Thome, V., and Marcelle, C. (2005). A common somitic origin for embryonic muscle progenitors and satellite cells. Nature *435*, 954-958.

Guo, Y., Ye, F., Sheng, Q., Clark, T., and Samuels, D.C. (2014). Three-stage quality control strategies for DNA re-sequencing data. Brief Bioinform *15*, 879-889.

Gupta, K., Sari-Ak, D., Haffke, M., Trowitzsch, S., and Berger, I. (2016). Zooming in on Transcription Preinitiation. J Mol Biol *428*, 2581-2591.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol *8*, R24.

Guttridge, D.C., Mayo, M.W., Madrid, L.V., Wang, C.Y., and Baldwin, A.S., Jr. (2000). NF-kappaB-induced loss of MyoD messenger RNA: possible role in muscle decay and cachexia. Science *289*, 2363-2366.

Hammond, C.L., Hinits, Y., Osborn, D.P., Minchin, J.E., Tettamanti, G., and Hughes, S.M. (2007). Signals and myogenic regulatory factors restrict pax3 and pax7 expression to dermomyotome-like tissue in zebrafish. Dev Biol *302*, 504-521.

Handy, D.E., Castro, R., and Loscalzo, J. (2011). Epigenetic modifications: basic mechanisms and role in cardiovascular disease. Circulation *123*, 2145-2156.

Hard, T., Kellenbach, E., Boelens, R., Maler, B.A., Dahlman, K., Freedman, L.P., Carlstedt-Duke, J., Yamamoto, K.R., Gustafsson, J.A., and Kaptein, R. (1990). Solution structure of the glucocorticoid receptor DNA-binding domain. Science *249*, 157-160.

Hardy, M.P., Gao, H.B., Dong, Q., Ge, R., Wang, Q., Chai, W.R., Feng, X., and Sottas, C. (2005). Stress hormone and male reproductive function. Cell Tissue Res *322*, 147-153.

Hasselgren, P.O. (1999). Glucocorticoids and muscle catabolism. Curr Opin Clin Nutr Metab Care *2*, 201-205.

Hasselgren, P.O., Alamdari, N., Aversa, Z., Gonnella, P., Smith, I.J., and Tizio, S. (2010). Corticosteroids and muscle wasting: role of transcription factors, nuclear cofactors, and hyperacetylation. Curr Opin Clin Nutr Metab Care *13*, 423-428.

Hasty, P., Bradley, A., Morris, J.H., Edmondson, D.G., Venuti, J.M., Olson, E.N., and Klein, W.H. (1993). Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. Nature *364*, 501-506.

He, B., Kemppainen, J.A., Voegel, J.J., Gronemeyer, H., and Wilson, E.M. (1999). Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH(2)-terminal domain. J Biol Chem *274*, 37219-37225.

He, Y., Yi, W., Suino-Powell, K., Zhou, X.E., Tolbert, W.D., Tang, X., Yang, J., Yang, H., Shi, J., Hou, L.*, et al.* (2014). Structures and mechanism for the design of highly potent glucocorticoids. Cell Res *24*, 713-726.

Heemers, H.V., and Tindall, D.J. (2007). Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. Endocr Rev *28*, 778-808.

Heery, D.M., Kalkhoven, E., Hoare, S., and Parker, M.G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature *387*, 733-736.

Heinlein, C.A., and Chang, C. (2002). Androgen receptor (AR) coregulators: an overview. Endocr Rev *23*, 175-200.

Heinlein, C.A., and Chang, C. (2004). Androgen receptor in prostate cancer. Endocr Rev *25*, 276-308.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W.*, et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Heldring, N., Isaacs, G.D., Diehl, A.G., Sun, M., Cheung, E., Ranish, J.A., and Kraus, W.L. (2011). Multiple sequence-specific DNA-binding proteins mediate estrogen receptor signaling through a tethering pathway. Mol Endocrinol *25*, 564-574.

Helsen, C., and Claessens, F. (2014). Looking at nuclear receptors from a new angle. Mol Cell Endocrinol *382*, 97-106.

Helsen, C., Kerkhofs, S., Clinckemalie, L., Spans, L., Laurent, M., Boonen, S., Vanderschueren, D., and Claessens, F. (2012). Structural basis for nuclear hormone receptor DNA binding. Mol Cell Endocrinol *348*, 411-417.

Heslop-Harrison, J.S., and Schwarzacher, T. (2013). Nucleosomes and centromeric DNA packaging. Proc Natl Acad Sci U S A *110*, 19974-19975.

Hinman, V., and Cary, G. (2017). The evolution of gene regulation. Elife *6*.

Hirayama, Y., and Sadar, M.D. (2018). Does increased expression of glucocorticoid receptor support application of antagonists to this receptor for the treatment of castration resistant prostate cancer? AME Med J *3*.

Hoang, D.T., Iczkowski, K.A., Kilari, D., See, W., and Nevalainen, M.T. (2017). Androgen receptor-dependent and -independent mechanisms driving prostate cancer progression: Opportunities for therapeutic targeting from multiple angles. Oncotarget *8*, 3724-3745.

Hobisch, A., Eder, I.E., Putz, T., Horninger, W., Bartsch, G., Klocker, H., and Culig, Z. (1998). Interleukin-6 regulates prostate-specific protein expression in prostate carcinoma cells by activation of the androgen receptor. Cancer Res *58*, 4640-4645.

Hooper, S.L., Hobbs, K.H., and Thuma, J.B. (2008). Invertebrate muscles: thin and thick filament structure; molecular basis of contraction and its regulation, catch and asynchronous muscle. Prog Neurobiol *86*, 72-127.

Horlein, A.J., Naar, A.M., Heinzel, T., Torchia, J., Gloss, B., Kurokawa, R., Ryan, A., Kamei, Y., Soderstrom, M., Glass, C.K.*, et al.* (1995). Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. Nature *377*, 397-404.

Hsia, E.Y., Goodson, M.L., Zou, J.X., Privalsky, M.L., and Chen, H.W. (2010). Nuclear receptor coregulators as a new paradigm for therapeutic targeting. Adv Drug Deliv Rev *62*, 1227-1237.

Hu, R., Dunn, T.A., Wei, S., Isharwal, S., Veltri, R.W., Humphreys, E., Han, M., Partin, A.W., Vessella, R.L., Isaacs, W.B.*, et al.* (2009a). Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. Cancer Res *69*, 16-22.

Hu, X., and Lazar, M.A. (2000). Transcriptional repression by nuclear hormone receptors. Trends Endocrinol Metab *11*, 6-10.

Hu, Z., Wang, H., Lee, I.H., Du, J., and Mitch, W.E. (2009b). Endogenous glucocorticoids and impaired insulin signaling are both required to stimulate muscle wasting under pathophysiological conditions in mice. J Clin Invest *119*, 3059-3069.

Hua, G., Ganti, K.P., and Chambon, P. (2016). Glucocorticoid-induced tethered transrepression requires SUMOylation of GR and formation of a SUMO-SMRT/NCoR1-HDAC3 repressing complex. Proc Natl Acad Sci U S A *113*, E635-643.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Huang, P., Chandra, V., and Rastinejad, F. (2010). Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. Annu Rev Physiol *72*, 247-272.

Hudson, W.H., Youn, C., and Ortlund, E.A. (2013). The structural basis of direct glucocorticoid-mediated transrepression. Nat Struct Mol Biol *20*, 53-58.

Imae, M., Fu, Z., Yoshida, A., Noguchi, T., and Kato, H. (2003). Nutritional and hormonal factors control the gene expression of FoxOs, the mammalian homologues of DAF-16. J Mol Endocrinol *30*, 253-262.

Inder, W.J., Jang, C., Obeyesekere, V.R., and Alford, F.P. (2010). Dexamethasone administration inhibits skeletal muscle expression of the androgen receptor and IGF-1--implications for steroid-induced myopathy. Clin Endocrinol (Oxf) *73*, 126-132.

Ing, N.H., Forrest, D.W., Riggs, P.K., Loux, S., Love, C.C., Brinsko, S.P., Varner, D.D., and Welsh, T.H., Jr. (2014). Dexamethasone acutely down-regulates genes involved in steroidogenesis in stallion testes. J Steroid Biochem Mol Biol *143*, 451-459.

Isikbay, M., Otto, K., Kregel, S., Kach, J., Cai, Y., Vander Griend, D.J., Conzen, S.D., and Szmulewitz, R.Z. (2014). Glucocorticoid receptor activity contributes to resistance to androgen-targeted therapy in prostate cancer. Horm Cancer *5*, 72-89.

Itani, O.A., Liu, K.Z., Cornish, K.L., Campbell, J.R., and Thomas, C.P. (2002). Glucocorticoids stimulate human sgk1 gene expression by activation of a GRE in its 5'-flanking region. Am J Physiol Endocrinol Metab *283*, E971-979.

Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Varnai, C., Thiecke, M.J.*, et al.* (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell *167*, 1369-1384 e1319.

Jen, Y., Weintraub, H., and Benezra, R. (1992). Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. Genes Dev *6*, 1466-1479.

Jeronimo, C., and Robert, F. (2017). A Spiking Strategy for ChIP-chip Data Normalization in S. cerevisiae. Methods Mol Biol *1528*, 211-227.

Jogo, M., Shiraishi, S., and Tamura, T.A. (2009). Identification of MAFbx as a myogenin-engaged F-box protein in SCF ubiquitin ligase. FEBS Lett *583*, 2715-2719.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497-1502.

Jones, A., Hwang, D.J., Narayanan, R., Miller, D.D., and Dalton, J.T. (2010). Effects of a novel selective androgen receptor modulator on dexamethasone-induced and hypogonadism-induced muscle atrophy. Endocrinology *151*, 3706-3719.

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L.*, et al.* (2005). Reactome: a knowledgebase of biological pathways. Nucleic Acids Res *33*, D428-432.

Kadmiel, M., and Cidlowski, J.A. (2013). Glucocorticoid receptor signaling in health and disease. Trends Pharmacol Sci *34*, 518-530.

Kadonaga, J.T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. Cell *116*, 247-257.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

Kang, H.Y., Cho, C.L., Huang, K.L., Wang, J.C., Hu, Y.C., Lin, H.K., Chang, C., and Huang, K.E. (2004). Nongenomic androgen activation of phosphatidylinositol 3-kinase/Akt signaling pathway in MC3T3-E1 osteoblasts. J Bone Miner Res *19*, 1181-1190.

Kassar-Duchossoy, L., Giacone, E., Gayraud-Morel, B., Jory, A., Gomes, D., and Tajbakhsh, S. (2005). Pax3/Pax7 mark a novel population of primitive myogenic cells during development. Genes Dev *19*, 1426-1431.

Kassel, O., and Herrlich, P. (2007). Crosstalk between the glucocorticoid receptor and other transcription factors: molecular aspects. Mol Cell Endocrinol *275*, 13-29.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res *12*, 996-1006.

Kerkhofs, S., Dubois, V., De Gendt, K., Helsen, C., Clinckemalie, L., Spans, L., Schuit, F., Boonen, S., Vanderschueren, D., Saunders, P.T.*, et al.* (2012). A role for selective androgen response elements in the development of the epididymis and the androgen control of the 5alpha reductase II gene. FASEB J *26*, 4360-4372.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Cheneby, J., Kulkarni, S.R., Tan, G.*, et al.* (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res *46*, D260-D266.

Khan, S.H., Arnott, J.A., and Kumar, R. (2011). Naturally occurring osmolyte, trehalose induces functional conformation in an intrinsically disordered activation domain of glucocorticoid receptor. PLoS One *6*, e19689.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kim, H.J., and Lee, W.J. (2009). Ligand-independent activation of the androgen receptor by insulin-like growth factor-I and the role of the MAPK pathway in skeletal muscle cells. Mol Cells *28*, 589-593.

Kino, T. (2000). Glucocorticoid Receptor. In Endotext, K.R. Feingold, B. Anawalt, A. Boyce, G. Chrousos, K. Dungan, A. Grossman, J.M. Hershman, G. Kaltsas, C. Koch, P. Kopp*, et al.*, eds. (South Dartmouth (MA)).

Kino, T., Su, Y.A., and Chrousos, G.P. (2009). Human glucocorticoid receptor isoform beta: recent understanding of its potential implications in physiology and pathophysiology. Cell Mol Life Sci *66*, 3435-3448.

Knudsen, K.E., and Penning, T.M. (2010). Partners in crime: deregulation of AR activity and androgen synthesis in prostate cancer. Trends Endocrinol Metab *21*, 315-324.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P.*, et al.* (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res *17*, 691-707.

Koochekpour, S. (2010). Androgen receptor signaling and mutations in prostate cancer. Asian J Androl *12*, 639-657.

Kostyo, J.L., and Redmond, A.F. (1966). Role of protein synthesis in the inhibitory action of adrenal steroid hormones on amino acid transport by muscle. Endocrinology *79*, 531-540.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kretz, O., Reichardt, H.M., Schutz, G., and Bock, R. (1999). Corticotropin-releasing hormone expression is the major target for glucocorticoid feedback-control at the hypothalamic level. Brain Res *818*, 488-491.

Kumar, R., and Litwack, G. (2009). Structural and functional relationships of the steroid hormone receptors' N-terminal transactivation domain. Steroids *74*, 877-883.

Kumar, R., and McEwan, I.J. (2012). Allosteric modulators of steroid hormone receptors: structural dynamics and gene regulation. Endocr Rev *33*, 271-299.

Kumar, R., and Thompson, E.B. (2012). Folding of the glucocorticoid receptor N-terminal transactivation function: dynamics and regulation. Mol Cell Endocrinol *348*, 450-456.

Lalevee, S., Ferry, C., and Rochette-Egly, C. (2010). Phosphorylation control of nuclear receptors. Methods Mol Biol *647*, 251-266.

Langlais, D., Couture, C., Balsalobre, A., and Drouin, J. (2012). The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. Mol Cell *47*, 38-49.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Laryea, G., Muglia, L., Arnett, M., and Muglia, L.J. (2015). Dissection of glucocorticoid receptor-mediated inhibition of the hypothalamic-pituitary-adrenal axis by gene targeting in mice. Front Neuroendocrinol *36*, 150-164.

Latres, E., Amini, A.R., Amini, A.A., Griffiths, J., Martin, F.J., Wei, Y., Lin, H.C., Yancopoulos, G.D., and Glass, D.J. (2005). Insulin-like growth factor-1 (IGF-1) inversely regulates atrophy-induced genes via the phosphatidylinositol 3-kinase/Akt/mammalian target of rapamycin (PI3K/Akt/mTOR) pathway. J Biol Chem *280*, 2737-2744.

Laudet, V., and Gronemeyer, H. (2002). The nuclear receptors factsbook. Academic Press.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet *32*, 42-56.

Lecker, S.H., Solomon, V., Mitch, W.E., and Goldberg, A.L. (1999). Muscle protein breakdown and the critical role of the ubiquitin-proteasome pathway in normal and disease states. J Nutr *129*, 227S-237S.

Lee, J.S., Smith, E., and Shilatifard, A. (2010). The language of histone crosstalk. Cell *142*, 682-685.

Lee, S., Seo, C.H., Lim, B., Yang, J.O., Oh, J., Kim, M., Lee, S., Lee, B., Kang, C., and Lee, S. (2011). Accurate quantification of transcriptome from RNA-Seq data by effective length normalization. Nucleic Acids Res *39*, e9.

Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. Cell *157*, 13-25.

Li, G., Cai, L., Chang, H., Hong, P., Zhou, Q., Kulakova, E.V., Kolchanov, N.A., and Ruan, Y. (2014). Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. BMC Genomics *15 Suppl 12*, S11.

Li, Y., Lambert, M.H., and Xu, H.E. (2003). Activation of nuclear receptors: a perspective from structural genomics. Structure *11*, 741-746.

Lin, B., Wang, J., Hong, X., Yan, X., Hwang, D., Cho, J.H., Yi, D., Utleg, A.G., Fang, X., Schones, D.E.*, et al.* (2009). Integrated expression profiling and ChIP-seq analyses of the growth inhibition response program of the androgen receptor. PLoS One *4*, e6589.

Lin, C., Garruss, A.S., Luo, Z., Guo, F., and Shilatifard, A. (2013a). The RNA Pol II elongation factor Ell3 marks enhancers in ES cells and primes future gene activation. Cell *152*, 144-156.

Lin, T.H., Lee, S.O., Niu, Y., Xu, D., Liang, L., Li, L., Yeh, S.D., Fujimoto, N., Yeh, S., and Chang, C. (2013b). Differential androgen deprivation therapies with anti-androgens casodex/bicalutamide or MDV3100/Enzalutamide versus anti-androgen receptor ASC-J9(R) Lead to promotion versus suppression of prostate cancer metastasis. J Biol Chem *288*, 19359-19369.

Lingbeck, J.M., Trausch-Azar, J.S., Ciechanover, A., and Schwartz, A.L. (2003). Determinants of nuclear and cytoplasmic ubiquitin-mediated degradation of MyoD. J Biol Chem *278*, 1817-1823.

Little, T.H., Zhang, Y., Matulis, C.K., Weck, J., Zhang, Z., Ramachandran, A., Mayo, K.E., and Radhakrishnan, I. (2006). Sequence-specific deoxyribonucleic acid (DNA) recognition by steroidogenic factor 1: a helix at the carboxy terminus of the DNA binding domain is necessary for complex stability. Mol Endocrinol *20*, 831-843.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of next-generation sequencing systems. J Biomed Biotechnol *2012*, 251364.

Liu, W., Zhang, B., He, W., Wang, Z., Li, G., and Liu, J. (2016). Characterization of in vivo phosphorylation modification of differentially accumulated proteins in cotton fiber-initiation process. Acta Biochim Biophys Sin (Shanghai) *48*, 756-761.

Lofberg, E., Gutierrez, A., Wernerman, J., Anderstam, B., Mitch, W.E., Price, S.R., Bergstrom, J., and Alvestrand, A. (2002). Effects of high doses of glucocorticoids on free amino acids, ribosomes and protein turnover in human muscle. Eur J Clin Invest *32*, 345-353.

Lonard, D.M., and O'Malley, B.W. (2012). Nuclear receptor coregulators: modulators of pathology and therapeutic targets. Nat Rev Endocrinol *8*, 598-604.

Losel, R.M., Falkenstein, E., Feuring, M., Schultz, A., Tillmann, H.C., Rossol-Haseroth, K., and Wehling, M. (2003). Nongenomic steroid action: controversies, questions, and answers. Physiol Rev *83*, 965-1016.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lu, N.Z., and Cidlowski, J.A. (2006). Glucocorticoid receptor isoforms generate transcription specificity. Trends Cell Biol *16*, 301-307.

Lu, R., Mucaki, E.J., and Rogan, P.K. (2017). Discovery and validation of information theory-based transcription factor and cofactor binding site motifs. Nucleic Acids Res *45*, e27.

Luse, D.S. (2014). The RNA polymerase II preinitiation complex. Through what pathway is the complex assembled? Transcription *5*, e27050.

Luther, P.K. (2009). The vertebrate muscle Z-disc: sarcomere anchor for structure and signalling. J Muscle Res Cell Motil *30*, 171-185.

Ma, K., Mallidis, C., Bhasin, S., Mahabadi, V., Artaza, J., Gonzalez-Cadavid, N., Arias, J., and Salehian, B. (2003). Glucocorticoid-induced skeletal muscle atrophy is associated with upregulation of myostatin gene expression. Am J Physiol Endocrinol Metab *285*, E363-371.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696-1697.

MacLean, H.E., Chiu, W.S., Notini, A.J., Axell, A.M., Davey, R.A., McManus, J.F., Ma, C., Plant, D.R., Lynch, G.S., and Zajac, J.D. (2008). Impaired skeletal muscle development and function in male, but not female, genomic androgen receptor knockout mice. FASEB J *22*, 2676-2689.

MacLean, H.E., and Handelsman, D.J. (2009). Unraveling androgen action in muscle: genetic tools probing cellular mechanisms. Endocrinology *150*, 3437-3439.

MacLean, H.E., Warne, G.L., and Zajac, J.D. (1997). Localization of functional domains in the androgen receptor. J Steroid Biochem Mol Biol *62*, 233-242.

Macpherson, P.C., Wang, X., and Goldman, D. (2011). Myogenin regulates denervation-dependent muscle atrophy in mouse soleus muscle. J Cell Biochem *112*, 2149-2159.

Malone, J.H., and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol *9*, 34.

Malovannaya, A., Lanz, R.B., Jung, S.Y., Bulynko, Y., Le, N.T., Chan, D.W., Ding, C., Shi, Y., Yucer, N., Krenciute, G.*, et al.* (2011). Analysis of the human endogenous coregulator complexome. Cell *145*, 787-799.

Mamchaoui, K., Trollet, C., Bigot, A., Negroni, E., Chaouch, S., Wolff, A., Kandalla, P.K., Marie, S., Di Santo, J., St Guily, J.L.*, et al.* (2011). Immortalized pathological human myoblasts: towards a universal tool for the study of neuromuscular disorders. Skelet Muscle *1*, 34.

Mammucari, C., Milan, G., Romanello, V., Masiero, E., Rudolf, R., Del Piccolo, P., Burden, S.J., Di Lisi, R., Sandri, C., Zhao, J.*, et al.* (2007). FoxO3 controls autophagy in skeletal muscle in vivo. Cell Metab *6*, 458-471.

Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P.*, et al.* (1995). The nuclear receptor superfamily: the second decade. Cell *83*, 835-839.

Mardis, E.R. (2013). Next-generation sequencing platforms. Annu Rev Anal Chem (Palo Alto Calif) *6*, 287-303.

Marguerat, S., Wilhelm, B.T., and Bahler, J. (2008). Next-generation sequencing: applications beyond genomes. Biochem Soc Trans *36*, 1091-1096.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z*., et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376-380.

Marino-Ramirez, L., Kann, M.G., Shoemaker, B.A., and Landsman, D. (2005). Histone structure and nucleosome stability. Expert Rev Proteomics *2*, 719-729.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res *18*, 1509-1517.

Maruvada, P., Baumann, C.T., Hager, G.L., and Yen, P.M. (2003). Dynamic shuttling and intranuclear mobility of nuclear hormone receptors. J Biol Chem *278*, 12425-12432.

Massari, M.E., and Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol *20*, 429-440.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet *7*, 29-59.

McEwan, I.J. (2004). Molecular mechanisms of androgen receptor-mediated gene regulation: structure-function analysis of the AF-1 domain. Endocr Relat Cancer *11*, 281-293.

McFarlane, C., Plummer, E., Thomas, M., Hennebry, A., Ashby, M., Ling, N., Smith, H., Sharma, M., and Kambadur, R. (2006). Myostatin induces cachexia by activating the ubiquitin proteolytic system through an NF-kappaB-independent, FoxO1-dependent mechanism. J Cell Physiol *209*, 501-514.

McMahon, D.K., Anderson, P.A., Nassar, R., Bunting, J.B., Saba, Z., Oakeley, A.E., and Malouf, N.N. (1994). C2C12 cells: biophysical, biochemical, and immunocytochemical properties. Am J Physiol *266*, C1795-1802.

Megeney, L.A., Kablar, B., Garrett, K., Anderson, J.E., and Rudnicki, M.A. (1996). MyoD is required for myogenic stem cell function in adult skeletal muscle. Genes Dev *10*, 1173-1183.

Meijsing, S.H., Pufall, M.A., So, A.Y., Bates, D.L., Chen, L., and Yamamoto, K.R. (2009). DNA binding site sequence directs glucocorticoid receptor structure and activity. Science *324*, 407-410.

Meinke, G., and Sigler, P.B. (1999). DNA-binding mechanism of the monomeric orphan nuclear receptor NGFI-B. Nat Struct Biol *6*, 471-477.

Meng, H., and Bartholomew, B. (2018). Emerging roles of transcriptional enhancers in chromatin looping and promoter-proximal pausing of RNA polymerase II. J Biol Chem *293*, 13786-13794.

Merika, M., Williams, A.J., Chen, G., Collins, T., and Thanos, D. (1998). Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. Mol Cell *1*, 277-287.

Meyer, M.B., Benkusky, N.A., Lee, C.H., and Pike, J.W. (2014). Genomic determinants of gene regulation by 1,25-dihydroxyvitamin D3 during osteoblast-lineage cell differentiation. J Biol Chem *289*, 19539-19554.

Mi, H., and Thomas, P. (2009). PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. Methods Mol Biol *563*, 123-140.

Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A*., et al.* (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet *47*, 598-606.

Millard, C.J., Watson, P.J., Fairall, L., and Schwabe, J.W. (2013). An evolving understanding of nuclear receptor coregulator proteins. J Mol Endocrinol *51*, T23-36.

Milne, J.C., Lambert, P.D., Schenk, S., Carney, D.P., Smith, J.J., Gagne, D.J., Jin, L., Boss, O., Perni, R.B., Vu, C.B*., et al.* (2007). Small molecule activators of SIRT1 as therapeutics for the treatment of type 2 diabetes. Nature *450*, 712-716.

Minoche, A.E., Dohm, J.C., and Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. Genome Biol *12*, R112.

Mitch, W.E., and Goldberg, A.L. (1996). Mechanisms of muscle wasting. The role of the ubiquitin-proteasome pathway. N Engl J Med *335*, 1897-1905.

Mitre-Aguilar, I.B., Cabrera-Quintero, A.J., and Zentella-Dehesa, A. (2015). Genomic and non-genomic effects of glucocorticoids: implications for breast cancer. Int J Clin Exp Pathol *8*, 1-10.

Moore, D.D. (1990). Diversity and unity in the nuclear hormone receptors: a terpenoid receptor superfamily. New Biol *2*, 100-105.

Mora, A., Sandve, G.K., Gabrielsen, O.S., and Eskeland, R. (2016). In the loop: promoter-enhancer interactions and bioinformatics. Brief Bioinform *17*, 980-995.

Moras, D., and Gronemeyer, H. (1998). The nuclear receptor ligand-binding domain: structure and function. Curr Opin Cell Biol *10*, 384-391.

Mori, S., Nada, S., Kimura, H., Tajima, S., Takahashi, Y., Kitamura, A., Oneyama, C., and Okada, M. (2014). The mTOR pathway controls cell proliferation by regulating the FoxO3a transcription factor via SGK1 kinase. PLoS One *9*, e88891.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.

Mullen, A.C., Orlando, D.A., Newman, J.J., Loven, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., DeKoter, R.P., and Young, R.A. (2011). Master transcription factors determine cell-type-specific responses to TGF-beta signaling. Cell *147*, 565-576.

Muses, S., Morgan, J.E., and Wells, D.J. (2011). A new extensively characterised conditionally immortal muscle cell-line for investigating therapeutic strategies in muscular dystrophies. PLoS One *6*, e24826.

Nabeshima, Y., Hanaoka, K., Hayasaka, M., Esumi, E., Li, S., Nonaka, I., and Nabeshima, Y. (1993). Myogenin gene disruption results in perinatal lethality because of severe muscle defect. Nature *364*, 532-535.

Nagy, L., and Schwabe, J.W. (2004). Mechanism of the nuclear receptor molecular switch. Trends Biochem Sci *29*, 317-324.

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M.C., Hirai, A., Takahashi, H.*, et al.* (2011). Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res *39*, e90.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110-113.

Ndibe, C., Wang, C.G., and Sonpavde, G. (2015). Corticosteroids in the management of prostate cancer: a critical review. Curr Treat Options Oncol *16*, 6.

Nguyen, P.L., Alibhai, S.M., Basaria, S., D'Amico, A.V., Kantoff, P.W., Keating, N.L., Penson, D.F., Rosario, D.J., Tombal, B., and Smith, M.R. (2015). Adverse effects of androgen deprivation therapy and strategies to mitigate them. Eur Urol *67*, 825-836.

Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., Montavon, T., Becker, J.S., Grindheim, J.M., Blahnik, K.*, et al.* (2019). H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. Science *363*, 294-297.

Niedringhaus, T.P., Milanova, D., Kerby, M.B., Snyder, M.P., and Barron, A.E. (2011). Landscape of next-generation sequencing technologies. Anal Chem *83*, 4327-4341.

Noonan, J.P., and McCallion, A.S. (2010). Genomics of long-range regulatory elements. Annu Rev Genomics Hum Genet *11*, 1-23.

Notini, A.J., Davey, R.A., McManus, J.F., Bate, K.L., and Zajac, J.D. (2005). Genomic actions of the androgen receptor are required for normal male sexual differentiation in a mouse model. J Mol Endocrinol *35*, 547-555.

Nystrom, G., Pruznak, A., Huber, D., Frost, R.A., and Lang, C.H. (2009). Local insulin-like growth factor I prevents sepsis-induced muscle atrophy. Metabolism *58*, 787-797.

O'Connor, T.R., and Bailey, T.L. (2014). Creating and validating cis-regulatory maps of tissue-specific gene expression regulation. Nucleic Acids Res *42*, 11000-11010.

Oakley, R.H., and Cidlowski, J.A. (2011). Cellular processing of the glucocorticoid receptor gene and protein: new mechanisms for generating tissue-specific actions of glucocorticoids. J Biol Chem *286*, 3177-3184.

Oakley, R.H., and Cidlowski, J.A. (2013). The biology of the glucocorticoid receptor: new signaling mechanisms in health and disease. J Allergy Clin Immunol *132*, 1033-1044.

Oakley, R.H., Revollo, J., and Cidlowski, J.A. (2012). Glucocorticoids regulate arrestin gene expression and redirect the signaling profile of G protein-coupled receptors. Proc Natl Acad Sci U S A *109*, 17591-17596.

Ordonez-Moran, P., and Munoz, A. (2009). Nuclear receptors: genomic and non-genomic effects converge. Cell Cycle *8*, 1675-1680.

Oshlack, A., Robinson, M.D., and Young, M.D. (2010). From RNA-seq reads to differential expression results. Genome Biol *11*, 220.

Otto, C., Reichardt, H.M., and Schutz, G. (1997). Absence of glucocorticoid receptor-beta in mice. J Biol Chem *272*, 26665-26668.

Oustanina, S., Hause, G., and Braun, T. (2004). Pax7 directs postnatal renewal and propagation of myogenic satellite cells but not their specification. EMBO J *23*, 3430-3439.

Panettieri, R.A., Schaafsma, D., Amrani, Y., Koziol-White, C., Ostrom, R., and Tliba, O. (2019). Non-genomic Effects of Glucocorticoids: An Updated View. Trends Pharmacol Sci *40*, 38-49.

Pareek, C.S., Smoczynski, R., and Tretyn, A. (2011). Sequencing technologies and genome sequencing. J Appl Genet *52*, 413-435.

Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet *10*, 669-680.

Pascual, G., and Glass, C.K. (2006). Nuclear receptors versus inflammation: mechanisms of transrepression. Trends Endocrinol Metab *17*, 321-327.

Pasini, D., Malatesta, M., Jung, H.R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O.N.*, et al.* (2010). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucleic Acids Res *38*, 4958-4969.

Patel, R., Williams-Dautovich, J., and Cummins, C.L. (2014). Minireview: new molecular mediators of glucocorticoid receptor activity in metabolic tissues. Mol Endocrinol *28*, 999-1011.

Pawlak, M., Lefebvre, P., and Staels, B. (2012). General molecular biology and architecture of nuclear receptors. Curr Top Med Chem *12*, 486-504.

Pepke, S., Wold, B., and Mortazavi, A. (2009). Computation for ChIP-seq and RNA-seq studies. Nat Methods *6*, S22-32.

Perlmann, T., Umesono, K., Rangarajan, P.N., Forman, B.M., and Evans, R.M. (1996). Two distinct dimerization interfaces differentially modulate target gene specificity of nuclear hormone receptors. Mol Endocrinol *10*, 958-966.

Phillips-Cremins, J.E., and Corces, V.G. (2013). Chromatin insulators: linking genome organization to cellular function. Mol Cell *50*, 461-474.

Pollard, T.D., and Weihing, R.R. (1974). Actin and myosin and cell movement. CRC Crit Rev Biochem *2*, 1-65.

Pownall, M.E., and Emerson, C.P., Jr. (1992). Sequential activation of three myogenic regulatory genes during somite morphogenesis in quail embryos. Dev Biol *151*, 67-79.

Privalsky, M.L. (2004). The role of corepressors in transcriptional regulation by nuclear hormone receptors. Annu Rev Physiol *66*, 315-360.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature *322*, 697-701.

Puhr, M., Hoefer, J., Eigentler, A., Ploner, C., Handle, F., Schaefer, G., Kroon, J., Leo, A., Heidegger, I., Eder, I.*, et al.* (2018). The Glucocorticoid Receptor Is a Key Player for Prostate Cancer Cell Survival and a Target for Improved Antiandrogen Therapy. Clin Cancer Res *24*, 927-938.

Puzianowska-Kuznicka, M., Pawlik-Pachucka, E., Owczarz, M., Budzinska, M., and Polosak, J. (2013). Small-molecule hormones: molecular mechanisms of action. Int J Endocrinol *2013*, 601246.

Qin, W., Pan, J., Wu, Y., Bauman, W.A., and Cardozo, C. (2010). Protection against dexamethasone-induced muscle atrophy is related to modulation by testosterone of FOXO1 and PGC-1alpha. Biochem Biophys Res Commun *403*, 473-478.

Ramamoorthy, S., and Cidlowski, J.A. (2013a). Exploring the molecular mechanisms of glucocorticoid receptor action from sensitivity to resistance. Endocr Dev *24*, 41-56.

Ramamoorthy, S., and Cidlowski, J.A. (2013b). Ligand-induced repression of the glucocorticoid receptor gene is mediated by an NCoR1 repression complex formed by long-range chromatin interactions with intragenic glucocorticoid response elements. Mol Cell Biol *33*, 1711-1722.

Ramaswamy, S., and Weinbauer, G.F. (2014). Endocrine control of spermatogenesis: Role of FSH and LH/testosterone. Spermatogenesis *4*, e996025.

Rana, K., Davey, R.A., and Zajac, J.D. (2014). Human androgen deficiency: insights gained from androgen receptor knockout mouse models. Asian J Androl *16*, 169-177.

Rastinejad, F., Huang, P., Chandra, V., and Khorasanizadeh, S. (2013). Understanding nuclear receptor form and function using structural biology. J Mol Endocrinol *51*, T1-T21.

Rastinejad, F., Perlmann, T., Evans, R.M., and Sigler, P.B. (1995). Structural determinants of nuclear receptor assembly on DNA direct repeats. Nature *375*, 203-211.

Reisz-Porszasz, S., Bhasin, S., Artaza, J.N., Shen, R., Sinha-Hikim, I., Hogue, A., Fielder, T.J., and Gonzalez-Cadavid, N.F. (2003). Lower skeletal muscle mass in male transgenic mice with muscle-specific overexpression of myostatin. Am J Physiol Endocrinol Metab *285*, E876-888.

Relaix, F., Montarras, D., Zaffran, S., Gayraud-Morel, B., Rocancourt, D., Tajbakhsh, S., Mansouri, A., Cumano, A., and Buckingham, M. (2006). Pax3 and Pax7 have distinct and overlapping functions in adult muscle progenitor cells. J Cell Biol *172*, 91-102.

Relaix, F., Rocancourt, D., Mansouri, A., and Buckingham, M. (2004). Divergent functions of murine Pax3 and Pax7 in limb muscle development. Genes Dev *18*, 1088-1105.

Richmond, T.J., and Davey, C.A. (2003). The structure of DNA in the nucleosome core. Nature *423*, 145-150.

Robinson-Rechavi, M., Carpentier, A.S., Duffraisse, M., and Laudet, V. (2001). How many nuclear hormone receptors are there in the human genome? Trends Genet *17*, 554-556.

Robinson-Rechavi, M., Escriva Garcia, H., and Laudet, V. (2003). The nuclear receptor superfamily. J Cell Sci *116*, 585-586.

Robinson-Rechavi, M., and Laudet, V. (2003). Bioinformatics of nuclear receptors. Methods Enzymol *364*, 95-118.

Robinson-Rechavi, M., Maina, C.V., Gissendanner, C.R., Laudet, V., and Sluder, A. (2005). Explosive lineage-specific expansion of the orphan nuclear receptor HNF4 in nematodes. J Mol Evol *60*, 577-586.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol *11*, R25.

Roemer, S.C., Donham, D.C., Sherman, L., Pon, V.H., Edwards, D.P., and Churchill, M.E. (2006). Structure of the progesterone receptor-deoxyribonucleic acid complex: novel interactions required for binding to half-site response elements. Mol Endocrinol *20*, 3042-3052.

Rosen, J., and Miner, J.N. (2005). The search for safer glucocorticoid receptor ligands. Endocr Rev *26*, 452-464.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M.*, et al.* (2011). An integrated semiconductor device enabling non-optical genome sequencing. Nature *475*, 348-352.

Rudnicki, M.A., Braun, T., Hinuma, S., and Jaenisch, R. (1992). Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. Cell *71*, 383-390.

Sacheck, J.M., Ohtsuka, A., McLary, S.C., and Goldberg, A.L. (2004). IGF-I stimulates muscle growth by suppressing protein breakdown and expression of atrophy-related ubiquitin ligases, atrogin-1 and MuRF1. Am J Physiol Endocrinol Metab *287*, E591-601.

Sahu, B., Laakso, M., Ovaska, K., Mirtti, T., Lundin, J., Rannikko, A., Sankila, A., Turunen, J.P., Lundin, M., Konsti, J.*, et al.* (2011). Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. EMBO J *30*, 3962-3976.

Sahu, B., Laakso, M., Pihlajamaa, P., Ovaska, K., Sinielnikov, I., Hautaniemi, S., and Janne, O.A. (2013). FoxA1 specifies unique androgen and glucocorticoid receptor binding events in prostate cancer cells. Cancer Res *73*, 1570-1580.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246-3248.

Sandri, M., Sandri, C., Gilbert, A., Skurk, C., Calabria, E., Picard, A., Walsh, K., Schiaffino, S., Lecker, S.H., and Goldberg, A.L. (2004). Foxo transcription factors induce the atrophy-related ubiquitin ligase atrogin-1 and cause skeletal muscle atrophy. Cell *117*, 399-412.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Sasse, S.K., and Gerber, A.N. (2015). Feed-forward transcriptional programming by nuclear receptors: regulatory principles and therapeutic implications. Pharmacol Ther *145*, 85-91.

Sassoon, D., Lyons, G., Wright, W.E., Lin, V., Lassar, A., Weintraub, H., and Buckingham, M. (1989). Expression of two myogenic regulatory factors myogenin and MyoD1 during mouse embryogenesis. Nature *341*, 303-307.

Sassoon, D.A. (1993). Myogenic regulatory factors: dissecting their role and regulation during vertebrate embryogenesis. Dev Biol *156*, 11-23.

Sato, T., Matsumoto, T., Yamada, T., Watanabe, T., Kawano, H., and Kato, S. (2003). Late onset of obesity in male androgen receptor-deficient (AR KO) mice. Biochem Biophys Res Commun *300*, 167-171.

Schaaf, M.J., Champagne, D., van Laanen, I.H., van Wijk, D.C., Meijer, A.H., Meijer, O.C., Spaink, H.P., and Richardson, M.K. (2008). Discovery of a functional glucocorticoid receptor beta-isoform in zebrafish. Endocrinology *149*, 1591-1599.

Schaaf, M.J.M. (2017). Nuclear receptor research in zebrafish. J Mol Endocrinol *59*, R65-R76.

Schacke, H., Berger, M., Rehwinkel, H., and Asadullah, K. (2007). Selective glucocorticoid receptor agonists (SEGRAs): novel ligands with an improved therapeutic index. Mol Cell Endocrinol *275*, 109-117.

Schakman, O., Gilson, H., de Coninck, V., Lause, P., Verniers, J., Havaux, X., Ketelslegers, J.M., and Thissen, J.P. (2005). Insulin-like growth factor-I gene transfer by electroporation prevents skeletal muscle atrophy in glucocorticoid-treated rats. Endocrinology *146*, 1789-1797.

Schakman, O., Gilson, H., and Thissen, J.P. (2008a). Mechanisms of glucocorticoid-induced myopathy. J Endocrinol *197*, 1-10.

Schakman, O., Kalista, S., Barbe, C., Loumaye, A., and Thissen, J.P. (2013). Glucocorticoid-induced skeletal muscle atrophy. Int J Biochem Cell Biol *45*, 2163-2172.

Schakman, O., Kalista, S., Bertrand, L., Lause, P., Verniers, J., Ketelslegers, J.M., and Thissen, J.P. (2008b). Role of Akt/GSK-3beta/beta-catenin transduction pathway in the muscle anti-atrophy action of insulin-like growth factor-I in glucocorticoid-treated rats. Endocrinology *149*, 3900-3908.

Scheschowitsch, K., Leite, J.A., and Assreuy, J. (2017). New Insights in Glucocorticoid Receptor Signaling-More Than Just a Ligand-Binding Receptor. Front Endocrinol (Lausanne) *8*, 16.

Schoch, G.A., D'Arcy, B., Stihle, M., Burger, D., Bar, D., Benz, J., Thoma, R., and Ruf, A. (2010). Molecular switch in the glucocorticoid receptor: active and passive antagonist conformations. J Mol Biol *395*, 568-577.

Schuler, M., Ali, F., Metzger, E., Chambon, P., and Metzger, D. (2005). Temporally controlled targeted somatic mutagenesis in skeletal muscles of the mouse. Genesis *41*, 165-170.

Schurch, N.J., Schofield, P., Gierlinski, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T*., et al.* (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA *22*, 839-851.

Seale, P., Sabourin, L.A., Girgis-Gabardo, A., Mansouri, A., Gruss, P., and Rudnicki, M.A. (2000). Pax7 is required for the specification of myogenic satellite cells. Cell *102*, 777-786.

Seene, T., and Viru, A. (1982). The catabolic effect of glucocorticoids on different types of skeletal muscle fibres and its dependence upon muscle activity and interaction with anabolic steroids. J Steroid Biochem *16*, 349-352.

Seto, E., and Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb Perspect Biol *6*, a018713.

Sever, R., and Glass, C.K. (2013). Signaling by nuclear receptors. Cold Spring Harb Perspect Biol *5*, a016709.

Shadrin, I.Y., Khodabukus, A., and Bursac, N. (2016). Striated muscle function, regeneration, and repair. Cell Mol Life Sci *73*, 4175-4202.

Shaffer, P.L., Jivan, A., Dollins, D.E., Claessens, F., and Gewirth, D.T. (2004). Structural basis of androgen receptor binding to selective androgen response elements. Proc Natl Acad Sci U S A *101*, 4758-4763.

Shah, O.J., Kimball, S.R., and Jefferson, L.S. (2000a). Acute attenuation of translation initiation and protein synthesis by glucocorticoids in skeletal muscle. Am J Physiol Endocrinol Metab *278*, E76-82.

Shah, O.J., Kimball, S.R., and Jefferson, L.S. (2000b). Among translational effectors, p70S6k is uniquely sensitive to inhibition by glucocorticoids. Biochem J *347*, 389-397.

Shang, Y., Myers, M., and Brown, M. (2002). Formation of the androgen receptor transcription complex. Mol Cell *9*, 601-610.

Sharifi-Zarchi, A., Gerovska, D., Adachi, K., Totonchi, M., Pezeshk, H., Taft, R.J., Scholer, H.R., Chitsaz, H., Sadeghi, M., Baharvand, H*., et al.* (2017). DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. BMC Genomics *18*, 964.

Sheffield-Moore, M., Urban, R.J., Wolf, S.E., Jiang, J., Catlin, D.H., Herndon, D.N., Wolfe, R.R., and Ferrando, A.A. (1999). Short-term oxandrolone administration stimulates net muscle protein synthesis in young men. J Clin Endocrinol Metab *84*, 2705-2711.

Shimizu, N., Maruyama, T., Yoshikawa, N., Matsumiya, R., Ma, Y., Ito, N., Tasaka, Y., Kuribara-Souta, A., Miyata, K., Oike, Y*., et al.* (2015). A muscle-liver-fat signalling axis is essential for central control of adaptive adipose remodelling. Nat Commun *6*, 6693.

Shin, Y.S., Fink, H., Khiroya, R., Ibebunjo, C., and Martyn, J. (2000). Prednisolone-induced muscle dysfunction is caused more by atrophy than by altered acetylcholine receptor expression. Anesth Analg *91*, 322-328.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272-286.

Showkat, M., Beigh, M.A., and Andrabi, K.I. (2014). mTOR Signaling in Protein Translation Regulation: Implications in Cancer Genesis and Therapeutic Interventions. Mol Biol Int *2014*, 686984.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet *38*, 1348-1354.

Sinha-Hikim, I., Artaza, J., Woodhouse, L., Gonzalez-Cadavid, N., Singh, A.B., Lee, M.I., Storer, T.W., Casaburi, R., Shen, R., and Bhasin, S. (2002). Testosterone-induced increase in muscle size in healthy young men is associated with muscle fiber hypertrophy. Am J Physiol Endocrinol Metab *283*, E154-164.

Sluder, A.E., Mathews, S.W., Hough, D., Yin, V.P., and Maina, C.V. (1999). The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. Genome Res *9*, 103-120.

Smith, E., and Shilatifard, A. (2010). The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. Mol Cell *40*, 689-701.

Smoak, K., and Cidlowski, J.A. (2006). Glucocorticoids regulate tristetraprolin synthesis and posttranscriptionally regulate tumor necrosis factor alpha inflammatory signaling. Mol Cell Biol *26*, 9126-9135.

Snel, B., Lehmann, G., Bork, P., and Huynen, M.A. (2000). STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res *28*, 3442-3444.

Song, I.H., and Buttgereit, F. (2006). Non-genomic glucocorticoid effects to provide the basis for new drug developments. Mol Cell Endocrinol *246*, 142-146.

Song, Y.H., Song, J.L., Delafontaine, P., and Godard, M.P. (2013). The therapeutic potential of IGF-I in skeletal muscle repair. Trends Endocrinol Metab *24*, 310-319.

Southgate, R.J., Neill, B., Prelovsek, O., El-Osta, A., Kamei, Y., Miura, S., Ezaki, O., McLoughlin, T.J., Zhang, W., Unterman, T.G.*, et al.* (2007). FOXO1 regulates the expression of 4E-BP1 and inhibits mTOR signaling in mammalian skeletal muscle. J Biol Chem *282*, 21176-21186.

Spicuglia, S., and Vanhille, L. (2012). Chromatin signatures of active enhancers. Nucleus *3*, 126-131.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet *13*, 613-626.

Srinivas-Shankar, U., and Wu, F.C. (2006). Drug insight: testosterone preparations. Nat Clin Pract Urol *3*, 653-665.

Stratton, M. (2008). Genome resequencing and genetic variation. Nat Biotechnol *26*, 65-66.

Sukari, A., Muqbil, I., Mohammad, R.M., Philip, P.A., and Azmi, A.S. (2016). F-BOX proteins in cancer cachexia and muscle wasting: Emerging regulators and therapeutic opportunities. Semin Cancer Biol *36*, 95-104.

Sun, L., Trausch-Azar, J.S., Muglia, L.J., and Schwartz, A.L. (2008). Glucocorticoids differentially regulate degradation of MyoD and Id1 by N-terminal ubiquitination to promote muscle protein catabolism. Proc Natl Acad Sci U S A *105*, 3339-3344.

Surjit, M., Ganti, K.P., Mukherji, A., Ye, T., Hua, G., Metzger, D., Li, M., and Chambon, P. (2011). Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor. Cell *145*, 224-241.

Suzuki, S., Murakami, Y., and Takahata, S. (2017). H3K36 methylation state and associated silencing mechanisms. Transcription *8*, 26-31.

Tan, M.H., Li, J., Xu, H.E., Melcher, K., and Yong, E.L. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. Acta Pharmacol Sin *36*, 3-23.

Taylor, W.E., Bhasin, S., Artaza, J., Byhower, F., Azam, M., Willard, D.H., Jr., Kull, F.C., Jr., and Gonzalez-Cadavid, N. (2001). Myostatin inhibits cell proliferation and protein synthesis in C2C12 muscle cells. Am J Physiol Endocrinol Metab *280*, E221-228.

te Pas, M.F., de Jong, P.R., and Verburg, F.J. (2000). Glucocorticoid inhibition of C2C12 proliferation rate and differentiation capacity in relation to mRNA levels of the MRF gene family. Mol Biol Rep *27*, 87-98.

Thomas-Chollier, M., Herrmann, C., Defrance, M., Sand, O., Thieffry, D., and van Helden, J. (2012). RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. Nucleic Acids Res *40*, e31.

Thomas, M., Langley, B., Berry, C., Sharma, M., Kirk, S., Bass, J., and Kambadur, R. (2000). Myostatin, a negative regulator of muscle growth, functions by inhibiting myoblast proliferation. J Biol Chem *275*, 40235-40243.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B.*, et al.* (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75-82.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell *10*, 1453-1465.

Tomas, F.M., Knowles, S.E., Owens, P.C., Chandler, C.S., Francis, G.L., Read, L.C., and Ballard, F.J. (1992). Insulin-like growth factor-I (IGF-I) and especially IGF-I variants are anabolic in dexamethasone-treated rats. Biochem J *282 (Pt 1)*, 91-97.

Tomas, F.M., Munro, H.N., and Young, V.R. (1979). Effect of glucocorticoid administration on the rate of muscle protein breakdown in vivo in rats, as measured by urinary excretion of N tau-methylhistidine. Biochem J *178*, 139-146.

Tomczak, A., Mortensen, J.M., Winnenburg, R., Liu, C., Alessi, D.T., Swamy, V., Vallania, F., Lofgren, S., Haynes, W., Shah, N.H.*, et al.* (2018). Interpretation of biological experiments changes with evolution of the Gene Ontology and its annotations. Sci Rep *8*, 5115.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Treuter, E., and Venteclef, N. (2011). Transcriptional control of metabolic and inflammatory pathways by nuclear receptor SUMOylation. Biochim Biophys Acta *1812*, 909-918.

Trivedi, U.H., Cezard, T., Bridgett, S., Montazam, A., Nichols, J., Blaxter, M., and Gharbi, K. (2014). Quality control of next-generation sequencing data without a reference. Front Genet *5*, 111.

Turner, J.D., and Muller, C.P. (2005). Structure of the glucocorticoid receptor (NR3C1) gene 5' untranslated region: identification, and tissue distribution of multiple new human exon 1. J Mol Endocrinol *35*, 283-292.

Ueda, T., Mawji, N.R., Bruchovsky, N., and Sadar, M.D. (2002). Ligand-independent activation of the androgen receptor by interleukin-6 and the role of steroid receptor coactivator-1 in prostate cancer cells. J Biol Chem *277*, 38087-38094.

Uhlenhaut, N.H., Barish, G.D., Yu, R.T., Downes, M., Karunasiri, M., Liddle, C., Schwalie, P., Hubner, N., and Evans, R.M. (2013). Insights into negative regulation by the glucocorticoid receptor from genome-wide profiling of inflammatory cistromes. Mol Cell *49*, 158-171.

Unsworth, A.J., Flora, G.D., and Gibbins, J.M. (2018). Non-genomic effects of nuclear receptors: insights from the anucleate platelet. Cardiovasc Res *114*, 645-655.

Urban, R.J., Bodenburg, Y.H., Gilkison, C., Foxworth, J., Coggan, A.R., Wolfe, R.R., and Ferrando, A. (1995). Testosterone administration to elderly men increases skeletal muscle strength and protein synthesis. Am J Physiol *269*, E820-826.

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K.*, et al.* (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res *18*, 1051-1063.

Van Balkom, R.H., Dekhuijzen, P.N., Folgering, H.T., Veerkamp, J.H., Van Moerkerk, H.T., Fransen, J.A., and Van Herwaarden, C.L. (1998). Anabolic steroids in part reverse glucocorticoid-induced alterations in rat diaphragm. J Appl Physiol (1985) *84*, 1492-1499.

van de Wijngaart, D.J., Dubbink, H.J., van Royen, M.E., Trapman, J., and Jenster, G. (2012). Androgen receptor coregulators: recruitment via the coactivator binding groove. Mol Cell Endocrinol *352*, 57-69.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. Trends Genet *30*, 418-426.

Van Nieuwerburgh, F., Thompson, R.C., Ledesma, J., Deforce, D., Gaasterland, T., Ordoukhanian, P., and Head, S.R. (2012). Illumina mate-paired DNA sequencing-library preparation using Cre-Lox recombination. Nucleic Acids Res *40*, e24.

Vastrik, I., D'Eustachio, P., Schmidt, E., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B., Lewis, S., Matthews, L*., et al.* (2007). Reactome: a knowledge base of biologic pathways and processes. Genome Biol *8*, R39.

Verdone, L., Caserta, M., and Di Mauro, E. (2005). Role of histone acetylation in the control of gene expression. Biochem Cell Biol *83*, 344-353.

Verrijdt, G., Haelens, A., and Claessens, F. (2003). Selective DNA recognition by the androgen receptor as a mechanism for hormone-specific regulation of gene expression. Mol Genet Metab *78*, 175-185.

Vilasco, M., Communal, L., Mourra, N., Courtin, A., Forgez, P., and Gompel, A. (2011). Glucocorticoid receptor and breast cancer. Breast Cancer Res Treat *130*, 1-10.

Wallberg, A.E., Wright, A., and Gustafsson, J.A. (2000). Chromatin-remodeling complexes involved in gene activation by the glucocorticoid receptor. Vitam Horm *60*, 75-122.

Waltering, K.K., Urbanucci, A., and Visakorpi, T. (2012). Androgen receptor (AR) aberrations in castration-resistant prostate cancer. Mol Cell Endocrinol *360*, 38-43.

Wang, C., Tian, L., Popov, V.M., and Pestell, R.G. (2011). Acetylation and nuclear receptor action. J Steroid Biochem Mol Biol *123*, 91-100.

Wang, H., Kubica, N., Ellisen, L.W., Jefferson, L.S., and Kimball, S.R. (2006). Dexamethasone represses signaling through the mammalian target of rapamycin in muscle cells by enhancing expression of REDD1. J Biol Chem *281*, 39128-39134.

Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res *41*, W77-83.

Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017). WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res *45*, W130-W137.

Wang, J.C., Derynck, M.K., Nonaka, D.F., Khodabakhsh, D.B., Haqq, C., and Yamamoto, K.R. (2004). Chromatin immunoprecipitation (ChIP) scanning identifies primary glucocorticoid receptor target genes. Proc Natl Acad Sci U S A *101*, 15603-15608.

Wang, L., Hsu, C.L., and Chang, C. (2005). Androgen receptor corepressors: an overview. Prostate *63*, 117-130.

Wang, Y., and Pessin, J.E. (2013). Mechanisms for fiber-type specificity of skeletal muscle atrophy. Curr Opin Clin Nutr Metab Care *16*, 243-250.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57-63.

Watson, L.C., Kuchenbecker, K.M., Schiller, B.J., Gross, J.D., Pufall, M.A., and Yamamoto, K.R. (2013). The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. Nat Struct Mol Biol *20*, 876-883.

Watson, M.L., Baehr, L.M., Reichardt, H.M., Tuckermann, J.P., Bodine, S.C., and Furlow, J.D. (2012a). A cell-autonomous role for the glucocorticoid receptor in skeletal muscle atrophy induced by systemic glucocorticoid exposure. Am J Physiol Endocrinol Metab *302*, E1210-1220.

Watson, P.J., Fairall, L., and Schwabe, J.W. (2012b). Nuclear hormone receptor co-repressors: structure and function. Mol Cell Endocrinol *348*, 440-449.

Wei, S., Li, C., Yin, Z., Wen, J., Meng, H., Xue, L., and Wang, J. (2018). Histone methylation in DNA repair and clinical practice: new findings during the past 5-years. J Cancer *9*, 2072-2081.

Weigel, N.L., and Zhang, Y. (1998). Ligand-independent activation of steroid hormone receptors. J Mol Med (Berl) *76*, 469-479.

Weikum, E.R., de Vera, I.M.S., Nwachukwu, J.C., Hudson, W.H., Nettles, K.W., Kojetin, D.J., and Ortlund, E.A. (2017). Tethering not required: the glucocorticoid receptor binds directly to activator protein-1 recognition motifs to repress inflammatory genes. Nucleic Acids Res *45*, 8596-8608.

Wetzel, J., Kingsford, C., and Pop, M. (2011). Assessing the benefits of using mate-pairs to resolve repeats in de novo short-read prokaryotic assemblies. BMC Bioinformatics *12*, 95.

Widom, J. (1997). Chromatin: the nucleosome unwrapped. Curr Biol *7*, R653-655.

Wilkinson, K.A., and Henley, J.M. (2010). Mechanisms, regulation and consequences of protein SUMOylation. Biochem J *428*, 133-145.

Wilson, E.M., and Rotwein, P. (2006). Control of MyoD function during initiation of muscle differentiation by an autocrine signaling pathway activated by insulin-like growth factor-II. J Biol Chem *281*, 29962-29971.

Wise, J.K., Hendler, R., and Felig, P. (1973). Influence of glucocorticoids on glucagon secretion and plasma amino acid concentrations in man. J Clin Invest *52*, 2774-2782.

Wittert, G.A., Chapman, I.M., Haren, M.T., Mackintosh, S., Coates, P., and Morley, J.E. (2003). Oral testosterone supplementation increases muscle and decreases fat mass in healthy elderly males with low-normal gonadal status. J Gerontol A Biol Sci Med Sci *58*, 618-625.

Wixon, J., and Kell, D. (2000). The Kyoto encyclopedia of genes and genomes--KEGG. Yeast *17*, 48-55.

Wu, Y., Zhao, W., Zhao, J., Pan, J., Wu, Q., Zhang, Y., Bauman, W.A., and Cardozo, C.P. (2007). Identification of androgen response elements in the insulin-like growth factor I upstream promoter. Endocrinology *148*, 2984-2993.

Wu, Y., Zhao, W., Zhao, J., Zhang, Y., Qin, W., Pan, J., Bauman, W.A., Blitzer, R.D., and Cardozo, C. (2010). REDD1 is a major target of testosterone action in preventing dexamethasone-induced muscle loss. Endocrinology *151*, 1050-1059.

Wurtz, J.M., Bourguet, W., Renaud, J.P., Vivat, V., Chambon, P., Moras, D., and Gronemeyer, H. (1996). A canonical structure for the ligand-binding domain of nuclear receptors. Nat Struct Biol *3*, 206.

Xavier, A.M., Anunciato, A.K., Rosenstock, T.R., and Glezer, I. (2016). Gene Expression Control by Glucocorticoid Receptors during Innate Immune Responses. Front Endocrinol (Lausanne) *7*, 31.

Xiao, X., Wang, P., and Chou, K.C. (2013). Recent progresses in identifying nuclear receptors and their families. Curr Top Med Chem *13*, 1192-1200.

Xie, N., Cheng, H., Lin, D., Liu, L., Yang, O., Jia, L., Fazli, L., Gleave, M.E., Wang, Y., Rennie, P.*, et al.* (2015). The expression of glucocorticoid receptor is negatively regulated by active androgen receptor signaling in prostate tumors. Int J Cancer *136*, E27-38.

Xu, H., Wei, C.L., Lin, F., and Sung, W.K. (2008). An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data. Bioinformatics *24*, 2344-2349.

Yablonka-Reuveni, Z. (2011). The skeletal muscle satellite cell: still young and fascinating at 50. J Histochem Cytochem *59*, 1041-1059.

Yaffe, D., and Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature *270*, 725-727.

Yang, H., Menconi, M.J., Wei, W., Petkova, V., and Hasselgren, P.O. (2005). Dexamethasone upregulates the expression of the nuclear cofactor p300 and its interaction with C/EBPbeta in cultured myotubes. J Cell Biochem *94*, 1058-1067.

Yang, H., Wei, W., Menconi, M., and Hasselgren, P.O. (2007). Dexamethasone-induced protein degradation in cultured myotubes is p300/HAT dependent. Am J Physiol Regul Integr Comp Physiol *292*, R337-334.

Yang, J., and Fuller, P.J. (2012). Interactions of the mineralocorticoid receptor--within and without. Mol Cell Endocrinol *350*, 196-205.

Yano, A., Fujii, Y., Iwai, A., Kawakami, S., Kageyama, Y., and Kihara, K. (2006). Glucocorticoids suppress tumor lymphangiogenesis of prostate cancer cells. Clin Cancer Res *12*, 6012-6017.

Yasir, M., and Sonthalia, S. (2019). Corticosteroid Adverse Effects. In StatPearls (Treasure Island (FL)).

Ye, F., McCoy, S.C., Ross, H.H., Bernardo, J.A., Beharry, A.W., Senf, S.M., Judge, A.R., Beck, D.T., Conover, C.F., Cannady, D.F.*, et al.* (2014). Transcriptional regulation of myotrophic actions by testosterone and trenbolone on androgen-responsive muscle. Steroids *87*, 59-66.

Ye, T., Krebs, A.R., Choukrallah, M.A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res *39*, e35.

Yeh, S., Tsai, M.Y., Xu, Q., Mu, X.M., Lardy, H., Huang, K.E., Lin, H., Yeh, S.D., Altuwaijri, S., Zhou, X.*, et al.* (2002). Generation and characterization of androgen receptor knockout (ARKO) mice: an in vivo model for the study of androgen functions in selective tissues. Proc Natl Acad Sci U S A *99*, 13498-13503.

Yemelyanov, A., Czwornog, J., Chebotaev, D., Karseladze, A., Kulevitch, E., Yang, X., and Budunova, I. (2007). Tumor suppressor activity of glucocorticoid receptor in the prostate. Oncogene *26*, 1885-1896.

Yin, D., Gao, W., Kearbey, J.D., Xu, H., Chung, K., He, Y., Marhefka, C.A., Veverka, K.A., Miller, D.D., and Dalton, J.T. (2003). Pharmacodynamics of selective androgen receptor modulators. J Pharmacol Exp Ther *304*, 1334-1340.

Yin, H., Price, F., and Rudnicki, M.A. (2013). Satellite cells and the muscle stem cell niche. Physiol Rev *93*, 23-67.

Yoon, M.S. (2017). mTOR as a Key Regulator in Maintaining Skeletal Muscle Mass. Front Physiol *8*, 788.

Yuan, X., and Balk, S.P. (2009). Mechanisms mediating androgen receptor reactivation after castration. Urol Oncol *27*, 36-41.

Zamir, O., Hasselgren, P.O., von Allmen, D., and Fischer, J.E. (1991). The effect of interleukin-1 alpha and the glucocorticoid receptor blocker RU 38486 on total and myofibrillar protein breakdown in skeletal muscle. J Surg Res *50*, 579-583.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952-1958.

Zanou, N., and Gailly, P. (2013). Skeletal muscle hypertrophy and regeneration: interplay between the myogenic regulatory factors (MRFs) and insulin-like growth factors (IGFs) pathways. Cell Mol Life Sci *70*, 4117-4130.

Zeitlinger, J., Simon, I., Harbison, C.T., Hannett, N.M., Volkert, T.L., Fink, G.R., and Young, R.A. (2003). Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. Cell *113*, 395-404.

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res *21*, 1273-1283.

Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res *33*, W741-748.

Zhang, L., Martini, G.D., Rube, H.T., Kribelbauer, J.F., Rastogi, C., FitzPatrick, V.D., Houtman, J.C., Bussemaker, H.J., and Pufall, M.A. (2018). SelexGLM differentiates androgen and glucocorticoid receptor DNA-binding preference over an extended binding site. Genome Res *28*, 111-121.

Zhang, Y., Shan, C.M., Wang, J., Bao, K., Tong, L., and Jia, S. (2017). Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. Sci Rep *7*, 43906.

Zhao, W., Pan, J., Zhao, Z., Wu, Y., Bauman, W.A., and Cardozo, C.P. (2008). Testosterone protects against dexamethasone-induced muscle atrophy, protein degradation and MAFbx upregulation. J Steroid Biochem Mol Biol *110*, 125-129.

Zhao, X.Y., Malloy, P.J., Krishnan, A.V., Swami, S., Navone, N.M., Peehl, D.M., and Feldman, D. (2000). Glucocorticoids can promote androgen-independent growth of prostate cancer cells through a mutated androgen receptor. Nat Med *6*, 703-706.

Zhao, Y., and Garcia, B.A. (2015). Comprehensive Catalog of Currently Documented Histone Modifications. Cold Spring Harb Perspect Biol *7*, a025064.

Results

PART I

Objectives of the part I of the thesis

Skeletal muscle is a dynamic tissue that has the capacity to modulate its size and mass through a balance between anabolic and catabolic pathways in response to external cues like nutrition and hormones. Glucocorticoids are hormones that regulate metabolism, circadian rhythm, immune functions and stress response. They exert their biological effects predominantly via glucocorticoid receptor (GR), a member of the nuclear receptor superfamily. GR is expressed in many cell types and regulates distinct set of genes in various tissues. In skeletal muscle, glucocorticoids regulate glucose, lipid and protein metabolism and they contribute to energy homeostasis.

The objective of the first part of the thesis is to provide insights into the regulation of anabolic and catabolic pathways in skeletal muscles controlled by myofiber GR at physiological glucocorticoid levels using a combination of transcriptomic and cistromic analyses.

To identify GR-regulated genes in skeletal myofibers, we performed global transcriptome analysis in control and mutant mice and we identified differentially expressed genes and enriched pathways. Moreover, to delineate GR target genes in skeletal muscle, we determined genome-wide GR chromatin occupancy under physiological glucocorticoid levels by chromatin immunoprecipitation, followed by massive parallel sequencing (ChIP-seq) and identified enriched binding sites and DNA response elements. Additionally, to characterize the chromatin landscape of GR binding sites, we performed ChIP-seq analyses for various histone marks. Lastly, to provide insights into the molecular determinants of GR transcriptional regulation in skeletal muscle, we performed a motif search on ChIP-seq peaks containing Glucocorticoid Response Elements, using collections of known motifs and we determined potential co-regulators of GR.

These results are part of a manuscript in preparation.

Available GR ChIP-Seq data in muscle tissue

Species	Biological source	Publication
Homo sapiens	airway smooth muscle 2	Sasse SK, et al. Am. J. Respir. Cell Mol. Biol. 2017

Myofiber glucocorticoid receptor coordinates the down regulation of anabolic pathways in skeletal muscles at physiological glucocorticoid levels

Vanessa Ueberschlag-Pitiot[1-4*], Daniela Rovito[1-4*], Anna Isavella Rerra[1-4*], Shilpy Joshi[1-4], Vanessa Dacleu-Siewe[2-4,7], Maxime Parisotto, Jean-Marc Bornert[1-4], Isabelle Hazemann[2-4,7], Arnaud Ferry[5-6], Bruno P. Klaholz[2-4,7], Isabelle M.L. Billas[2-4,7], Gilles Laverny[1-4], Delphine Duteil[1-4#], Daniel Metzger[1-4#]

[1]Department of Functional genomics and cancer, IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), Illkirch, 67404, France

[2]Inserm, U1258, Illkirch, 67404, France

[3]CNRS, UMR7104, Illkirch, 67404, France

[4]Université de Strasbourg, Strasbourg, 67000, France

[5]Institut National de la Santé et de la Recherche Médicale U974, Centre National de la Recherche Scientifique Unité Mixte de Recherche 7215, Université Pierre et Marie Curie-Paris 6, 75013 Paris, France

[6]Université Paris Descartes, 75006 Paris, France

[7]Centre for Integrative Biology (CBI), Department of Integrated Structural Biology, IGBMC (Institute of Genetics and of Molecular and Cellular Biology), Illkirch, France

* Equal participation
# Correspondence: duteild@igbmc.fr, metzger@igbmc.fr

Corresponding Author**:**

Daniel Metzger: metzger@igbmc.fr.

Abstract

Skeletal muscle is essential for posture and locomotion, and is a major organ for nutrient storage and supply. It has the capacity to modulate its size, in response to various stimuli, including glucocorticoids, the effects of which are mediated by the ubiquitously expressed glucocorticoid receptor (GR). As molecular determinants of GR-mediated transcriptional regulation in skeletal muscle remain elusive, we generated GR$^{(i)skm-/-}$ mice in which GR is selectively ablated in myofibers at adulthood. GR loss in skeletal muscle did not affect catabolic pathways, but enhanced the expression of anabolic factors and reduced that of anti-anabolic ones. As a consequence, muscle fiber size, mass and strength were increased in GR$^{(i)skm-/-}$ mice. Genome-wide GR chromatin occupancy in skeletal muscles identified 23196 GR binding sites, associated to 11302 genes. Our data show that myofiber GR mainly binds to glucocorticoid response elements (GREs) located at active enhancers enriched in H3K27ac, H3K4me1 and devoid of H3K4me3. Detailed characterization of the genes encoding the anti-anabolic factors *Eif4ebp2* and *Pik3r1* (also known as *p85a*) revealed that GRE-bound GR cooperates with Myod1 and the chromatin-associated factor Foxf2 at enhancers, and interacts with Nrf1 bound at promoter regions, to stimulate gene transcription.

Thus, physiological glucocorticoid levels have a negative impact on muscle mass in adult mice, by coordinating the down-regulation of anabolic pathways. Importantly, the cooperation between GR and transcription factors such as Myod1, Foxf2 and Nrf1 plays a key role in glucocorticoid-induced myofiber-specific gene regulation.

Introduction

Skeletal muscle accounts for about half of the body mass and is essential for posture, locomotion, and energy balance in mammals (Hawley et al., 2018). In response to a variety of external cues, including mechanical load, nutritional status and hormones, this dynamic tissue has the capacity to modulate its size(Lecker et al., 2004) via a balance between anabolic and catabolic pathways(Hoffman and Nader, 2004).

Glucocorticoids, such as cortisol and corticosterone, are cholesterol-derived steroid hormones that are essential regulators of energy homeostasis in various tissues in mammals, including skeletal muscle(Tanaka et al., 2017). Synthetic glucocorticoid analogues are among the most worldwide prescribed drugs because of their anti-inflammatory and immunosuppressive properties(Ito et al., 2006). Despite the induction of various side effects, including diabetes, osteoporosis and muscle atrophy(Braun et al., 2011), they remain the main treatment of various diseases including rheumatoid arthritis and asthma(Rosen and Miner, 2005). Both natural and synthetic glucocorticoids exert their biological effects predominantly via the glucocorticoid receptor (GR, Nr3c1), a member of the nuclear receptor superfamily(Kino, 2000), that acts as a ligand-dependent transcription factor(Meijsing, 2015). Upon ligand binding, GR is translocated to the nucleus to activate or repress gene expression in a cell-type-specific manner. GR positive regulation is mediated by its recruitment to specific DNA segments termed glucocorticoid response elements (GRE). The consensus GRE, 5'-AGAACAnnnTGTTCT-3', is an inverted palindrome separated by 3 base pairs(Lieberman et al., 1993; Presman et al., 2014), on which two GR molecules bind as a homodimer. Recent studies indicate that sequence variation in GREs, including the 3 non-specific spacer, and in the flanking nucleotides, influences the 3-dimensional structure of the GR DNA binding domain (DBD) and modulates GR transcriptional activity(Meijsing et al., 2009; Watson et al., 2013). Negative influence on gene transcription is achieved by GR binding to negative GRE(Hua et al., 2016b; Surjit et al., 2011) or by interacting with DNA-bound transcription factors, such as AP-1 or NF-κB(Meijsing, 2015; Tan and Wahli, 2016). There has been a widespread view for many years that the beneficial anti-inflammatory effects result from GR-mediated transrepression(Uhlenhaut et al., 2013), whilst the adverse side-effects of prolonged glucocorticoid treatment result from GR-mediated gene activation and/or direct repression(Beck et al., 2009; Surjit et al., 2011). GR is expressed in many cell types in mammals, and despite intense efforts, the molecular mechanisms underlying glucocorticoid-dependent cell-specific transcriptional regulation remain unclear.

In skeletal muscle, pharmacological glucocorticoid levels affect glucose and protein metabolism by activating proteasome and autophagy systems, as well as the anti-anabolic factors Ddit4 (also known as Redd1), an inhibitor of mTOR activity(DeYoung et al., 2008; Wang et al., 2006), and Pik3r1(Kuo et al., 2012) (also known as p85α), a regulatory subunit of Pi3k, thereby limiting protein synthesis(Hu et al., 2009b; Shimizu et al., 2011; Watson et al., 2012a). Moreover, a recent study showed that loss of GR in developing mouse skeletal muscle leads to increased muscle mass with reduced adipose tissue, accompanied by major modifications of the transcriptional repertoire of

muscle, liver and fat depots(Shimizu et al., 2015). Nevertheless, the molecular determinants controlling increased muscle mass were not investigated.

To delineate the physiological and molecular function of GR in mature skeletal muscles, we generated GR[(i)skm-/-] mice in which GR is selectively ablated in myofibers at adulthood. Our results show that the fiber size, mass and strength of skeletal muscles are increased upon GR loss. Combination of transcriptome and cistrome analyses revealed that physiological glucocorticoids coordinate the down-regulation of anabolic pathways. In addition, we provide mechanistic insights into glucocorticoid-regulated gene expression in skeletal myofibers.

Results

Myofiber GR down-regulates muscle mass and strength

To determine the role of GR in mature skeletal muscles, we generated GR[(i)skm-/-] mice in which GR is selectively ablated in myofibers at adulthood (Supplementary Fig. 1a-c). The body weight of GR[(i)skm-/-] mice was increased by 5 to 13.5 % between 10 and 30 weeks of age (Supplementary Fig. 1d). Moreover, body mass repartition of 30 week-old mice, determined by quantitative nuclear magnetic resonance (qNMR), revealed a 14 % increase in lean mass, but no difference in fat content (Fig. 1a). In accordance, loss of GR was associated with increased mass of gastrocnemius, tibialis and quadriceps limb muscles (Fig. 1b and Supplementary Fig. 1e, f). Histological analyses revealed that the number of muscle fibers in these muscles was similar in 4 month-old control and GR[(i)skm-/-] mice (Fig. 1c, d and Supplementary Fig. 1g, h). However, whereas fiber cross sectional area (CSA) distribution was centred around 2000 μm² in gastrocnemius muscle of control mice, it was shifted to 2500 μm² in that of GR[(i)skm-/-] mice (Fig. 1e), resulting in an increased average fiber CSA (Fig. 1f). Similar shifts in CSA occurred in tibialis and quadriceps muscles (Supplementary Fig. 1i-l). Limb muscle strength assessed by grip test was increased by 6 and 10 % in GR[(i)skm-/-] mice at 4 and 5 months of age, respectively (Fig. 1g). Moreover, at 4 months of age, tibialis maximal tetanic force was 24 % higher in GR[(i)skm-/-] mice than in control mice (Fig. 1h), whereas its specific force was similar in control and GR[(i)skm-/-] mice (Fig. 1i), showing that increased muscle strength results from increased muscle mass. Together, our results show that physiological glucocorticoid levels negatively regulate hindlimb muscle mass and strength in adult mice by restricting fiber size via myofiber GR.

Physiological glucocorticoid levels down-regulate the anabolic pathway via myofiber GR

To identify GR-regulated genes in gastrocnemius muscle, we performed global transcriptome analysis in control and GR[(i)skm-/-] mice, one week after GR ablation. We found 1335 differentially expressed genes, of which 677 were up- and 658 were down-regulated (Fig. 2a). Pathway analysis revealed that down-regulated genes were related to muscle metabolism, and in particular genes encoding enzymes involved in glycogen metabolism (Fig. 2b). RT-qPCR experiments showed that loss of GR led to a 50 % decrease in transcripts of enzymes promoting glycogen synthesis, such as Ugp2 and Gyg, as well as in those of enzymes involved in glycogen catabolism, such as Phka1, Gbe1 and Agl (Supplementary Fig. 2a, b) in gastrocnemius muscle. Glycogen content determined by

Periodic acid–Schiff (PAS) staining was however not significantly altered in GR-depleted gastrocnemius muscles (Fig. 2c), and basal blood glucose levels (Supplementary Fig. 2c) and glucose uptake (Supplementary Fig. 2d) were similar in control and GR$^{(i)skm-/-}$ mice. Altogether, our data show that, even though several genes encoding enzymes involved in glycogen metabolism were dysregulated by the loss of GR in myofibers, serum glucose and muscle glycogen levels were not affected.

Additional enriched pathways of down-regulated genes unravelled the terms translation factors and insulin signaling (Fig. 2b). In particular, we found that transcripts encoding two known GR targets, *Pik3r1 and Ddit4*(Kuo et al., 2012; Shimizu et al., 2011), were significantly decreased in the absence of GR (Fig. 2e, f and Supplementary Fig. 2e). In addition, transcripts encoding the translation inhibitors *Eif2ak1*, *Eif4ebp1* and *Eif4ebp2* were less expressed in GR$^{(i)skm-/-}$ mice (Fig. 2e, f and Supplementary Fig. 2e). In line with these data, up-regulated transcripts in GR-deficient gastrocnemius revealed pathways related to cytoplasmic ribosomal protein and mRNA processing (Fig. 2d), and included the anabolic factors *Akt3*, *Rps6*, and *Pik3ca* (Fig. 2e, f and Supplementary Fig. 2e). In accordance with these data, protein levels of Akt3 were increased in gastrocnemius muscle of GR$^{(i)skm-/-}$ mice, whereas those of Pik3r1 and Ddit4 were strongly decreased (Fig. 2g, h). Moreover, phosphorylation levels of mTOR at Ser2448 were higher in the absence of GR (Fig. 2g, i) indicative of increased mTOR activity. In addition, protein levels of the translation inhibitors Eif4ebp1 and Eif4ebp2 were much lower in GR$^{(i)skm-/-}$ mice than in control mice (Fig. 2g, h), whereas their phosphorylation at threonine residues located at positions 37 and 46 (Thr37/46), corresponding to their inactive form, was increased (Fig. 2g, i). Note that transcript and/or protein levels of Igf1, Akt1, Raptor, and Rictor were similar in gastrocnemius muscles of GR$^{(i)skm-/-}$ and control mice (Supplementary Fig. 2f-h), and that levels of Akt1/2/3 phosphorylated at serine 473 (Ser473) or threonine 308 (Thr308) were not affected (Supplementary Fig. 2g, i). Of note, our transcriptomic data did not reveal differences in mRNA levels of genes involved in muscle catabolism upon GR loss. In agreement with these data, transcript levels of *myostatin (Mstn)* and of genes involved in ubiquitin proteasome system (e.g. *Ubc, Fbxo32* also known as *Mafbx* or *Atrogin-1*, *Trim63* also known as *Murf1*), calpain pathway (*Capn1 and* 2) or autophagy program (*Atg3, Atg5, Atg12, Gabaralp1*, *Bnip3*, *Map1lc3a*, *Ctsl*, and *Becn1*) were similar in control and GR$^{(i)skm-/-}$ mice (Supplementary Fig. 2f). Moreover, protein levels of Foxo1 and Foxo3a, two key regulators of muscle proteolytic pathways, were not altered in GR-depleted myofibers, and even if Ser256-phosphorylated Foxo1 levels (inactive form) were decreased, those of Foxo3a were similar in control and GR$^{(i)skm-/-}$ mice (Supplementary Fig. 2g, i).

Thus, our results demonstrate that physiological glucocorticoids, via myofiber GR, reduce the expression of several anabolic factors and induce that of anti-anabolic factors, leading to a decreased anabolic pathway, thereby limiting muscle fiber size and mass, without stimulating catabolic pathways (Supplementary Fig. 2e).

GR is located at active enhancers of genes encoding anti-anabolic factors in skeletal muscle in the presence of physiological glucocorticoid levels

To delineate GR target genes in skeletal muscle, we determined genome-wide GR chromatin occupancy under physiological glucocorticoid levels in limb muscles by chromatin immunoprecipitation, followed by massive parallel sequencing (ChIP-seq). We identified 23196 high-confidence peaks located in 11302 genes, including those encoding the anti-anabolic factors *Ddit4*, *Eif2ak1*, *Eif4ebp1*, *Eif4ebp2* and *Pik3r1*, and the anabolic factors *Rps6*, *Rps6ka2*, *Rps6kc1*, *Akt3*, *Pik3c2a*, *Eif2a*, *Eif2d* and *Pabpc1* (see below). They were distributed along the genome, with 37.5 % in the transcription start site (TSS) region (-1000 bp; +100 bp), 31.4 % in introns and 24.8 % in intergenic locations (Fig. 3a). *De novo* motif search using hypergeometric optimization of motif enrichment (HOMER) analysis revealed that GR binds to 5'-AGRACAraaTGTTCY-3' and 5'-NGNRCAnnnTGTNCT-3' GREs at intergenic and intronic regions, respectively (Fig. 3b). These results were confirmed by Motif-based sequence analysis tools (MEME suite) and Regulatory Sequence Analysis Tools (RSAT) analyses (Supplementary Fig. 3a). However, no Nf-kb and AP-1 binding sites were identified in these regions. Of note, one third of genes up-regulated in GR[(i)skm-/-] mice presented a GR binding site. In particular, GR was bound at the promoter region of anabolic factors up-regulated in GR[(i)skm-/-] mice (Supplementary Fig. 3b). Genomatix analysis of these binding sites revealed shared motifs corresponding to cAMP-responsive element binding proteins (Atf and Creb factors) and AP-1 related factors (Bach, Maf, Nfe2), as well as ETS1 factors (Elf, Elk, Etf, Etv factors), Krueppel like transcription factors (Klfs) and MAF.

To characterize the genomic landscape of GR binding sites, we performed a ChIP-seq analysis for various histone marks. We found 21377 peaks for histone H3 acetylated at lysine 27 (H3K27ac, a mark of active promoter and enhancer regions(Creyghton et al., 2010)), 75523 for H3 monomethylated at lysine 4 (H3K4me1, a chromatin hallmark of enhancers(Sharifi-Zarchi et al., 2017)), 19818 for trimethylated H3K4 (H3K4me3, a mark enriched at promoter regions(Sharifi-Zarchi et al., 2017)) and 13053 peaks for the polymerase 2 (Pol2) (Supplementary Fig. 3c). SeqMINER-generated heatmaps revealed two GR peak clusters, one at active promoters (11038 peaks) defined by the presence of H3K27ac, H3K4me3 and Pol2, and low H3K4me1 levels, one at active enhancers (12158 peaks) defined by the presence of H3K27ac, H3K4me1 and Pol2, and low H3K4me3 levels (Fig. 3c, d).

Interestingly, more than 90 % of the genes on which GR is recruited in wild-type mice and are down-regulated in GR[(i)skm-/-] mice show peaks for H3K27ac and H3K4me1 (Fig. 4a), and the vast majority (~90 %) of such genes were bound by H3K4me3 and Pol2 at their promoter (Fig. 4b). Of note the anti-anabolic factors down-regulated in GR[(i)skm-/-] mice (i.e. *Ddit4*, *Eif2ak1*, *Eif4ebp1*, *Eif4ebp2* and *Pik3r1*) were included in these 90 %. Their loci are depicted in Fig. 4c and Supplementary Fig. 4a. *De novo* motif search revealed that GR binds to 5'-RGNACAnnnTGTNCY-3' GRE motifs at these glucocorticoid-induced genes, indicating that physiological glucocorticoid levels stimulate their transcription in myofibers mainly via GR bound to enhancer GREs.

To further investigate the molecular basis of GR-mediated gene activation, we focused on representative enhancer and promoter regions of the translation repressor *Eif4ebp2*, on which we identified 5 MACS peaks, 3 located at an upstream intergenic region (MACS peaks 1885, 1884 and 1883, located at -21.1, -15.5, and -12.5 kbp from the TSS, respectively) and 2 located at the promoter region (MACS peaks 1882 and 1881, -390 bp to +9 bp and +137 bp to +282 bp from the TSS, respectively) (Fig. 4c). *De novo* motif search identified putative GREs under the enhancer-located MACS peaks 1884 (GRE1, 5'-AGAACActcAGTCCT-3', -15,500 bp to -15,485 bp) and 1883 (GRE2, 5'-GGTACAcagAGTGCC-3', -12,595 bp to -12,580 bp), but not at MACS peak 1885, nor at promoter-located MACS peaks 1881 and 1882. ChIP followed by qPCR (ChIP-qPCR) analysis of additional limb muscles confirmed GR binding at both enhancer and promoter regions for control mice, whereas no specific amplification was detected for GR[(i)skm-/-] mice (Fig. 4d), demonstrating that GR recruitment to these loci is myofiber-specific. As expected, ChIP-seq data revealed that the genomic regions encompassing the enhancer GR peaks were enriched in H3K4me1 and devoid of H3K4me3 (Fig. 4c), whereas the promoter region encompassing GR peaks had an opposite profile (Fig. 4c). These data were confirmed by ChIP-qPCR analysis (Supplementary Fig. 4b). Of note, the DNA segments with the promoter mark H3K9ac correlated with those with H3K4me3, and histone H3 levels were similar at all investigated DNA segments (Supplementary Fig. 4b). Thus, these data show that GR peaks 1883 to 1885 are located at an active enhancer and that GR peaks 1881 and 1882 are at an active promoter region.

De novo motif search of the region located 250 kb upstream of the *Pik3r1* TSS identified a close-to-consensus GRE (5'-AGAACAtcgTGTTCC-3') under the GR MACS peak 6346, located at the enhancer, whereas no GRE was present at the promoter region (MACS peaks 6344-6345) (Supplementary Fig. 4a and Supplementary Fig. 6c). Combined ChIP-seq and ChIP-qPCR analyses demonstrated that GR was bound at both active promoter and enhancer regions of *Pik3r1*, specifically in myofibers (Supplementary Fig. 4a, c, d). Thus, myofiber GR-recruited DNA elements are located at active enhancers and promoter elements to promote the expression of anti-anabolic factors.

GR binds as a homodimer at enhancer GREs of the anti-anabolic factor Eif4ebp2

Since GRE1 (5'-AGAACActcAGTCCT-3') and GRE2 (5'-GGTACAcagAGTGCC-3') of the *Eif4ebp2* locus differed from the consensus GRE (5'-AGAACAnnnTGTTCT-3') by 2 and 5 nucleotides, respectively (Fig. 5a), we determined GR binding to such elements by native polyacrylamide gel analysis using the human GR DNA binding domain (DBD) purified from *E. coli*. GR DBD migration profiles in the presence of GRE1, GRE2 and the canonical *GILZ* GRE (5'-AGAACAttgGGTTCC-3')(Meijsing et al., 2009) were similar (Fig. 5b). In contrast, only few complexes, with different migration profiles, were formed when one base pair in each of the two GRE1 and GRE2 half-sites was mutated (GRE1 mut and GRE2 mut) or in the presence of unrelated probes, located either in the vicinity of GRE2 (non-specific probe, NSP) or corresponding to a response element of the oestrogen-related receptor (ERRE, 5'-TGAAGGTCA-3') (Fig. 5b).

The stoichiometry of complexes formed between purified GR DBD (11 kDa as a monomer) and various DNA probes (16 kDa), present in excess, was determined by size exclusion chromatography-multi angle laser light scattering (SEC-MALLS) analysis. The elution profile of the GR DBD in presence of GRE1, GRE2 or a consensus GRE revealed two peaks, with relative masses of 36 kDa and 16 kDa, indicative of two GR DBD monomers bound to one GRE, and free DNA, respectively, showing that GR DBD is recruited as a dimer to such elements (Fig. 5c and Supplementary Fig. 5a). In contrast, mutations of the half-sites resulted in an asymmetric SEC peak, and the measured molecular mass (30 kDa) indicates a mixture of monomers (26 kDa) and dimers (36 kDa) on DNA, as well as free DNA (16 kDa) (Fig. 5c and Supplementary Fig. 5a). In contrast, in the presence of a non-specific probe in the vicinity of the GRE, an average mass of 22 kDa was observed, which might correspond to unbound dimers of GR DBD, a monomer on DNA or free DNA. Of note, GR DBD in the presence of the ERRE probe resulted in free DNA and unbound monomers (16 kDa in average), showing that no complex was formed (Fig. 5c and Supplementary Fig. 5a). Microscale thermophoresis (MST) analyses revealed that the dissociation constant (Kd) for GRE1 and GRE2 (195 nM and 204 nM, respectively) was similar to that of the consensus GRE (154 nM) (Fig. 5d), whereas Kd values of non-specific probes were at least 30 times above these values. Thus, GR DBD binds as a homodimer to the two identified *Eif4ebp2* GREs at high affinity.

To determine whether full length GR protein also binds to GRE1, we transfected monkey kidney COS-1 cells with expression vectors encoding either mouse or human GR, or an empty vector, and performed an electrophoretic mobility shift assay (EMSA) with a consensus GRE (Fig. 5e, lanes 1 to 4, and Supplementary Fig. 5b, lanes 1 to 6), GRE1 (Fig. 5e, lanes 5 to 8, and Supplementary Fig. 5b, lanes 7 to 12), or their mutated versions (GRE mut and GRE1 mut, Fig. 5e, lanes 9 to 15, and Supplementary Fig. 5b, lanes 13 to 24) as probes. Both mouse (Fig. 5e) and human (Supplementary Fig. 5b) full length GR proteins bound to the consensus GRE and to GRE1, whereas no binding was observed when half-sites were mutated, or when cells expressed a mutated human GR DNA binding domain (Supplementary Fig. 5b).

Altogether, our data show that GR binds as a homodimer with a high affinity to non-consensus GREs of distinct sequences located at the enhancer of the anti-anabolic factor *Eif4ebp2*.

Myod1 and Foxf2 are bound to their cognate response elements located in the vicinity of enhancer-GRE containing regions in myofibers

To provide insights into the molecular determinants of GR positive transcriptional regulation in skeletal muscle, we performed HOMER known motif search at GRE-containing enhancers (+/- 100 bp from centre of peak). We found Myod1 E-boxes (5'-CAGCTG-3') as the most enriched motif (intergenic p=1e-257, intron p=1e-236), as well as binding site of Ctcf (5'-GCCCTCTTCTGG-3', intergenic p=1e-147, intron p=1e-74) and members of the Signal transducer and activator of transcription (e.g. Stat1, Stat3, 5'-yTTCCa/tGGAAr-3', intergenic p=1e-78, intron p=1e-64) (Supplementary Fig. 6a). The overlap of genes bound by GR in mouse skeletal muscle and those bound by Myod1 in C2C12 myotubes from three independent data sets(Mousavi et al., 2013;

Mullen et al., 2011; Umansky et al., 2015) revealed that Myod1 is recruited at half of GR target genes (Supplementary Fig. 6b). In addition, half of the genes down-regulated in GR[(i)skm-/-] mice were bound by GR and Myod1. SeqMINER analysis showed the co-occurrence of GR and Myod1 at 3139 sites out of 23196 GR peaks (15 %) (Fig. 6a), including several within *Eif4ebp2*, *Pik3r1*, *Ddit4*, *Eif4ebp1* and *Eif2ak1* loci (Fig. 6b and Supplementary Fig. 6c). At the *Eif4ebp2* locus, we found 7 E-boxes (-21.5 to -12.4 kb from the TSS), 4 Ctcf binding sites (-19 kb, -13.9 kb, -7.3 kb, and -0.87 kb from the TSS), but no Stat DNA binding sequences (Fig. 6c). To determine whether Myod1 and Ctcf are recruited at the predicted sites, we performed chromatin immunoprecipitation with antibodies directed against these factors (Fig. 6d, e). Myod1 was recruited at the DNA segments encompassing the E-boxes (primer pairs #1, 2, 4, 7, 9, 11, 13, Fig. 6c, d) located in the enhancer region, but not at an unrelated region located within the promoter region (primer pair #16, Fig. 6c, d). In addition, we confirmed Ctcf recruitment at predicted sites (primer pairs # 6, 10, 15, 17, Fig. 6e). Myod1 and Ctcf binding sites were also found at the enhancer region of the *Pik3r1* locus (Supplementary Fig. 6d-f). Together, these data show that Myod1 and Ctcf are recruited at genomic regions located in the vicinity of GR response elements.

To determine GR interacting partners, gastrocnemius muscle nuclear extracts were immunoprecipitated with a GR antibody directed against its C-terminal domain, followed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Immunoprecipitated gastrocnemius nuclear extracts from wild-type mice with a rIgG or from GR[(i)skm-/-] mice with an anti-GR antibody were used as negative controls. The intersection of the interactomes obtained with anti-GR immunoprecipitated samples from wild-type mice and negative controls uncovered 360 GR partners (Fig. 6f), of which 204 were nuclear proteins and 41 chromatin-associated factors (Table S1), including previously identified GR partners (e.g. Stat1 and Stat3)(Aittomaki et al., 2000; Langlais et al., 2012; Zhang et al., 1997), as well as additional interacting proteins, such as Foxf2, a member of Forkhead box (FOX) transcription factors (Aitola et al., 2000). The interaction of GR with Foxf2 and Stat3 was confirmed by co-immunoprecipitation of muscle nuclear extracts (Fig. 6g). Interestingly, four Foxf2 putative binding sites (5'-c/gg/aTAAACA-3'; Jaspar database) surround the *Eif4ebp2* enhancer localized GR peaks (Fig. 6c), and chromatin immunoprecipitation with antibodies directed against Foxf2 revealed that this factor was bound to these response elements (primer pairs # 1, 5, 9, 14, Fig. 6h), but not at unrelated sequences (as exemplified by the E-box #2, Fig. 6h). Foxf2 was also recruited at two sites located in the *Pik3r1* enhancer region (Supplementary Fig. 6g). Thus, these results indicate that myofiber GR might cooperates with MyoD and Foxf2 to enhance glucocorticoid target gene expression.

Myod1 and Foxf2 enhance GR binding to *Eif4ebp2* and *Pik3r1* GREs

To investigate the possible interplay between GR, Myod1 and Foxf2 on *Eif4ebp2* and *Pik3r1* gene regulation, we performed small interfering RNA (siRNA)-mediated knock-down of each of these factors in C2C12 myotubes (Supplementary Fig. 7a). GR silencing led to a 50 % decrease in *Eif4ebp2* and *Pik3r1* mRNA levels (Supplementary Fig. 7b), in agreement with data obtained after GR ablation in mouse skeletal muscles (Fig. 2f).

ChIP-qPCR analysis of C2C12 myotubes transfected with a scrambled siRNA showed GR binding at DNA sequences corresponding to the regions encompassing GR MACS peaks 1884 at *Eif4ebp2* locus and 6346 at *Pik3r1* locus identified in skeletal muscle, whereas no DNA was amplified upon GR silencing or at an unrelated region (Fig. 7a, b and Supplementary Fig. 7c, d). Similarly, Myod1 and Foxf2 were recruited to the identified cognate DNA regions of the *Eif4ebp2* and the *Pik3r1* locus, and their binding was abolished after siRNA-mediated knockdown (Fig. 7c, d, and Supplementary Fig. 7e, f). Reduced expression of Myod1 or Foxf2 had no effect on GR protein levels (Supplementary Fig. 7g, h), but decreased GR recruitment by at least 50 % (Fig. 7e and Supplementary Fig. 7i). Note that Myod1 silencing decreased *Eif4ebp2* transcript levels by 50 %, whereas those of *Pik3r1* were unaffected, and that Foxf2 knockdown did not significantly affect the expression of these genes (Supplementary Fig. 7b). Importantly, GR expression was required for Foxf2 binding (Fig. 7f and Supplementary Fig. 7j, k), and Myod1 recruitment to M_E2 and M_P1 and M_P2 E boxes located at proximity of Foxf2 binding sites at *Eif4ebp2* and *Pik3r1* loci was strongly decreased after GR silencing (Fig. 7a,g and Supplementary Fig. 7c, l, m), indicating that GR and Myod1 proteins may cooperate to control gene expression. Note that Foxf2 binding to the *Eif4ebp2* F_E1 site was slightly increased after Myod1 silencing, whereas binding to the *Pik3r1* Fp1 and Fp2 sites was slightly decreased (Fig. 7h and Supplementary Fig. 7g, n). Moreover, Myod1 binding was slightly enhanced at its cognate sites after Foxf2 silencing (Fig. 7i and Supplementary Fig. 7h, o). Thus, Myod1 and Foxf2 are required for efficient GR binding at *Eif4ebp2* and *Pik3r1* enhancer regions. Moreover, Foxf2 and Myod1 binding to their cognate elements is fully or partially GR-dependent, respectively. In addition, Foxf2 has a slight negative impact on Myod1 binding, whereas Myod1 either facilitates or slightly impairs Foxf2 binding, depending on the genomic location. Of note, Ctcf binding was reduced by 20-30 % in GR siRNA transfected cells (Fig. 7j and Supplementary Fig. 7p).

Thus, even though Foxf2 and Myod1 might affect the binding of each other, they enhance GR binding, and GR promotes Myod1 and Foxf2 binding to their cognate response elements located in the vicinity of GREs.

GR mediates enhancer to promoter communication
Whereas GR was bound at GREs in more than 30 % of the enhancer sites, promoter occupancy of GR correlated with less than 2 % of GREs. *De novo* motif search at GR-occupied promoter regions identified Nrf1 binding sites as one the most frequent motifs (Fig. 3b and Supplementary Fig. 3a). An Nrf1 binding sequence (5'-GCGCatGCGC-3') was located next to MACS 1882 at -1,568 bp of the *Eif4ebp2* TSS and within the MACS peak 6344 at -260 bp of the *Pik3r1* TSS. ChIP with an antibody directed against Nrf1 followed by qPCR analysis revealed that Nrf1 specifically binds to these regions (Fig. 4e). Interestingly, Nrf1 was co-immunoprecipitated with GR in muscle nuclear extracts (Fig. 4f). Thus, our data indicate that GR bound at enhancers communicates with Nrf1 at the promoter region of GR-regulated genes.

Discussion

Glucocorticoids are pleiotropic hormones that regulate metabolism(Tanaka et al., 2017), immune functions, and stress response(Cain and Cidlowski, 2015). In this study, we investigated their role in skeletal muscle, by analysing GR$^{(i)skm-/-}$ mice in which GR is selectively ablated in myofibers at adulthood. Our results demonstrate that physiological glucocorticoids have a negative impact on muscle fiber size and mass at adulthood via a myofiber GR. Even though increased muscle mass was previously observed in two mouse lines in which GR was ablated in muscle fibers during embryonic development(Shimizu et al., 2015; Watson et al., 2012a), fiber cross sectional area increase was only reported in the study by Shimizu et al. (2015). As GR ablation only induced a mild fiber hypertrophy, it might have been overlooked by Watson et al. (2012a). Alternatively, the discrepancy between the two studies might result from the use of different transgenic lines (MCK-Cre versus Acta1-Cre) to ablate GR. Increased adiposity observed in mice in which GR is ablated in muscle during development, but not in our study, further supports that the stage and/or extent at which GR is ablated in muscle fibers has functional implications.

Our data demonstrate that myofiber GR coordinates the expression of many genes at physiological glucocorticoid levels. Interestingly, the transcript levels of numerous genes promoting protein synthesis, such as Akt3, Rps6, Rps6kc1, Pabpc1 and Pi3kc2a, were increased in skeletal muscles of in GR$^{(i)skm-/-}$ mice, whereas those of anti-anabolic factors, such as Eif2ak1, Eif4ebp1, Eif4ebp2, Ddit4 and Pik3r1, were decreased, leading to a global activation of the anabolic Pik3/mTOR pathway. In contrast, genes induced by prolonged high circulating glucocorticoid levels (Schakman et al., 2013) controlling muscle catabolism were not affected by GR loss. Thus, GR orchestrates in a glucocorticoid dose-dependent manner myofiber-specific gene expression.

By determining the genomic landscape of GR binding sites in skeletal muscle, we provide evidence that GR positive control of gene expression is mainly mediated by enhancer-located GREs, the consensus sequence of which is 5'-RGNACAnnnTGTNCY-3'. These results are in agreement with previous studies performed in C2C12 cells that showed that GR binds to 5'-RGNACAnnnTGTNCY-3' motifs on dexamethasone-induced genes (Kuo et al., 2012).

Our bioinformatics analyses revealed that Myod1 E-boxes are frequently located in the vicinity of enhancer GRE-containing regions. Moreover, we have shown that GR interacts with the Forkhead box family member Foxf2 in skeletal muscles, indicating that these factors might cooperate to control GR target gene expression. The characterisation of the identified myofiber GR target gene Eif4ebp2 revealed that GR efficiently binds to the two enhancer elements GRE1 (5'-AGAACActcAGTCCT-3') and GRE2 (5'-GGTACAcagAGTGCC-3'), even though they differed from consensus GREs by at least 2 nucleotides. Moreover, MyoD1 and FoxF2 were bound to several E-boxes and Fox binding sites located in proximity of these GREs, respectively, and both factors enhanced GR recruitment. We show that GR interacts with Foxf2 and binding of Foxf2 to its binding sites was dependent on GR. In contrast, even though we could not evidence a direct interaction between GR and Myod1, Myod1 binding to E-boxes located in the vicinity of Eif4ebp2 and Pik3r1

enhancer GREs was promoted by GR. Thus, GR, MyoD and Foxf2 are likely to coordinately modulate the expression of various genes in skeletal muscle fibers.

Interestingly, it was previously shown that GR and Foxo1 synergistically activate the skeletal muscle atrophy-associated genes upon dexamethasone treatment(Waddell et al., 2008) and that GR cooperates with Foxa2 to promote hepatic gluconeogenic program(Zhang et al., 2005b). As we identified Foxf2 but not these Fox family members in the GR interactome in skeletal muscles of untreated mice, the association of GR with various Fox proteins might contribute to its promoter/cell specificity.

Our cistrome analyses also revealed that more than 35% of the sites bound by GR were located in the TSS region of genes, and that only 2 % of them contained GREs. In contrast most of these binding sites correspond to Nrf1 response elements, and our data show that Nrf1 indeed binds to its cognate site located in the promoter regions of Eif4ebp2 and Pik3r1. In addition, as GR interacts with Nrf1, our results indicate that enhancer GRE-bound GR communicates with Nrf1 bound to its response element located in promoter regions to regulate gene transcription.

Altogether, our analyses unravelled a large set of genes in muscle fibers that are coordinately regulated by myofiber GR at physiological glucocorticoid levels, leading to a downregulation of anabolic pathways and reduced muscle fiber size. Moreover, we provide molecular insights into the genomic landscape at enhancer and promoter regions of glucocorticoid-regulated genes, and the analysis of the two GR target genes encoding the anti-anabolic factors Eif4ebp2 and Pik3r1 revealed that GR cooperates with various transcription factors, including MyoD, Foxf2 and Nrf1.

Experimental procedures

Mice

Mice were maintained in a temperature and humidity-controlled animal facility, with a 12 hours light/dark cycle. Standard rodent chow (2800 kcal/kg, Usine d'Alimentation Rationelle, Villemoisson-sur-Orge, France) and water were provided ad libitum. Breeding and maintenance of mice were performed according to institutional guidelines. Animals were killed by cervical dislocation and tissues were immediately collected, weighed, and frozen in liquid nitrogen or processed for biochemical and histological analysis. All experiments were done in an accredited animal house, in compliance with French and EU regulations on the use of laboratory animals for research. Intended manipulations were submitted to the Ethical committee (Com'Eth, Strasbourg, France) for approval and to the French Research Ministry (MESR) for ethical evaluation and authorization according to the 2010/63/EU directive.

Generation of GR[(i)skm-/-] mice in which GR is selectively ablated in skeletal muscle myofibres at adulthood.

All experiments were performed in C57/Bl6J background. The targeting strategy for the conditional deletion of GR is available upon request. Briefly, mice in which exons 3 and 4 encoding the GR DNA binding domain were flanked with 2 LoxP sites (Surjit et al., 2011) were crossed with HSA-Cre-ER[T2] mice expressing the Cre-ER[T2] recombinase selectively in skeletal muscle myofibers (Schuler et al., 2005). Seven week-old mice are intraperitoneally injected with Tamoxifen (Tam, 1 mg/mouse/day) to generate control (GR floxed) and GR[(i)skm-/-] mutant mice. Primers used for genotyping are described in Supplementary Table 2.

Body lean and fat content.

Body lean and fat content were recorded in anaesthetized mice by qNMR (PIXIMUS, GE Medical Systems) according to the manufacturer's instructions. The study was performed at the Mouse Clinical Institute (Illkirch, France).

Muscle strength

Grip strength: A Grip Strength Meter (Bioseb) was used to measure forelimb and hindlimb grip strength. The test was repeated 3 consecutive times within the same session, and the mean value was recorded as the maximal grip strength for each mouse.

Contractile measurements: in situ isometric tibialis anterior muscle contraction in response to nerve stimulation was performed as described (Lahoute et al., 2008). Mice were anaesthetized using a pentobarbital solution (ip, 60 mg/kg). Muscle distal tendons were attached to an isometric transducer (Harvard Bioscience). Sciatic nerves were proximally crushed and distally stimulated by a bipolar silver electrode using supramaximal square wave pulses of 0.1 ms. All data provided by the isometric transducer were recorded and analysed on a microcomputer using a PowerLab system (4SP, AD Instruments). All isometric measurements were made at an initial length corresponding to

the maximal tension obtained during the twitch. Responses to tetanic stimulation (pulse frequency from 6.25, 12.5, 25, 50, 100, and 143 Hz) were recorded and maximal tetanic force was determined. Muscle mass was measured to calculate specific force. Fatigue resistance was assessed by repeated contractions (75 hz for 500 ms, evoked once every second for 100 s). After measurements, mice were sacrificed with an overdose of anaesthetic solution.

Histology analysis

Hematoxylin & eosin staining. For frozen sections, muscles were quickly frozen in dry ice-cooled isopentane. Deparaffinized and rehydrated or flashfrozen tissue sections (5 or 10 µm, respectively) were stained according to a standard protocol with haematoxylin (Gill No. 3, Sigma, GHS332) and eosin Y solution (Sigma, HT110332) and mounted.

Periodic acid–Schiff staining. Deparaffinized and rehydrated tissue sections (5 µm) were treated with 0.5 % periodic acid solution (Sigma, 3951), stained with Schiff's reagent (Sigma, 3952016), dehydrated, and mounted.

Fiber cross-sectional area measurements

Gastrocnemius, quadriceps, and tibialis muscle cross-sections were stained for dystrophin to mark the sarcolemma. The cross-sectional area was quantified by an automated method using the image processing software, FIJI. Individual fibers were identified based on the intensity and continuity of the dystrophin-stained sarcolemma surrounding each fiber by segmentation. The area was measured after background subtraction, automated thresholding and by using the FIJI analyse particles function(Gali Ramamoorthy et al., 2015). The calculated area and number of fibers were converted to a text file format and the results expressed as percentage of fibers distributed over a different range of fiber area.

Glucose tolerance test

Intraperitoneal glucose tolerance test (IPGTT) was performed after a 6 h fasting. Following measurement of the basal glucose level (time 0), mice were intraperitoneally injected with 20 % glucose in sterile saline solution (0.9% NaCl) at a dose of 2 g per kg body weight. Blood was collected from the tail vein after 15, 30, 45, 60, 90, and 120 min for glucose determination.

RNA preparation and analysis

Muscles were homogenized in TRIzol reagent (Life Technologies, Darmstadt, Germany) using a Minilys personal homogenizer (Bertin, Montigny, France) and 0.5 or 2.0 ml CK14 lysing kits (Precellys, Montigny, France). RNA was isolated using a standard phenol/chloroform extraction protocol. cDNA was prepared by reverse transcription of total RNA using SuperScript II (Life Technologies) and oligo(dT) primer according to the supplier's protocol. Quantitative RT-PCR was performed with a Lightcycler 480 II (Roche) using QuantiTectTM SYBR® Green PCR kit (Roche) and the primers described in Supplementary Table 3. Hprt, Gapdh, and 36b4 were used as internal controls. Data were analysed using the standard curve method (Bookout et al., 2006).

Microarray analysis

Gene expression profiling was performed on total RNA isolated from control and GR(i)skm-/- gastrocnemius muscle. Biotinylated single strand cDNA targets were prepared using the Ambion WT Expression Kit and the Affymetrix GeneChip® WT Terminal Labeling Kit according to Affymetrix recommendations. Following fragmentation and end-labeling, cDNA was hybridized on GeneChip® Mouse Gene 1.0 ST arrays (Affymetrix). Chips were washed, stained, and scanned with the GeneChip® Scanner 3000 7G (Affymetrix) Raw data CEL files were processed with Affymetrix Expression Console to calculate probe set signal intensities using Robust Multi-array Average (RMA) algorithms with default settings. Differentially regulated genes (reads > 50, p < 0.01) were further used for pathway analysis in WebGestalt (Wang et al., 2013). Heatmaps were generated by centring and normalizing expression values with Cluster 3.0 (de Hoon et al., 2004) and importing them to MultiExperiment Viewer (MeV) (Saeed et al., 2006).

Chromatin immunoprecipitation

Nuclei isolation from skeletal muscle was performed as described (Joshi et al., 2017). In brief, muscle was homogenized in cytosolic lysis buffer, treated with 1 % formaldehyde in PBS for 10 min, incubated in 125 mM glycine, and resuspended in nuclear lysis buffer. Sonicated samples were processed for chromatin immunoprecipitation (ChIP) followed by qPCR analysis (ChIP-qPCR) or by massive parallel sequencing (ChIP-seq) analyses. ChIP experiments were performed using anti-GR (C-terminal, homemade, IGBMC, 2 µg per 10 µg de chromatin), anti-Nrf1 (Abcam, ab55744), anti-H3 (Cell signaling, 9715), anti-H3K4me3 (Abcam, 1012-100), anti-H3K4me1 (Active Motif, 39297), anti-H3K9ac (Cell signaling, 9671.), Myod1 (Cell signaling, 13812), anti-Foxf2 (#H00002295-M04, Abnova), or anti-Ctcf (Sigma-Aldrich, 07-729) antibodies, or a rabbit IgG negative control on protein G-Sepharose 4B (GE Healthcare) essentially as described (Metzger et al., 2008).

For ChIP-seq analysis, libraries were prepared from GR-, H3K27ac- (Active Motif, 39133), or Pol2- (Santa Cruz H-224, SC9001) immunoprecipitated DNA as described (Joshi et al., 2017). ChIP-seq libraries were sequenced using an Illumina Hiseq 4000 and mapped to the mm10 reference genome using Bowtie 2 (Langmead and Salzberg, 2012). Only uniquely mapped reads were retained for further analysis. Data were analysed using the peak calling algorithm MACS2 (Zhang et al., 2008) using input as control. All peaks with a FDR greater than 0.01 were excluded from further analysis. The uniquely mapped reads were used to generate the genome-wide intensity profiles, which were visualized using the IGV genome browser (Thorvaldsdottir et al., 2013). HOMER (Heinz et al., 2010) was used to annotate peaks and for motif searches. Genomic features (promoter, 5' UTR, exon, intron, 3' UTR, and intergenic regions) were defined and calculated using Refseq and HOMER. Further binding site analyses were performed using the MEME-ChIP from MEME-Suite (Bailey et al., 2009) and the RSAT. Clustering analyses were performed with the seqMINER software (Ye et al., 2011). Venn diagrams were generated with the help of Venny (Oliveros, 2007-2015). Myod1 chromatin association in C2C12 myotubes was analysed using previously deposited GEO data sets (GSE21614) (Mullen et al., 2011). De novo identified motifs were referred to as follow: R = G or A; Y = T or C. Primers used for ChIP-qPCR are described in Supplementary Table 4.

Protein analysis

Muscles were homogenized in RIPA buffer [50 mM Tris pH 7.5, 1 % Nonident p40, 0.5 % Sodium Deoxycholate, 0.1 % SDS, 150 mM NaCl, 5 mM EDTA, 1 mM PMSF and protease inhibitor cocktail (45 µg/mL)]. Homogenates were separated in 6 % to 12 % Bis-acrylamid gels and blotted to Hybond nitrocellulose membranes (Amersham Biosciences). Membranes were decorated using the following antibodies: anti-GR (M20, sc-1004, Santa Cruz, 1/500), anti-Ddit4 (10638-1-AP, Proteintech, 1/500), anti-phospho-mTOR (Ser2448, 5536, Cell Signaling, 1/1000), anti-mTOR (2983, Cell Signaling, 1/500), anti-phospho-4E-BP1 (Thr37/46, 2855, Cell Signaling, 1/1500), anti-4E-BP1 (9644, Cell Signaling, 1/1500), anti-4E-BP2 (2845, Cell Signaling, 1/200), Akt3 (14982, Cell Signaling, 1/200), anti-alpha-Tubulin (homemade, IGBMC, 1/5000), anti-phospho-Akt (Ser473, 9271, Cell Signaling, 1/1000), anti-phospho-Akt (Thr308, Cell Signaling, 1/1000), anti-Akt (4691, Cell Signaling, 1/500), anti-phospho-FOXO1 (Ser256, 9461, Cell Signaling, 1/1000), anti-FOXO1 (2880, Cell Signaling, 1/1000), anti-phospho-FOXO3a (Ser318/321, 9465, Cell Signaling, 1/1000), anti-FOXO3a (2497, Cell Signaling, 1/1000), anti-Pi3 kinase p85 (ab71925, abcam, 1/500), anti-Igf1ea (20214-1-AP, Proteintech, 1/500), anti-Rictor (2114, Cell Signaling, 1/500), anti-Raptor (2280, Cell Signaling, 1/500), anti-Nrf1 (ab55744, Abcam, 1/500), anti-Myod (Cell Signaling, 13812, 1/200), anti-Foxf2 (#H00002295-M04, Abnova, 1/500) and anti-Gapdh (clone 6C5, MAB374, Millipore). Secondary antibodies conjugated to horseradish peroxidase (Amersham Biosciences) were detected using an enhanced chemiluminescence detection system (Pierce, Rockford, IL, 1/10000). Protein quantification was assessed by the ImageJ software (Schneider et al., 2012).

For immunoprecipitation assay, 200 mg of muscle nuclear extracts were incubated with 5 µg of antibody. Mass spectrometry experiments were performed as follow. After immunoprecipitation with GR antibody (homemade), gel bands were reduced, alkylated, and digested with trypsin at 37°C overnight. Extracted peptides were then analysed using an Ultimate 3000 nano-RSLC (Thermo Scientific, San Jose California) coupled in line with an Orbitrap ELITE (Thermo Scientific, San Jose Califronia). Each sample was analysed in triplicate. Briefly, peptides were separated on a C18 nano-column with a linear gradient of acetonitrile and analysed in a Top 20 CID (Collision-induced dissociation) data-dependent mass spectrometry. Data were processed by database searching against Mus Musculus Uniprot Proteome database using Proteome Discoverer 2.1 software (Thermo Fisher Scientific). Precursor and fragment mass tolerance were set at 7 ppm and 0.6 Da respectively. Trypsin was set as enzyme, and up to 2 missed cleavages were allowed. Oxidation (M), N-term acetylation were set as variable modification and Carbamidomethylation (C) as fixed modification. Proteins were identified with a minimum of two unique peptides and were filtered with False Discovery Rate < 1 %. Lastly quantitative values were obtained from Extracted Ion Chromatogram (Precursor Ions Area Detector node). Cellular Component GO term analysis was performed using the Panther algorithm (Ashburner et al., 2000).

Recombinant protein expression and purification for biophysical characterization

The cDNA encoding a His_6-tagged hGR DBD (A412-G506)-SUMO fusion protein cloned in the pETite vector (Lucigen Corp.) was expressed in *E. coli* BL21 (DE3) pRARE strain. Bacteria were re-suspended

in 20 mM Tris-HCl pH 8, 400 mM NaCl, 5% glycerol, 4 mM 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate (CHAPS), 2 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 20 mM imidazole and protease inhibitor cocktail, sonicated and centrifuged. The supernatant was loaded on 5 ml HisTrap FF crude column (GE Healthcare). The protein was eluted at 250 mM imidazole and dialyzed in imidazole-free buffer in the presence of SUMO protease (1 U/1000 µg protein). Further purification involved a Heparin purification step on a 5 mL Heparin prepacked column (GE Healthcare), where the protein was eluted using a salt gradient (20 to 1000 mM) and further purified by SEC on Superdex S75 (16/60 and 10/300, GE Healthcare) using 10 mM Hepes KOH pH 7.5, 150 mM KAc, 3 mM MgAc, 1 mM TCEP buffer. Protein samples were concentrated using Amicon-Ultra centrifugal filter units (Millipore). Purity and homogeneity of the protein were assessed by SDS-PAGE.

Microscale Thermophoresis

Microscale thermophoresis (MST) was performed as described (Takacs et al., 2013). The apparent K_D value for the GR DBD binding to DNA to was measured using the Monolith NT 115 from NanoTemper Technologies GmbH. GR DBD was fluorescently labeled with the fluorescent dye NT-647 (NanoTemper Technologies) using the Monolith NT^{TM} Protein Labeling kit (amine reactive). The labeling procedure and the subsequent removal of free dye were performed within 1 hour. The solution of unlabeled DNA was serially diluted from a concentration of about 100 µM down to 1 nM in the presence of 1667 nM labeled receptor. The serially diluted samples were loaded into Premium capillaries (NanoTemper Technologies reference MOK025). Measurements were performed at 20°C in 10 mM Hepes KOH pH=7.4, 150 mM KOH, 3 mM MgAc, 1 mM TCEP and 1 % BSA, by using 50 % LED power and 20 % IR-laser power. Data were analyzed using MO. Affinity Analysis v2.3 software.

GR DBD-DNA complex formation and native polyacrylamide gel electrophoresis

Oligonucleotides (Supplementary table 5) were annealed at 1 mM in annealing buffer (10 mM Tris HCl pH 8.0, 100 mM NaCl, 0.1 mM EDTA). GR DBD protein was incubated with double-stranded DNA (dsDNA) at 1:1.2 protein dimer:DNA molar ratio. When required, GR DBD–DNA complexes were concentrated slowly at 4 °C using a 10 kDa MWCO centrifugal filter (Amicon). 5 µg of GR DBD–DNA complexes were loaded on an 8 % polyacrylamide gel in TBE [89 mM Tris base, 89 mM boric acid (pH 8.3) and 2 mM Na$_2$EDTA] and run at 100 V at 4°C. The polyacrylamide gels were stained using Instant Blue Protein Stain (Expedeon Protein Solutions) for 15 min and rinsed in water.

Size-exclusion chromatography coupled to Multi-Angle Laser Light Scattering

Size-exclusion chromatography (SEC) coupled to Multi-Angle Laser Light Scattering (MALLS) experiments were performed on a multi-angle laser light scattering detector (miniDAWN TREOS, Wyatt Technologies) coupled in-line with SEC and an interferometric refractometer (Optilab T-rEX, Wyatt Technologies). A Superdex S75 10/300 GL column (total volume 24 mL, GE Healthcare) with a flow rate of 0.5 mL/min was used to separate the sample before performing the MALLS/QELS measurement. Experiments were done with 50 µL receptor–DNA complex samples at

concentrations between 1 and 2 mg/mL in 10 mM Hepes KOH pH 7.5, 150 mM KAc, 5 mM MgAc, 1 mM TCEP buffer. The molar mass was determined by construction of Debye plot using Zimm formalism [plot of K*c/R(θ) as a function of sin2(θ/2)] at 1 second data interval. Data analysis was performed using the ASTRA 6.1.7 software (Wyatt Technologies).

Cell culture and transfection assays

C2C12 cells were obtained from ATCC (CRL-1772). For Electrophoretic mobility shift assays, 5.10^4 C2C12 myoblasts were seeded in 24-well plates and grown for 24 hr in proliferation medium (Dulbecco's modified Eagle medium [DMEM]; glucose, 1 g/l, supplemented with 20 % FCS). pSG5-mGR, pSG5-hGR and pSG5-hGRmut expression vectors were obtained by cloning the corresponding cDNAs into pSG5 (Green et al., 1988; Leid et al., 1992). 1 µg pSG5, pSG5-mGR, pSG5hGR, or pSG5hGRmut was mixed with 3 µl of Fugene9 transfection reagent (Roche Diagnostics) in 100 µl DMEM, according to the supplier's protocol. After 15 min at room temperature, 30 µl transfection mix was added to each well. Sixteen hours later, cells were harvested, lysed, and assayed for Electrophoretic mobility shift assays.

To induce myogenesis, cells were differentiated in DMEM 1 g/l glucose, supplemented with 2 % horse serum for 4 days. Two days prior to differentiation, C2C12 cells were transfected with 1 mM siRNA against GR, Myod1, Foxf2, or a scrambled control (Invitrogen) using Lipofectamine RNAimax (Invitrogen) according to the manufacturer's instructions. Cells were transfected a second time one day after myogenic induction. siRNA oligonucleotide sequences were as follows:

GR siRNA: 5'- GCUUUGCUCCUGAUCUGAUUAUUAA -3';

Myod1 siRNA: 5'- UUAUCAGGUGCUUUGAGAGAUCGAC -3';

Foxf2 siRNA: 5'- AUCACCAGAGCGUGUGCCAAGAUAU -3';

scrambled siRNA: 5'- AGGUUCCGUGUACGUAAGACAAACU -3'

Electrophoretic mobility shift assays (EMSA)

Gel retardation assays(Metzger et al., 1995) contained 3 µg of protein extract, 0.5 µg of poly(dI-dC) and 0.05 pmol of end-labelled oligonucleotide pairs in 10 mM Tris pH 7.5, 30 mM KCl, 0.75 mM $MgCl_2$, 0.5 mM DTT, 2.5 % glycerol. The consensus GRE, GRE1 probe, and their mutated forms were obtained by annealing the following oligonucleotides:

consensus GRE: forward: 5'-GTGAGCTGAGAACATTGTGTTCTGGCT-3'

reverse: 5'-AGCCAGAACACAATGTTCTCAGCTCAC-3'

GRE1: forward: 5'-TGAGTCAGGACTGAGTGTTCTCACGG-3'

reverse: 5'-CCGTGAGAACACTCAGTCCTGACTCA-3'

mutated GRE: forward: 5'-TGAGTCAGAATATTGGATTCCCACGG-3'

reverse: 5'-CCGTGGGAACCCAATGTTCTGACTCA-3'

mutated GRE1: forward: 5'- TGAGTCAGGATTGAGTATTCTCACGG-3'

reverse: 5'-CCGTGAGAACACTCAGTCCTGACTCA-3'

Receptor-DNA complexes were separated on 5% polyacrylamide gels in 0.5 x TBE at 150 V. Gels were dried before autoradiography.

Data analysis

Data are represented as mean + SEM. Significance was calculated by

(1) two-tailed Student's t test;

(2) one-way ANOVA;

(3) two-way ANOVA; and

(4) Wilcoxon rank-sum test.

Data availability

Microarray and ChIP-seq data reported in this study are available at GEO database under the accession numbers GSEXXXXX and GSEXXXXX, respectively. The accession number for the mass spectrometry proteomics data reported in this paper is ProteomeXchange Consortium/PRIDE (Vizcaino et al., 2016) partner repository: PXDXXXX. Myod1 ChIP-seq dataset 1 was obtained from GSE21621(Mullen et al., 2011), dataset 2 from GSE49313(Mousavi et al., 2013) and dataset 3 from GSE56077(Umansky et al., 2015).

Author contributions

D.M. generated the original hypothesis. D.D., D.R., V.U-P., G.L. and J-M.B. performed the experiments. S.J., M.P. and D.R. performed the ChIP-seq analyses. A-I.R. and D.D. performed the bioinformatics analyses. D.D. performed the mass spectrometry analysis in collaboration with the platform. I.B., I.H., B.P.K. and V.D-S. performed biophysic analyses. The team of A.F. performed in situ muscle strength measurements. D.D. and D.M. took primary responsibility for writing the manuscript. All authors edited the manuscript.

References

1   Hawley, J. A., Lundby, C., Cotter, J. D. & Burke, L. M. Maximizing Cellular Adaptation to Endurance Exercise in Skeletal Muscle. *Cell Metab* **27**, 962-976, doi:10.1016/j.cmet.2018.04.014 (2018).

2   Lecker, S. H. *et al.* Multiple types of skeletal muscle atrophy involve a common program of changes in gene expression. *FASEB J* **18**, 39-51, doi:10.1096/fj.03-0610com (2004).

3   Hoffman, E. P. & Nader, G. A. Balancing muscle hypertrophy and atrophy. *Nat Med* **10**, 584-585, doi:10.1038/nm0604-584 (2004).

4   Tanaka, H., Shimizu, N. & Yoshikawa, N. Role of skeletal muscle glucocorticoid receptor in systemic energy homeostasis. *Exp Cell Res* **360**, 24-26, doi:10.1016/j.yexcr.2017.03.049 (2017).

5   Ito, K., Chung, K. F. & Adcock, I. M. Update on glucocorticoid action and resistance. *J Allergy Clin Immunol* **117**, 522-543, doi:10.1016/j.jaci.2006.01.032 (2006).

6   Braun, T. P. *et al.* Central nervous system inflammation induces muscle atrophy via activation of the hypothalamic-pituitary-adrenal axis. *J Exp Med* **208**, 2449-2463, doi:10.1084/jem.20111020 (2011).

7   Rosen, J. & Miner, J. N. The search for safer glucocorticoid receptor ligands. *Endocr Rev* **26**, 452-464, doi:10.1210/er.2005-0002 (2005).

8   Kino, T. in *Endotext*   (eds L. J. De Groot *et al.*) (2000).

9   Meijsing, S. H. Mechanisms of Glucocorticoid-Regulated Gene Transcription. *Adv Exp Med Biol* **872**, 59-81, doi:10.1007/978-1-4939-2895-8_3 (2015).

10   Lieberman, B. A., Bona, B. J., Edwards, D. P. & Nordeen, S. K. The constitution of a progesterone response element. *Mol Endocrinol* **7**, 515-527, doi:10.1210/mend.7.4.8388996 (1993).

11   Presman, D. M. *et al.* Live cell imaging unveils multiple domain requirements for in vivo dimerization of the glucocorticoid receptor. *PLoS Biol* **12**, e1001813, doi:10.1371/journal.pbio.1001813 (2014).

12   Meijsing, S. H. *et al.* DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* **324**, 407-410, doi:10.1126/science.1164265 (2009).

13   Watson, L. C. *et al.* The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. *Nat Struct Mol Biol* **20**, 876-883, doi:10.1038/nsmb.2595 (2013).

14   Hua, G., Paulen, L. & Chambon, P. GR SUMOylation and formation of an SUMO-SMRT/NCoR1-HDAC3 repressing complex is mandatory for GC-induced IR nGRE-mediated transrepression. *Proc Natl Acad Sci U S A* **113**, E626-634, doi:10.1073/pnas.1522821113 (2016).

15   Surjit, M. *et al.* Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor. *Cell* **145**, 224-241, doi:10.1016/j.cell.2011.03.027 (2011).

16   Tan, C. K. & Wahli, W. A trilogy of glucocorticoid receptor actions. *Proc Natl Acad Sci U S A* **113**, 1115-1117, doi:10.1073/pnas.1524215113 (2016).

17   Uhlenhaut, N. H. *et al.* Insights into negative regulation by the glucocorticoid receptor from genome-wide profiling of inflammatory cistromes. *Mol Cell* **49**, 158-171, doi:10.1016/j.molcel.2012.10.013 (2013).

18   Beck, I. M. *et al.* Crosstalk in inflammation: the interplay of glucocorticoid receptor-based mechanisms and kinases and phosphatases. *Endocr Rev* **30**, 830-882, doi:10.1210/er.2009-0013 (2009).

19   Wang, H., Kubica, N., Ellisen, L. W., Jefferson, L. S. & Kimball, S. R. Dexamethasone represses signaling through the mammalian target of rapamycin in muscle cells by enhancing expression of REDD1. *J Biol Chem* **281**, 39128-39134, doi:10.1074/jbc.M610023200 (2006).

20   DeYoung, M. P., Horak, P., Sofer, A., Sgroi, D. & Ellisen, L. W. Hypoxia regulates TSC1/2-mTOR signaling and tumor suppression through REDD1-mediated 14-3-3 shuttling. *Genes Dev* **22**, 239-251, doi:10.1101/gad.1617608 (2008).

21   Kuo, T. *et al.* Genome-wide analysis of glucocorticoid receptor-binding sites in myotubes identifies gene networks modulating insulin signaling. *Proc Natl Acad Sci U S A* **109**, 11160-11165, doi:10.1073/pnas.1111334109 (2012).

22   Hu, Z., Wang, H., Lee, I. H., Du, J. & Mitch, W. E. Endogenous glucocorticoids and impaired insulin signaling are both required to stimulate muscle wasting under pathophysiological conditions in mice. *J Clin Invest* **119**, 3059-3069, doi:10.1172/JCI38770 (2009).

23   Shimizu, N. *et al.* Crosstalk between glucocorticoid receptor and nutritional sensor mTOR in skeletal muscle. *Cell Metab* **13**, 170-182, doi:10.1016/j.cmet.2011.01.001 (2011).

24   Watson, M. L. *et al.* A cell-autonomous role for the glucocorticoid receptor in skeletal muscle atrophy induced by systemic glucocorticoid exposure. *Am J Physiol Endocrinol Metab* **302**, E1210-1220, doi:10.1152/ajpendo.00512.2011 (2012).

25   Shimizu, N. *et al.* A muscle-liver-fat signalling axis is essential for central control of adaptive adipose remodelling. *Nat Commun* **6**, 6693, doi:10.1038/ncomms7693 (2015).

26   Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

27   Sharifi-Zarchi, A. *et al.* DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *BMC Genomics* **18**, 964, doi:10.1186/s12864-017-4353-7 (2017).

28   Mousavi, K. *et al.* eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* **51**, 606-617, doi:10.1016/j.molcel.2013.07.022 (2013).

29   Mullen, A. C. *et al.* Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* **147**, 565-576, doi:10.1016/j.cell.2011.08.050 (2011).

30   Umansky, K. B. *et al.* Runx1 Transcription Factor Is Required for Myoblasts Proliferation during Muscle Regeneration. *PLoS Genet* **11**, e1005457, doi:10.1371/journal.pgen.1005457 (2015).

31   Aittomaki, S. *et al.* Cooperation among Stat1, glucocorticoid receptor, and PU.1 in transcriptional activation of the high-affinity Fc gamma receptor I in monocytes. *J Immunol* **164**, 5689-5697, doi:10.4049/jimmunol.164.11.5689 (2000).

32   Langlais, D., Couture, C., Balsalobre, A. & Drouin, J. The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. *Mol Cell* **47**, 38-49, doi:10.1016/j.molcel.2012.04.021 (2012).

33   Zhang, Z., Jones, S., Hagood, J. S., Fuentes, N. L. & Fuller, G. M. STAT3 acts as a co-activator of glucocorticoid receptor signaling. *J Biol Chem* **272**, 30607-30610, doi:10.1074/jbc.272.49.30607 (1997).

34   Aitola, M., Carlsson, P., Mahlapuu, M., Enerback, S. & Pelto-Huikko, M. Forkhead transcription factor FoxF2 is expressed in mesodermal tissues involved in epithelio-mesenchymal interactions. *Dev Dyn* **218**, 136-149, doi:10.1002/(SICI)1097-0177(200005)218:1<136::AID-DVDY12>3.0.CO;2-U (2000).

35   Cain, D. W. & Cidlowski, J. A. Specificity and sensitivity of glucocorticoid signaling in health and disease. *Best Pract Res Clin Endocrinol Metab* **29**, 545-556, doi:10.1016/j.beem.2015.04.007 (2015).

36   Schakman, O., Kalista, S., Barbe, C., Loumaye, A. & Thissen, J. P. Glucocorticoid-induced skeletal muscle atrophy. *Int J Biochem Cell Biol* **45**, 2163-2172, doi:10.1016/j.biocel.2013.05.036 (2013).

37   Waddell, D. S. *et al.* The glucocorticoid receptor and FOXO1 synergistically activate the skeletal muscle atrophy-associated MuRF1 gene. *Am J Physiol Endocrinol Metab* **295**, E785-797, doi:10.1152/ajpendo.00646.2007 (2008).

38   Zhang, L., Rubins, N. E., Ahima, R. S., Greenbaum, L. E. & Kaestner, K. H. Foxa2 integrates the transcriptional response of the hepatocyte to fasting. *Cell Metab* **2**, 141-148, doi:10.1016/j.cmet.2005.07.002 (2005).

39   Schuler, M., Ali, F., Metzger, E., Chambon, P. & Metzger, D. Temporally controlled targeted somatic mutagenesis in skeletal muscles of the mouse. *Genesis* **41**, 165-170, doi:10.1002/gene.20107 (2005).

40   Lahoute, C. *et al.* Premature aging in skeletal muscle lacking serum response factor. *PLoS One* **3**, e3910, doi:10.1371/journal.pone.0003910 (2008).

41   Gali Ramamoorthy, T. *et al.* The transcriptional coregulator PGC-1beta controls mitochondrial function and anti-oxidant defence in skeletal muscles. *Nat Commun* **6**, 10210, doi:10.1038/ncomms10210 (2015).

42   Bookout, A. L., Cummins, C. L., Mangelsdorf, D. J., Pesola, J. M. & Kramer, M. F. High-throughput real-time quantitative reverse transcription PCR. *Curr Protoc Mol Biol* **Chapter 15**, Unit 15 18, doi:10.1002/0471142727.mb1508s73 (2006).

43   Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* **41**, W77-83, doi:10.1093/nar/gkt439 (2013).

44   de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453-1454, doi:10.1093/bioinformatics/bth078 (2004).

45   Saeed, A. I. *et al.* TM4 microarray software suite. *Methods Enzymol* **411**, 134-193, doi:10.1016/S0076-6879(06)11009-5 (2006).

46   Joshi, S., Ueberschlag-Pitiot, V., Metzger, D. & Davidson, I. Improved Protocol for Chromatin Immunoprecipitation from Mouse Skeletal Muscle. *J Vis Exp*, doi:10.3791/56504 (2017).

47   Metzger, E. *et al.* Phosphorylation of histone H3 at threonine 11 establishes a novel chromatin mark for transcriptional regulation. *Nat Cell Biol* **10**, 53-60, doi:10.1038/ncb1668 (2008).

48   Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

49   Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

50   Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).

51   Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

52   Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-208, doi:10.1093/nar/gkp335 (2009).

53   Ye, T. *et al.* seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res* **39**, e35, doi:10.1093/nar/gkq1287 (2011).

54   Oliveros, J. C. Venny. An interactive tool for comparing lists with Venn's diagrams. . (2007-2015).

55   Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).

56   Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).

57   Takacs, M. *et al.* The asymmetric binding of PGC-1alpha to the ERRalpha and ERRgamma nuclear receptor homodimers involves a similar recognition mechanism. *PLoS One* **8**, e67810, doi:10.1371/journal.pone.0067810 (2013).

58   Metzger, D., Berry, M., Ali, S. & Chambon, P. Effect of antagonists on DNA binding properties of the human estrogen receptor in vitro and in vivo. *Mol Endocrinol* **9**, 579-591, doi:10.1210/mend.9.5.7565805 (1995).

Figure legends

Fig. 1: Loss of GR in myofibers leads to increased skeletal muscle mass and strength

(a) qNMR analysis of total lean, fat and free body fluid (FBF) content of 16 week old control and GR[(i)skm-/-] mice. (b-f) Mass (b), haematoxylin and eosin (H&E) staining (c), number of fibers (d), distribution of fiber cross section area (CSA) (e) and mean CSA (f) of gastrocnemius muscle from control and GR[(i)skm-/-] mice at indicated ages. (g) Grip strength of 8 to 20 week-old control and GR[(i)skm-/-] mice. (h-i) *In vivo* absolute maximal isometric tetanic force (h) and specific maximal isometric force (i) of tibialis anterior (TA) muscle from control and GR[(i)skm-/-] mice at 16 weeks.

a-c and g: n = 10 mice, d-f: n = 4 mice, h-i: n = 5 mice.

Mean + SEM. *p<0.05, **p<0.01, ***p < 0.001. Scale bar: 100 μm

Fig. 2: GR is an anti-anabolic factor in skeletal muscle

(a) Pie chart depicting the number of differentially expressed up- and down-regulated genes (DEGs) in gastrocnemius muscle of GR[(i)skm-/-] mice obtained by microarray analysis performed one week after gene ablation. (b) Enriched pathways obtained from GO term analysis for down-regulated genes in gastrocnemius muscle of GR[(i)skm-/-] mice. (c) Representative periodic acid–Schiff (PAS) staining of gastrocnemius muscle of 16 week-old control and GR[(i)skm-/-] mice. (d) Enriched pathways obtained from GO term analysis for up-regulated genes in gastrocnemius muscle of GR[(i)skm-/-] mice. (e) Heatmap depicting the mean centred normalized expression of indicated anti-anabolic and anabolic factors obtained from microarray analysis performed in gastrocnemius muscle of 9 week-old control and GR[(i)skm-/-] mice. (f) Relative mRNA levels of representative genes differentially expressed in microarray analysis in gastrocnemius muscle of 16 week-old control and GR[(i)skm-/-] mice. (g-h) Representative Western blot analysis (g) and relative levels of the indicated proteins (h) in gastrocnemius muscle of 16 week-old control and GR[(i)skm-/-] mice. α-Tubulin was used as a loading control. (I) Evaluation of the ratio between the phosphorylated and total mTOR and Eif4ebp1/2 protein content.

c, e: n = 3 mice, f: n = 10 mice, h-i: n=10 mice.

Mean + SEM. *p<0.05, **p<0.01, ***p < 0.001. Scale bar: 200 μm

Fig. 3: Localization of GR binding sites in the genome of skeletal muscle

(a) Pie chart depicting the position of GR binding sites on the genome. (b) HOMER motif analysis on peaks located at intergenic, intronic or TSS (-1000 to +100 bp) regions. (c) Tag density map of GR, H3K27ac, H3K4me1, H3K4me3 and Pol2 binding sites +/- 5 kb from the GR peak centre. (d) Average tag density profiles of the two clusters.

Fig. 4: GR is bound to GREs at active enhancers

(a) Overlap between genes with GR, H3K27ac and H3K4me1 peaks, and down-regulated upon GR loss. (b) Overlap between genes with GR peaks and down-regulated upon GR loss, and genes with H3K4me3 and Pol2 binding sites. (c) Localization of GR, H3K27ac, H3K4me1, H3K4me3 and Pol2 at the *Eif4ebp2* locus. (d) Chromatin immunoprecipitation (ChIP) followed by qPCR analysis (ChIP-

qPCR) performed with an anti-GR antibody in skeletal muscle of control and GR[(i)skm-/-] mice at GR-binding sites identified at the *Eif4ebp2* locus. (e) ChIP-qPCR analysis to detect *Eif4ebp2* and *Pik3r1* promoter occupancy performed with anti-Nrf1 or a rIgG in skeletal muscle of wild-type mice. (f) Immunoprecipitation (IP) of skeletal muscle nuclear extracts with anti-GR antibodies. Membranes were decorated with anti-GR and anti-Nrf1 antibodies. rIgG served as a control for immunoprecipitation. Non-immunoprecipitated extracts (10% input) were also analysed.

d, e: n=3 mice.

Mean + SEM. *p<0.05, **p<0.01.

Fig. 5: Characterization of GR binding to Eif4ebp2 GRE1 and GRE2

(a) Scheme depicting the position of the probes used to characterize the GR-response elements (GRE) identified at the *Eif4ebp2* locus. A non-specific probe (NSP) was selected in the vicinity of GRE2. (b) Native gel electrophoresis of GR DNA binding domain (DBD) and indicated DNA binding sites. (c) Size-exclusion chromatography (SEC)-MALLS analysis of GR DBD in the presence of indicated DNA probes. (d) Microscale thermophoresis analysis and binding affinities. Consensus GRE identified at the *Gilz* locus was used as a positive control. The NS probe and the consensus oestrogen-related receptor response element (ERRE) were used as a negative control. (e) Electrophoretic mobility shift assay (EMSA) of indicated radiolabeled probes in the presence (pSG5-mGR) or absence (pSG5) of murine GR.

Fig. 6: Identification of transcription factors bound to the *Eif4ebp2* and *Pik3r1* loci

(a) Tag density map of GR and Myod1 binding sites +/- 5 kb from the GR peak centre and corresponding average tag density profiles. (b) Genomic localization of GR in skeletal muscle and Myod1 in C2C12 myotubes at the *Eif4ebp2* locus. Myod1 ChIP-seq dataset 1 was obtained from GSE21621(Mullen et al., 2011), dataset 2 from GSE49313(Mousavi et al., 2013) and dataset 3 from GSE56077(Umansky et al., 2015). (c) Scheme depicting the genomic localization of predicted GREs, E-boxes, Foxf2 and Ctcf binding sites at the *Eif4ebp2* locus and the primers used for ChIP-qPCR experiments. (d-e) ChIP-qPCR analysis performed with anti-Myod1 (d) and anti-Ctcf (e) antibodies, or rabbit immunoglobulin G (IgG) (rIgG) in skeletal muscle of wild-type mice. (f) Venn diagram depicting the overlap between GR interacting proteins identified by immunoprecipitation using anti-GR antibody followed by mass spectrometry in gastrocnemius muscle when compared to a rIgG or a GR immunoprecipitation in GR[(i)skm-/-] mice. (g) Immunoprecipitation with anti-GR antibody from gastrocnemius muscle nuclear extracts. Membranes were decorated with anti-Stat3 and anti-Foxf2 antibodies. rIgG served as a control for immunoprecipitation. Non-immunoprecipitated extracts (10 and 30 % input) were also analysed. (h) ChIP-qPCR analysis performed with an anti-Foxf2 antibody or a rIgG in gastrocnemius muscle of wild-type mice.

d-f, H: n=3 mice.

Mean + SEM. *p<0.05, **p<0.01, ***p < 0.001.

Fig. 7: Characterisation of the interaction between GR, Myod1 and Foxf2 for binding to their cognate elements

(a) Schematic representation of the predicted Myod1, Foxf2 and Ctcf binding sites located in the proximity of the *Eif4ebp2* GR MACS peak 1884. The localisation of the primer pairs for ChIP analyses (M_E1, M_E2, F_E1, C_E1 and C_E2) are indicated. (b-d) ChIP-qPCR analysis performed at the *Eif4ebp2* locus with anti-GR (b), anti-Myod1 (c) and anti-Foxf2 (d) antibody in C2C12 myotubes transfected with siRNA directed against GR (siGR) (b), Myod1 (siMyod1) (c), Foxf2 (siFoxf2) (d) or scrambled siRNA (siCtrl), an anti-GR antibody in C2C12 myotubes transfected with siMyod1, siFoxf2 or siCtrl (e), anti Foxf2 (f) or anti-Myod1 (g) antibodies in C2C12 myotubes transfected with siGR or siCtrl, an anti-Foxf2 antibody or a rIgG in C2C12 myotubes transfected with siMyod or siCtrl (h), an anti-Myod1 antibody or a rIgG in C2C12 myotubes transfected with siFoxf2 or siCtrl (i), and an anti-Ctcf antibody or a rIgG in C2C12 myotubes transfected with siGR or siCtrl (j).

n=3 independent experiments in triplicate.

Mean + SEM. *p<0.05, **p<0.01, ***p < 0.001.

Supplementary Figures

Supplementary Fig. 1: Loss of GR in myofibers leads to a progressive increase in skeletal muscle mass and strength

(a) Schematic representation of the HSA-CreER[T2] transgene and of the wild type (WT allele, upper panel), floxed (L2 allele, middle panel), and the Cre-mediated DBD encoding exon deleted (L- allele, lower panel) GR alleles. Primers used to characterize the various alleles are materialized with arrows and sequences are available in Supplementary Table 2. LoxP sites are shown by arrowheads. (b-c) Relative GR transcript (b) and protein (c) levels in indicated tissues isolated from control and GR[(i)skm-/-] mice 3 weeks after GR ablation. α-Tubulin was used as a loading control. (d-f) Body (d), tibialis (e) and quadriceps (f) mass of control and GR[(i)skm-/-] mice at indicated ages. (g-l) Number of fibers (g-h), distribution of fiber cross-section area (CSA) (i-j) and average of the CSA (k-l) of tibialis (e, g, I, k) and quadriceps (f, h, j, l) muscles of 16 week-old control and GR[(i)skm-/-] mice.

b, e: n = 10 mice, C: n=3 mice, D: n=20 mice, F-N: n = 20 mice.

Mean + SEM. *p<0.05, **p<0.01, ***p < 0.001.

Supplementary Fig. 2: GR is an anti-anabolic factor in skeletal muscle

(a) Scheme depicting the genes encoding enzymes involved in glycogen synthesis or catabolism. Genes differentially expressed between control and GR[(i)skm-/-] mice are in green. (b) Relative mRNA levels of indicated genes in 16 week-old control and GR[(i)skm-/-] mice. (c-d) Basal glucose levels (c) and glucose tolerance test (IPGTT) (d) of control and GR[(i)skm-/-] mice evaluated at 16 weeks of age. (e) Scheme depicting the genes encoding the factors involved in the Pi3k/Akt/Mtor pathway that are up- (green) or down-regulated in GR[(i)skm-/-] mice vs control mice. (f) Relative transcript levels of indicated genes belonging to anabolic, catabolic, proteasome, calpain, and autophagy pathways determined in gastrocnemius muscle of 16 week-old control and GR[(i)skm-/-] mice. (g-h) Representative Western blot analysis (g) and relative quantification of indicated proteins (h) in

gastrocnemius muscle of 16 week-old control and GR$^{(i)skm-/-}$ mice. Gapdh and α-Tubulin were used as a loading controls. (i) Ratio between the phosphorylated and total Akt, Foxo1 or Foxo3 protein content.

b-d, f, h-i: n = 10 mice.

Mean + SEM. *p<0.05, **p<0.01, ***p < 0.001.

Supplementary Fig. 3: Localization of GR binding sites in the genome of skeletal muscle

(a) MEME and RSAT motif analysis on peaks located at intergenic, intronic and promoter regions. (b) Localisation of GR at genes encoding the indicated anabolic factors. (c) Pie charts depicting the position of histone H3 acetylated at lysine 27 (H3K27ac), mono- (H3K4me1) or trimethylated at lysine 4 (H3K4me3) and Pol2 binding sites in the genome.

Supplementary Fig. 4: GR is bound to GREs at active enhancers

(a) Localization of GR, H3K27ac, H3K4me1, H3K4me3 and Pol2 at indicated genes encoding anti-anabolic factors down-regulated in GR$^{(i)skm-/-}$ mice. (b) ChIP-qPCR analysis performed with anti-H3K4me1, anti-H3K4me3, H3K9ac and anti-H3 antibodies or rabbit immunoglobulin G (IgG) (rIgG) in skeletal muscle of wild-type mice at the *Eif4ebp2* locus. (c-d) ChIP-qPCR analysis performed with an anti-GR antibody (c), or anti-H3K4me1, anti-H3K4me3, H3K9ac and anti-H3 antibodies or rIgG (d) in skeletal muscle of wild-type mice at GR binding sites identified at the *Pik3r1* locus.

b-d, n=3 mice. Mean + SEM. *p<0.05, **p<0.01, ***p < 0.001.

Supplementary Fig. 5: Characterization of GR binding to Eif4ebp2 GRE1 and GRE2

(a) Size-exclusion chromatography (SEC)-MALLS analysis of GR DBD in the presence of the indicated DNA probes. Consensus GRE identified at the *Gilz* locus was used as a positive control. A non-specific probe (NSP) selected in the vicinity of GRE2 and the consensus oestrogen-related receptor response element (ERRE) were used as a negative control. (b) Electrophoretic mobility shift assay (EMSA) of indicated radiolabeled probes in the presence or absence (pSG5) of a wild-type (pSG5-hGR) or mutated (pSG5-hGRmut) human GR.

Supplementary Fig. 6: Identification of transcription factors bound to genes encoding anti-anabolic factors

(a) HOMER motif analysis on peaks located at intergenic and intronic regions. (b) Overlap of the genes bound by GR with those bound by Myod1. (c) Genomic localization of GR and Myod1 at the *Pik3r1*, *Ddit4*, *Eif4ebp1* and *Eif2ak1* loci. (d) Scheme depicting the genomic localization of identified E-boxes, Foxf2 and Ctcf motives, and GR-response elements (GRE) at the *Pik3r1* locus and the primers used for ChIP-qPCR experiments. (e-g) ChIP-qPCR analysis performed with anti-Myod1 (e), anti-Ctcf (f) or anti-Foxf2 (g) antibodies or rIgG in skeletal muscle of wild-type mice at the *Pik3r1* locus.

c,d,f-i: n=3 mice. Mean + SEM. *p<0.05, **p<0.01, ***p < 0.001.

Supplementary Fig. 7: Characterisation of the interaction between GR, Myod1 and Foxf2 for binding to their cognate elements

(a) Representative Western blot analysis of GR, Myod1 and Foxf2 protein in C2C12 myotubes transfected with siRNA directed against GR (siGR), Myod1 (siMyod1), Foxf2 (siFoxf2) or scrambled siRNA (siCtrl). Gapdh was used as a loading control. (b) Relative mRNA levels of *Eif4ebp2* and *Pik3r1* in C2C12 myotubes transfected with siCtrl, siGR, siMyod1 or siFoxf2. (c) Schematic representation of the predicted E-boxes, Foxf2 and Ctcf binding sites located in the proximity of the *Pik3r1* GR MACS peak 6346. The localisation of the primer pairs for ChIP analyses (M_P1, M_P2, F_P1, F_P2, and C_P1) are indicated. (d-f) ChIP-qPCR analysis performed at the *Pik3r1* locus with an anti-GR antibody in C2C12 myotubes transfected with siCtrl or siGR (d), an anti-Myod1 antibody in C2C12 myotubes transfected with siCtrl or siMyod1 (e), and an anti-Foxf2 antibody in C2C12 myotubes transfected with siCtrl or siFoxf2 (f). (g-h) Representative Western blot analysis of GR, Myod1 and Foxf2 protein levels in C2C12 myotubes transfected with siCtrl (g, h), siMyod1 (g) or siFoxf2 (h). Gapdh was used as a loading control. (i-j) ChIP-qPCR analysis performed at the *Pik3r1* locus with (i) an anti-GR antibody in C2C12 myotubes transfected with siCtrl, siMyod1 or siFoxf2, (j) an anti Foxf2 antibody in C2C12 myotubes transfected with siCtrl or siGR. (k-l) Representative Western blot analysis of GR, (k) Foxf2 and (l) Myod1 protein levels in C2C12 myotubes transfected with siCtrl or siGR. Gapdh was used as a loading control. (m-p) ChIP-qPCR analysis performed at the *Pik3r1* locus with (m) an anti-Myod1 antibody in C2C12 myotubes transfected with siCtrl or siGR, (n) an anti-Foxf2 antibody or a rIgG in C2C12 myotubes transfected with siCtrl or siMyod1, (o) an anti-Myod1 antibody or a rIgG in C2C12 myotubes transfected with siCtrl or siFoxf2, (p) an anti-Ctcf antibody or a rIgG in C2C12 myotubes transfected with siCtrl or siGR.

n=3 independent experiments in triplicate.

Mean + SEM. *p<0.05, **p<0.01, ***p < 0.001.

Supplemental information

Supplemental figures and tables

Supplemental table 1: 41 GR-interacting partners identified from mass-spectrometry analysis.

Gene symbol	Accession	Description	# Peptides
Dnajb1	Q9QYJ3	DnaJ homolog subfamily B member 1	2
Ewsr1	Q5SUS9	RNA-binding protein EWS	2
Foxf2	O54743	Forkhead box protein F2	1
Gstp1	P19157	Glutathione S-transferase P 1	2
Gtf2i	Q9ESZ8	General transcription factor II-I	10
H2afy	Q9QZQ8	Core histone macro-H2A.1	1
H2afz	P0C0S6	Histone H2A.Z	3
Hist1h2ab	P22752	Histone H2A type 1	3
Hist1h4a	P62806	Histone H4	5
Hist2h2be	Q64524	Histone H2B type 2-E	3
Hnrnpd	Q60668	Heterogeneous nuclear ribonucleoprotein D0	2
Hnrnpu	Q8VEK3	Heterogeneous nuclear ribonucleoprotein U	2
Khdrbs1	Q60749	KH domain-containing, RNA-binding, signal transduction-associated protein 1	4
Kif5b	Q61768	Kinesin-1 heavy chain	2
Mif	P34884	Macrophage migration inhibitory factor	1
Naca	P70670	Nascent polypeptide-associated complex subunit alpha, muscle-specific form	6
Ncl	P09405	Nucleolin	3
Npm1	Q61937	Nucleophmin	2
Nr3c1	E9PUR6	Glucocorticoid receptor	11
Pa2g4	P50580	Proliferation-associated protein 2G4	3
Pcbp2	Q61990	Poly(rC)-binding protein 2	3
Phb	P67778	Prohibitin	1
Psma6	Q9QUM9	Proteasome subunit alpha type-6	1
Psmc1	P62192	26S protease regulatory subunit 4	1
Psmc3	B7ZCF1	26S protease regulatory subunit 6A	1
Ptbp1	Q922I7	MCG13402, isoform CRA_c	1
Ptges3	Q9R0Q7	Prtaglandin E synthase 3	1
Ptma	A0A087WQN2	Prothymin alpha (Fragment)	1
Rpl23	P62830	60S ribomal protein L23	3
Rps3	P62908	40S ribomal protein S3	7
Srsf1	H7BX95	Serine/arginine-rich-splicing factor 1	1
Srsf2	Q62093	Serine/arginine-rich splicing factor 2	1
Srsf3	P84104	Serine/arginine-rich splicing factor 3	1
Stat1	A0A087WSP5	Signal transducer and activator of transcription	1
Stat3	P42227	Signal transducer and activator of transcription 3	1
Sub1	P11031	Activated RNA polymerase II transcriptional coactivator p15	2
Syncrip	Q7TMK9	Heterogeneous nuclear ribonucleoprotein Q	4
Vcp	Q01853	Transitional endoplasmic reticulum ATPase	16

Ybx1	P62960	Nuclease-sensitive element-binding protein 1	4
Ywhab	Q9CQV8	14-3-3 protein beta/alpha	5
Ywhaz	P63101	14-3-3 protein zeta/delta	11

Supplemental table 2: Primers used for genotyping

Name	Sequence forward	Sequence reverse
Cre recombinase	TTCCCGCAGAACCTGAAGATGTTCG	GGGTGTTATAAGCAATCCCCAGAAATGC
GR allele primer 2	AGATCATTTGCCTAGCAGGCATGAG	
GR allele primer 3	GTCAACACATGATCACCTTGCAGTC	
GR allele primer 1	CCAGAGAACTAATTGGCTCTTGCAC	

Supplemental table 3: Primers used for RT-qPCR analysis

Name	Sequence forward	Sequence reverse
36b4	5'-AGATTCGGGATATGCTGTTGG-3'	5'-AAAGCCTGGAAGAAGGAGGTC-3'
Hprt	5'-GTTGGATACAGGCCAGACTTTGTTG-3'	5'-GATTCAACTTGCGCTCATCTTAGGC-3'
18S	5'-TCGTCTTCGAAACTCCGACT-3'	5'-CGCGGTTCTATTTTGTTGGT-3'
GR	5'-CGCTGCCAATTCTGACTGGAGTTT-3'	5'-ACACCTGGATGACCAAATGACCCT-3'
Phka1	5'-CGTAGGCTGTCTGTCTCGAT-3'	5'-CATCTAGCCTTCTCCTGCGT-3'
Ugp2	5'-AGACTGGTGGAAATCGCTCA-3'	5'-TTCAGGCCTCCATCCAATGT-3'
Gys2	5'-TTTGTAAACAGTCACGCCGG-3'	5'-CGGAGAAGGTGGTACTGAGG-3'
Gyg	5'-TTATCAGCAGCACCAGACCC-3'	5'-CGTTGCCCAGCCACTAAAAT-3'
Gbe1	5'-AGGATGTATCAGGGATGCCG-3'	5'-CAAGGTAGCGTCGATTGGTG-3'
Agl	5'-AGACCGAAGAATGACCTGGG-3'	5'-ATGAGGTAGCGTGGGATCTG-3'
Eif2ak1	5'-AGCTCGGAATTGGAAGGGAA-3'	5'-TCCGCTTGTTCCTCTCAGTT-3'
Eif4ebp1	5'-GATGAGCCTCCCATGCAA-3'	5'-CCATCTCAAATTGTGACTCTTCA-3'
Eif4ebp2	5'-GTTGGACCGTCGCAATTCTC-3'	5'-AAACTGAGCCTCATCCCCAA-3'
Ddit4	5'-CTGTGCCCACCTTTCAGTTG-3'	5'-GTCAGGGACTGGCTGTAACC-3'
Akt3	5'-GTTGGGTTCAGAAGAGGGGA-3'	5'-TGGCTTTGGTCGTTCTGTTT-3'
Rsp6kb1	5'-ACTAGTGTGAACAGAGGGCC-3'	5'-TTCCTCCAGAATGTTCCGCT-3'
Pi3kr1	5'-CACCCAAGCCCACTACTGTA-3'	5'- GAGTGTAATCGCCGTGCATT-3'
Pik3cg	5'-CCAGAGTCGACCAAGTGCTT-3'	5'-TGAGCTCCATGGAAGACAGG-3'
Akt1	5'-CTGCCCTTCTACAACCAGGA-3'	5'-CATACACATCCTGCCACACG-3'
Igf1	5'-AGCAGCCTTCCAACTCAATTAT-3'	5'-GAAGACGACATGATGTGTATCTTTATC-3'
Mtor	5'-TCGTCTCCATCAAGCTGTTAGC-3'	5'-CAATCGGAGGCAACAACAAGT-3'
Foxo1	5'-AACCAAAGCTTCCCACACAG-3'	5'-TGGACTGCTCCTCAGTTCCT-3'
Foxo3	5'-CAAACGGCTCACTTTGTCCC-3'	5'-TCATTCTGAACGCGCATGAA-3'
Mstn	5'-GCTACCACGGAAACAATCAT-3'	5'-CAATACTCTGCCAAATACCA-3'
Ubc	5'-TCTTCGTGAAGACCCTGACC-3'	5'-CAGGTGCAGGGTTGACTCTT-3'
Fbxo32	5'-TCACAGCTCACATCCCTGAG-3'	5'-TCAGCCTCTGCATGATGTTC-3'
Trim63	5'-TGAGGTGCCTACTTGCTCCT-3'	5'-GTGGACTTTTCCAGCTGCTC-3'
Capn1	5'-AAGCGTGATTTCTTCCTGGC-3'	5'-GTCCCAGCCTTCTTCTCTGA-3'
Capn2	5'-TCCTCCCAACCTGTTCAAG-3'	5'-GCCTCCAGTTCCCATCCA-3'
Atg3	5'-ATGTTCCATGCTACAAGCGGT-3'	5'-TCCTTGCTTTCCAGTGTAATCTC-3'
Atg5	5'-ATGCGGTTGAGGCTCACTTTA-3'	5'-GCCCAAAACTGGTCAAATCTGTC-3'
Atg12	5'-AACAAAGAAATGGGCTGTGG-3'	5'-ATGCCTGGGATTTGCAGTAA-3'
Gabarapl1	5'-CATCGTGGAGAAGGCTCCTA-3'	5'-ATACAGCTGGCCCATGGTAG-3'
Bnip3	5'-TTGGGGCATTTTACTAACCTTG-3'	5'-TGCAGGTGACTGGTGGTACTAA-3'
Map1lc3a	5'-CATGAGCGAGTTGGTCAAGA-3'	5'-TTGACTCAGAAGCCGAAGGT-3'
Ctsl	5'-GTGGACTGTTCTCACGCTCAAG-3'	5'-TCCGTCCTTCGCTTCATAGG-3'
Becn1	5'-GAGCCATTTATTGAAACTCGCCA-3'	5'-CCTCCCCGATCAGAGTGAA-3'

Supplemental table 4: Primers used for ChIP-qPCR analysis

Name	Sequence forward	Sequence reverse
Eif4ebp2_MACS1884-5	CGGGCCGACTTCCTAATTTG	GGATGGGGTATGGATGGGAG
Eif4ebp2_MACS1884_1	GGCTGCTGAGAAAGTGTGAG	GTACGGGGCTCTGAGATTGA
Eif4ebp2_MACS1884_2	TCTGGGTGTTGGCAGAATCA	AGTGGGAGAGAACTTCGCAG
Eif4ebp2_MACS1883-4_1	GAGGGAGAGAGGGAGAGACA	TCTGAATTGCCCATGACCCT
Eif4ebp2_MACS1883-4_2	AAGCAAGGAGAGAGGCATGT	ACCCCATGACAATCACGACT
Eif4ebp2_MACS1883	ACAACCTTGACATCCACCCA	GGTGCATCTGGTGGAATGTG
Eif4ebp2_MACS1882_1	GCTCCACCCTTCAACACTTC	AGCAAGGGGTAGTAGAGGGA
Eif4ebp2_MACS1882_2	CTTCCGGTAGTCATCGTTGC	CTCTCAACTCGCCTGCTCT
Eif4ebp2_Intron1_fw	CGGATTTGGAGTTCAGCCTG	CCCCTTCCTTGTTTGGTTGG

Supplemental table 5: DNA sequences corresponding to putative GREs identified under MACS peaks 1883 and 1884. Oligonucleotide sequences are as follows: half-sites are shown in uppercase; flanking and separating nucleotides are shown in lowercase.

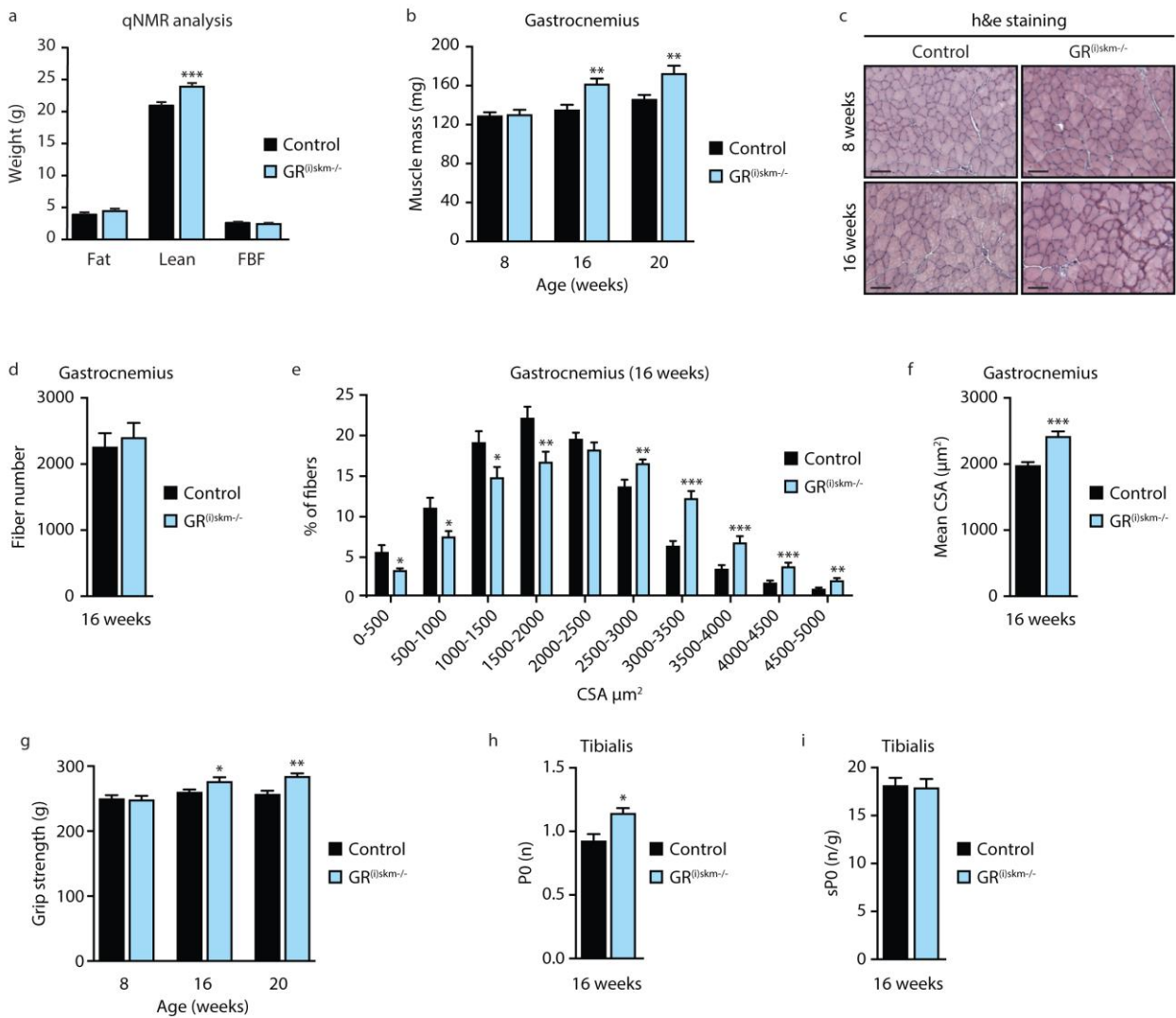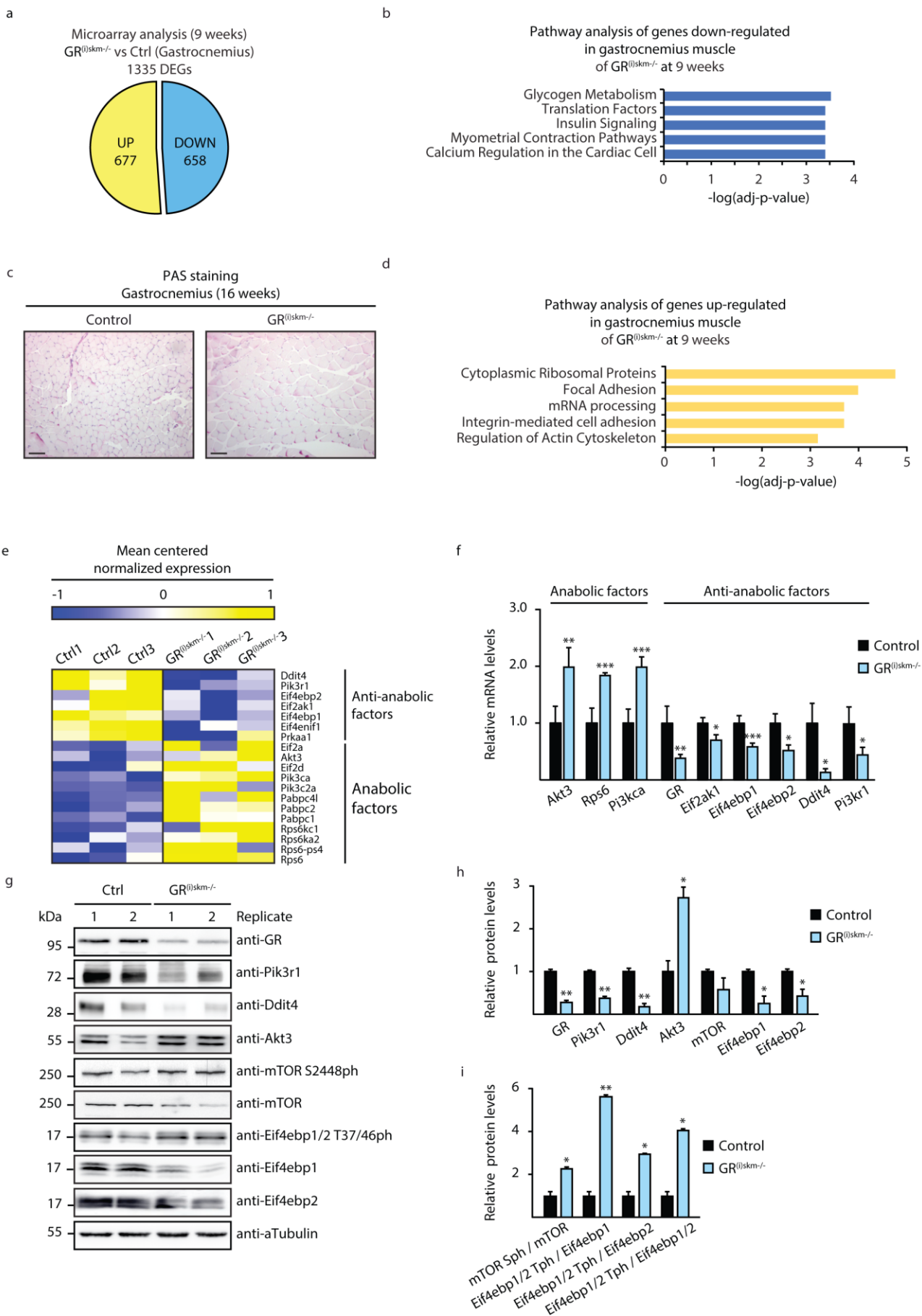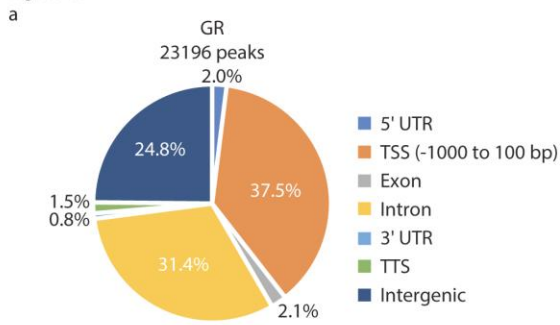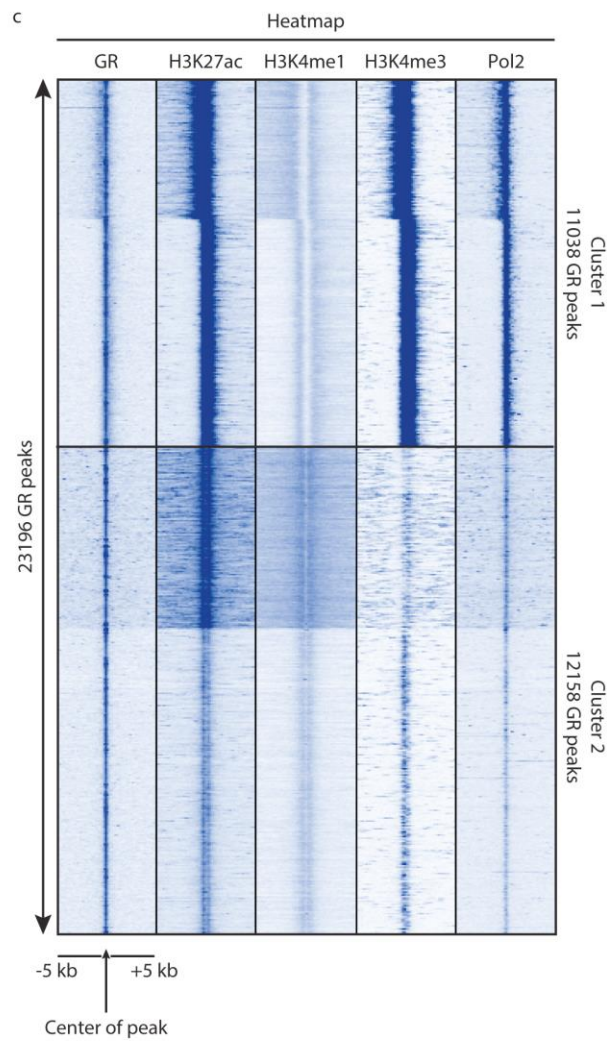| DNA probe	Sequence 5' to 3'
GRE (Gilz)	tgagtcAGAACAttgGGTTCCcacgg
GRE1	tgagtcAGGACTgagTGTTCTcacgg
GRE1 mut	tgagtcAGGATTgagTATTCTcacgg
GRE2	aggctgGGTACAcagAGTGCCctgcc
GRE2 mut	aggctgGGTATAcagAATGCCctgcc
Unrelated	tgggctGAGGCTgggTGTGGCccgac
ERRE	tgaaggtca

Figure 1.

Figure 2.

a

Microarray analysis (9 weeks)
GR$^{(i)skm-/-}$ vs Ctrl (Gastrocnemius)
1335 DEGs

UP 677
DOWN 658

b

Pathway analysis of genes down-regulated
in gastrocnemius muscle
of GR$^{(i)skm-/-}$ at 9 weeks

Glycogen Metabolism
Translation Factors
Insulin Signaling
Myometrial Contraction Pathways
Calcium Regulation in the Cardiac Cell

-log(adj-p-value)

c

PAS staining
Gastrocnemius (16 weeks)

Control   GR$^{(i)skm-/-}$

d

Pathway analysis of genes up-regulated
in gastrocnemius muscle
of GR$^{(i)skm-/-}$ at 9 weeks

Cytoplasmic Ribosomal Proteins
Focal Adhesion
mRNA processing
Integrin-mediated cell adhesion
Regulation of Actin Cytoskeleton

-log(adj-p-value)

e

Mean centered
normalized expression

-1   0   1

Ctrl1 Ctrl2 Ctrl3 GR$^{(i)skm-/-}$1 GR$^{(i)skm-/-}$2 GR$^{(i)skm-/-}$3

Ddit4
Pik3r1
Eif4ebp2
Eif2ak1
Eif4ebp1   Anti-anabolic
Eif4enif1   factors
Prkaa1
Eif2a
Akt3
Eif2d
Pik3ca
Pik3c2a   Anabolic
Pabpc4l   factors
Pabpc2
Pabpc1
Rps6kc1
Rps6ka2
Rps6-ps4
Rps6

f

Anabolic factors   Anti-anabolic factors

Relative mRNA levels

■ Control
■ GR$^{(i)skm-/-}$

Akt3 Rps6 Pi3kca GR Eif2ak1 Eif4ebp1 Eif4ebp2 Ddit4 Pi3kr1

g

Ctrl   GR$^{(i)skm-/-}$

kDa   1   2   1   2   Replicate

95   anti-GR
72   anti-Pik3r1
28   anti-Ddit4
55   anti-Akt3
250   anti-mTOR S2448ph
250   anti-mTOR
17   anti-Eif4ebp1/2 T37/46ph
17   anti-Eif4ebp1
17   anti-Eif4ebp2
55   anti-aTubulin

h

Relative protein levels

■ Control
■ GR$^{(i)skm-/-}$

GR Pik3r1 Ddit4 Akt3 mTOR Eif4ebp1 Eif4ebp2

i

Relative protein levels

■ Control
■ GR$^{(i)skm-/-}$

mTOR Sph / mTOR
Eif4ebp1/2 Tph / Eif4ebp1
Eif4ebp1/2 Tph / Eif4ebp2
Eif4ebp1/2 Tph / Eif4ebp1/2

Figure 3.

a

GR
23196 peaks

- 5' UTR
- TSS (-1000 to 100 bp)
- Exon
- Intron
- 3' UTR
- TTS
- Intergenic

b

Genome location	HOMER de novo motif analysis		
	Motif	Logo	p-value
Intergenic	GRE		1e-644
Intron	GRE		1e-504
TSS (-1000 to 100 bp)	NRF1		1e-192

c

Heatmap

GR   H3K27ac   H3K4me1   H3K4me3   Pol2

23196 GR peaks

Cluster 1
11038 GR peaks

Cluster 2
12158 GR peaks

-5 kb   +5 kb

Center of peak

d

Merged dataset profile of cluster 1
Active promoters

Mean density (tag/50 bp)

-2.5 kb   +2.5 kb

- GR
- H3K27ac
- H3K4me3
- H3K4me1
- Pol2

Merged dataset profile of cluster 2
Active enhancers

Mean density (tag/50 bp)

-2.5 kb   +2.5 kb

- GR
- H3K27ac
- H3K4me3
- H3K4me1
- Pol2

Figure 4.

a

H3K27ac
(12953 genes)   H3K4me1
(17420 genes)

344   4271
304   2940   70
923   8908   71   120
773   376
11   7   3
0

GR
(11302 genes)   Downregulated
(658 genes)

b

GR-bound downregulated
(376 genes)

19
20   10
327
3402   8589   739

H3K4me3 peaks
at promoter
(12338 genes)   Pol2 peaks
at promoter
(9665 genes)

c

Eif4ebp2

45 kb

Input	0-250
GR MACS2_peak	0-250
	1881 1882 1883 1884 1885
H3K27ac MACS2_peak	0-120
H3K4me1 MACS2_peak	0-50
H3K4me3 MACS2_peak	0-320
Pol2 MACS2_peak	
Refseq	

d

Eif4ebp2 locus
anti-GR

% Input

0.20
0.16
0.12
0.08
0.06
0.04
0.02

■ Control
■ GR$^{(i)skm-/-}$

MACS peak   1885   1884   1883   1882   1881   Unrel
                    GRE1   GRE2   (+1.5 kb)

e

Eif4ebp2 locus

% Input

0.6

0.4

0.2

0

■ Nrf1
□ mIgG

16   19
Primers

Pik3r1 locus

% Input

0.6

0.4

0.2

0

■ Nrf1
□ mIgG

32   21
Primers

f

kDa   10 %
       Input   GR   rIgG   IP

95   anti-GR

50   anti-Nrf1

Figure 5.

a

Eif4ebp2

b

c

SEC-MALLS

d

Microscale thermophoresis

Probe	Kd (nM)
GRE	154
GRE1	195
GRE2	204
NSP	5638
ERRE	14102

e

Figure 6.

a

Heatmap

GR muscle   Myod1 myotubes datasets
1   2   3

23196 GR peaks

Cluster 1

Cluster 2

Cluster 3

-5kb ↑ +5kb
Center of peak

Merged dataset profile
of cluster 1

96.0

Mean density
(tag/50bp)

-2 kb   +2 kb

— GR
— Myod1 dataset 1
— Myod1 dataset 2
— Myod1 dataset 3

b

Eif4ebp2

33 kb

GR
MACS2_peak
1881 1882   1883   1884   1885

Myod1
MACS2_peak

Refseq

c

Eif4ebp2 locus

Primers   1   2   3   4   5   6   7   8   9   10   11   12   13   14   15   16   17   18   19
Foxf2 Ebox Ebox GR Ebox Foxf2 Ctcf Ebox GR Ebox Foxf2 Ctcf Ebox GR Ebox Foxf2 Ctcf Nrf1 Ctcf GR unrelated
MACS1885   MACS1884   MACS1883

-21.5 -21.5 -21.3 -21.1 -21.0 -20.4 -19.0 -16.1 -15.5 -14.1 -14 -13.9 -13.1 -12.5 -12.4 -10.2 -7.3 -1.57 -0.87 TSS 1.35

d

Eif4ebp2 locus

0.8

% Input

anti-Myod1
rIgG

** *** * ** *   **

1 2 4 7 9 11 13 16
Primers

e

Eif4ebp2 locus

1

% Input

anti-Ctcf
rIgG

**   *   * **

6 10 15 17 19
Primers

f

GR interacting partners
identified by mass spectrometry

37   360   39

anti-GR vs rIgG   Ctrl vs GR(i)skm-/-
Ctrl (397)   anti-GR (399)

g

kDa   10 %   IP   30 %
Input   GR rIgG   Input

95   anti-Stat3

50   anti-Foxf2

h

Eif4ebp2 locus

0.8

% Input

anti-Foxf2
rIgG

**   ** ** **

1 2 5 9 14
Primers

154

Figure 7.

Suplementary Figure 1.

Suplementary Figure 2.

a

b

Glycogen metabolism

c

d

IPGTT on 16 week-old mice

e

- Increased in GR$^{(i)skm-/-}$ mice
- Decreased in GR$^{(i)skm-/-}$ mice
- Activation of protein activity
- Inhibition of protein activity

f

Suplementary Figure 2.

g

kDa	Ctrl		GR(i)skm-/-		Replicate
	1	2	1	2	
17					anti-Igf1
55					anti-Akt1
130					anti-Raptor
130					anti-Rictor
55					anti-Akt1/2/3 S473ph
55					anti-Akt1/2/3 T308ph
72					anti-Rps6kb1
72					anti-Foxo1 S256ph
72					anti-Foxo1
72					anti-Foxo3a S318-321ph
72					anti-Foxo3a
37					anti-Gapdh
55					anti-aTubulin

h

i

Supplementary Figure 3.

a

Genome location	MEME de novo motif analysis			RSAT de novo motif analysis		
	Motif	Logo	e-value	Motif	Logo	p-value
Intergenic	GRE		4.0e-204	GRE		1e-171
Intron	GRE		4.3e-274	GRE		3e-151
TSS (-1000 to 100 bp)	NRF1		3.1e-11	NRF1		2.3e-122

b
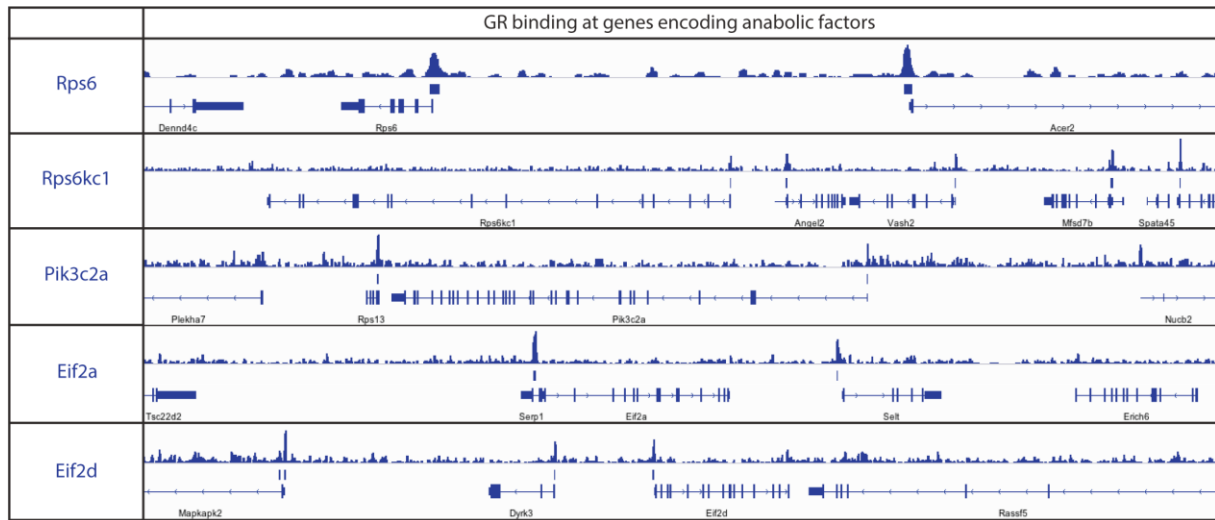
c
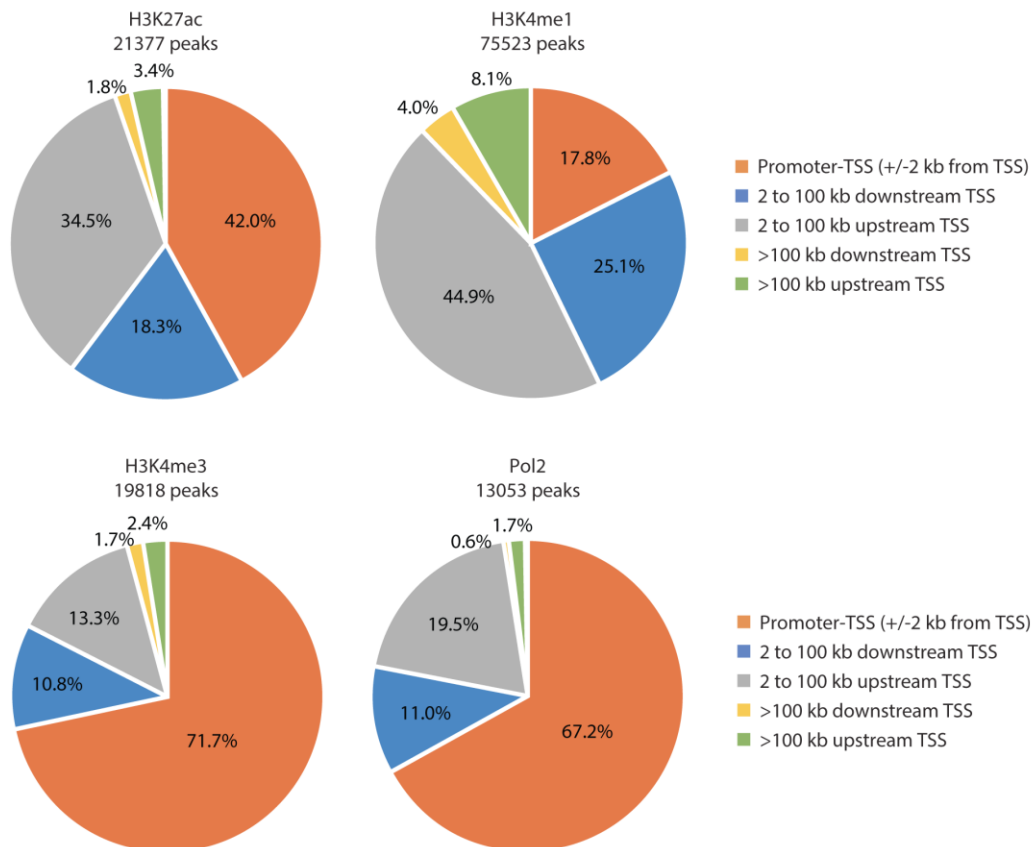
159

Supplementary Figure 4.

a

b

c

d

Pik3r1 locus
anti-H3K4me1

Pik3r1 locus
anti-H3K4me3

Pik3r1 locus
anti-H3K9ac

Pik3r1 locus
anti-H3

Pik3r1 locus
rIgG

Supplementary Figure 5.

a

b

Supplementary Figure 6.

a

Genome location	HOMER known motif analysis (total peaks)		
	Motif	Logo	p-value
Intergenic	MYOD		1e-167
	CTCF		1e-147
	STAT3		1e-73
Intron	MYOD		1e-155
	CTCF		1e-74
	STAT3		1e-64

b

c

d

Pik3r1 locus
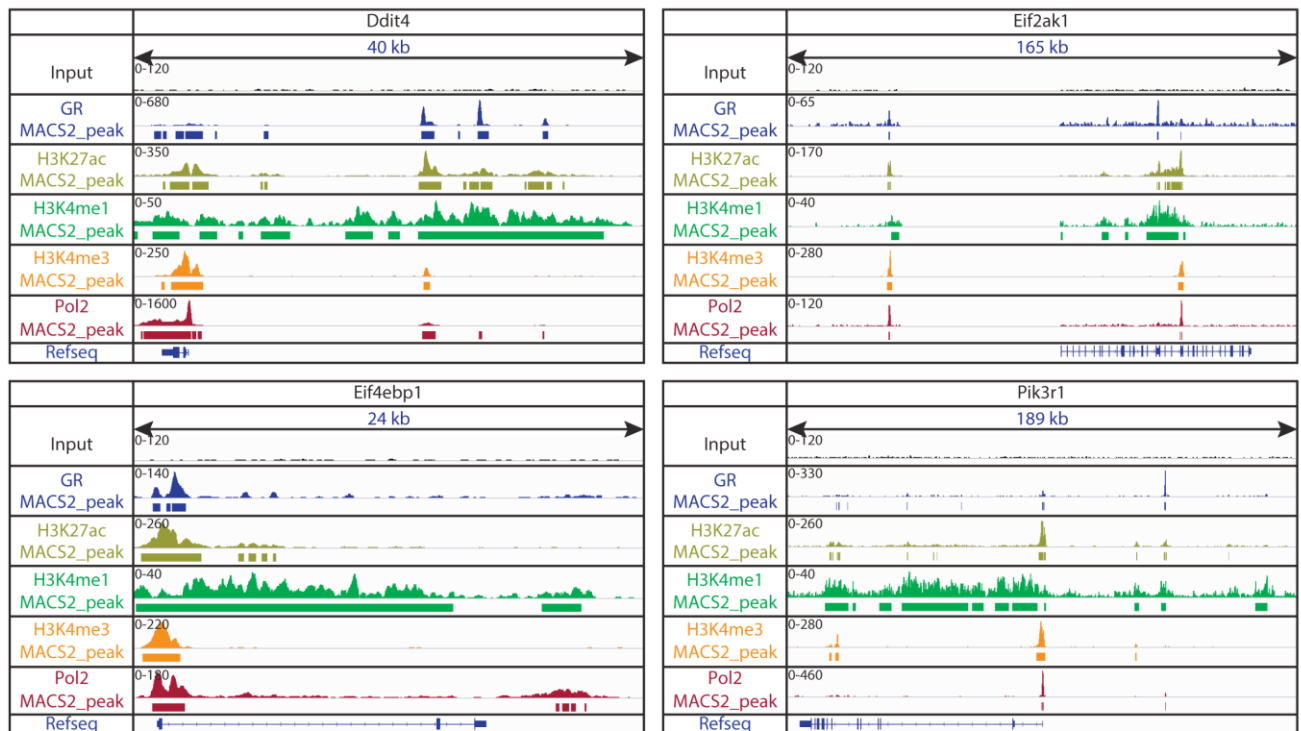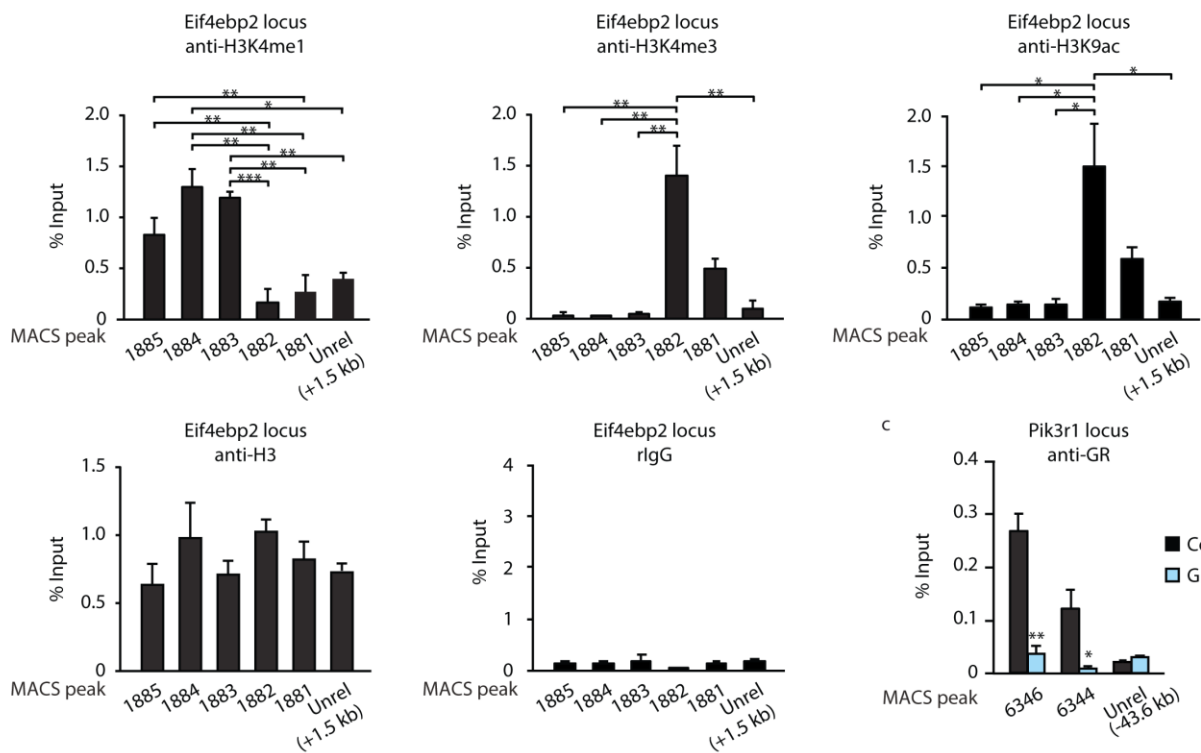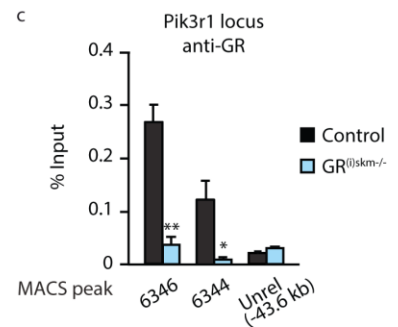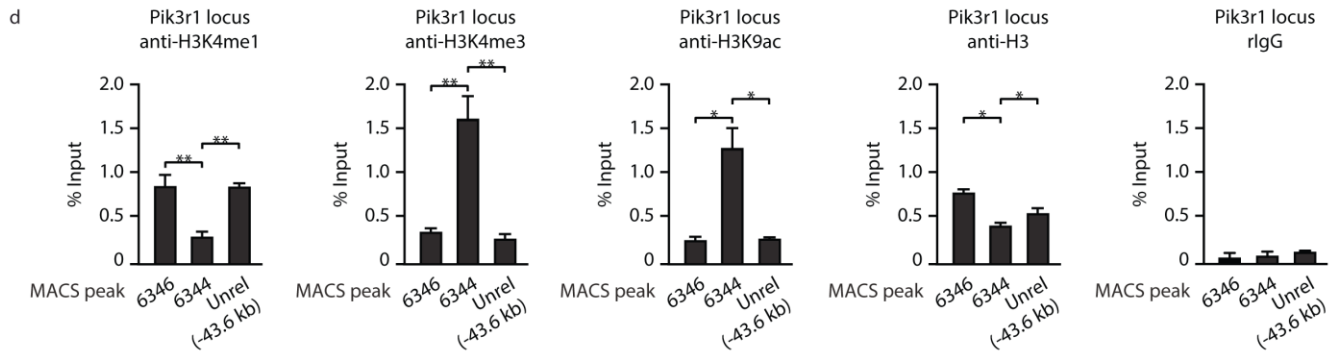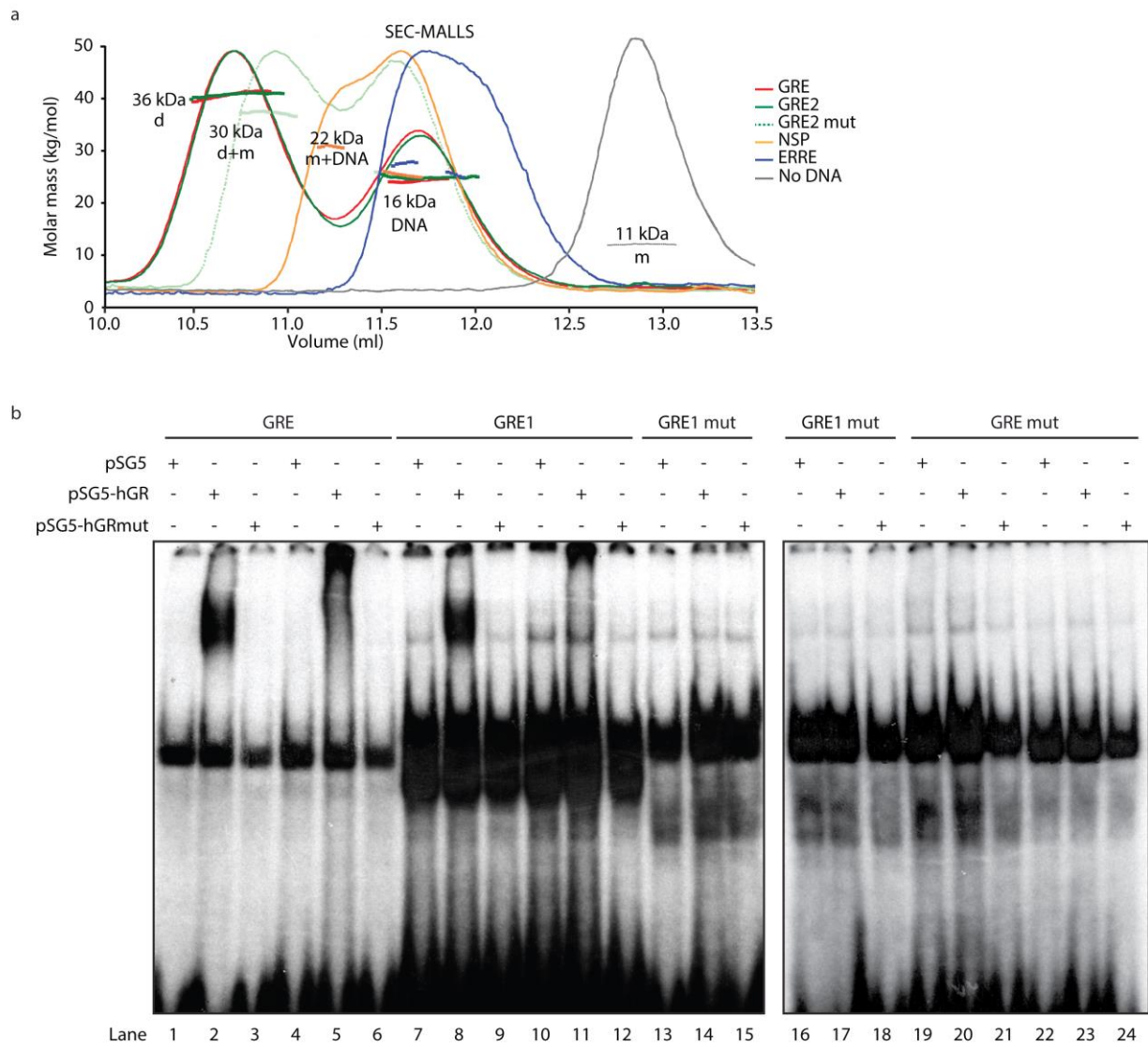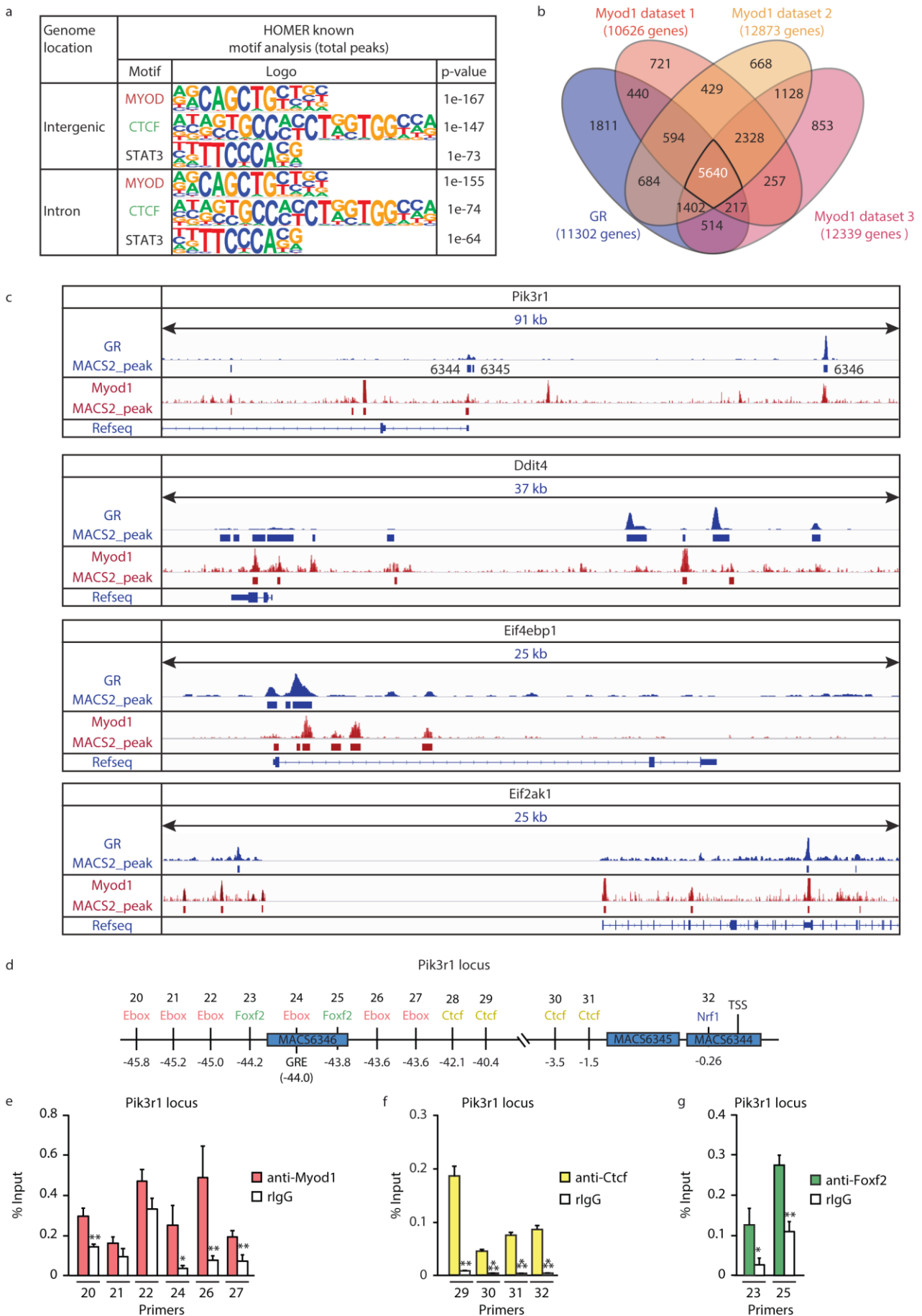
Supplementary Figure 7.
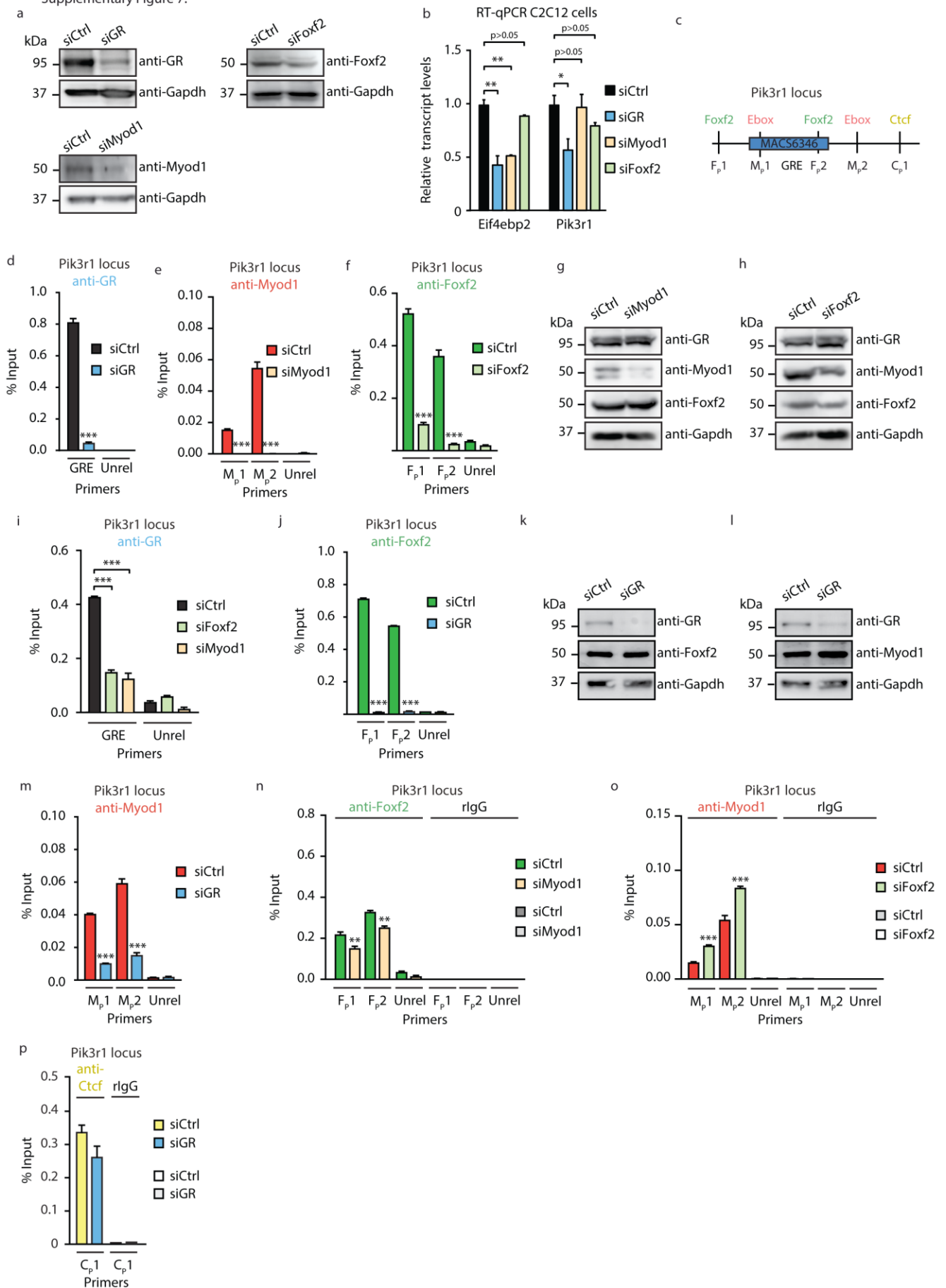
PART II

Objectives of the part II of the thesis

GR is expressed in many cell types, but it regulates distinct set of genes in various tissues. In skeletal muscle, glucocorticoids regulate glucose, lipid and protein metabolism and thus contribute to energy homeostasis (Munck et al., 1984). Given the fact that GR is also expressed in prostate, we investigated the tissue-specificity of GR signaling in prostate and compared it to skeletal muscle.

Moreover, as AR and GR recognize very similar response elements, composed of 5'-AGAACA-3' consensus sequence organized as an inverted repeat with a 3n spacer, and as androgens induce the growth and proliferation of prostatic epithelial cells, we determined AR cistrome and transcriptome in murine prostate. Furthermore, to investigate what confers to these receptors their selectivity under physiological conditions, we performed a genome-wide comparative analysis between GR and AR cistromes and transcriptomes in prostate.

Available GR ChIP-Seq data in prostate

Species	Biological source	Publication
Homo sapiens	LNCaP cells	Sahu B, et al. EMBO J. 2011
Homo sapiens	LNCaP cells	Sahu B, et al. Cancer Res. 2013
Homo sapiens	VCaP cells	Sahu B, et al. Cancer Res. 2013

Available AR ChIP-Seq data in prostate

Species	Biological source	Publication
Homo sapiens	LNCaP cells	Cato L, et al. Elife
Homo sapiens	LNCaP cells	Paltoglou S, et al. Cancer Res. 2017
Homo sapiens	VCaP cells	Mounir Z, et al. Elife 2016
Homo sapiens	LNCaP cells	Zhao JC, et al. Oncogene 2016
Homo sapiens	C4-2B cells	Zhang A, et al. Cell Rep 2015

Homo sapiens	Prostate	Pomerantz MM, et al. Nat. Genet. 2015
Homo sapiens	LNCaP cells	Takayama K, et al. Cancer Res. 2014
Homo sapiens	PC-3	Sutinen P, et al. Nucleic Acids Res. 2014
Homo sapiens	LNCaP cells	Ramos-Montoya A, et al. EMBO Mol Med 2014
Homo sapiens	LNCaP cells	Sahu B, et al. Cancer Res. 2013
Homo sapiens	LNCaP cells	McNair C, et al. Oncogene 2016
Homo sapiens	C4-2B cells	Wang J, et al. Nat. Med. 2016
Homo sapiens	LNCaP cells	Takayama K, et al. Oncotarget 2015
Homo sapiens	LNCaP cells	Barfeld SJ, et al. EBioMedicine 2017
Homo sapiens	LNCaP cells	Zhao Y, et al. Cell Rep 2016
Homo sapiens	LNCaP cells	Stelloo S, et al. Oncogene 2018
Homo sapiens	LNCaP cells	Malinen M, et al. Nucleic Acids Res. 2016
Homo sapiens	R1-D567; Epithelium; Prostate	Chan SC, et al. Nucleic Acids Res. 2015
Mus musculus	Epithelium; Prostate	Chen Y, et al. Nat. Med. 2013
Mus musculus	Epithelium; Prostate	Pihlajamaa P, et al. EMBO J. 2014
Mus musculus	Epithelium; Prostate	Sahu B, et al. Nucl Acids Res 2014

I. Genome-wide comparative analysis of glucocorticoid receptor's cistrome and transcriptome in prostate and skeletal muscle

Results

To determine GR cistrome in prostate, we performed chromatin immunoprecipitation (ChIP) followed by massively parallel DNA sequencing (Chip-Seq) of prostate tissue of 10-week-old wild-type mice and using an anti-GR antibody. 15,247,343 50-nucleotide (nt) sequence tags were obtained in prostate that were mapped uniquely to the mouse genome (mm10).

Peak calling by MACS2 (Feng et al., 2011) revealed 7940 GR-binding sites (GRBS), using a minimum FDR 0.05 (from one biological sample), which were associated to 5237 genes in prostate. The genome-wide distribution of the prostate GRBS revealed that 6% (472/7940) of them were located in proximal promoter regions (-1 kb to +100 bp), 40% (3197/7940) in intronic regions and 49% (3916/7940) in intergenic regions (Figure 1A). Similar analysis of the 23196 skm GRBS, associated to 11302 genes, revealed that 38% (8694/23196) were located in proximal promoter regions, 31% (7271/23196) in intronic regions and 25% (5672/23196) in intergenic regions (Figure 1B) (manuscript in preparation). Moreover, there is an accumulation of skm binding sites around the TSS (Figure 1D), compared to prostate (Figure 1C).

Thus, GRBS in prostate are three times less compared to GRBS in muscle and they are associated to two times less genes in prostate than in muscle. Furthermore, the genomic distribution of the GRBS in prostate is mostly enhancer regions compared to muscle which is mostly promoter regions.

De novo motif analysis of the prostate GRBS, using the software MEME-ChIP (Machanick and Bailey, 2011) from the MEME-Suite (Bailey et al., 2009), revealed 15-bp canonical Glucocorticoid Response Elements (GREs) in intergenic (434/2587 sites, e-value: 4.0e-073, 16.7%) and intron regions (427/2468 sites, e-value: 9.3e-088, 17.3%) (Figure 1E). Additionally, TP53/63/73 motifs were identified in intergenic (236/2587 sites, e-value: 5.9e-050, 9.1%) and intron regions (529/2468 sites, e-value: 5.1e-110, 21.4%) (Figure 1E). Proximal promoter regions were not enriched for GREs and TP53/63/73 motifs, but for motifs such as SP2, ZNF263, etc. (Figure 1E).

Similar analysis of the skm GRBS, had revealed 15-bp canonical GREs in intergenic (1341/5672 sites, e-value: 1e-171, 23.6%) and intronic regions (1587/7271 sites, e-value: 4.3e-274, 21.8%) (Figure 1F). Moreover, CTCF motifs were present in intergenic (155/5672 sites, e-value: 4.4e-008, 2.7%) and intronic regions (133/7271 sites, e-value: 1.4e-006, 1.8%) (Figure 1F). Proximal promoter regions were not enriched for GREs and CTCF motifs, but for SP1, YY1, NFY and NRF1 motifs (Figure 1F).

Thus, these results indicate that GR binds DNA to GREs in intergenic and intron regions in the two tissues and it interacts with distinct factors located at the promoters.

By overlapping the data from the two ChIP-Seq assays, more than 3500 genes were bound by GR in both prostate and skm (Figure 2A). Among these common genes we identified Ddit4 (Gordon et al., 2017) (Figure 2C), Pik3r1 (Antonetti et al., 1996), Foxo1/3 (Sanchez et al., 2014), Eif4ebp1 (Tsai et al.,

2015), Myog (Flynn et al., 2010), Dusp6 (Pourteymour et al., 2017), as well as Fkbp5 (Ni et al., 2010), Pmepa1 (Liu et al., 2011), Plpp1 (Ppap2a), Plpp3(Ppap2b) and Cdkn1a (Jain et al., 2013) (Figure 2D).

In addition, using bedtools (Quinlan and Hall, 2010), we identified almost 1700 overlapping regions (for at least 1 bp) among the GRBS in skm and prostate (Figure 2B). A selection of genes bound by GR in the two tissues and their number of GRBS in each tissue are shown in Table 1.

Table 1. Genes bound by GR in skm and prostate and their number of GRBS.

Genes	GRBS skm	GRBS prostate
Ddit4 - DNA damage inducible transcript 4	10	4
Pik3r1 - phosphoinositide-3-kinase regulatory subunit 1	16	10
Foxo1 - forkhead box O1	8	2
Foxo3 - forkhead box O3	10	4
Eif4ebp1 - eukaryotic translation initiation factor 4E binding protein 1	3	1
Myog - myogenin	5	1
Dusp6 - dual specificity phosphatase 6	3	1
Fkbp5 - FK506 binding protein 5	7	1
Pmepa1 - prostate transmembrane protein, androgen induced 1	5	1
Plpp1 - phospholipid phosphatase 1	6	1
Plpp3 - phospholipid phosphatase 3	11	4
Cdkn1a - cyclin-dependent kinase inhibitor 1A (P21)	4	2
Sgk1 -serum/glucocorticoid regulated kinase 1	7	4

The 5627 prostate GRBS located in the genes bound by GR in both tissues were at 42% (2349/5627) in intronic regions, 47% (2645/5627) in intergenic regions, and only 7% (397/5627) in proximal promoter regions (Figure 2E). The 10135 skm GRBS located in genes bound by GR in both tissues were at 37% (3751/10135) in intronic regions, 33% (3351/10135) in intergenic regions and 24% (2461/10135) in proximal promoter regions (Figure 2F). Interestingly, the skm GRBS, located in genes bound by GR in both tissues, decreased in promoter regions compared to the bulk GRBS in skm, whereas the ones in intergenic and intronic regions increased (Table 2). In addition, the prostate GRBS, located in genes bound by GR in both tissues, increased in promoter and intronic regions compared to the bulk GRBS in prostate, whereas the ones in intergenic regions decreased (Table 2).

Table 2. Recapitulative table of the genomic distribution of the GRBS in skm and prostate.

Genomic location	Bulk GRBS skm	GRBS in both tissues Skm coordinates	Bulk GRBS prostate	GRBS in both tissues Prostate coordinates
promoter	38%	24%	6%	7%
intergenic	25%	33%	49%	47%
intron	31%	37%	40%	42%

De novo motif analysis of the GRBS bound by GR in both tissues, using the skm GR coordinates, revealed 15-bp canonical GREs in intergenic (1019/3351 sites, e-value: 1.4e-285, 30.4%) and intronic regions (1171/3751 sites, e-value: 1.5e-313, 31.2%) (Figure 3B). Moreover, CTCF motifs were identified only in intergenic regions (105/3351 sites, e-value: 1.3e-035, 3.1%) (Figure 3B). In proximal promoter regions no GREs or CTCF motifs were enriched, but NRF1, YY1, C2H2 zinc finger factors, NFY, STAT3, bZIP and bHLH motifs were identified.

The same procedure, using the prostate GR coordinates, revealed 15-bp canonical GREs in intergenic (286/2051 sites, e-value: 1.5e-087, 13.9%) and intronic regions (329/2018 sites, e-value: 3.0e-113, 16.3%) (Figure 3C). Furthermore, TP53/63/73 motifs were identified in intergenic (199/2051 sites, e-value: 6.3e-072, 9.7%) and intronic regions (421/2018 sites, e-value: 2.0e-118, 20.9%) (Figure 3C). In proximal promoter regions no GREs or TP53/63/73 motifs were enriched, but ZNF263, SP1, SP2, KLF5 motifs.

Of note, the CTCF motifs were found only in skm and not in prostate. They were located either close to GREs (e.g. Tmem45b, 1 bp between the two motifs) either from 7 bp to more than 145 kb from each other (e.g. Tsc22d3, Per1, Hpacl1, Plxdc1, Tmem107, Txn2, Tbcb, Mrps31, Nnmt, Wee1, Rfx8, Morn3, Cux1 ([Arthur et al., 2017](); [Whalen et al., 2016]()), Usp3) (For more detailed representation of GREs and CTCF motifs localization see Results part I).

De novo motif analysis of the skm-specific GRBS revealed 15-bp canonical GREs in intergenic (510/2321 sites, e-value: 1.6e-129, 21.9%) and intronic regions (914/3520 sites, e-value: 1.6e-225, 25.9%) (Figure 4B). Furthermore, CTCF motifs were identified in intergenic (87/2321 sites, e-value: 5.2e-050, 3.7%) and intron regions (48/3520 sites, e-value: 1.5e-002, 1.4%) (Figure 4B). In promoter regions no GREs or CTCF motifs were enriched, but YY1, NRF1, NFYs, STATs, bZIP, bHLH and ETS-related motifs.

De novo motif analysis of the prostate-specific GRBS also revealed 15-bp canonical GREs in intergenic (74/570 sites, e-value: 2.7e-034, 12.9%) and intronic regions (91/500 sites, e-value: 9.8e-021, 18.2%) (Figure 4C). Moreover, TP53/63/73 motifs were identified in intergenic (141/570 sites, e-value: 1.7e-064, 24.7%) and intronic regions (139/500 sites, e-value: 5.2e-070, 27.8%) (Figure 4C). In proximal promoter regions no GREs or TP53/63/73 or other significant motifs where identified.

Of note, again the CTCF motifs were found only in skm and not in prostate. They were located either side by side to GREs (e.g. Trip10, no bp between the two motifs), either from 6 bp to more than 100 kb from each other (e.g. Sept5, Mustn1, Ptk2b, Cacna1s, Hsp90ab1, Samd4 re, Eif4g1, Ttc8, Fam124b, Mrp145, Ltb4r2, Trim47, Des).

Thus, GR binds DNA to GREs in intergenic and intron regions and might interact or cooperate with other distinct factors depending on the tissue.

Functional annotation of the genes bound by GR in both tissues, skm-specific genes and prostate-specific genes, revealed many enriched common pathways, but also specific ones (in bold) (Table 3).

Table 3. Enriched shared and unique pathways and their FDR values.

KEGG Pathways	Genes bound by GR in both tissues (FDR)	Skm-specific genes (FDR)	Pro-specific genes (FDR)
MAPK signaling pathway	2.44e-07	2.2130e-14	1.08e-02
FoxO signaling pathway	4.24e-06	1.6076e-8	7.62e-03
p53 signaling pathway	3.5e-05	0.0006	2e-04
Insulin signaling pathway	3.48e-04	1.2838e-9	0.0006
Apoptosis signaling pathway	2.1e-02	0.000005	4.22e-02
Adipogenesis genes	5.65e-03	0.0014	2.41e-02
HIF-1 signaling pathway	9.07e-06	-	3.11e-02
Wnt signaling pathway	6.82e-05	-	0.00003
PI3K-Akt signaling pathway	2.3e-02	-	0.0003
Rap1 signaling pathway	0.0002	-	0.0003
Prostate cancer	-	6.0655e-7	0.0018
Integrin signaling pathway	2.53e-02	-	-
Regulation of lipid metabolism by PPARalpha	1.26e-02	-	-
Ras signaling pathway	-	-	0.0021
Circadian rhythm	-	0.000001	-
Glucagon signaling pathway	-	2.9237e-8	-
Cell cycle	-	2.9237e-8	-
mTOR signaling pathway	-	2.8967e-8	-

To determine which of the genes bound directly by GR in skm and prostate are expressed, we overlapped the ChIP-Seq data with RNA-Seq data from skm and prostate (Figure 5A). The RNA-Seq was performed on extracted muscles of 10-week-old wild-type mice and on prostates of 26-28 weeks old wild-type mice. The overlap among genes with at least one GRBS in skm and prostate and expressed in both tissues (number of reads greater than 100) revealed almost 2100 genes (Figure 5A). Using a heatmap to visualize the mean centered normalized expression of the 2100 shared genes, we observed that 62% of the genes (1306/2102) are "prostate-specific" and 38% (796/2102) are "skm-specific (Figure 5B). Pathway analysis of the "skm-specific" genes revealed terms such as actin filament-based process, regulation of muscle adaptation, mitochondrion organization, regulation of smooth muscle cell proliferation and others (Figure 5C). Similar analysis of the "prostate-specific" genes revealed terms such as apoptotic signaling pathway, gland development, cellular response to peptide hormone stimulus, negative regulation of cell differentiation and others (Figure 5D). *De novo* motif analysis of the "skm-specific" genes revealed canonical and half-site GREs, MEF2A/B/C/D and bHLH motifs (Myod1, Myog) (Figure 5E). Similar analysis of the "prostate-specific" genes revealed also canonical and half-site GREs, Forkhead box factors, TEAD2 and SMAD/NF-1 DNA binding domain factors. No CTCF and TP53/63/73 motifs were

identified this time and the response elements of GR are very similar between the two tissues. Thus, the tissue-specificity mainly depends on the distinct surrounding factors.

After the gene analysis, we performed a binding-sites analysis in order to get rid of the annotation biases. We used seqMINER in order to visualize the tag density maps of the GR ChIP-seq reads from skm and prostate, within +/-5kb, and using all the skm GRBS as reference. The heatmap revealed **shared** sites between skm and prostate, but also **unique** to skm (Figure 6A). The GR cistrome in skm exhibit major overlap with that of prostate and 56% of the GRBS (11893/21318) are shared by the two tissues (cluster 1, 2, 3). The rest of the binding sites are unique to GR in skm (cluster 4). The average tag density profiles of the clusters are depicted in Figure 6B.

De novo motif analysis of the sequences below all the binding sites of the clusters revealed canonical and half-sites of GREs (Figure 6C, D). Moreover, similar analysis of the shared clusters revealed additional motifs from factors like STAT3, NFYA/B, NRF1, YY1 motifs and ETS-related factors (Figure 6C). *De novo* motif analysis of the sequences below the binding sites of the unique cluster revealed additional motifs from factors such as bHLH (Myod1, Myog), MEF2A/B/C/D, TEAD2 and CTCF (Figure 6D).

Thus, GR binds to canonical and half-site GREs in skm and prostate and the different motifs found close to GREs are probably imposing the tissue specificity of the receptor.

The tag density maps of GR, H3K4me1, H3K27ac, H3K4me3 and Pol2 ChIP-Seq reads from skm, using the seqMINER, within +/-5kb, and using all the skm GRBS as reference, revealed a genomic co-localisation of GRBS with the histones marks and the Pol2 (Figure 7A). H3K27ac and H3K4me1 are marks of active enhancer and promoters and H3K4me3 is a mark of active promoter. The average tag density profiles of H3K4me1 are enriched around the center of the binding sites, where the signal is depleted (Figure 7B). The average tag density profiles of H3K27ac and H3K4me3 are also enriched around the center of the binding sites (Figure 7B). The average tag density profiles of Pol2 and GR are enriched at the center of binding sites (Figure 7B). The genome-wide distribution of the binding sites was mostly promoter regions for the first two clusters and enhancer regions for the last two clusters (Figure 7C).

De novo motif analysis of the sequences below the binding sites of the 1st and the 2nd cluster revealed NFYA/B, YY1, STAT3, NRF1 motifs and ETS-related factors (Figure 7D). No GREs were identified under the binding sites of the first two clusters. Similar analysis of the sequences below the binding sites of the 3rd and the 4th cluster revealed canonical and half-site GREs, MEF2A/B/C/D motifs and only at the 4th cluster CTCF motifs (Figure 7D).

Thus, the first two clusters, contain the binding sites at active promoter regions, validated by the presence of NRF1 motifs, the presence of Pol2 and H3K4me3 that co-localize, the low H3K4me1 levels and the genomic distribution of the binding sites which are mostly located at promoter regions in both clusters. The last two clusters contain binding sites at active enhancer regions, validated by the fact that GR binds to GREs, there are low H3K4me3 levels, H3K27ac, H3K4me1 and

Pol2 are present and the genomic distribution of the binding sites that are mostly located at enhancer regions in both clusters.

Discussion

This comparative genome-wide analysis demonstrated that under physiological conditions GR is bound to three times more sites in muscle than in prostate, and these sites are located in two times more genes. This fact is probably due to the efficacy of the ChIP-Seq in prostate, as such experiments in tissues can be really challenging, or due to the depth of sequencing.

The distribution of the binding sites in prostate is mostly in enhancer regions compared to muscle that is mostly in promoter regions. Moreover, *de novo* motif analysis of GRBS in both tissues, revealed GREs in intergenic and intronic regions but not in proximal promoter regions, where GR might interact with distinct factors. Of note, we also identified distinct factors close to GREs, like TP53/63/73 in prostate and CTCF in skm that probably define the tissue specificity of GR.

GR was bound to almost 3500 genes in both prostate and skm. *De novo* motif analysis of the GRBS located in the 3500 genes, revealed GREs in intergenic and intronic regions but not in proximal promoter regions, where GR might interact with distinct factors depending on the tissue. Of note, in intergenic and intronic regions, we also identified distinct factors close to GREs, such as TP53/63/73 in prostate and CTCF in skm. Similarly, *de novo* motif analysis of the tissue-specific GRBS revealed GREs in intergenic and intronic regions, as well as TP53/63/73 motifs in prostate and CTCF, MEF2A/B/C/D motifs in skm, but not in proximal promoter regions. Depending on the tissue, the co-factors of GR are different.

Detailed analysis of the response elements in the different tissues and genomic repartitions revealed that the identified GREs bound by GR do not exhibit any selectivity between the two tissues and that the tissue-specificity of GR is defined by the distinct surrounding co-factors.

CTCF is known to anchor chromatin loops (Rao et al., 2014; Splinter et al., 2006). In the manuscript in preparation (Results part I), the lab already demonstrated that CTCF is recruited at genomic regions located in the vicinity of GREs, in intergenic and intronic regions. However, in prostate, no CTCF motifs were identified.

P53 is a tumor suppressor and potent inhibitor of cell growth. P73 and P63 similar to P53 in amino acid sequence and structure (Courtois et al., 2004). There is evidence for negative cross-talk between GR and p53 (Sengupta et al., 2000; Sengupta and Wasylyk, 2001) and there is a study examining the ability of p53 and p73 to interact with and inhibit GR transcriptional activity (Zhang et al., 2006). Thus, probably GR interacts with p53/63/73, in intergenic and intronic regions, to modulate transcription.

The cistromic overlap of GR in prostate and skm, by using the skm GRBS as reference, revealed that 56% of them are shared between the two tissues, although there are unique sites to skm. Similar analysis, using the prostate GRBS as reference, could provide insights into the unique sites in prostate.

De novo motif analysis of all the binding sites revealed canonical and half-site GREs. The shared binding sites were also enriched for motifs of distinct factors that GR interacts and they probably define the tissue specificity of the receptor, such as STAT3; a chromatin associated factor and tethering partner of GR (Langlais et al., 2012; Petta et al., 2016; Zhang et al., 1997), YY1 (Breslin and Vedeckis, 1998), NRF1, NFYA/B and ETS-related factors (Starick et al., 2015). Motifs of MEF2A/B/C/D, CTCF, TEA domain factors; "partnering" proteins of GR in a composite binding site (Starick et al., 2015), and of bHLH factors, such as Myod (Oakley et al., 2017), were also identified close to GREs, but only in muscle. In the manuscript in preparation (Results part I), it was already demonstrated that GR cooperates with Stat3 and Myod1 in muscle and that their cognate response elements are in the vicinity of enhancer-GRE containing regions. In addition, the lab demonstrated that enhancer GRE-bound GR interacts and communicates with Nrf1, bound to its response element located in promoter regions, to regulate gene transcription.

Finally, we report that in skm GR binds at active promoter regions, validated by the absence of GREs at these sites, the presence of NRF1, the high levels of Pol2 and H3K4me3, the co-localization of the two, the low H3K4me1 levels and the genomic distribution of the binding sites that are mostly located at promoter regions. GR in skm also binds at active enhancer regions, validated by the presence of enriched GREs at these sites, there are low H3K4me3 levels, H3K27ac, Pol2 and H3K4me1 are present and the genomic distribution of the binding sites that is mostly enhancer regions. In addition, GR interacts with distinct factors in promoter regions (NRF1, NFYA/B, YY1, STAT3, ETS-related factors) and cooperates with others in enhancers regions (MEF2A/B/C/D, CTCF) that are probably imposing the genomic-specificity of the binding in skm.

Further studies are needed to elucidate the mechanism(s) of the specificity of GR binding in the different tissues, as there are similarities but also distinct differences. Importantly, our results indicate that the specificity of the responses is not only based ion the tissue-specific expression of the receptor, but also on the cooperation with distinct surrounding transcription factors, rather than the sequence of the response elements.

Taken together, these results provide the basis of a molecular understanding of tissue-specific activity of glucocorticoids, and thus open new avenues to design screens for analogs, inducing genes selectively.

Materials and methods

Mice

Experiment	Type	Age	Phenotype	Samples
GR ChIP-Seq skm	C57BL/6	10 weeks	Wild-type	1
GR ChIP-Seq prostate	C57BL/6	10 weeks	Wild-type	1
RNA-Seq skm	C57BL/6	10 weeks	Wild-type	4
RNA-Seq prostate	C57BL/6	26-28 weeks	Wild-type	5

Chromatin immunoprecipitation assays and ChIP-sequencing

ChIP assays were performed as described previously (Joshi et al., 2017) and (Ueberschlag-Pitiot, Rovito, Rerra, et al. manuscript in preparation).

Sequencing was performed by the IGBMC Microarray and Sequencing platform, a member of the "France Génomique" consortium (ANR-10-INBS-0009). Immunoprecitated DNA samples were processed for library preparation on Illumina Hiseq 4000 sequencer as a single-read 50 base reads following Illumina's instructions. Image analysis and base calling were performed using RTA 2.7.7 and bcl2fastq 2.17.1.14. Adapter dimer reads were removed using DimerRemover. The FastQC 0.11.2 was used to evaluate the quality of the sequencing. Sequenced reads were mapped to the mus musculus genome assembly 10 (mm10) using Bowtie 2 (Langmead and Salzberg, 2012). Uniquely mapped reads were used for further analysis. The peak calling was performed with the MACS2 algorithm (Feng et al., 2011) using appropriate inputs. Peaks were annotated relative to genomic features using Homer (Heinz et al., 2010) according to the distance to the nearest TSS. Distance to TSS was calculated using the online software GREAT (McLean et al., 2010). Data visualization was carried out using Integrative Genomics Viewer (IGV) (Robinson et al., 2011). *De novo* motif analysis of the binding sites by adding 100 nucleotides at both side of the peak summit was performed by using the online software MEME-ChIP (Machanick and Bailey, 2011) from MEME-Suite (Bailey et al., 2009) after extracting their nucleotide sequences. The *de novo* motifs were then compared to a database of known motifs and ranked by the TOMTOM tool (Gupta et al., 2007) of MEME Suite. The tag density maps were produced using the software SeqMiner (Ye et al., 2011) and the clustering normalization was done using the KMeans linear method. The intersection of intervals was performed with the intersect function of bedtools (Quinlan and Hall, 2010).

Antibodies

- anti-GR IGBMC 3249, fraction no4 (used in skm and pro assays)
- anti-Pol2 Santa Cruz (H224, SC9001) (used in skm assay)
- anti-H3K4me1 Active Motif (39297) (used in skm assay)
- anti-H3K27ac Active Motif (39133) (used in skm assay)
- anti-H3K4me3 Abcam (1012-100) (used in skm assay)

RNA extraction and RNA sequencing

Total RNA was isolated from control gastrocnemius muscle samples using Trizol and from control prostate samples using RNeasy Micro Kit (74004) from Qiagen, reverse transcribed (RT) using SuperScript II Reverse Transcriptase (Invitrogen) and amplified by quantitative PCR with the SYBER Green kit (Roche) and LightCycler 480 (Roche Diagnostics) according to the manufacturer's instructions.

Sequencing was performed by the IGBMC Microarray and Sequencing platform, a member of the "France Génomique" consortium (ANR-10-INBS-0009). The library was prepared on Illumina Hiseq 4000 sequencer as a single-read 50 base reads following Illumina's instructions. Image analysis and base calling were performed using RTA 2.7.7 and bcl2fastq 2.17.1.14. Adapter dimer reads were removed using DimerRemover. The FastQC 0.11.2 was used to evaluate the quality of the sequencing. Reads were mapped onto the mm10 assembly of mouse genome using Tophat 2.1.1 (Kim et al., 2013) and the Bowtie2 2.3.4.3 (Langmead and Salzberg, 2012). Only uniquely aligned reads have been retained for further analyses. Quantification of gene expression was performed using HTSeq-0.11.0 (Anders et al., 2015). If the raw read counts for one gene were greater than 100, then the gene was considered expressed. Hierarchical clustering was performed using Cluster 3.0 (de Hoon et al., 2004) and the heatmap was visualized using the Java TreeView (Saldanha, 2004).

The pathway analyses were performed using the online softwares WebGestalt GSAT (Wang et al., 2013), and more specifically the Over-Representation Analysis (ORA) method, and the Metascape (Zhou et al., 2019).

References

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Antonetti, D.A., Algenstaedt, P., and Kahn, C.R. (1996). Insulin receptor substrate 1 binds two novel splice variants of the regulatory subunit of phosphatidylinositol 3-kinase in muscle and brain. Mol Cell Biol *16*, 2195-2203.

Arthur, R.K., An, N., Khan, S., and McNerney, M.E. (2017). The haploinsufficient tumor suppressor, CUX1, acts as an analog transcriptional regulator that controls target genes through distal enhancers that loop to target promoters. Nucleic Acids Res *45*, 6350-6361.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res *37*, W202-208.

Breslin, M.B., and Vedeckis, W.V. (1998). The human glucocorticoid receptor promoter upstream sequences contain binding sites for the ubiquitous transcription factor, Yin Yang 1. J Steroid Biochem Mol Biol *67*, 369-381.

Courtois, S., Caron de Fromentel, C., and Hainaut, P. (2004). p53 protein variants: structural and functional similarities with p63 and p73 isoforms. Oncogene *23*, 631-638.

de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453-1454.

Feng, J., Liu, T., and Zhang, Y. (2011). Using MACS to identify peaks from ChIP-Seq data. Curr Protoc Bioinformatics *Chapter 2*, Unit 2 14.

Flynn, J.M., Meadows, E., Fiorotto, M., and Klein, W.H. (2010). Myogenin regulates exercise capacity and skeletal muscle metabolism in the adult mouse. PLoS One *5*, e13535.

Gordon, B.S., Steiner, J.L., Rossetti, M.L., Qiao, S., Ellisen, L.W., Govindarajan, S.S., Eroshkin, A.M., Williamson, D.L., and Coen, P.M. (2017). REDD1 induction regulates the skeletal muscle gene expression signature following acute aerobic exercise. Am J Physiol Endocrinol Metab *313*, E737-E747.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol *8*, R24.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Jain, A.K., Raina, K., and Agarwal, R. (2013). Deletion of p21/Cdkn1a confers protective effect against prostate tumorigenesis in transgenic adenocarcinoma of the mouse prostate model. Cell Cycle *12*, 1598-1604.

Joshi, S., Ueberschlag-Pitiot, V., Metzger, D., and Davidson, I. (2017). Improved Protocol for Chromatin Immunoprecipitation from Mouse Skeletal Muscle. J Vis Exp.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Langlais, D., Couture, C., Balsalobre, A., and Drouin, J. (2012). The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. Mol Cell *47*, 38-49.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Liu, R., Zhou, Z., Huang, J., and Chen, C. (2011). PMEPA1 promotes androgen receptor-negative prostate cell proliferation through suppressing the Smad3/4-c-Myc-p21 Cip1 signaling pathway. J Pathol *223*, 683-694.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696-1697.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Munck, A., Guyre, P.M., and Holbrook, N.J. (1984). Physiological functions of glucocorticoids in stress and their relation to pharmacological actions. Endocr Rev *5*, 25-44.

Ni, L., Yang, C.S., Gioeli, D., Frierson, H., Toft, D.O., and Paschal, B.M. (2010). FKBP51 promotes assembly of the Hsp90 chaperone complex and regulates androgen receptor signaling in prostate cancer cells. Mol Cell Biol *30*, 1243-1253.

Oakley, R.H., Busillo, J.M., and Cidlowski, J.A. (2017). Cross-talk between the glucocorticoid receptor and MyoD family inhibitor domain-containing protein provides a new mechanism for generating tissue-specific responses to glucocorticoids. J Biol Chem *292*, 5825-5844.

Petta, I., Dejager, L., Ballegeer, M., Lievens, S., Tavernier, J., De Bosscher, K., and Libert, C. (2016). The Interactome of the Glucocorticoid Receptor and Its Influence on the Actions of Glucocorticoids in Combatting Inflammatory and Infectious Diseases. Microbiol Mol Biol Rev *80*, 495-522.

Pourteymour, S., Hjorth, M., Lee, S., Holen, T., Langleite, T.M., Jensen, J., Birkeland, K.I., Drevon, C.A., and Eckardt, K. (2017). Dual specificity phosphatase 5 and 6 are oppositely regulated in human skeletal muscle by acute exercise. Physiol Rep *5*.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.
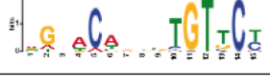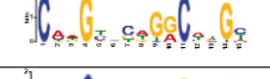
Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S*., et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246-3248.

Sanchez, A.M., Candau, R.B., and Bernardi, H. (2014). FoxO transcription factors: their roles in the maintenance of skeletal muscle homeostasis. Cell Mol Life Sci *71*, 1657-1671.

Sengupta, S., Vonesch, J.L., Waltzinger, C., Zheng, H., and Wasylyk, B. (2000). Negative cross-talk between p53 and the glucocorticoid receptor and its role in neuroblastoma cells. EMBO J *19*, 6051-6064.

Sengupta, S., and Wasylyk, B. (2001). Ligand-dependent interaction of the glucocorticoid receptor with p53 enhances their degradation by Hdm2. Genes Dev *15*, 2367-2380.

Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev *20*, 2349-2354.

Starick, S.R., Ibn-Salem, J., Jurk, M., Hernandez, C., Love, M.I., Chung, H.R., Vingron, M., Thomas-Chollier, M., and Meijsing, S.H. (2015). ChIP-exo signal associated with DNA-binding motifs provides insight into the genomic binding of the glucocorticoid receptor and cooperating transcription factors. Genome Res *25*, 825-835.

Tsai, S., Sitzmann, J.M., Dastidar, S.G., Rodriguez, A.A., Vu, S.L., McDonald, C.E., Academia, E.C., O'Leary, M.N., Ashe, T.D., La Spada, A.R*., et al.* (2015). Muscle-specific 4E-BP1 signaling activation improves metabolic parameters during aging and obesity. J Clin Invest *125*, 2952-2964.

Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res *41*, W77-83.

Whalen, S., Truty, R.M., and Pollard, K.S. (2016). Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. Nat Genet *48*, 488-496.

Ye, T., Krebs, A.R., Choukrallah, M.A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res *39*, e35.

Zhang, L., Nie, L., and Maki, C.G. (2006). P53 and p73 differ in their ability to inhibit glucocorticoid receptor (GR) transcriptional activity. Mol Cancer *5*, 68.

Zhang, Z., Jones, S., Hagood, J.S., Fuentes, N.L., and Fuller, G.M. (1997). STAT3 acts as a co-activator of glucocorticoid receptor signaling. J Biol Chem *272*, 30607-30610.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun *10*, 1523.

Figure 1.

A   prostate GRBS

B   muscle GRBS

C   prostate GRBS

D   muscle GRBS
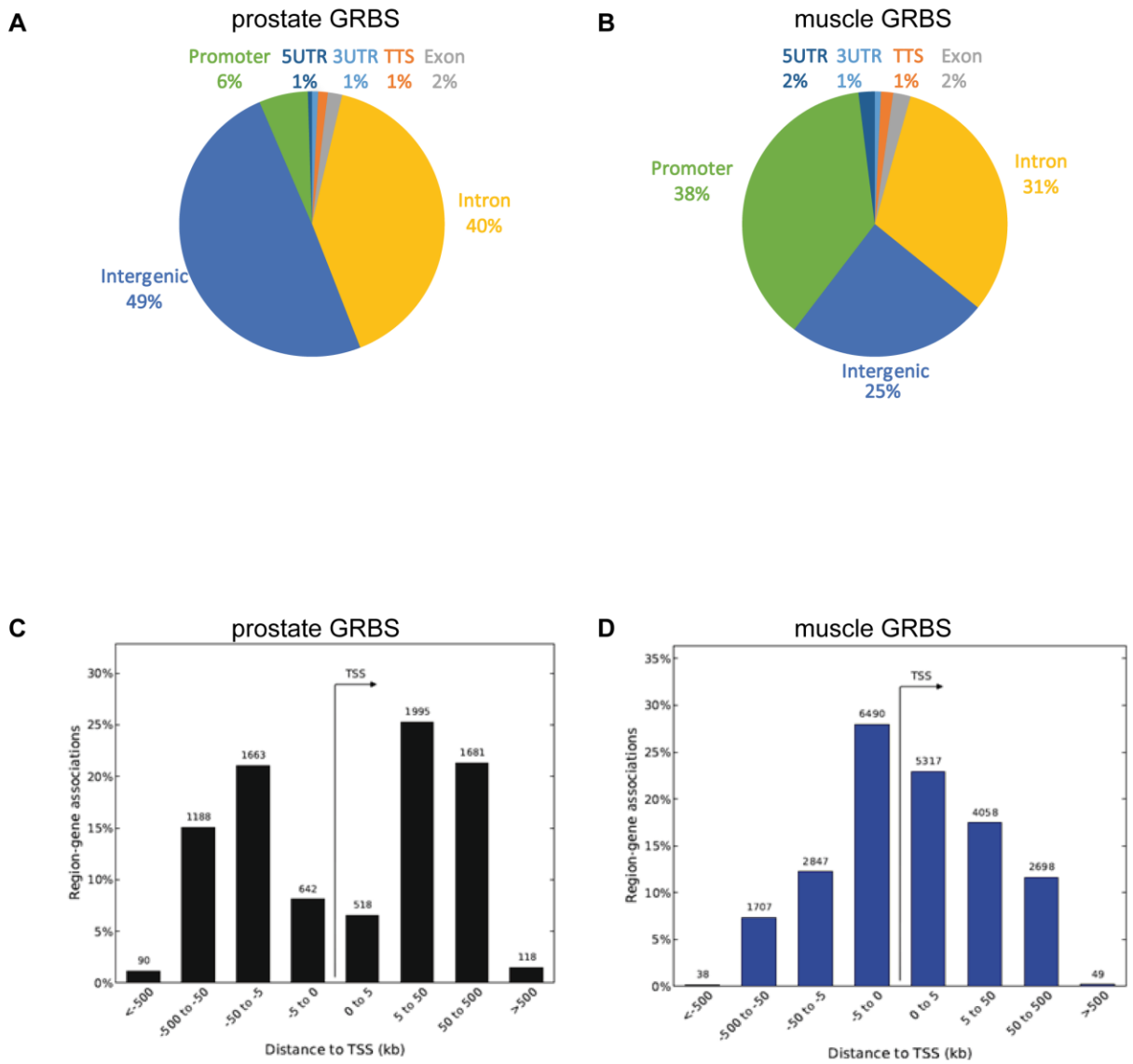
E   prostate GR coordinates

Genomic location	Logo	Motif	Sites	E-value
promoter		ZNF263, E2F6/4, SP2	82/447	1.8e-012
intergenic		GRE	434/2587	4.0e-073
intergenic		TP53/63/73	236/2587	5.9e-050
intron		GRE	427/2468	9.3e-088
intron		TP53/63/73	529/2468	5.1e-110

F   muscle GR coordinates
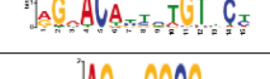
Genomic location	Logo	Motif	Sites	E-value
promoter		SP1	3944/8694	1.3e-165
promoter		YY1	338/8694	1.9e-016
promoter		NFYA	276/8694	5.2e-004
promoter		NRF1	672/8694	5.3e-024
intergenic		GRE	1341/5672	4.0e-204
intergenic		CTCF	155/5672	4.4e-008
intron		GRE	1587/7271	4.3e-274
intron		CTCF	133/7271	1.4e-006

Figure 1. GRBS in skm and prostate

A. Genome-wide distribution of the prostate GR binding sites (GRBS).

B. Genome-wide distribution of the skm GR binding sites (GRBS).

C. Distance to TSS in kb of the prostate GRBS.

D. Distance to TSS in kb of the skm GRBS.

E. Significant cis-elements of the prostate GRBS as identified by *de novo* motif searches.

F. Significant cis-elements of the skm GRBS as identified by *de novo* motif searches.

Figure 2.

A genes bound by GR in skm / genes bound by GR in prostate

7803 | 3499 | 1738

B GRBS skm / GRBS pro

23337 | 1666 | 7799

C

156 kb

Input skm

GR skm

Ddit4 Ddit4 Ddit4 Ddit4 Ddit4 Ascc1 Ascc1

Input pro

GR pro

Ddit4 Ddit4 Ddit4 Ascc1 Ascc1

Ddit4 Anapc16 Ascc1

D

114 kb

Input skm

GR skm

Trp53cor1 Cdkn1a Rab44

Input pro

GR pro

Trp53cor1 Trp53cor1 Cdkn1a

Trp53cor1 Cdkn1a Rab44 Cpne5

E GRBS correspoding to genes bound by GR in prostate and skm - prostate coordinates

F GRBS correspoding to genes bound by GR in prostate and skm - skm coordinates

Figure 2. Overlap of genes bound by GR in skm and prostate.

A. Venn diagram of genes bound by GR in skm and prostate.

B. Overlapping regions among GRBS in skm and prostate.

C. Location of GRBS in the regulatory regions of **Ddit4** in skm and prostate. The input is in black, the skm track is in red and the prostate track is in green.

D. Location of GRBS in the regulatory regions of **Cdkn1a** in skm and prostate. The input is in black, the skm track is in red and the prostate track is in green.

E. Genome-wide distribution of the GRBS corresponding to the genes bound by GR in skm and prostate using prostate GR coordinates.

F. Genome-wide distribution of the GRBS corresponding to the genes bound by GR in skm and prostate using skm GR coordinates.

Figure 3.

A

B   skm GR coordinates

Genomic location	Logo	Motif	Sites	E-value
intergenic		**GRE**	**1019/3351**	**1.4e-285**
intergenic		CTCF	105 /3351	1.3e-035
intron		**GRE**	**1171/3751**	**1.5e-313**

C   prostate GR coordinates

Genomic location	Logo	Motif	Sites	E-value
intergenic		**GRE**	**286/2051**	**1.5e-087**
intergenic		TP63/73/53	199/2051	6.3e-072
intron		**GRE**	**329/2018**	**3.0e-113**
intron		TP63/73/53	421/2018	2.0e-118
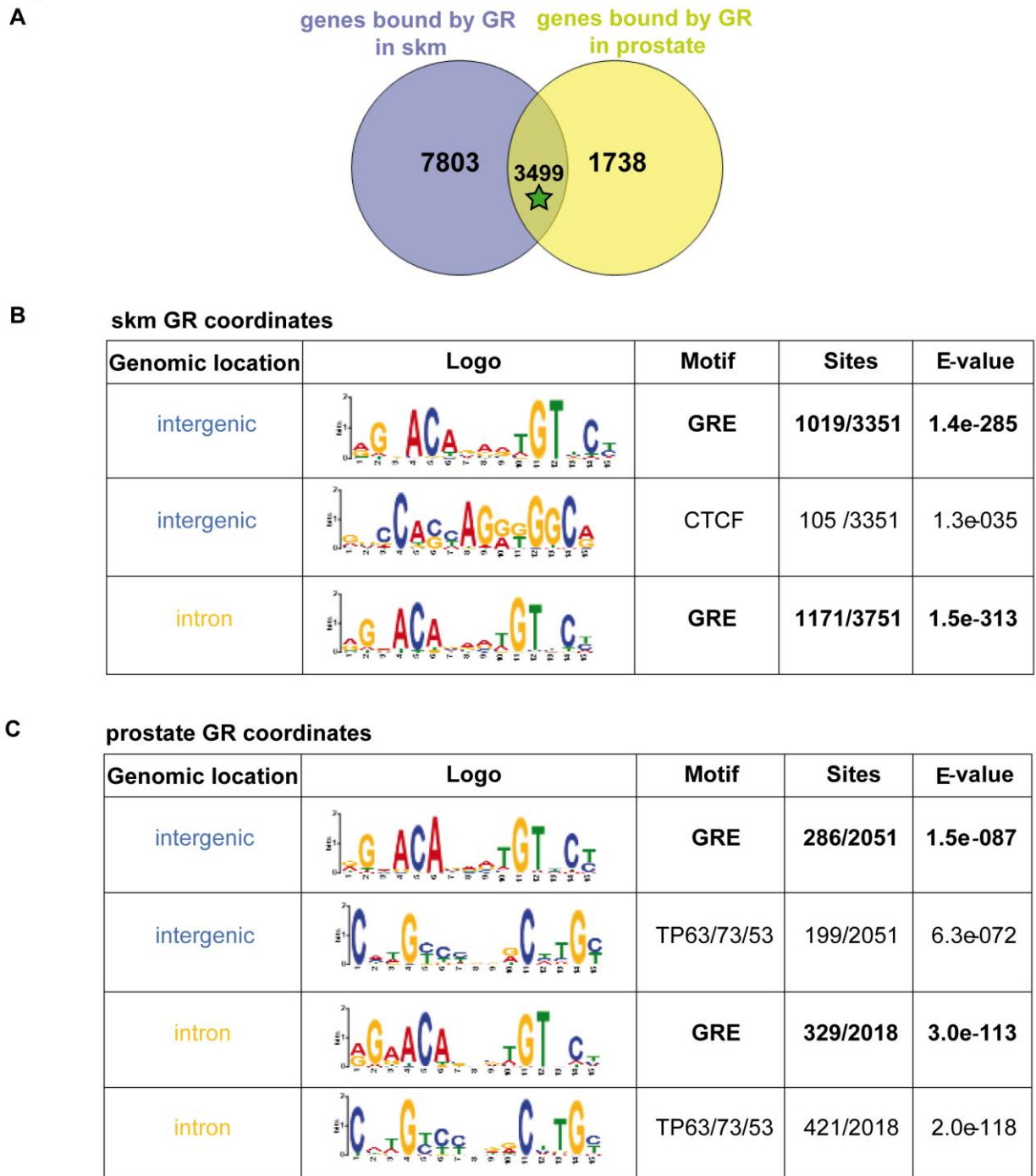
Figure 3. Shared genes between GR in skm and prostate.

A. Venn diagram of genes bound by GR in skm and prostate. The star is showing which genes are used for the further analysis in B and C.

B. Significant cis-elements of the binding sites located in the genes bound by GR in the two tissues, using the skm GR coordinates, as identified by *de novo* motif searches.

C. Significant cis-elements of the binding sites located in the genes bound by GR in the two tissues, using the prostate GR coordinates, as identified by *de novo* motif searches.

Figure 4.

A

genes bound by GR in skm   genes bound by GR in prostate

7803   3499   1738

B   **skm-specific GR coordinates**

Genomic location	Logo	Motif	Sites	E value
intergenic		**GRE**	**510/2321**	**1.6e-129**
intergenic		CTCF	87/2321	5.2e-050
intron		**GRE**	**914/3520**	**1.6e-225**
intron		CTCF	45/3520	9.5e-007
intron		MEF2A/B/C/D	48/3520	1.5e-002

C   **prostate-specific GR coordinates**

Genomic location	Logo	Motif	Sites	E-value
intergenic		**GRE**	**74/570**	**2.7e-034**
intergenic		TP63/73/53	141/570	1.7e-064
intron		**GRE**	**91/500**	**9.8e-021**
intron		TP63/73/53	139/500	5.2e-070

Figure 4. Specific genes bound by GR in skm and prostate.
A. Venn diagram of genes bound by GR in skm and prostate. The stars are showing which genes are used for the further analysis in B and C.
B. Significant cis-elements of the GRBS located in the skm-specific genes as identified by *de novo* motif searches.
C. Significant cis-elements of the GRBS located in the prostate-specific genes as identified by *de novo* motif searches.

Figure 5.

A
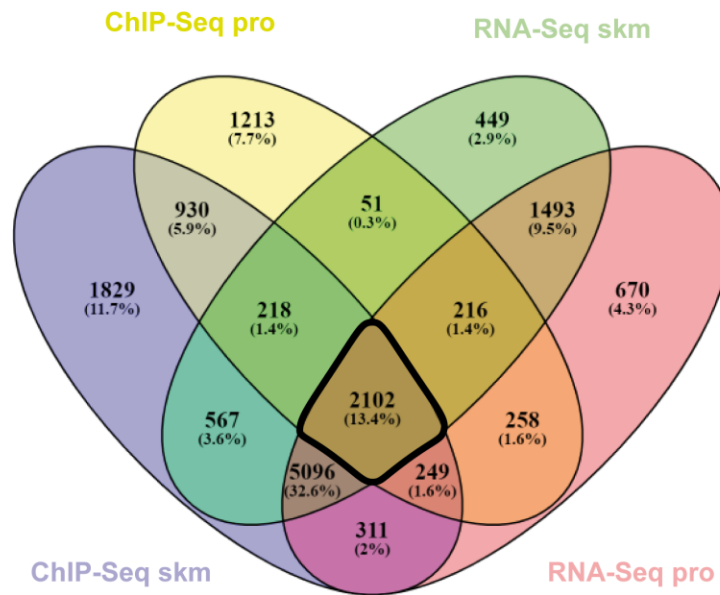
B
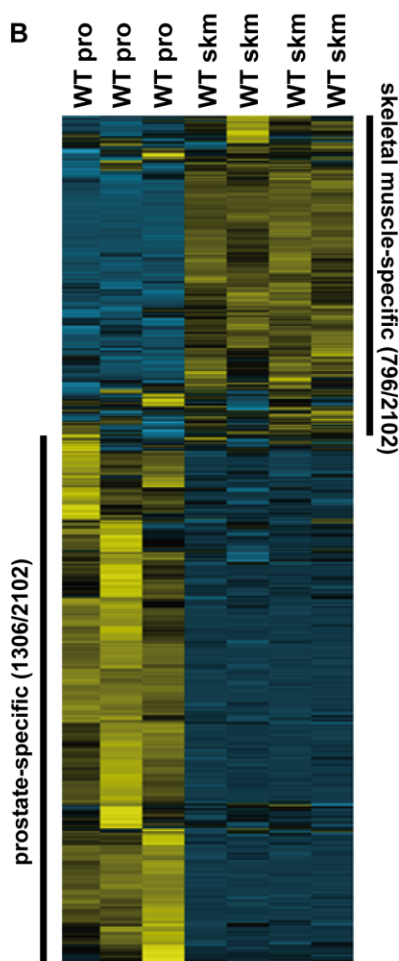
C

Pathways analysis skeletal muscle-specific genes
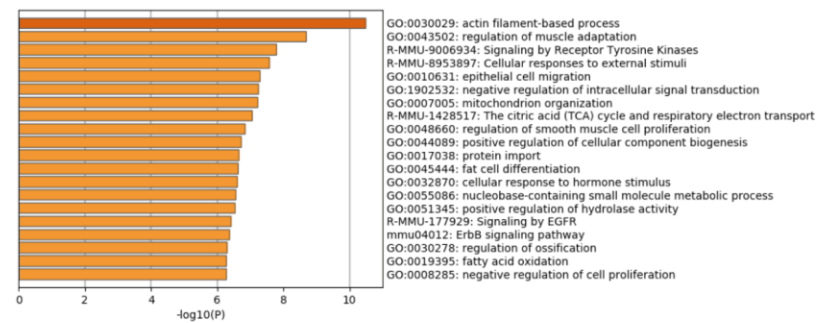
D

Pathways analysis prostate-specific genes
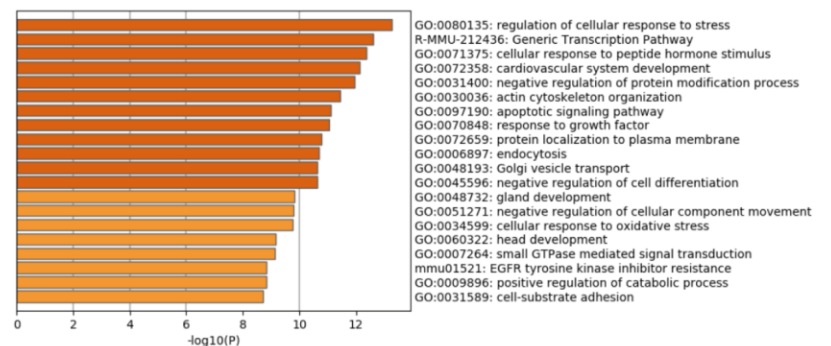
Mean centered normalized expresion colorbar
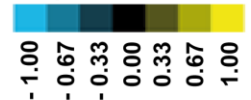
E skm-specific genes

Logo	Motif	E-value
	GRE	1.0e-118
	MEF2A/B/C/D	3.7e-010
Logo	**Short motif**	**E-value**
	half-site GRE	2.6e-059
	Myod1, Myog	1.1e-006

F prostate-specific genes
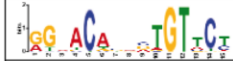
Logo	Motif	E-value
	GRE	4.4e-075
Logo	**Short motif**	**E-value**
	half-site GRE	1.1e-040
	Forkhead box (FOX)	2.0e-021
	TEAD2	1.3e-012
	SMAD/NF-1 DNA-binding domain	8.1e-010

Figure 5. **GR targets and expressed genes in skm and prostate.**

A. Venn diagram of the genes shared between ChIP-Seq and RNA-Seq in skm and prostate.

B. Heatmap of the mean centered normalized expression of the 2102 shared genes between ChIP-Seq and RNA-Seq in skm and prostate. The colorbar is depicted.

C. Pathway analysis of the skeletal muscle-specific genes.

D. Pathway analysis of the prostate-specific genes.

E. Table of significant cis-elements as identified by *de novo* searches of the skm-specific genes.

F. Table of significant cis-elements as identified by *de novo* searches of the prostate-specific genes.

Figure 6.

A

B

C Cluster 1 - Shared GR skm/pro

Logo	Motifs	Sites	E-value
	GRE	133/2620	1.0e-016
Logo	**Short motifs**	-	**E-value**
	Half-site GRE	-	3.7e-017
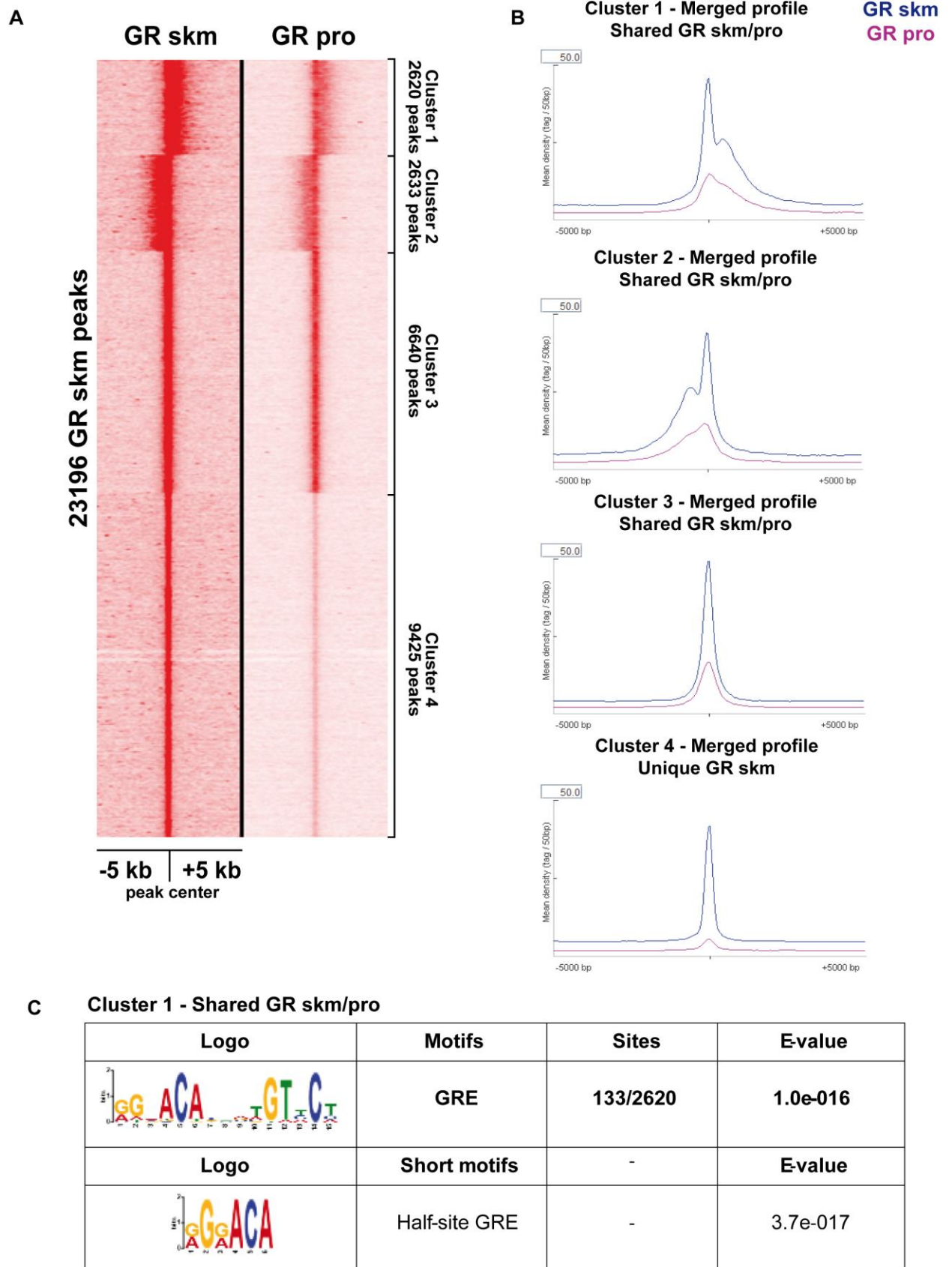
Cluster 2 - Shared GR skm/pro

Logo	Motifs	Sites	E-value
	GRE	**200/2633**	**4.0e-047**
Logo	Short motifs	-	E-value
	STAT3	-	5.8e-024
	NFYA/B	-	5.9e-013
	Half-site GRE	-	3.4e-005

Cluster 3 - Shared GR skm/pro

Logo	Motifs	Sites	E-value
	GRE	**488/6640**	3.2e-039
	NRF1	540/6640	1.2e-030
Logo	Short motifs	-	E-value
	ETS-related factors	-	4.1e-100
	NFYA/B	-	4.0e-056
	YY1	-	3.5e-041
	Half-site GRE	-	1.2e-036
	STAT3	-	2.4e-024

D

Cluster 4 - Unique GR skm
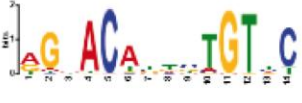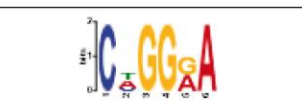
Logo	Motifs	Sites	E-value
	GRE	**2040/9425**	**1.9e-206**
	Myog, Myod1	1429/9425	6.6e-026
Logo	**Short motifs**	**-**	**E-value**
	Half-site GRE	-	2.4e-301
	MEF2A/B/C/D	-	3.3e-100
	TEAD2	-	2.4e-063
	CTCF	-	1.5e-018

Figure 6. **Cistromic overlap of GR in skm and prostate.**

A. Tag density maps showing GR binding events centered within a +/-5 kb window in skm and prostate. The GR ChIP-Seq reads in skm are used as reference. The shared and the unique clusters are depicted.
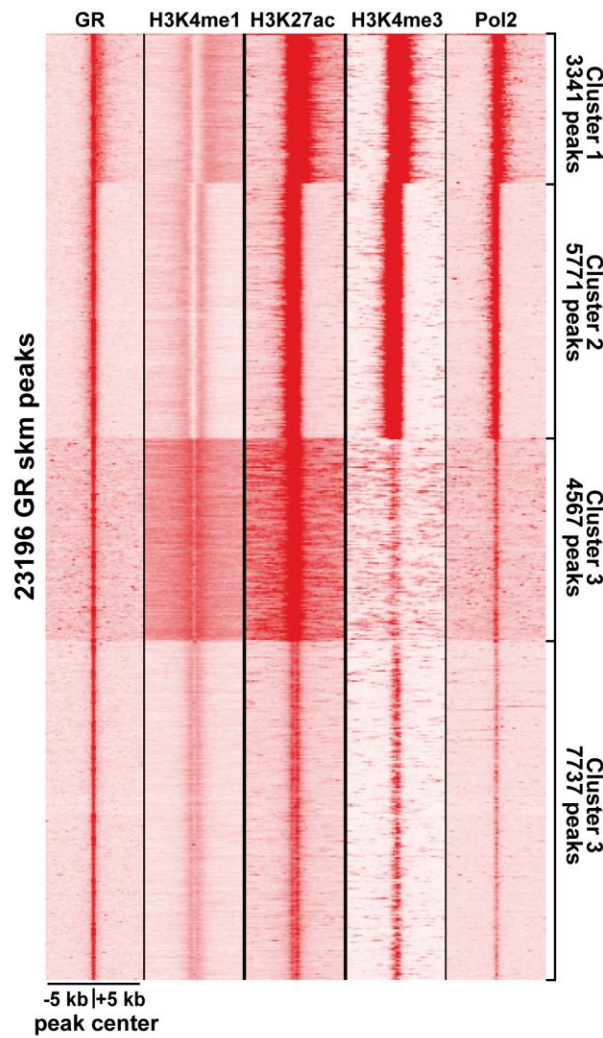
B. Mean tag density merged profiles for GR in skm at unique and shared GRBS with those in prostate in a window of +/-5 kb centered around the summit of the binding sites.

C. Tables of significant cis-elements as identified by *de novo* motif searches of the shared clusters.
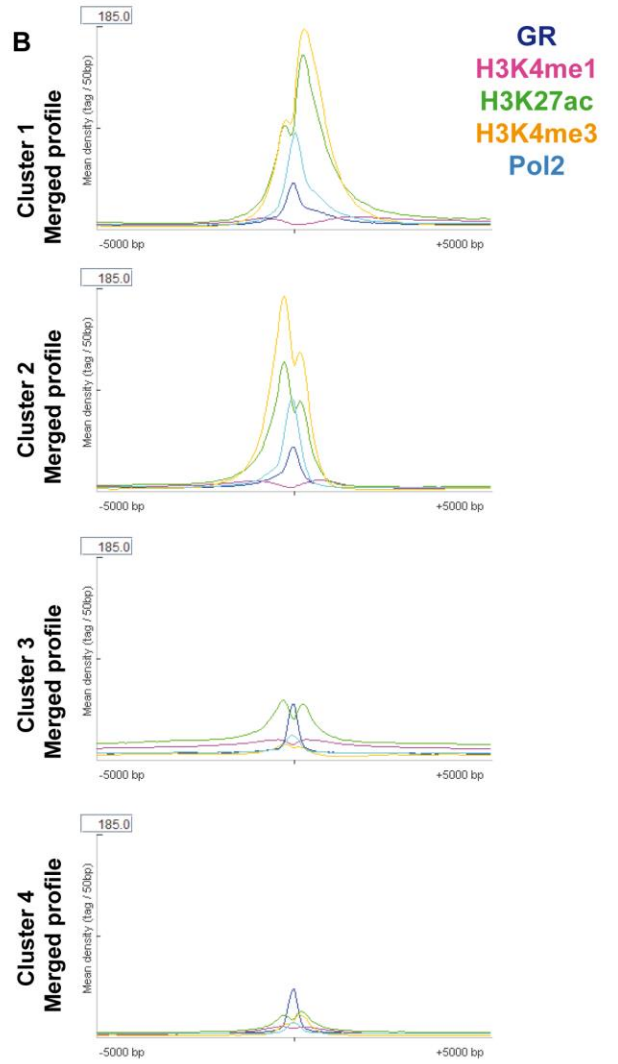
D. Table of significant cis-elements as identified by *de novo* searches of the unique cluster.

Figure 7.
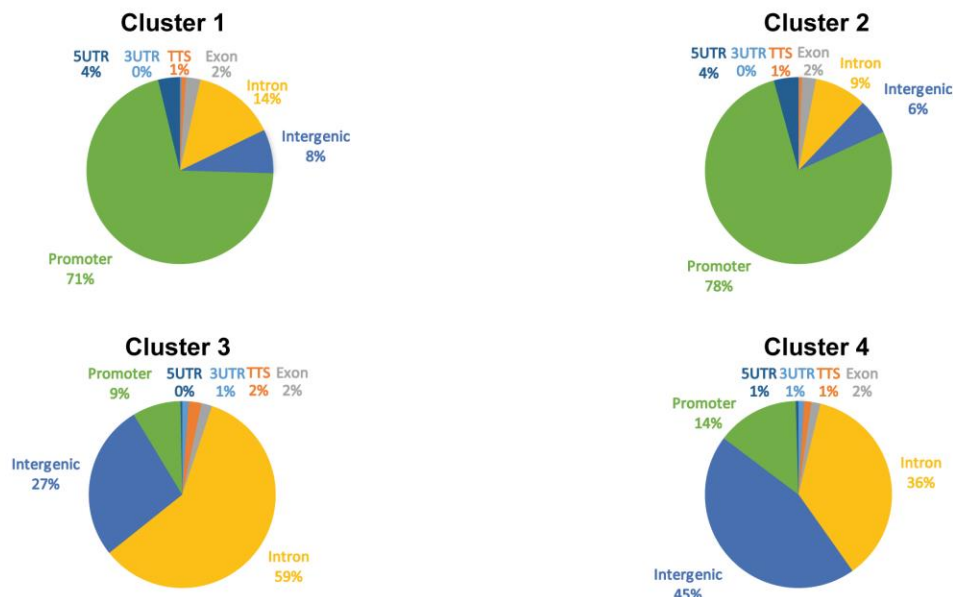
D **Cluster 1**

Logo	Motifs	Sites	E-value
	YY1	123/3341	3.6e-022
Logo	**Short motifs**	-	**E-value**
	NFYA/B	-	4.4e-029
	ETS-related factors	-	1.5e-028
	STAT3	-	2.1e-019
	NRF1	-	7.0e-014

Cluster 2

Logo	Motifs	Sites	E-value
	ETS-related factors	975/5771	7.3e-074
	NFYA/B	423/5771	3.1e-032
	NRF1	479/5771	2.3e-026
Logo	**Short motifs**	-	**E-value**
	YY1	-	7.6e-036
	STAT3	-	5.8e-016

Cluster 3

Logo	Motifs	Sites	E-value
	GRE	**1026/4567**	**5.6e-163**
	MEF2A/B/C/D	281/4567	1.0e-007
Logo	**Short motifs**	**-**	**E-value**
	Half-site GRE	-	1.1e-117

Cluster 4

Logo	Motifs	Sites	E-value
	GRE	**1983/7737**	**1.6e-181**
Logo	**Short motifs**	**-**	**E-value**
	Half-site GRE	-	1.3e-246
	MEF2A/B/C/D	-	4.7e-049
	CTCF	-	2.5e-024

Figure 7. **Cistromic overlap of GR, H3K4me1, H3K27ac, H3K4me3 and Pol2 in skm.**

A. Tag density maps showing GR, H3K4me1, H3K27ac, H3K4me3 and Pol2 binding events centered within a +/-5 kb window in skm. The GR ChIP-Seq reads in skm are used as reference. Clusters of genomic co-localization are depicted.

B. Mean tag density merged profiles for GR at GRBs shared with H3K4me1, H3K27ac, H3K4me3 and Pol2 sites in skm in a window of +/-5 kb centered around the summit of the binding sites.

C. Pie-charts of the genomic distribution of the binding sites for each cluster.

D. Tables of significant cis-elements as identified by *de novo* motif searches for each cluster.

II. Genome-wide comparative analysis of androgen & glucocorticoid receptors' cistomes and transcriptomes in prostate

Results

To determine the AR cistrome in prostate, we performed chromatin immunoprecipitation (ChIP) followed by massively parallel DNA sequencing of prostate tissues of 10-week-old wild-type mice using an anti-AR antibody. 14,930,616 50-nucleotide (nt) sequence tags were obtained and mapped uniquely to the mouse genome (mm10).

Peak calling by MACS2 (Feng et al., 2011), revealed 3857 AR-binding sites (ARBS), using a minimum False Discovery Rate (FDR) 0.1 (from one biological sample), which were associated to 2766 genes. The genome-wide distribution of the ARBS revealed that 64% (2468/3857) were located in intergenic regions, 33% (1281/3857) in intronic regions and only 1% (39/3857) in proximal promoter regions (-1 kb to +100 bp) (Figure 1A). Similar analysis of the 7940 GRBS, associated to 5237 genes, revealed that 49% (3916/7940) were located in intergenic regions, 40% (3197/7940) in intronic regions and 6% (472/7940) in proximal promoter regions (Figure 1B). Moreover, we observed an accumulation of ARBS 50 kb to 500 kb from the TSS (Figure 1C) compared to an accumulation of GRBS 5 kb to 50 kb form the TSS (Figure 1D).

Thus, in prostate, ARBS are two times less compared to GRBS and they are associated to two times less genes. In addition, the genome-wide distribution patterns, mostly in intergenic and intronic regions, are similar between ARBS and GRBS.

De novo motif analysis of the ARBS, using the software MEME-ChIP (Machanick and Bailey, 2011) from MEME-Suite (Bailey et al., 2009), revealed 15-bp canonical Androgen Response elements (AREs) in intergenic (31/113 sites, e-value: 5.3e-018, 27.4%) and intronic regions (30/109 sites, e-value: 2.5e-016, 27.5%), but not in proximal promoter regions, where no motifs were significantly enriched (Figure 1E).

Similar analysis for the GRBS, revealed 15-bp GREs in intergenic (434/2587 sites, e-value: 4.0e-073, 16.7%) and intronic regions (427/2468 sites, e-value: 9.3e-088, 17.3%). Moreover, TP53/63/73 motifs were identified in intergenic (236/2587 sites, e-value: 5.9e-050, 9.1%) and intron regions (529/2468 sites, e-value: 5.1e-110, 21.4%), but not in proximal promoter regions where SP2, ZNF263 and other motifs where identified (Figure 1F).

Thus, these results show that AR and GR bind DNA to AREs and GREs in intergenic and intron regions, but not in the promoter regions, where they might interact with distinct factors depending on the receptor.

By overlapping the data from the two ChIP-Seq assays, almost 2000 genes were bound by AR and GR (Figure 2A). Among these common genes, we identified Ddit4 (Britto et al., 2018)

(Figure 2C), Pik3r1 ([Kuo et al., 2017](#)), Foxo3 ([Lutzner et al., 2012](#)), as well as Foxa1 ([Jones et al., 2015](#)), Nkx2-5 ([Chung et al., 2008](#)) (Figure 2D), Klk8 and Tmprss11c.

In addition, using bedtools ([Quinlan and Hall, 2010](#)), we identified more than 2000 overlapping regions (for at least 1 bp) among ARBS and GRBS in prostate (Figure 2B). A selection of genes bound by AR and GR in prostate and their number of binding sites are shown in Table 1.

Table 1. Genes bound by AR and GR in prostate and their number of binding sites.

Genes	ARBS	GRBS
Ddit4 - DNA damage inducible transcript 4	3	4
Pik3r1 - phosphoinositide-3-kinase regulatory subunit 1	1	10
Foxo3 - forkhead box O3	1	4
Foxa1 - forkhead box A1	1	2
Nkx2-5 - NK2 homeobox 5	1	1
Klk8 - kallikrein related peptidase 8	1	1
Tmprss11c - transmembrane protease, serine 11c	1	1
Nfkb1 - nuclear factor of kappa light polypeptide gene enhancer in B cells 1, p105	3	4
Sesn1 - sestrin 1	1	5
Sgk1 - serum/glucocorticoid regulated kinase 1	1	4

The 2910 ARBS located in the genes bound by AR and GR in prostate, were at 64% (1874/2910) in intergenic regions, 34% (975/2910) in intronic regions and 1% (27/2910) in proximal promoter regions (Figure 2E). The 3322 GRBS located in the genes bound by AR and GR in prostate were at 59% (1949/3322) in intergenic regions, 37% (1229/3322) in intronic regions and 2% (79/3322) in proximal promoter regions (Figure 2F). Interestingly, the ARBS located in the genes bound by both receptors stayed stable in promoter and intergenic regions and increased in intronic regions compared to the bulk ARBS in prostate (Table 2). In addition, the GRBS located in the genes bound by both receptors decreased in promoter and intronic regions and increased in intergenic regions compared to the bulk GRBS in prostate (Table 2).

Table 2. Recapitulative table of the genomic distribution of the ARBS and GRBS in prostate.

Genomic location	Bulk ARBS prostate	Genes bound by AR & GR ARBS	Bulk GRBS prostate	Genes bound by AR & GR GRBS
promoter	1%	1%	6%	2%
intergenic	64%	64%	49%	59%
intron	33%	34%	40%	37%

De novo motif analysis of the ARBs and GRBs located in the genes bound by both AR and GR in prostate identified 15-bp canonical AREs and GREs in intergenic and intron regions, but not in proximal promoter regions (Figure 3). More precisely, by using the AR coordinates, 15-bp canonical AREs were revealed in intergenic regions (37/124 sites, e-value: 5.1e-023, 29.8%), in intronic regions (48/108 sites, e-value: 2.3e-024, 44.4%) but not in proximal promoter regions (Figure 3B). By using the GR coordinates, 15-bp canonical GREs were identified in intergenic regions (180/763 sites, e-value: 2.3e-139, 23.6%), in intronic regions (142/605 sites, e-value: 5.6e-122, 23.5%) but not in proximal promoter regions (Figure 3C). Moreover, TP53/63/73 motifs were identified in intergenic regions (90/763 sites, e-value: 1.7e-024, 11.8%), in intronic regions (92/605 sites, e-value: 1.9e-024, 15.2%), but not in proximal promoter regions (Figure 3C).

De novo motif analysis of the AR-specific binding sites did not reveal AREs in none of the intergenic, intron and proximal promoter regions, but only repetitive elements (TC or AG). On the other hand, *de novo* motif analysis of the GR-specific binding sites revealed GREs in intergenic regions (312/1860 sites, e-value: 6.2e-057, 16.8%), in intronic regions (381/1968 sites, e-value: 1.5e-074, 19.3%), but not in proximal promoter regions (Figure 4B). Moreover, TP53/63/73 motifs were identified in intergenic regions (406/1860 sites, e-value: 3.4e-080, 21.8%), in intron regions (540/1968 sites, e-value: 6.1e-102, 27.4%), but not in proximal promoter regions (Figure 4B).

Thus, AR and GR bind DNA to AREs and GREs in intergenic and intron regions and probably interact or cooperate with distinct factors depending on the receptor.

Functional annotation of the genes bound by both AR and GR, AR-specific genes and GR-specific genes, revealed many enriched shared pathways, but also specific ones (in bold) (Table 3).

Table 3. Enriched shared and unique pathways and their FDR values.

KEGG Pathways	Genes bound by AR and GR (FDR)	AR-specific genes (FDR)	GR-specific genes (FDR)
Prostate cancer	1.06e-05	2.25e-05	3.79e-13
p53 signaling pathway	1.32e-06	7.24e-06	2e-04
Cytokine-cytokine receptor interaction	4.79e-02	1.25e-14	1.89e-13
Chemokine signaling pathway	0.0003	5.64e-05	4.81e-12
Wnt signaling pathway	1.72e-05	0.0001	9.07e-24
Jak-STAT signaling pathway	4.76e-05	4.70e-06	2.07e-10
Cell cycle	0.0074	0.0020	9.98e-07
Toll-like receptor signaling pathway	3.11e-05	2.53e-05	2.55e-08
MAPK signaling pathway	2.23e-08	8.11e-11	1.08e-02

Insulin signaling pathway	5.03e-06	3.03e-07	3.13e-20
TNF signaling pathway	-	3.59e-02	5.09e-03
FoxO signaling pathway	-	-	7.62e-03
PPAR signaling pathway	-	-	7.94e-07
mTOR signaling pathway	-	-	1.38e-06
Glycolysis / gluconeogenesis	-	-	3.32e-06
Adipocytokine signaling pathway	-	-	7.30e-12
Natural killer cell mediated cytotoxicity	-	-	2.29e-07
Apoptosis	-	-	1.24e-08

To determine which of the genes bound directly by AR and GR in prostate are expressed, we overlapped the ChIP-Seq data with RNA-Seq data of prostate (Figure 5A). The RNA-Seq was performed on extracted prostates of 26-28-week-old wild-type mice. The overlap among genes with at least one ARBS and GRBS in prostate and genes expressed in prostate (number of reads greater than 100) revealed almost 800 genes (Figure 5A). Pathway analysis of the common genes revealed pathways such as p53 signaling, prostate cancer, TNF signaling, cellular senescence, MAPK signaling, PI3K-Akt signaling, oxytocin signaling, apoptosis, Wnt signaling, Hippo signaling (Salem and Hansen, 2019). *De novo* motif analysis of the binding sites located in the 800 common genes revealed AREs/GREs using separately AR and GR coordinates (Figure 5B, C).

After the gene analysis, we performed a binding-sites analysis to get rid of the annotation biases. We used seqMINER to visualize the tag density maps of the AR and GR ChIP-seq reads from prostate, within +/-5kb, and using the GRBS as reference. The heatmap revealed shared binding events between AR and GR in prostate, but also unique to GR (Figure 6A). The AR cistrome and the GR cistrome in prostate exhibit moderate overlap (cluster 1 & 2) (2551/7435, 34%) and 66% of the GRBS (4884/7435) are not bound by AR (cluster 3 & 4) (Figure 6A). The average tag density profiles of GR and AR are enriched at the center of the binding sites, at the same sites, in a window of +/-5 kb centered around the summit of the binding sites (Figure 6B).

De novo motif analysis of the sequences below the binding sites of the shared clusters revealed canonical and half-site AREs/GREs and motifs of C2H2 zinc finger factors (PRDM1) (Figure 6C)*. De novo* motif analysis of the sequences below the binding sites of the unique cluster revealed TP53/63/73 and TEAD2 motifs and motifs of Forkhead box (FOX) and Homeo domain factors (Figure 6D).

The tag density maps of AR, GR and H3K4me1 (Chen et al., 2013) ChIP-Seq reads from prostate, using the seqMINER, within +/-5kb, and using the GRBS as reference, revealed that more than the half of the GRBS (4729/7947, 60%) in prostate overlap with the chromatin mark in the respective tissue in clusters 1 & 2 (Figure 7A). H3K4me1 marks active enhancers

and promoters. This implies that these binding sites are located at active enhancer and promoter regions. The ARBS overalp with GRBS for 40% of the total sites (3218/7947) in cluster 3. The absence of H3K4me1 signal in cluster 3 can be a problem of specificity of AR antibody (background binding, random binding), or might indicate that AR binds to non-active promoter and enhancer regions or that AR is a repressor. The average tag density profiles of H3K4me1 are enriched around the center of the binding sites, where the signal is depleted (Figure 7B). The average tag density profiles of AR and GR are enriched at the center of binding sites (Figure 7B). The genomic distribution of the binding sites of the three clusters was mostly intergenic and intronic regions (Figure 7C).

De novo motif analysis of the shared clusters between GR and H3K4me1 revealed canonical and half-site GREs, STAT1 motifs and motifs of Forkhead box (FOX) factors, ETS-related factors, SMAD/NF-1 DNA-binding domain factors, bHLH factors (ASCL1) and Homeo domain factors (Figure 7D). *De novo* motif analysis of the shared cluster between AR and GR revealed motifs of Homeo domain factors (PBX3) and TEA domain factors (TEAD2/3) (Figure 7C).

Thus, AR and GR binds to canonical and half-site AREs and GREs in prostate and the distinct motifs found close to them are probably imposing the specificity of the receptor.

Discussion

This genome-wide comparative analysis in prostate demonstrated that under physiological conditions GR binds to two times more sites than AR and these sites are located in two times more genes. This fact is probably due to the assay, chromatin immunoprecipitation of a tissue can be experimentally challenging, due to the depth of the sequencing or finally due to the specificity of the AR antibody used.

The genomic distribution of the ARBS is similar to that of the GRBS, mostly in intergenic and intronic regions. *De novo* motif analysis of ARBS in prostate, revealed AREs in intergenic and intronic regions, but not in proximal promoter regions. *De novo* motif analysis of GRBS in prostate, revealed GREs and TP53/63/73 motifs in intergenic and intronic regions, but not in proximal promoter regions. Thus, both receptors bind DNA to AREs and GREs in intergenic and intronic regions in prostate, but not in promoter regions where might interact with distinct factors depending on the receptor.

AR and GR were bound to almost 2000 target genes in prostate. *De novo* motif analysis of GRBS located in the genes bound by AR and GR, revealed GREs and TP53/63/73 motifs in intergenic and intronic region, but not in proximal promoter regions. *De novo* motif analysis of ARBS located in the genes bound by AR and GR, revealed AREs in intergenic and intronic regions, but not in the promoter regions. Thus, both receptors bind to AREs and GREs in intergenic and intronic regions located in the common genes and probably interact or cooperate with distinct factors that impose the specificity of the receptor in the given tissue.

P53 is a tumor suppressor and potent inhibitor of cell growth. P73 and P63 similar to P53 in amino acid sequence and structure (Courtois et al., 2004). There is evidence for negative cross-talk between GR and p53 (Sengupta et al., 2000; Sengupta and Wasylyk, 2001) and there is a study examining the ability of p53 and p73 to interact with and inhibit GR transcriptional activity (Zhang et al., 2006). Thus, probably GR interacts with p53/63/73, in intergenic and intronic regions, to modulate transcription

Interestingly, we observed that the response elements of the genes bound by both receptors in intergenic and intronic regions are composed elements consisting of the 1st half site of a canonical 15-bp ARE, as described by the Jaspar motif database, with a very conserved G at the position 2 or 3 and the same probability for the A and C in positions 4 or 5 and 5 or 6, respectively, and the 2nd half site of a canonical 15-bp GRE, as described by the Jaspar motif database, with a very conserved G at the position 11, a conserved C at the position 14 and two T, the first less conserved that the other, at the positions 10 and 12, respectively (Figure 8A).

The cistromic overlap of GR and AR in prostate, using the GRBS as reference, revealed that 66% the GRBs are not bound by AR, although there are shared sites between the receptors. Similar analysis, using the ARBS as reference, could provide insights into the unique ARBS.

De novo motif analysis of the shared binding sites revealed canonical and half-site AREs/GREs and motifs of C2H2 zinc finger factors. However, *de novo* motif analysis of the unique binding sites did not reveal GREs, but TP53/63/73 motifs and motifs of Forkhead box (FOX), Homeo domain and TEA domain factors. Thus, we report that in prostate, AR and GR bind to AREs and GREs and cooperate with same (C2H2 zinc finger factors) but also different factors (p53 domain, Forkhead box, Homeo domain, TEA domain factors). The absence of enriched GREs under the unique binding sites probably indicates the presence of binding sites at promoter regions.

Finally, we demonstrated that in prostate more than half of the GRBs (60%) are located at active enhancer regions indicated by the high H3K4me1 signal and the genomic distribution of the binding sites that was mostly intergenic and intronic regions. *De novo* motif analysis of these binding sites revealed canonical and half-site GREs, STAT1 motifs; whose deregulation has been implicated in prostate cancer cell growth and survival (Hatziieremia et al., 2016), motifs of SMAD/NF-1 DNA-binding domain factors, Forkhead box (FOX) factors (van der Heul-Nieuwenhuijsen et al., 2009), Homeo domain factors; important for the development of the normal prostate gland (Javed and Langley, 2014) and ETS-related factors (Shaikhibrahim and Wernert, 2012). The ARBS overlap with GRBS for 40% of the total sites in absence of H3K4me1 signal that is probably due to the non-specificity of AR antibody (background or random binding), might indicate that AR binds to poised promoter and enhancer regions or that AR is a repressor.

Although the genomic distribution of all the binding sites was mostly at enhancer regions, additional histone marks and Pol2 are needed in order to investigate the presence of binding sites at active enhancer and promoter regions. In addition, we speculate that in prostate GR cooperates with p53 domain, Forkhead box (FOX), Homeo domain and ETS-related factors but for AR is more difficult to define with the present data.

Receptor specificity depends on the surrounding transcription factors that are already bound to the chromatin and the ones that will be recruited later, rather than the sequence of the response elements.

Further studies are needed to elucidate the mechanism(s) of the AR/GR specificity in a given tissue as there are similarities but also distinct differences that should be validated experimentally. In this way, it will be possible to design selective receptors that bind to discriminating response elements with selective activities and reduced side effects.

Materials and methods

Mice

Experiment	Type	Age	Phenotype	Samples
AR ChIP-Seq prostate	C57BL/6	10 weeks	Wild-type	1
GR ChIP-Seq prostate	C57BL/6	10 weeks	Wild-type	1
RNA-Seq prostate	C57BL/6	26-28 weeks	Wild-type	5

Chromatin immunoprecipitation assays and ChIP-sequencing

ChIP assays were performed as described previously (Joshi et al., 2017).

Sequencing was performed by the IGBMC Microarray and Sequencing platform, a member of the "France Génomique" consortium (ANR-10-INBS-0009). Immunoprecipitated DNA samples were processed for library preparation on Illumina Hiseq 4000 sequencer as a single-read 50 base reads following Illumina's instructions. Image analysis and base calling were performed using RTA 2.7.7 and bcl2fastq 2.17.1.14. Adapter dimer reads were removed using DimerRemover. The FastQC 0.11.2 was used to evaluate the quality of the sequencing. Sequenced reads were mapped to the mus musculus genome assembly 10 (mm10) using Bowtie 2 (Langmead and Salzberg, 2012). Uniquely mapped reads were used for further analysis. The peak calling was performed with the MACS2 algorithm (Feng et al., 2011) using appropriate inputs. Peaks were annotated relative to genomic features using Homer (Heinz et al., 2010) according to the distance to the nearest TSS. Distance to TSS was calculated using the online software GREAT (McLean et al., 2010). Data visualization was carried out using Integrative Genomics Viewer (IGV) (Robinson et al., 2011). *De novo* motif analysis of all the binding sites (100 nucleotides both side of the peak summit) was performed by using the online software MEME-ChIP (Machanick and Bailey, 2011) from MEME-Suite (Bailey et al., 2009) after extracting their nucleotide sequences. The *de novo* motifs were then compared to a database of known motifs and ranked by the TOMTOM tool (Gupta et al., 2007) of MEME Suite. The Jaspar database (Sandelin et al., 2004) was used to check for transcription factor binding site profiles. The tag density maps were produced using the software SeqMiner (Ye et al., 2011) and the clustering normalization was done using the KMeans linear method. The intersection of intervals was performed with the intersect function of bedtools (Quinlan and Hall, 2010).

Antibodies
- anti-GR IGBMC 3249, fraction no4
- Anti-AR Abcam ChIP Grade (ab74272)
- Anti-H3K4me1 (Chen et al., 2013) Abcam (ab8895)

RNA extraction and RNA sequencing

Total RNA was isolated from control prostate samples using RNeasy Micro Kit (74004) from Qiagen, reverse transcribed (RT) using SuperScript II Reverse Transcriptase (Invitrogen) and amplified by quantitative PCR with the SYBER Green kit (Roche) and LightCycler 480 (Roche Diagnostics) according to the manufacturer's instructions.

Sequencing was performed by the IGBMC Microarray and Sequencing platform, a member of the "France Génomique" consortium (ANR-10-INBS-0009). The library was prepared on Illumina Hiseq 4000 sequencer as a single-read 50 base reads following Illumina's instructions. Image analysis and base calling were performed using RTA 2.7.7 and bcl2fastq 2.17.1.14. Adapter dimer reads were removed using DimerRemover. The FastQC 0.11.2 was used to evaluate the quality of the sequencing. Reads were mapped onto the mm10 assembly of mouse genome using Tophat 2.1.1 (Kim et al., 2013) and the Bowtie2 2.3.4.3 (Langmead and Salzberg, 2012). Only uniquely aligned reads have been retained for further analyses. Quantification of gene expression was performed using HTSeq-0.11.0 (Anders et al., 2015). If the raw read counts were greater than 100, then the genes were considered expressed.

The pathway analysis was performed using the online software WebGestalt GSAT (Wang et al., 2013) and the Over-Representation Analysis (ORA) method.

References

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res *37*, W202-208.

Britto, F.A., Cortade, F., Belloum, Y., Blaquiere, M., Gallot, Y.S., Docquier, A., Pagano, A.F., Jublanc, E., Bendridi, N., Koechlin-Ramonatxo, C*., et al.* (2018). Glucocorticoid-dependent REDD1 expression reduces muscle metabolism to enable adaptation under energetic stress. BMC Biol *16*, 65.

Chen, Y., Chi, P., Rockowitz, S., Iaquinta, P.J., Shamu, T., Shukla, S., Gao, D., Sirota, I., Carver, B.S., Wongvipat, J*., et al.* (2013). ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. Nat Med *19*, 1023-1029.

Chung, W., Kwabi-Addo, B., Ittmann, M., Jelinek, J., Shen, L., Yu, Y., and Issa, J.P. (2008). Identification of novel tumor markers in prostate, colon and breast cancer by unbiased methylation profiling. PLoS One *3*, e2079.

Courtois, S., Caron de Fromentel, C., and Hainaut, P. (2004). p53 protein variants: structural and functional similarities with p63 and p73 isoforms. Oncogene *23*, 631-638.

Feng, J., Liu, T., and Zhang, Y. (2011). Using MACS to identify peaks from ChIP-Seq data. Curr Protoc Bioinformatics *Chapter 2*, Unit 2 14.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol *8*, R24.

Hatziieremia, S., Mohammed, Z., McCall, P., Willder, J.M., Roseweir, A.K., Underwood, M.A., and Edwards, J. (2016). Loss of signal transducer and activator of transcription 1 is associated with prostate cancer recurrence. Mol Carcinog *55*, 1667-1677.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Javed, S., and Langley, S.E. (2014). Importance of HOX genes in normal prostate gland formation, prostate cancer development and its early detection. BJU Int *113*, 535-540.

Jones, D., Wade, M., Nakjang, S., Chaytor, L., Grey, J., Robson, C.N., and Gaughan, L. (2015). FOXA1 regulates androgen receptor variant activity in models of castrate-resistant prostate cancer. Oncotarget *6*, 29782-29794.

Joshi, S., Ueberschlag-Pitiot, V., Metzger, D., and Davidson, I. (2017). Improved Protocol for Chromatin Immunoprecipitation from Mouse Skeletal Muscle. J Vis Exp.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kuo, T., Chen, T.C., Lee, R.A., Nguyen, N.H.T., Broughton, A.E., Zhang, D., and Wang, J.C. (2017). Pik3r1 Is Required for Glucocorticoid-Induced Perilipin 1 Phosphorylation in Lipid Droplet for Adipocyte Lipolysis. Diabetes *66*, 1601-1610.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Lutzner, N., Kalbacher, H., Krones-Herzig, A., and Rosl, F. (2012). FOXO3 is a glucocorticoid receptor target and regulates LKB1 and its own expression based on cellular AMP levels via a positive autoregulatory loop. PLoS One *7*, e42166.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696-1697.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Salem, O., and Hansen, C.G. (2019). The Hippo Pathway in Prostate Cancer. Cells *8*.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res *32*, D91-94.

Sengupta, S., Vonesch, J.L., Waltzinger, C., Zheng, H., and Wasylyk, B. (2000). Negative cross-talk between p53 and the glucocorticoid receptor and its role in neuroblastoma cells. EMBO J *19*, 6051-6064.

Sengupta, S., and Wasylyk, B. (2001). Ligand-dependent interaction of the glucocorticoid receptor with p53 enhances their degradation by Hdm2. Genes Dev *15*, 2367-2380.

Shaikhibrahim, Z., and Wernert, N. (2012). ETS transcription factors and prostate cancer: the role of the family prototype ETS-1 (review). Int J Oncol *40*, 1748-1754.

van der Heul-Nieuwenhuijsen, L., Dits, N.F., and Jenster, G. (2009). Gene expression of forkhead transcription factors in the normal and diseased human prostate. BJU Int *103*, 1574-1580.

Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res *41*, W77-83.

Ye, T., Krebs, A.R., Choukrallah, M.A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res *39*, e35.

Zhang, L., Nie, L., and Maki, C.G. (2006). P53 and p73 differ in their ability to inhibit glucocorticoid receptor (GR) transcriptional activity. Mol Cancer *5*, 68.

Figure 1.

A   ARBS

B   GRBS

C   ARBS

D   GRBS
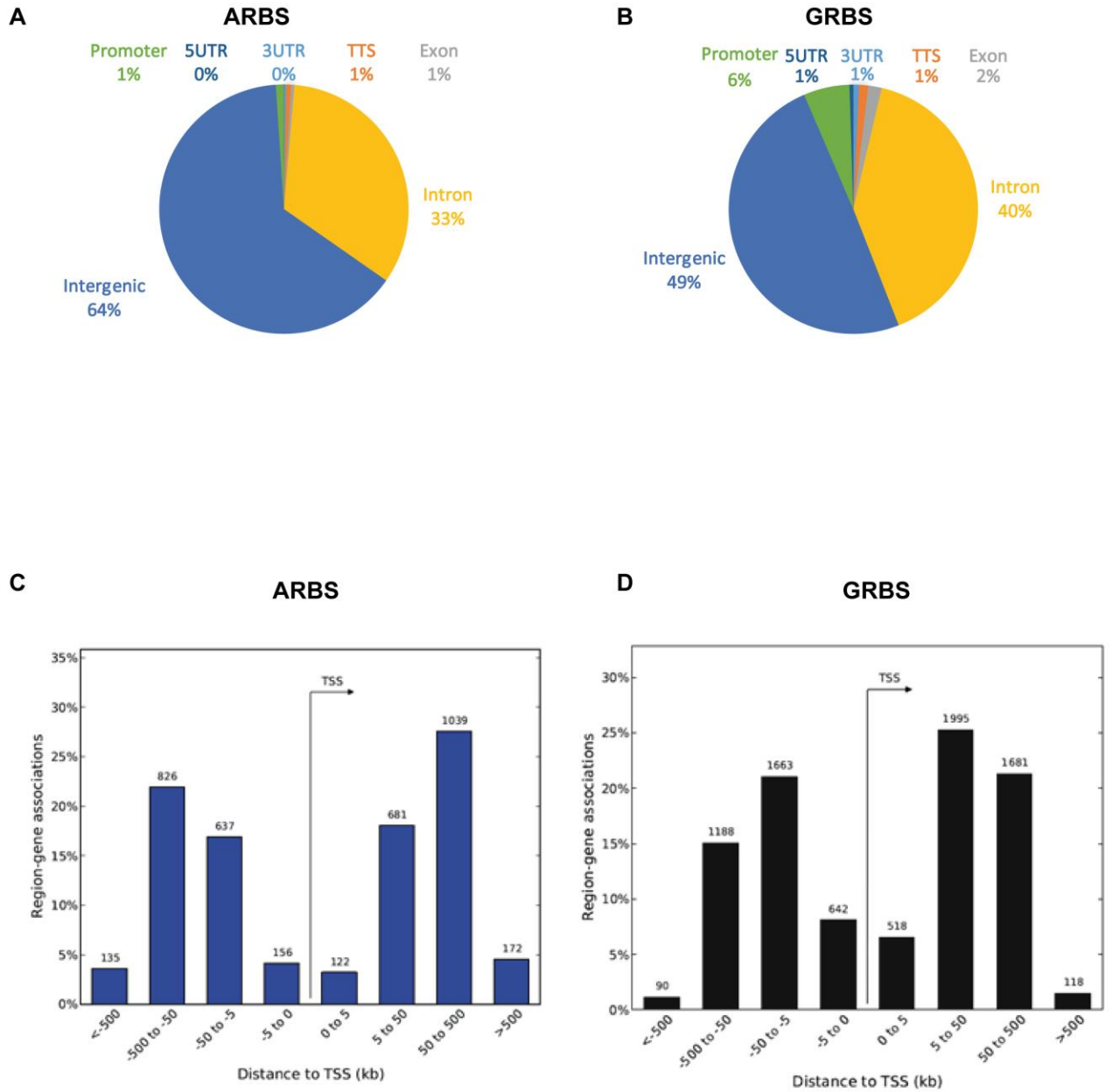
E   ARBS

Genomic location	Logo	Motif	Sites	E-value
promoter	-	-	-	-
intergenic	GAACA...TGT.Cc	ARE	31/113	5.3e-018
intron	GGAAC...GTGT.C.	ARE	30/109	2.5e-016

F   GRBS

Genomic location	Logo	Motif	Sites	E-value
promoter	GGAGGA.G.G	ZNF263, E2F6/4, SP2	82/447	1.8e-012
intergenic	G.aCa...TGTtC.t	GRE	434/2587	4.0e-073
intergenic	Ca.G.t.c.GGC..G.	TP53/63/73	236/2587	5.9e-050
intron	G.ACa..TGT.Cc	GRE	427/2468	9.3e-088
intron	Ca.G.cc..C..TGt	TP53/63/73	529/2468	5.1e-110

Figure 1. ARBS and GRBS in prostate.

A. Genome-wide distribution of the prostate AR binding sites (ARBS).

B. Genome-wide distribution of the prostate GR binding sites (GRBS).

C. Distance to TSS in kb of the prostate ARBS.

D. Distance to TSS in kb of the prostate GRBS.

E. Significant cis-elements of the prostate ARBS as identified by *de novo* motif searches.

F. Significant cis-elements of the prostate GRBS as identified by *de novo* motif searches.

Figure 2.

A

genes bound by AR in prostate   genes bound by GR in prostate

B

GRBS prostate   ARBS prostate

C

D

E

binding sites correspoding to genes
bound by AR and GR in prostate -
AR coordinates

Promoter 3' UTR TTS Exon 5' UTR
1%   0%   1%   0%   0%

Intron
34%

Intergenic
64%

F

binding sites correspoding to genes
bound by AR and GR in prostate -
GR coordinates

Promoter TTS 3' UTR Exon 5' UTR
2%   1%   0%   1%   0%

Intron
37%

Intergenic
59%

Figure 2. Overlap of genes bound by AR and GR in prostate.

A. Venn diagram of genes bound by AR and GR in prostate.

B. Overlapping regions among ARBS and GRBS in prostate.

C. Location of ARBS and GRBS in the regulatory regions of **Ddit4** in prostate. The input is depicted in black, green for GR and yellow for AR.

D. Location of ARBS and GRBS in the regulatory regions of **Nkx2-5** in prostate. The input is depicted in black, green for GR and yellow for AR.

E. Genome-wide distribution of the binding sites located in the genes bound by AR and GR in prostate, using AR coordinates.

F. Genome-wide distribution of the binding sites located in the genes bound by AR and GR in prostate, using GR coordinates.

Figure 3.

A

genes bound by AR
in prostate

genes bound by GR
in prostate

833   1933   3304
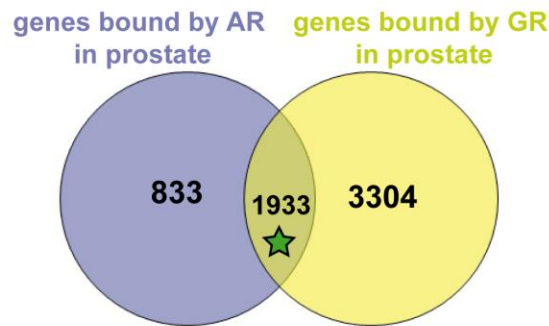
B   **AR coordinates**

Genomic location	Logo	Motif	SItes	E-value
intergenic		ARE/GRE	37/124	5.1e-023
intron		ARE/GRE	48/108	2.3e-024

C   **GR coordinates**

Genomic location	Logo	Motif	SItes	E-value
intergenic		ARE/GRE	180/763	2.3e-139
intergenic		TP53/63/73	90/763	1.7e-024
Intron		ARE/GRE	142/605	5.6e-122
intron		TP53/63/73	92/605	1.9e-024

Figure 3. **Shared genes between AR and GR in prostate.**

A. Venn diagram of genes bound by AR and GR in prostate. The star is showing which genes are used for the further analysis in B and C.

B. Significant cis-elements of the binding sites located in the genes bound by AR and GR in prostate, using the AR coordinates, as identified by *de novo* motif searches.

C. Significant cis-elements of the binding sites located in the genes bound by AR and GR in prostate, using the GR coordinates, as identified by *de novo* motif searches.

Figure 4.

A

genes bound by AR
in prostate

genes bound by GR
in prostate

833   1933   3304
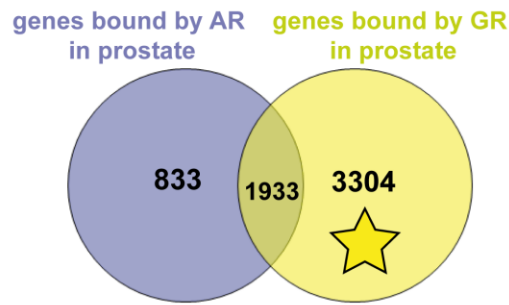
B   **GR-specific coordinates**

Genomic location	Logo	Motif	SItes	E-value
intergenic		GRE	312/1860	6.2e-057
intergenic		TP53/63/73	406/1860	3.4e-080
Intron		GRE	381/1968	1.5e-074
intron		TP53/63/73	540/1968	6.1e-102

Figure 4. **GR specific genes in prostate.**

A. Venn diagram of genes bound by AR and GR in prostate. The star is showing which genes are used for the further analysis in B.

B. Significant cis-elements of the binding sites located in the GR-speicific genes in prostate as identified by *de novo* motif searches.

Figure 5.

A

GR ChIP-Seq prostate   AR ChIP-Seq prostate

RNA-Seq prostate
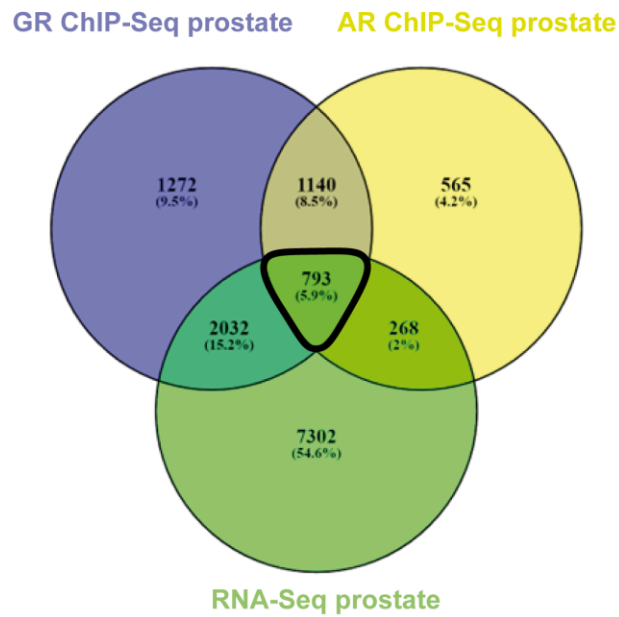
B   **AR coordinates**

Logo	Motif	E-value
	ARE/GRE	1.0e-024

C   **GR coordinates**

Logo	Motif	E-value
	ARE/GRE	9.2e-157
Logo	**Short motif**	**E-value**
	half-site ARE/GRE	9.8e-024

Figure 5. AR & GR direct targets and expressed genes in prostate.

A. Venn diagram of the common genes between ChIP-Seq and RNA-Seq in prostate.

B. Significant cis-elements of the 793 common genes between ChIP-Seq and RNA-Seq in prostate using AR coordinates.

C. Significant cis-elements of the 793 common genes between ChIP-Seq and RNA-Seq in prostate using GR coordinates.

Figure 6.

A

GR prostate | AR prostate

7940 GR pro peaks

Cluster 1 — 1056 peaks
Cluster 2 — 1495 peaks
Cluster 3 — 2277 peaks
Cluster 4 — 2607 peaks

-5 kb | +5 kb
peak center

B

GR pro
AR pro

Cluster 1 - Merged profile
Shared AR/GR pro

Cluster 2 - Merged profile
Shared AR/GR pro

Cluster 3 - Merged profile
Unique GR pro

Cluster 4 - Merged profile
Unique GR pro
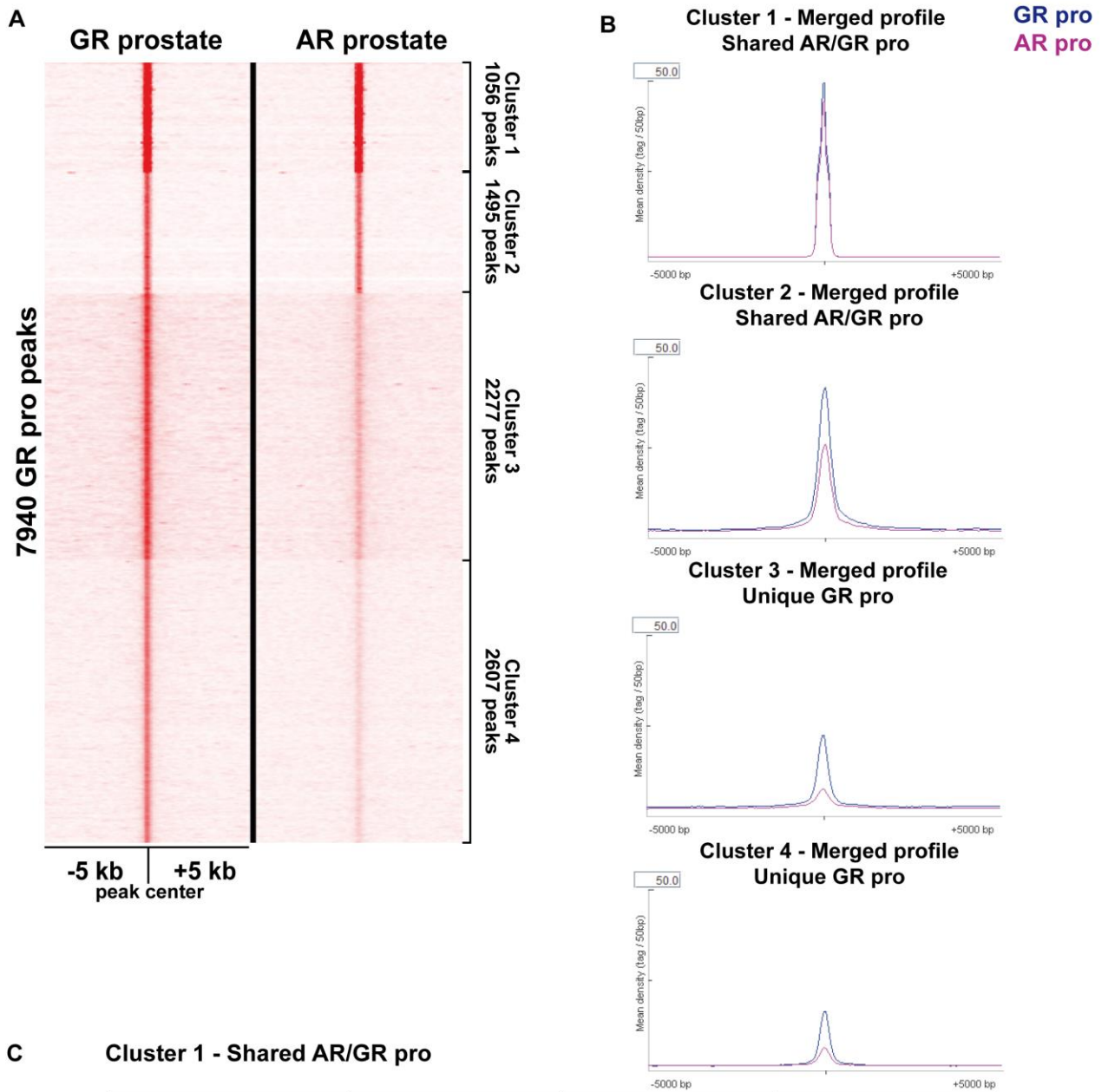
C   Cluster 1 - Shared AR/GR pro

Logo	Motif	Sites	E-value
GaACa TGT	**ARE/GRE**	**204/1056**	**1.3e-110**
Logo	Short motif	-	E-value
aGaACA	Half site ARE/GRE	-	9.8e-024

Cluster 2 - Shared AR/GR pro

Logo	Short motif	E-value
GAAAgAGA	PRDM1	2.0e-023

218

D　　　**Cluster 3 - Unique GR pro**

Logo	Motif	Sites	E-value
	TP53/63/73	448/2277	2.6e-069
Logo	**Short motif**	**-**	**E-value**
	Forkheadbox (FOX)	-	5.1e-028
	TEAD2	-	2.7e-017

Cluster 4 - Unique GR pro

Logo	Short motif	E-value
	Homeo domain factors	9.6e-021
	TP53/63/73	9.1e-004

Figure 6. **Cistromic overlap of AR & GR in prostate.**

A. Tag density maps showing AR and GR binding events centered within a +/-5 kb window in prostate. The shared and the unique clusters are depicted.

B. Mean tag density merged profiles for GR at unique and shared GRBS with ARBS in prostate in a window of +/-5 kb centered around the summit of the binding sites.

C. Tables of significant cis-elements as identified by *de novo* motif searches in shared clusters.

D. Tables of significant cis-elements as identified by *de novo* motif searched in unique clusters.

Figure 7.

A

GR prostate   AR prostate   H3K4me1 prostate

Cluster 1
1754 peaks

7940 GR pro peaks

Cluster 2
2975 peaks

Cluster 3
3218 peaks

-5 kb | +5 kb
peak center

B

GR
AR
H3K4me1

Cluster 2 - Shared GR/H3K4me1 Merged profile

35.0

Mean density (tag / 50bp)

-5000 bp   +5000 bp

Cluster 3 - Shared GR/AR Merged profile

35.0

Mean density (tag / 50bp)

-5000 bp   +5000 bp

C

Cluster 1

Promoter 8%   5UTR 1%   3UTR 1%   TTS 1%   Exon 1%
Intergenic 31%
Intron 57%

Cluster 2

Promoter 8%   5UTR 1%   3UTR 1%   TTS 2%   Exon 2%
Intergenic 47%
Intron 39%

Cluster 3

Promoter 3%   5UTR 0%   3UTR 0%   TTS 1%   Exon 2%
Intergenic 61%
Intron 33%

220

D **Cluster 1 - Shared GR/H3K4me1**

Logo	Motif	Sites	E-value
	GRE	**366/1754**	**1.3e-059**
Logo	**Short motif**	**-**	**E-value**
	Half-site GRE	-	5.4e-049
	Forkheadbox(FOX)	-	1.2e-029
	STAT1	-	4.1e-020
	SMAD/NF-1 DNA binding domain factors	-	7.3e-017
	Homeo domain factors	-	1.3e-010

Cluster 2 - Shared GR/H3K4me1

Logo	Motif	Sites	E-value
	GRE	**372/2975**	**2.4e-075**
Logo	**Short motif**	**-**	**E-value**
	Forkhead box(FOX)	-	2.8e-028
	ETS-related factors	-	2.7e-025
	ASCL1	-	3.9e-014

Cluster 3 - Shared GR/AR

Logo	Short motif	E-value
	PBX3	2.0e-017
	TEAD2	2.5e-011
	TEAD3	2.2e-005

221

Figure 7. Cistromic overap of AR, GR and H3K4me1 in prostate.

A. Tag density maps showing AR, GR and H3K4me1 binding events centered within a +/-5 kb window in prostate. Clusters of genomic co-localization are depicted.

B. Mean tag density merged profiles for GR at GRBs shared with ARBS and H3K4me1 sites in prostate in a window of +/-5 kb centered around the summit of the binding sites.

C. Pie-charts of the genomic distribution of the binding sites for each cluster.

D. Tables of significant cis-elements as identified by *de novo* motif searches in each cluster.

Figure 8.

A
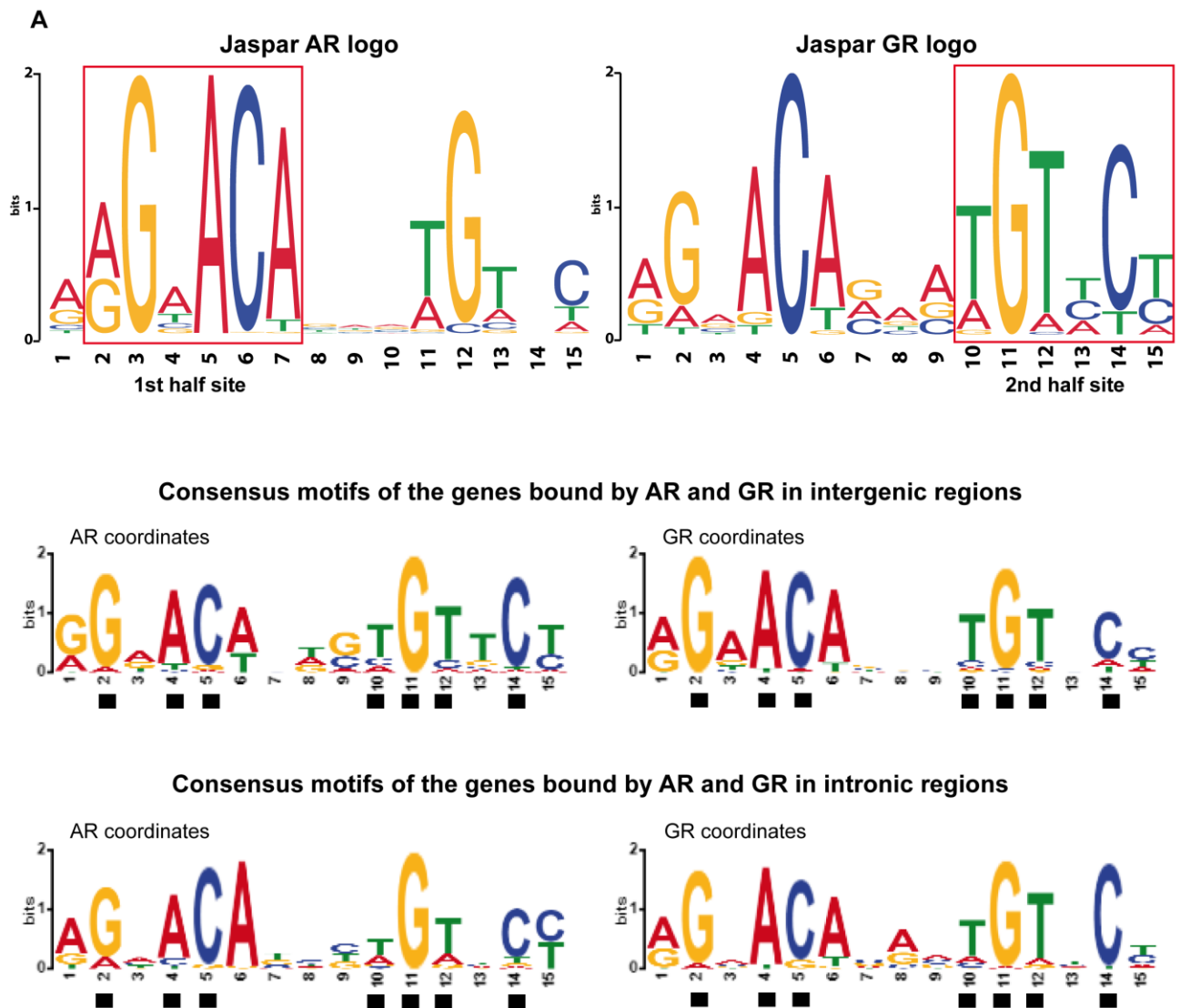
Figure 8. Comparison of AREs and GREs in prostate.

A. Comparison of the consensus ARE/GRE motifs of the genes bound by AR and GR in intronic and intergenic regions with the consensus motifs of the Jaspar database.

Part III

Objectives of the part III of the thesis

Myogenesis is a two-step process including determination of the muscle lineage committed from satellite cells and differentiation of committed myoblasts to myotubes (Moran et al., 2002; Pownall et al., 2002).

Myoblast differentiation is required for skeletal muscle formation during embryonic muscle development or postnatal muscle regeneration (Bentzinger et al., 2012; Dumont et al., 2015). Upon adult muscle injury, the quiescent adult muscle stem cells are activated and differentiate to myoblasts, undergo proliferative expansion and differentiate into myocytes that ultimately fuse to form new myofibers (Peng et al., 2017).

C2C12 cells, derived from murine skeletal muscle cells, is a well-established model which mimics the development of skeletal muscle (Burattini et al., 2004; Cheema et al., 2003; Manabe et al., 2012; Moran et al., 2002; Nedachi et al., 2008; Schoneich et al., 2014). Upon withdrawal of the serum in a culture medium, proliferating myoblasts exit the cell cycle and activate the differentiation program to differentiate to myocytes that fuse (Maglara et al., 2003). During this process, many myogenic regulatory factors (MRFs) including PAX7, MYF5, MYOD, MYOG, and MRF4 act in a sequential manner to activate the transcriptional reprogramming (Bentzinger et al., 2012; Dumont et al., 2015; Rudnicki and Jaenisch, 1995).

To provide insights into the molecular and transcriptional mechanisms underlying muscle differentiation, we performed a genome-wide comparative analysis of C2C12 myoblasts and myotubes and murine skeletal muscle tissue in terms of gene expression, levels of expression, functional annotation, families of transcription factors, binding motifs and histone modifications.

Genome-wide comparative transcriptomic and epigenomic analyses of cultured murine myoblasts and murine skeletal muscle

Results

Comparative transcriptomic and functional enrichment analysis between C2C12 myoblasts and myotubes and murine skeletal muscle

To compare the expression profiles between C2C12 cells and skeletal muscle, we analyzed transcriptomic (RNA-Seq) data of proliferating C2C12 myoblasts (MB) (Mousavi et al., 2012), of 3 days differentiated myotubes (MT) (Doynova et al., 2017) and of gastrocnemius muscles (GM) from 10 week-old wild-type (WT) mice. By setting the number of reads greater than 100, we identified 7978, 8098 and 10192 genes expressed in MB, MT, and GM, respectively.

<u>Common genes</u>

The overlap between the 3 datasets revealed that 57% of the genes expressed in either cell type were expressed in MB, MT, and GM (6496/11395 genes) (Figure 1A). To determine the enriched biological pathways within these 6496 common genes we performed an Overrepresentation Enrichment Analysis (OEA), which revealed pathways such as **translation factors, transcription Initiation**, **metabolism** and **cell cycle** (Figure 1B). To investigate further the common expressed genes, we associated them with their levels of expression by visualizing their mean-centered normalized expression using a heatmap. Hierarchical clustering defined three main clusters, genes over expressed in GM (cluster 1), in MB (cluster 2), and in MT (cluster 3), respectively (Figure 1C). Pathway/GO analysis of the first cluster revealed terms associated with **mitochondrial function**, **transport**, and **autophagy** (Figure 1D). Pathway/GO analysis of the second cluster reported terms associated with **RNA functions**, **cell cycle**, **chromatin organization** and **DNA Damage Response (DDR)** (Figure 1E). Pathway/GO analysis of the third cluster indicated terms associated with **autophagy**, **transport** and **protein modifications** (Figure 1F).

<u>Shared genes</u>

Moreover, 4.9% of the total number of expressed genes (562/11395) were common between MB and GM, but not with MT (Figure 2A). OEA of these genes revealed pathways involved in **telomeres**, **DNA repair**, **homologous recombination** and **cell cycle** (Figure 2B). Using a heatmap we visualized the mean-centered normalized expression of the shared genes and hierarchical clustering revealed 3 main clusters (Figure 2C). The first one consists of sets of genes over expressed in GM. Pathway/GO analysis revealed terms associated with **mitochondrial function** (Figure 2D). The heatmap of the genes involved in the mitochondrial functions showed that their expression is higher in GM as expected and the expression in MB is closer to GM than the expression in MT (Figure 2E). The second cluster contains genes highly expressed in both MB and GM (Figure 2C). The third cluster, the most prominent one, includes genes highly expressed in MB (Figure 2C). Pathway/GO analysis revealed terms

related to **cell cycle regulation**, **DNA damage** and **Homology Directed Repair (HDR)** (Figure 2D). Heatmaps of the genes involved in these pathways, depicted in Figure 2E, show that indeed most of them are highly expressed in MB and few genes are also highly expressed in GM. The expression of all these genes is low in MT (Figure 2C).

Furthermore, 4.1% of the total expressed genes (463/11395) overlapped between MB and MT, but not with GM (Figure 3A). These shared genes were involved in pathways such as **axon guidance**, **kinesins** and **mitotic cell cycle** (Figure 3B). The heatmap of the 463 genes shared between MB and MT revealed 3 main clusters (Figure 3C). The first cluster encompasses genes that are highly expressed in both MB and MT (Figure 3C). The second cluster contains genes that are over expressed in MB (Figure 3C). Pathway/GO analysis revealed terms associated with **mitotic cell cycle**, **cytoskeletal organization**, **cell shape and morphogenesis, axon guidance, DNA replication** and c**hromosome condensation** (Figure 3D). Heatmaps of the genes involved the main pathways (cytoskeletal organization, cell morphogenesis, and mitotic cell cycle) showed high expression in MB and that the expression in MT is closer to MB compared to GM (Figure 3E). Additionally, the genes in the third cluster are over expressed in MT and they are associated with **axon guidance** (Figure 3D). The heatmap of the genes involved in axon guidance from both MB and MT clusters is depicted in Figure 3E. The expression of all these genes is low in GM.

7.5% of the total expressed genes (856/11395) were shared between MT and GM, but not with MB (Figure 4A). OEA of these genes revealed pathways such as **striated muscle contraction**, **cytoskeletal regulation**, and **calcium signaling pathway** (Figure 4B). The heatmap of the 856 genes shared between GM and MT is divided into 3 main clusters (Figure 4C). The first cluster encompasses genes highly expressed in both MT and GM (Figure 4C). No significant pathways were associated with this cluster. The second cluster contains genes over expressed in GM (Figure 4C) that are associated with **muscle contraction, differentiation and development**, **ion homeostasis**, **cation transport** and **mitochondrial function** terms (Figure 4D). A heatmap showing the high expression profiles in GM of the genes involved in mitochondrial function is depicted in Figure 4E. In the third cluster the genes are over expressed in MT (Figure 4C). These genes are enriched for pathways and GO terms involved in **muscle contraction**, **muscle development**, **ion homeostasis** and **extracellular matrix organization** (Figure 4D). The heatmaps of the genes involved in muscle contraction and ion homeostasis from GM and MT clusters are depicted in Figure 4E. The enriched in MT muscle contraction genes are involved in embryonic development (Tnnt1, Fgf11, Myh8, Myh3, Myl4, Tnni1) whereas GM muscle contraction genes are associated with calcium regulation (Cacnb1, Kcnj2, Cacna1s, Atp2a1, Atp1a2, Camk2b, Cox8b, Tnni2, Tnnt3, Casq1/2, Cox6a2, Mylk/2, Ryr1, Scn4a, Fxyd6, Myl2). As expected, the expression of all these genes is low in MB.

<u>Specific genes</u>

Focusing on the specific genes of the Venn diagram, 457 genes (4% of the total expressed genes) were MB-specific (Figure 5A). The 457 MB-specific genes were associated with terms like **telomeres**, **homologous recombination**, **kinesins** as well as **cell cycle** and **mitosis** (Figure 5B). The heatmap of the mean-centered normalized expression of these genes was divided into two main clusters (Figure 5C). The first cluster, the most prominent one, is enriched for pathways and GO terms associated with **mitotic cell cycle**, **DNA replication**, **HDR** and **cytoskeletal organization** (Figure 5D). Heatmaps of the genes of the main pathways (cytoskeletal organization, HDR, DNA replication and mitotic cell cycle) are depicted in Figure 5E. In the second cluster, most of the genes were highly expressed in MB and the expression in MT was closer to MB compared to GM (Figure 5C). No significant pathways were identified for this cluster.

In addition, the 283 MT-specific genes (2.5% of the total expressed genes) (Figure 6A) were not enriched for any significant pathway. The heatmap of the mean-centered normalized expression of these genes was organized in two main clusters (Figure 6B). In the first cluster, most of the genes were highly expressed in MT, and the expression in MB was closer to MT compared to GM (Figure 6B). No significant pathways and GO terms were associated with this cluster. The second cluster, highly selective for MT, contains genes involved in **glycosylation** (Figure 6C) and the heatmap of the genes involved in glycosylation is depicted in Figure 6D.

Finally, the 2278 skeletal muscle-specific genes (20% of the total expressed genes) (Figure 7A), after OEA, were enriched for terms related to **complement activation**, **lipid metabolism and toxicity**, and **adipogenesis genes** (Figure 7B). Most of these genes are probably expressed in blood vessels and preadipocytes in skeletal muscles. The heatmap of the mean-centered normalized expression of the 2278 GM-specific genes was organized in two main clusters (Figure 7C). The first cluster, highly selective for GM, contains genes involved in **complement activation/cascade** and **peroxisome** (Figure 7D). The heatmaps of the genes associated with the two categories show a clear enrichment of these genes in GM (Figure 7E). In the second cluster mixed expression patterns were observed, most of the genes were highly expressed in GM and few genes also in MB and MT, probably explained by the low expression threshold of the analysis; more strict thresholds could give more specificity to the clustering (Figure 7C). Pathway/GO term analysis of the second cluster revealed terms associated with **break repair**, **Notch signaling** and **inflammation** (cytokines and chemokines) (Figure 7D).

<u>Differential expression analysis</u>

To quantify the differences in expression levels we performed differential expression analysis and we identified 13210 differentially expressed genes (DEGs) between GM and MT (6453 over expressed genes, 5957 under expressed genes), 13900 DEGs between GM and MB (6587 over expressed genes, 6631 under expressed genes) and 7650 DEGs between MT

and MB (3669 over expressed genes, 3545 under expressed genes) (Supplementary Figure 1A, B, C).

Among the 13900 DEGs between GM and MB, the highly over expressed genes in MB are involved in **cell proliferation**, **differentiation**, **development** and **morphogenesis**, as well as **cell cycle control** and **mitosis**, **homologous recombination** and **DNA repair**, correlating with our previous observations on the MB stage (Supplementary Figure 1A). On the other hand, the highly over expressed genes in GM have a role in **skeletal muscle contraction**, **calcium channels**, which are important for **muscle contraction**, **cytoskeletal organization**, and **mitochondrial function**, matching with our previous results on GM (Supplementary Figure 1A).

In addition, among the 13210 DEGs between GM and MT, the highly over expressed genes in MT are involved in **calcium/sodium channels**, in **embryonic development**, **morphogenesis**, **muscle contraction**, and **calcium channels** fitting with our previous findings on the MT stage (Supplementary Figure 1B). The highly over expressed genes in GM are also associated with **calcium/sodium channels**, **embryonic development**, **muscle contraction**, as well as **Notch signaling** and **mitochondrial function**, matching with our previous outcomes for GM (Supplementary Figure 1B).

Among the 7650 DEGs between MT and MB, the highly over expressed genes in MB have functions relevant to **RNA**, **embryonic development**, **cytoskeletal organization**, **cell shape**, **motility** as well as **DNA replication**, **cell cycle**, and **mitosis**, correlating with our previous conclusions for the MB stage (Supplementary Figure 1C). On the other hand, the highly over expressed genes in MT are important for **muscle contraction** and **calcium channels**, as well as **embryonic development**, **cytoskeletal organization**, **myogenesis**, and **post-translational modifications**, matching with our previous postulations on the MT stage (Supplementary Figure 1C).

To summarize, although there are genes expressed in all 3 stages (MB, MT, GM), their expression levels vary from one stage to the other. MT and GM express more genes in common compared to MB. In addition, there are 13900 differentially expressed genes between GM and MB reporting the differences between the two stages. Almost ⅓ of the genes are specifically expressed in either MB, MT or GM.

Interestingly, 3 main signatures were observed in the 6496 common genes according to the stage of differentiation: i. *MB-signature*: **cell cycle**, **RNA**, **DDR**, ii. *MT-signature*: **autophagy** and **protein modifications**, iii. *GM-signature*: **autophagy** and **mitochondrial function** (Figure 8A). MT and GM have more signatures in common compared to MB.

More specifically, the 562 shared genes between MB and GM revealed two main signatures. The MB-signature is defined by genes involved in **cell cycle control** and **mitosis**, **DDR** and **HDR** (Figure 8B). The GM-signature is defined by genes involved in **mitochondrial functions** (Figure 8B). The 463 shared genes between MB and MT revealed also two main signatures,

the MB-signature defined by genes involved in **mitosis** and **cell cycle**, **cell shape** and **cytoskeleton organization** as well as **axon guidance** and the MT-signature defined by genes involved in **axon guidance** (Figure 8B). Finally, the 856 shared genes between MT and GM revealed similar signatures defined by **muscle contraction** and **ion homeostasis**, except for the GM-signature which is also defined by **mitochondrial functions** (Figure 8B).

Lastly, the MB-specific signature was determined by genes involved in **mitosis** and **cell cycle**, **HDR** and **DNA replication** and **cytoskeleton organization** (Figure 8C). The MT-specific signature was defined by **glycosylation** and the GM-specific signature as specified by genes involved in **complement activation**, **peroxisome**, **break repair** and **inflammation** (Figure 8C).

Enriched classes and families of Transcription Factors in the specific and shared genes during myogenesis

We then focused on potentially enriched transcription factor target genes among the genes of the specific and shared genes using the software **PASTAA** and setting a threshold of less than 0.05 for the enrichment to be considered as significant.

Figure 9 recapitulates the enriched transcription factors in the common, shared and specific genes.

Common genes

The 6496 common genes between MB, MT, and GM (Figure 9A) are enriched for factors of the class of basic leucine zipper (bZIP) and more precisely factors of the Jun-related family (Nrf1, Jun, Nfe2l2) and the Creb-related family (Creb1, Atf1/3). Moreover, factors of the class of basic helix-span-helix (bHSH) are enriched and especially the ones of AP2 family (Tfap2a), as well as of the class of basic helix-loop-helix factors (bHLH) and more precisely of the PAS domain family (Ahr, Hif1a, Arnt, Clock), the Hairy-related family (Hes1, Bhlhe40) and the bHLH-ZIP family (Myc, Srebf1). Factors of the class of C2H2 zinc finger (Hic1, Znf143, Zic1, Kif12, Egr1, Mzf1, Sp1, and Yy1), of the class of nuclear receptors with C4 zinc fingers (Ppara) and of the class of C4 zinc finger-type factors (Gata1) are also enriched among the common genes. Additional enriched factors belong to the families of Homeo (Nkx2-1), NFY (Nfya), E2f-related factors (E2f1), ETS-related factors (Elk1) and to the classes of paired box factors (Pax4), TEA domain factors (Tead2), GCM domain factors (Gcm1), Tryptophan cluster factors (Myb), p53 domain factors (P53) and STAT domain factors (Stat1).

Shared genes

The 562 shared genes between MB and GM (Figure 9B) are enriched for factors of the class of bZIP and especially factors of the Creb-related family (Creb1, Atf1/2), the Jun-related family (Jun, Bac1/2, Nrf1) and the Fos-related family (Fosb, Fosl1). Moreover, the factors of the class of bHSH are enriched and especially factors of the AP2 family (Tfap2a), as well as of the class of bHLH factors and more precisely of the PAS domain family (Arnt, Clock), the

Hairy-related family (Hes1, Bhlhe40), the bHLH-ZIP family (Myc, Srebf1, Usf1/2), the E2A-related family (Tcf3) and the MYOD/ASC-related family (Myod1). Factors of the class of C2H2 zinc finger (Hic1, Znf143, Sp1, Zbtb18, Mtf1) are also enriched among the shared genes. Additional enriched factors belong to the families of E2F-related factors (E2f1, Tfdp1), ETS-related factors (Elk1, Ets1) and to the classes of paired box factors (Pax5), Rel homology region factors (Rbpj) and SMAD/NF1 DNA binding domain factors (Smad4).

The 463 shared genes between MB and MT (Figure 9B) are enriched for factors of the class of bZIP and especially for factors of the Jun-related family (Bach1/2, Nfe2l2), the class of bHSH and more precisely of the AP2 family (Tfap2a) and the class of bHLH and especially of the families bHLH-ZIP (Srebf1, Tfap4) and PAS domain (Ahr, Hif1a). Furthermore, factors of the class of C2H2 zinc finger (Egr1, Mtf1, Hic1, Sp1) are enriched among the shared genes as well as of the class of nuclear receptors with C4 zinc fingers (Nr2f2). Additional enriched factors belong to the families of NFY (Nfya/b), E2F-related factors (E2f1) and to the classes of paired box factors (Pax4) and TEA domain factors (Tead2).

The 856 shared genes between MT and GM (Figure 9B) are enriched for factors of the class of bZIP and especially for factors of the Fos-related family (Fosb, Fosl1), the Jun-related family (Bach2) and the Creb-related family (Creb1). In addition, factors of the class of bHSH are enriched and especially of the AP2 family (Tfap2a), as well as of the class of bHLH and more precisely of the PAS domain family (Arnt, Clock), the Hairy-related family (Bhlhe40), the bHLH-ZIP family (Myc, Srebf1, Usf1/2, Tfap4), the E2A-related family (Tcf3), the MYOD/ASC-related family (Myod1) and the Tal-related family (Nhlh1). Factors of the class of C2H2 zinc finger factors (Egr1, Mtf1, Hic1, Sp1, Mzf1, Zic3, Rest, Rreb1) are also enriched among the shared genes as well as of the class of nuclear receptors with C4 zinc fingers (Nr2f2). Additional enriched factors belong to the family of E2F-related factors (E2f1) and to the classes of paired box factors (Pax1), TEA domain factors (Tead2), MADS box factors (Mef2c, Srf) and tryptophan cluster (Irf1, Myb).

Specific genes

The 457 MB-specific genes (Figure 9C) are enriched for factors of the class of bZIP and especially for factors of the Creb-related family (Creb1). In addition, factors of the class of bHSH are enriched and especially of the AP2 family (Tfap2c) as well as of the class of bHLH factors and more precisely of the PAS domain family (Arnt, Ahr), the Hairy-related family (Bhlhe40, Hes1) and the bHLH-ZIP family (Myc, Srebp1). Factors of the class of C2H2 zinc finger factors (Znf143, Egr1, Mtf1, Sp1), the class of nuclear receptors with C4 zinc fingers (Nr1h3/2) and the class of C4 zinc finger-type factors (Gata1) are also enriched among the specific genes. Additional enriched factors belong to the families of NFY (Nfya), E2F-related factors (E2f1), ETS-related factors (Elk1) and to the classes of Homeo domain factors and especially to the family of POU domain factors (Pou3f2, Pou2f1), paired box factors (Pax9), Rel homology region and especially the NF-kappaB-related factors (Nfkb1/2, Rela) and SAND domain factors and especially the Deaf family factors (Deaf1).

The 283 MT-specific genes (Figure 9C) are enriched for factors of the class of bZIP and especially for factors of the Fos-related family (Fosb, Fosl1) and the Jun-related family (Bach2, Jun). In addition, factors of the class of bHLH are enriched and more precisely of the Hairy-related family (Bhlhe40), the bHLH-ZIP family (Tfap4), the E2A-related family (Tcf3, Tcf4) and the MYOD/ASC-related family (Myod1) as well as of the class of C2H2 zinc finger factors (Egr1, Egr2). Additional enriched factors belong to the classes of Homeo domain factors (Zeb, Cux1), paired box factors (Pax4), STAT domain factors (Stat1a/b), Rel homology region factors and especially the family of NF-kappaB-related factors (Nfkb1/2), MADS box factors (Mef2a, Srf) and SMAD/NF1 DNA binding domain factors (Smad4, Smad1).

Lastly, the 2278 GM-specific genes (Figure 9C) are enriched for factors of the class of bZIP and especially for factors of the family of C/EBP-related factors (Cebpa). Furthermore, factors of the class of bHSH are enriched and especially of the AP2 family (Tfap2a/c) as well as of the class of bHLH factors and more precisely of the PAS domain family (Arnt, Hif1a), the Hairy-related family (Bhlhe40), the bHLH-ZIP family (Max, Myc, Sreb11), the E2A-related family (Tcf3, Tcf4) and the MYOD/ASC-related family (Myod1). Factors of the class of C2H2 zinc finger factors (Zic2, Zbtb7a, Egr1, Mzf1, Sp1, Hic1, Rreb1, Rest) and of the class of nuclear receptors with C4 zinc fingers (Vdr) are also enriched among the specific genes. Additional enriched factors belong to the families of NFY (Nfya/b), E2F-related factors (E2f1) and to the classes of paired box factors (Pax4, Pax5, Pax9), Rel homology region and especially the NF-kappaB-related factors (Nfkb1/2, Rela), SMAD/NF1 DNA binding domain factors (Smad4) and MADS box (Mef2a).

Clustering of Transcription Factors according to expression profiling and functional annotation during myogenesis

We were also interested in the expression levels of the known mouse TFs and their evolution through myogenesis. Therefore, we focused on specific TF families and TFs involved in certain biological processes in order to monitor their evolution through the myogenic process and their preferences for a particular stage (Figure 10).

The majority of the **Zinc finger proteins (Zfp)** are highly expressed in GM, some are highly expressed in both MT and GM, others in MB and some few in MT and in both MB and MT (Figure 10A).

Among the **NF-kb subunits**, Nfkbil1, NFkbib and Relb are highly expressed in GM, Nfkbie, Nfkb2, Rest and Rela are over expressed in both MB and MT and Nfkb1 and Rel are over expressed in MB (Figure 10A).

The TFs and TFs families involved in **mitochondrial functions** are spread throughout the differentiation process, Stat3 is over expressed in MT stage, Stat1, Tfam, Mef2d, Nfkbib, Nfkbil1 are highly expressed in GM, Pparg and Nfkb1 are over expressed in MB and Nfkb2 and Nfkbie are highly expressed in both MB and MT (Figure 10A).

Moreover, among the **AP-1 subunits**, Fosb, Fosl1 and Junb are highly expressed in MB, Fos and Jun are over expressed in GM and only Fosl2, which is a regulator of cell proliferation, differentiation, and transformation, is selectively expressed in the MT stage (Figure 10B).

The **homeobox (Hox)** factors are mostly expressed in MB stage and GM, and all of them are under expressed in MT stage (Figure 10B).

Some **SIX homeobox (Six)** genes are highly expressed in GM (Six2, Six5) and in MB (Six3, Six4) and only Six1 is over expressed in MT and MB (Figure 10B).

The class of **paired box (Pax)** factors is over expressed in MB (Pax1, Pax2, Pax6, Pax7, and Pax8) and GM (Pax3, Pax5, and Pax9) (Figure 10B).

The **SRY-box transcription (Sox)** factors are over expressed in MB (Sox1, Sox12, Sox4, Sox9) and GM except for Sox15, Sox11 and Sox8 which are involved in the regulation of embryonic development and in the determination of the cell fate and they are selectively enriched in MT (Figure 10B).

The family of the **ETS-related** factors was mostly enriched in MB (Elf4, Elk3, Erf, Etv4), in GM (Elf3, Elf2, Erg, Fli1, Spdef, Etv3) and in both MB and GM (Elk4, Ets1, Spib/c), with the exception of Elf1, Elk1, Gabpa and Etv1 (involved in proliferation and differentiation) that are over expressed selectively in MT and Etv6 and Ets2 (involved in proliferation and differentiation) that are highly expressed in both MB and MT (Figure 10B).

The **Notch receptors** and other factors involved in **Notch signaling** such as Hes1, Notch3 and Heyl are over expressed in MT, Notch1, which is expressed in satellite cells, is over expressed also in MB together with the Notch2 and Notch4 and Hey1 are over expressed in GM (Figure 10C).

The signal **transducer and activator of transcription (Stat)** factors such as Stat4, Stat1, Sta5b and Stat6 are highly expressed in GM and Stat2, Stat3 and Stat5a are over expressed in MT stage (Figure 10C).

The **myocyte enhancer factors 2 (Mef2)** are over expressed between the MT stage (Mef2a) and GM (Mef2c/d) (Figure 10C).

The TF families involved in **cell cycle regulation** like Kruppel-like factors are highly expressed in MB (Klf1, Klf16, Klf13, Klf7), MT (Klf4, Kl3, Klf5) and GM (Klf15, Klf12, Klf2), and the majority of the E2f transcription factors are highly expressed in MB (E2f1, E2f5, E2f3, E2f4), with the exception of the E2f2, which is over expressed selectively in MT and the E2f6, which is over expressed uniquely in GM (Figure 10D).

The TFs and TFs families involved in **DNA replication** like Brca1 and the minichromosome maintenance complex components are highly expressed in the MB stage and the SWI/SNF related, matrix associated, actin-dependent regulator of chromatin, subfamily A members are spread among MB (Smarca4, Smarcb1), MT (Smarce1) and GM (Smarca1/2) (Figure 10E).

Some **Transcription Factors (Tcf)** are over expressed in the MB stage (Tcf20, Tcf7, Tcf3, Tcf19, Tcf7l2), whereas Tcf15 is highly expressed in GM, and Tcf4 and Tcf12 are highly expressed in MT (Figure 10E).

Moreover, the **Forkhead box (Fox)** genes are over expressed mostly in MB stage (Figure 10E). Foxo1, which may play a role in myogenic growth and differentiation, and Foxo3 are selectively over expressed in MT, Foxa3, Foxn2, Foxc1 (regulation of embryonic development), Foxp3 and Foxq1 (embryonic development) are highly expressed in both MB and MT stages and Foxd1, Foxk1, Foxd3, Foxj2 and Foxp2 are over expressed in GM (Figure 10E).

The **TEA domain transcription (Tead)** factors like Tead2 and Tead3 are highly expressed in both MB and MT stages, Tead1 is over expressed only in GM and Tead4 is over expressed only in MT (Figure 10F).

The heatmap of the **myogenic regulatory factors (MRFs)** is showing the evolution of these factors through the differentiation process (Figure 10F). Myod1 and Myf5 are important for the formation or survival of the MB, and they are over expressed in MB and MT stages, Myog has a role in the terminal differentiation of MT, and it is highly expressed in MT stage and only Myf6, which is essential for the last stage of differentiation, is highly expressed in GM (Figure 10F).

Most of the **POU domain transcription (Pou)** factors are over expressed in GM (Figure 10G). Pou2af1 and Pou4f1 also expressed in MT stage and Pou2f1 and Pou3f2 over expressed only in MB (Figure 10G).

The majority of the **nuclear receptors (Nr)** are highly expressed in GM, although some are also highly expressed in MB (Nr2f2, Nr4a2, Ppard, Pparg, Rara), MT (Nr1i2, Nr2c1, Nr2c2) and both MB and MT stages (Nr1i3, Nr2f6, Nr2f1, Rarg, Vdr, Rorb) (Figure 10G). Interestingly, nuclear receptors from the same subfamily are not clustering together as they are not highly expressed at the same stage of differentiation (Rora and Rorc are over expressed in GM, whereas Rorb is over expressed in MB/MT). Of note, Ar and Gr (Nr3c1) are highly expressed only in GM.

As far as the myosins are concerned, the **myosin heavy chain (Myh)** genes are mostly expressed in GM (Figure 10H). Myh15, Myh9, Myh10 which are involved in cytokinesis, cell motility and maintenance of cell shape are highly expressed in MB and Myh6, Myh3, Myh8 and Myh7b are over expressed selectively in MT (Figure 10H). The **myosin light chain (Myl)** genes are highly expressed in GM, few in MB (Myl7, Myl9) and Myl12a/b, Myl4, Mylk3 and Myl6 are over expressed in MT (Figure 10H). Most of the **myosins (Myo)** are highly expressed in MB and GM and some in MT such as Myo10, Myo9a, and Myo1c/e (Figure 10H).

Comparative epigenetic profiling between murine skeletal muscle and C2C12 cells

In order to compare the epigenetic loci between C2C12 cells and skeletal muscle, we used seqMINER to visualize the tag density maps of H3K27ac ChIP-Seq reads from the 3 stages (MB, MT and GM), within +/- 5 kb and using the appropriate reference for each stage (Figure 11). H3K27ac is a marker of active enhancers and promoters.

Interestingly, 7 distinct clusters were identified: a cluster shared between MB and MT and an MB-specific cluster in the tag density map of MB (Figure 11A), a cluster shared between MB and MT and two MT-specific clusters in the tag density map of MT (Figure 11B), two GM-specific clusters in the tag density map of GM (Figure 11C). The number of peaks and genes of each cluster, as well as the boundaries of the clusters, are indicated.

By further investigating the genes in the specific and shared clusters, we found genes involved in myogenesis and muscle differentiation, TFs and TFs families with specific biological roles as well as some nuclear receptors. Interestingly, we observed that the expression levels of these genes were correlating with the epigenetic pattern of each cluster.

Shared clusters

In the shared MB/MT cluster of the MB tag density map, we found that the 34% of the genes in this cluster (1443/4252) are under expressed in GM and that the 16% of the genes (696/4252) are over expressed in GM (Figure 11A). For example, the genes Myod1 and Pax7 are under expressed in GM compared to MB and MT (genes in green), correlating with the absence of H3K27ac signal in GM (Figure 11A). However, few genes were also over expressed in GM compared to MB or MT (genes in red), without major effect on the epigenetic locus (Figure 11A).

In the shared MB/MT cluster of the MT tag density map, we identified that the 34% of the genes in this cluster (1187/3536) are under expressed in GM and that the 14% of the genes (503/3536) are over expressed in GM (Figure 11B). For example, the genes Pax7 and E2f8 are under expressed in GM compared to MB and MT (genes in green), correlating with the absence of the H3K27ac signal in GM (Figure 11B). As previously, some genes were over expressed in GM compared to MB or MT (genes in red), without major effect on the epigenetic locus.

Specific clusters

In the MT-specific cluster of the MT tag density map, the 29% of the genes in this cluster (833/2916) are under expressed in MT and the 24% of the genes (707/2916) are over expressed in MT (Figure 11B). For instance, the genes Pax7 and Myog are over expressed in MT compared to GM and Pax7 was also under expressed in MT compared to MB (Figure 11B).

Finally, in the GM-specific cluster of the GM tag density map, 20% of the genes of this cluster (521/2544) are under expressed in GM and 47% of the genes (1204/2544) are over

expressed in GM (Figure 3C). For instance, Myog was over expressed in GM compared to MB, matching with the absence of H3K27ac signal from MB (Figure 11C). Myog was also under expressed in GM compared to MT, but without major effect on the epigenetic locus (Figure 11C).

Moreover, we performed *de novo* motif analysis of the sequences under the peaks of the selected clusters and we identified common but also specific motifs present during the muscle differentiation (Figure 11D). Motifs of the classes of bHLH (Myod1, Myog, Tcf12), bZIP (JUN, FOS, FOSL2), C2H2 zinc finger (Bcl6, YY1, MTF1), Forkhead box (FOXH1, Foxd3, Foxj3) and Homeo domain (TGIF1, HOXB13, PBX3) factors are enriched throughout differentiation.

At the stage of MB, motifs of the classes of TEA domain (TEAD2), Runt domain (RUNX1), SOX-related (SOX10) and the family of ETS-related factors (Elk3) are enriched (Figure 11D). At the MT stage, motifs of the classes of the TEA domain (TEAD2), Runt domain (RUNX1), ARID domain (Arid3a), MADS box (MEF2A/B/C/D), SMAD/NF1 DNA binding domain (SMAD2), paired box (Pax2) factors, and the families of ETS-related factors (Elk1), POU domain factors (POU2F2) and steroid hormone receptors (ESR2) are highly enriched (Figure 11D). Finally, at GM, motifs of receptors like the steroid hormone (ESR2), FTZ-F1-related (Nr5a2), NGFI-B-related (NR4A2), RXR-related (NR2F2), Thyroid hormone-related (RORA) receptors are overrepresented together with the ones from the class of Rel homology region (NFAT5) factors (Figure 11D).

Genome-wide chromatin state maps of murine skeletal muscle and C2C12 cells associated with myogenic differentiation

In order to compare the epigenomes between skeletal muscle, and differentiated and undifferentiated C2C12 cells, we generated genome-wide chromatin state maps using data of chromatin immunoprecipitation followed by massive parallel sequencing (ChIP-Seq), profiling histone modifications (H3K27ac, H3K4me1, H3K4me3) and the polymerase 2 (Pol2) for the 3 stages. H3K27ac and H3K4me1 are indicating active enhancers and promoters or in general "open" chromatin regions. H3K4me3 is enriched in transcriptionally active promoters and Pol2 is known to be involved in active transcription. We chose genes from the different transcription factor families that were highly expressed according to the transcriptomic analysis and specific for each stage, and we visualized them using the IGV browser (Figure 12).

The **Foxc2** (forkhead box C2) gene (Figure 12A), an MB-specific gene involved in muscle regeneration and proliferation of multipotent muscle stem cells, is marked by broad H3K27ac mark in MB (enhancer region, gene body), less broad in MT (enhancer region) and narrow marks in GM (gene body). H3K4me1 marks are broad in MB (enhancer region) and GM (gene body) and narrow in MT (enhancer region). The H3K27ac and H3K4me1 marks are

indicating active enhancer regions in MB and MT. The presence of the Pol2 at the promoter region, only in MB, indicates active transcription at this stage. The H3K4me3 marks at the promoter of the gene, in the 3 stages, indicates active promoter region.

The **Myog** (myogenin) gene (Figure 12B), highly transcribed in MT, is marked by broad H3K27ac in MT (enhancer region, gene body) and by narrow marks in GM (enhancer region, promoter region, intronic region) and there is no deposition of mark in MB. The H3K4me1 marks are broad in GM (enhancer region, gene body) and less broad in MT (enhancer region, intronic region) and MB (enhancer region). The H3K27ac and H3K4me1 marks are indicating active enhancer regions in MT and GM. The Pol2 was present at the promoter region of Myog, indicating active transcription, but also at enhancer regions in MT. There is also a Pol2 mark at the promoter of Myog in GM. The H3K4me3 strong mark indicates the active promoter region of Myog in MT.

Furthermore, we found that the locus of the highly transcribed gene in skeletal muscle **Mef2d** (myocyte enhancer factor 2D) (Figure 12C), which is involved in control of muscle and neuronal cell differentiation and development, is broadly marked by H3K27ac and H3K4me1, indicating active enhancer regions, in GM. These marks are less broad in MB and MT. Interestingly, the H3K4me1 mark is not covering the promoter region in the 3 stages. The mark of the Pol2 at the promoter of the gene, only in GM, indicates the active transcription. The promoter of Mef2d was marked by H3K4me3 in the 3 stages.

We also identified a commonly regulated gene, **Notch3** (notch 3) (Figure 12D), which is marked by active chromatin marks at enhancer regions in the 3 stages. Of note, the H3K27ac and H3K4me1 marks in MB and MT are not covering the promoter region compared to GM which is located at enhancer and promoter regions. Moreover, the promoter of the gene is marked by Pol2 in GM, indicating active transcription. The Pol2 marks in MB and MT are located in intronic region. Lastly, the promoter of Notch3 is also marked by H3K4me3 marks in the 3 stages, indicative of its active status.

Overall, the expression levels of the cell-specific genes are correlated with the open chromatin states and the active transcription. Moreover, we observed more differences between murine skeletal muscle and C2C12 cells at the enhancer level compared to promoter regions, indicating that distant enhancers provide selectivity in the differentiation process.

Together these datasets constitute comprehensive and comparative reference maps of the epigenome of MB, MT and skeletal muscle.

Discussion

By performing a genome-wide comparative analysis between murine skeletal muscle and C2C12 cells, comparing transcriptomic and cistromic datasets, in terms of gene expression, levels of expression, functional annotation, families of transcription factors, binding motifs and histone modifications, we identified shared and specific characteristics.

The gene expression analysis revealed that 57% of the expressed genes were common among the three datasets. Moreover, 20% of the genes are selectively expressed in GM-specific genes. Among the expressed genes of the shared stages, the MT/GM shows 7.5% overlap of genes. There are more than 6400 common expressed genes among MB, MT and GM, but at various levels. In addition, there are 13900 differentially expressed genes between GM and MB reporting the differences between the two stages. Almost ⅓ of the genes are specifically expressed in either MB, MT or GM.

Pathway analysis revealed that the three stages exhibit specific signatures:
  i. *MB-signature*: mitosis, cell cycle, DNA replication, cytoskeleton regulation, DDR, HDR, axon guidance,
  ii. *MT-signature*: protein modifications, autophagy, muscle contraction, ion homeostasis, axon guidance and
  iii. GM-signature: mitochondrial functions, autophagy, muscle contraction, ion homeostasis, complement activation, peroxisome, break repair, inflammation.

The differentiated C2C12 cells share more common characteristics with skeletal muscle, rather than the undifferentiated ones.

It is well known, that upon serum deprivation, proliferating MB asynchronously activate the expression of myogenin and undergo irreversible mitotic cell cycle arrest, yet remain capable of replicating DNA and rearranging cytoskeleton. Then they undergo phenotypic differentiation and cell fusion. The outcome of this fusion is the generation of genomic DNA strand breaks (Sancho and Ouchi, 2015). Proliferating cells can enter cell-cycle arrest and repair DNA damage. In contrast, terminally differentiated cells do not replicate their genomic DNA and they show accumulation of DNA breaks and a low capacity for DNA repair mechanisms.

For differentiating MT, neural development and muscle contraction are important during the differentiation process. Additionally, ions and calcium are associated with muscle contractility. Post-translational modifications are tightly regulated over time and can be characterized by changes in gene expression associated with the myogenesis and they can be even myogenic-specific.

Muscle contraction and calcium regulation are also essential for skeletal muscle. Inflammation in response to muscle injury or disease is intimately associated with muscle regeneration and many chemokines and cytokines are involved. Moreover, complement activation promotes inflammation and muscle regeneration. Mitochondria are enriched in skeletal muscle and they regulate many critical cellular processes for skeletal muscle physiology like metabolism, energy supply, calcium homeostasis and regulation of apoptosis. Peroxisome also plays an important role in lipid metabolism and muscle physiology.

During the muscle differentiation process specific and shared classes of transcription factors are enriched. Factors of the classes of bHLH, bZIP, C2H2 zinc finger and paired box are enriched throughout this process. In addition, the shared stages are additionally enriched for factors of the class of the bHSH and the family of E2F-related factors. The specific stages were also enriched for factors of the class of Rel homology region.

In addition, *de novo* motif analysis of the MB/MT, MB-specific, MT-specific and GM-specific binding sites revealed various motifs. Those corresponding to the classes of bHLH, bZIP, C2H2 zinc finger, Forkhead and Homeo domain factors are enriched throughout the differentiation. MB and MT are additionally enriched for motifs of the classes of TEA domain, Runt domain and the family of ETS-related factors. MT and GM are also enriched for motifs of the family of steroid hormone receptors. The motifs of the family of SOX-related factors are selectively enriched in MB stage, motifs of the families ARID domain, MADS box, SMAD/NF1 DNA binding domain, POU domain factors and the class of paired box factors are uniquely enriched in MT stage and motifs of most of the factors of the class of nuclear receptors with C4 zinc fingers and the class of the Rel homology region factors are specifically enriched in GM.

Importantly, our study highlighted the expression selectivity of transcription factors and families involved in myogenesis and in other biological processes. Of course, there are shared, and specific patterns of expression and all the members of the family are not similarly expressed during the muscle differentiation process. Of note, members of the same subfamilies are not over or under expressed all at the same myogenic stage. For example, the members of the family of nuclear receptors and especially the members of the Nerve Growth Factor IB-like subfamily that are highly expressed in GM (Nr4a1, Nr4a3) but also in MB (Nr4a2).

Moreover, the correlation of the epigenetic regulation with gene expression is crucial as the former can alter chromatin dynamics during myogenesis and it is essential for skeletal muscle stem cell identity and subsequent cell development. We report that by correlating the epigenetic marks with gene expression, we found that the over expression or under expression of genes in a specific cluster correlates with the absence or presence of H3K27ac, an open chromatin mark, in the epigenetic locus. It would be interesting to make associations with additional chromatin marks and Pol2.

Finally, the chromatin state maps across genome (genome-wide changes in the epigenetic landscape) validate this correlation and they offer useful information for better understanding of the epigenetic regulation of skeletal muscle development. We report that there are specific characteristics for each stage and more differences between murine skeletal muscle and C2C12 cells at the enhancer level compared to promoter regions, indicating that distant enhancers provide selectivity in the differentiation process.

To conclude, this study provides a useful database of genes, signaling pathways and transcription factors that are differentially expressed during myogenic proliferation and differentiation. The similarities of the differentiated C2C12 with the skeletal muscle indicate that the former can be used to some extend as an *in vitro* model of skeletal muscle tissue that recapitulates myogenic differentiation. Nevertheless, as the publicly available datasets used in this study originate from different labs, we cannot exclude the presence of biases in the analysis.

Materials and Methods

Chromatin immunoprecipitation assays and ChIP-sequencing

The H3K27ac ChIP assays in C2C12 myoblasts and myotubes were performed as described in (Dell'Orso et al., 2016) using anti-histone H3 (acetyl K27) (Abcam, Cat# ab4729, Lot# GR184332-1) antibody. The data were deposed in GEO datasets under the accession code **GSE76010** and in order to be used for this analysis they were re-aligned to mm10 genome version.

The H3K27ac ChIP assay in skeletal muscle was performed as described previously (Joshi, Ueberschlag-Pitiot, et al. 2017) (Ueberschlag-Pitiot, Rovito, Rerra, et al. in preparation) using an anti-H3K27ac Active Motif (39133) antibody. Sequencing was performed by the IGBMC Microarray and Sequencing platform, a member of the "France Génomique" consortium (ANR-10-INBS-0009). Immunoprecipitated DNA samples were processed for library preparation on Illumina Hiseq 4000 sequencer as a single-read 50 base reads following Illumina's instructions. Image analysis and base calling were performed using RTA 2.7.7 and bcl2fastq 2.17.1.14. Adapter dimer reads were removed using DimerRemover. FastQC 0.11.2 was used to evaluate the quality of the sequencing. Sequenced reads were mapped to the mus musculus genome assembly 10 (mm10) using Bowtie 2 (Langmead and Salzberg, 2012). Uniquely mapped reads were used for further analysis. The peak calling was performed using the MACS2 algorithm (Feng et al., 2011). Peaks were annotated relative to genomic features using Homer (Heinz et al., 2010) according to the distance to the nearest TSS. Data visualization was carried out using Integrative Genomics Viewer (IGV) (Robinson et al., 2011). De novo motif analysis of the binding sites by adding 100 nucleotides at both sides of the peak summit was performed by using the online software MEME-ChIP (Machanick and Bailey, 2011) from MEME-Suite (Bailey et al., 2009) after extracting their nucleotide sequences. The de novo motifs were then compared to a database of known motifs and ranked by the TOMTOM tool (Kumar et al., 2007) of MEME Suite. The tag density maps were produced using the software SeqMiner (Ye et al., 2011) and the clustering normalization was done using the KMeans linear method.

RNA extraction and RNA sequencing

Total RNA was isolated from C2C12 myoblasts (2 replicates) as described in (Mousavi et al., 2012) and from the 3 days differentiated C2C12 myotubes (2 replicates) as described in (Doynova et al., 2017). The data were deposed in the GEO datasets under the accession numbers **GSE25549** and **GSE84158**, respectively and the C2C12 myoblasts data were re-aligned in mm10 genome version in order to be used for this analysis.

Total RNA was isolated from control gastrocnemius muscle samples (4 samples) using Trizol. Sequencing was performed by the IGBMC Microarray and Sequencing platform, a member of the "France Génomique" consortium (ANR-10-INBS-0009). The library was prepared on Illumina Hiseq 4000 sequencer as a single-read 50 base reads following Illumina's

instructions. Image analysis and base calling were performed using RTA 2.7.7 and bcl2fastq 2.17.1.14. Adapter dimer reads were removed using DimerRemover. FastQC 0.11.2 was used to evaluate the quality of the sequencing. Reads were mapped onto the mm10 assembly of the mouse genome using Tophat 2.1.1 (Kim et al., 2013) and the Bowtie2 2.3.4.3 (Langmead and Salzberg, 2012). Only uniquely aligned reads have been retained for further analyses. Quantification of gene expression was performed using HTSeq-0.11.0 (Anders et al., 2015). For the comparisons among the datasets, we considered arbitrarily as expressed the genes that have more than 100 raw reads. Read counts were normalized across libraries with the method proposed by Anders and Huber (Anders and Huber, 2010). Comparisons of interest were performed using the method proposed by Love et al. (Love et al., 2014) implemented in the DESeq2 Bioconductor library (DESeq2 v1.0.19). Resulting p-values were further adjusted for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg 1995). A gene is differentially expressed if the adjusted p-value is less than 0.05 and the |log2 Fold-change| > 0.5.

Hierarchical clustering was performed using Cluster 3.0 (de Hoon et al., 2004) and the heatmaps were visualized using the Java TreeView (Saldanha, 2004). The pathway analysis was performed using the online softwares WEB-based Gene SeT AnaLysis Toolkit (WebGestalt GSAT), and more specifically the method Overrepresentation Enrichment Analysis using FDR less or equal to 0.05 (Liao et al., 2019), and Metascape (Zhou et al., 2019). PASTAA program predicts Transcription Factors (TFs) regulating a user defined set of genes and it was used to rank all TF matrices according to how strongly they associate with the input set (Roider et al. 2009).

References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res *37*, W202-208.

Bentzinger, C.F., Wang, Y.X., and Rudnicki, M.A. (2012). Building muscle: molecular regulation of myogenesis. Cold Spring Harb Perspect Biol *4*.

Burattini, S., Ferri, P., Battistelli, M., Curci, R., Luchetti, F., and Falcieri, E. (2004). C2C12 murine myoblasts as a model of skeletal muscle development: morpho-functional characterization. Eur J Histochem *48*, 223-233.

Cheema, U., Yang, S.Y., Mudera, V., Goldspink, G.G., and Brown, R.A. (2003). 3-D in vitro model of early skeletal muscle development. Cell Motil Cytoskeleton *54*, 226-236.

de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453-1454.

Dell'Orso, S., Wang, A.H., Shih, H.Y., Saso, K., Berghella, L., Gutierrez-Cruz, G., Ladurner, A.G., O'Shea, J.J., Sartorelli, V., and Zare, H. (2016). The Histone Variant MacroH2A1.2 Is Necessary for the Activation of Muscle Enhancers and Recruitment of the Transcription Factor Pbx1. Cell Rep *14*, 1156-1168.

Doynova, M.D., Markworth, J.F., Cameron-Smith, D., Vickers, M.H., and O'Sullivan, J.M. (2017). Linkages between changes in the 3D organization of the genome and transcription during myotube differentiation in vitro. Skelet Muscle *7*, 5.

Dumont, N.A., Bentzinger, C.F., Sincennes, M.C., and Rudnicki, M.A. (2015). Satellite Cells and Skeletal Muscle Regeneration. Compr Physiol *5*, 1027-1059.

Feng, J., Liu, T., and Zhang, Y. (2011). Using MACS to identify peaks from ChIP-Seq data. Curr Protoc Bioinformatics *Chapter 2*, Unit 2 14.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kumar, A., Bhatti, S.S., Sharma, S., Gupta, S.D., and Kumar, R. (2007). Inflammatory pseudotumor of urinary bladder - a diagnostic and management dilemma. Int Urol Nephrol *39*, 799-802.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res *47*, W199-W205.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696-1697.

Maglara, A.A., Vasilaki, A., Jackson, M.J., and McArdle, A. (2003). Damage to developing mouse skeletal muscle myotubes in culture: protective effect of heat shock proteins. J Physiol *548*, 837-846.

Manabe, Y., Miyatake, S., Takagi, M., Nakamura, M., Okeda, A., Nakano, T., Hirshman, M.F., Goodyear, L.J., and Fujii, N.L. (2012). Characterization of an acute muscle contraction model using cultured C2C12 myotubes. PLoS One *7*, e52592.

Moran, J.L., Li, Y., Hill, A.A., Mounts, W.M., and Miller, C.P. (2002). Gene expression changes during mouse skeletal myoblast differentiation revealed by transcriptional profiling. Physiol Genomics *10*, 103-111.

Mousavi, K., Zare, H., Wang, A.H., and Sartorelli, V. (2012). Polycomb protein Ezh1 promotes RNA polymerase II elongation. Mol Cell *45*, 255-262.

Nedachi, T., Fujita, H., and Kanzaki, M. (2008). Contractile C2C12 myotube model for studying exercise-inducible responses in skeletal muscle. Am J Physiol Endocrinol Metab *295*, E1191-1204.

Peng, X.L., So, K.K., He, L., Zhao, Y., Zhou, J., Li, Y., Yao, M., Xu, B., Zhang, S., Yao, H.*, et al.* (2017). MyoD- and FoxO3-mediated hotspot interaction orchestrates super-enhancer activity during myogenic differentiation. Nucleic Acids Res *45*, 8785-8805.

Pownall, M.E., Gustafsson, M.K., and Emerson, C.P., Jr. (2002). Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. Annu Rev Cell Dev Biol *18*, 747-783.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Rudnicki, M.A., and Jaenisch, R. (1995). The MyoD family of transcription factors and skeletal myogenesis. Bioessays *17*, 203-209.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246-3248.

Sancho, S.C., and Ouchi, T. (2015). Cell Differentiation and Checkpoint. Int J Cancer Res Mol Mech *1*.

Schoneich, C., Dremina, E., Galeva, N., and Sharov, V. (2014). Apoptosis in differentiating C2C12 muscle cells selectively targets Bcl-2-deficient myotubes. Apoptosis *19*, 42-57.

Ye, T., Krebs, A.R., Choukrallah, M.A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res *39*, e35.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun *10*, 1523.

Figure 1.

A

Venn diagram of the 3 RNA-Seq datasets
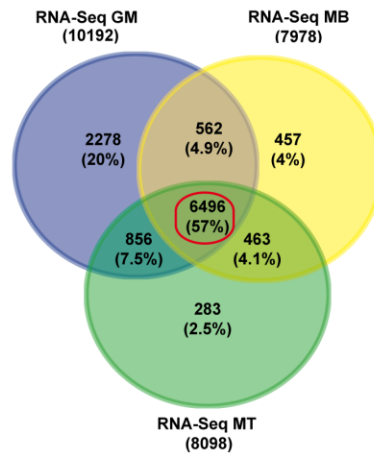
B Pathway analysis of the common genes among MB, MT and GM (6496)

C MB MT GM

6496 common genes MB/MT/GM

GM

MB

MT

Mean centered normalized expression colorbar

-1,00
-0,67
-0,33
0,00
0,33
0,67
1,00
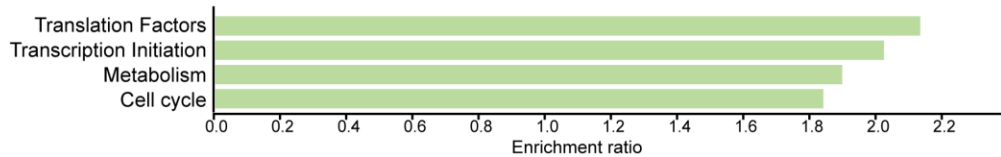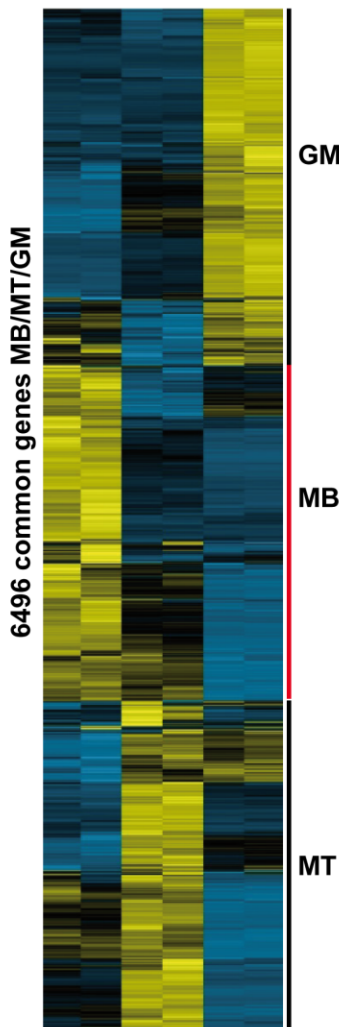
D GM cluster: mitochondrion + autophagy + transport
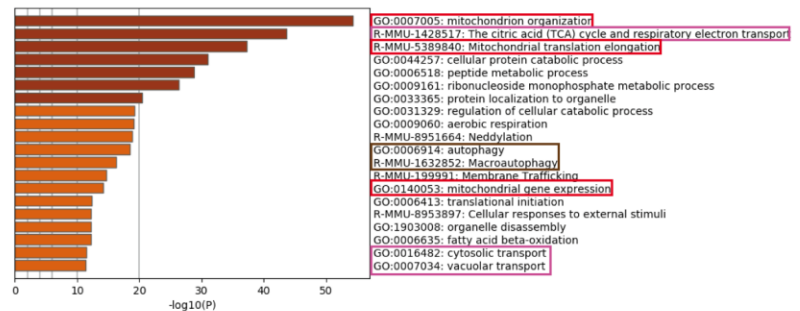
E MB cluster: RNA + mitotic cell cycle + DDR
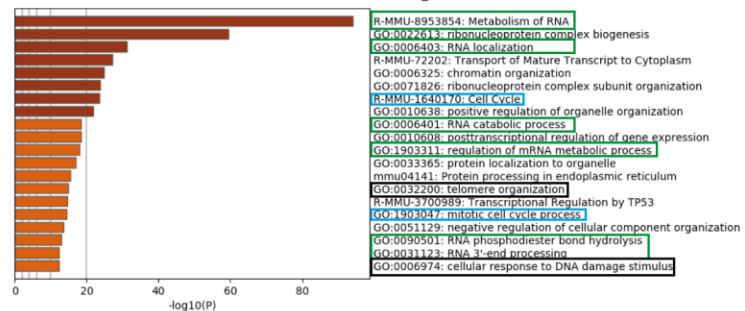
F MT cluster: protein modifications + autophagy + transport

Figure 1. **Common expressed genes.**

A. Venn diagram of expressed genes from RNA-Seq of gastrocnemius muscles (GM) of 10 week-old wild-type mice (purple), proliferating C2C12 myoblasts (MB) (yellow) and 3-day differentiated C2C12 myotubes (MT) (green). The total number of genes for each dataset as well as the % of the single represented total number of genes are indicated. N=2 for GM, n=2 for MB and n=2 for MT.

B. Pathway analysis of the common expressed genes among MB, MT and GM. The pathways depicted have an FDR < = 0.05. N=2 for GM, n=2 for MB and n=2 for MT. The color code is suggested in Figure 1A.

C. Heatmap of the mean-centered normalized expression of the common expressed genes among MB, MT and GM. The different clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

D. Pathway analysis of the GM cluster. The main pathways are boxed.

E. Pathway analysis of the MB cluster. The main pathways are boxed.

Figure 2.

A

Venn diagram of the 3 RNA-Seq datasets

RNA-Seq GM
(10192)

RNA-Seq MB
(7978)

2278
(20%)

562
(4.9%)

457
(4%)

6496
(57%)

856
(7.5%)

463
(4.1%)

283
(2.5%)

RNA-Seq MT
(8098)

B

Pathway analysis of the common genes between MB and GM (562)

Telomere C-strand Synthesis
HDR through Homologous Recombination
DNA Double-Strand Break Repair
Cell Cycle

Enrichment ratio

C

MB   MT   GM

562 shared genes MB/GM

GM

Shared
MB/GM

MB

D

GM cluster: mitochondrion

GO:0006399: tRNA metabolic process
GO:0140053: mitochondrial gene expression
GO:0043039: tRNA aminoacylation
R-MMU-189445: Metabolism of porphyrins
R-MMU-1614635: Sulfur amino acid metabolism
mmu00514: Other types of O-glycan biosynthesis
GO:0051186: cofactor metabolic process
GO:0045839: negative regulation of mitotic nuclear division
GO:0070129: regulation of mitochondrial translation
GO:0007005: mitochondrion organization
GO:0010939: regulation of necrotic cell death
GO:0006974: cellular response to DNA damage stimulus
GO:0034142: toll-like receptor 4 signaling pathway
R-MMU-674695: RNA Polymerase II Pre-transcription Events
R-MMU-5696400: Dual Incision in GG-NER

-log10(P)

MB cluster: cell cycle + HDR/DDR

GO:0006974: cellular response to DNA damage stimulus
R-MMU-5693567: HDR through Homologous Recombination (HRR) or Single Strand Annealing (SSA)
GO:0071897: DNA biosynthetic process
mmu03430: Mismatch repair
R-MMU-1640170: Cell Cycle
GO:0003002: regionalization
R-MMU-446193: Biosynthesis of the N-glycan precursor (dolichol lipid-linked oligosaccharide, LLO) and transfer to
GO:0001522: pseudouridine synthesis
GO:0051053: negative regulation of DNA metabolic process
R-MMU-5357769: Caspase activation via extrinsic apoptotic signalling pathway
mmu00240: Pyrimidine metabolism
GO:0006284: base-excision repair
GO:0034655: nucleobase-containing compound catabolic process
GO:0021532: neural tube patterning
GO:0051303: establishment of chromosome localization
GO:0098917: retrograde trans-synaptic signaling
GO:0045786: negative regulation of cell cycle
GO:0090305: nucleic acid phosphodiester bond hydrolysis
GO:0035019: somatic stem cell population maintenance

-log10(P)

**Mean centered
normalized
expression colorbar**

-1,00
-0,67
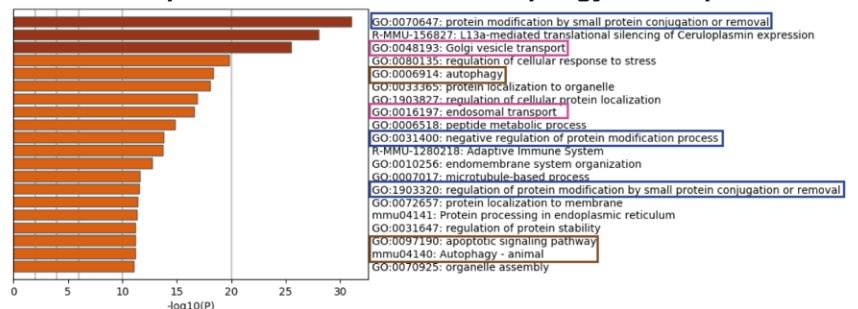-0,33
0,00
0,33
0,67
1,00

E

Figure 2. Shared expressed genes between MB and GM.

A. Venn diagram of expressed genes from RNA-Seq of gastrocnemius muscles (GM) of 10 week-old-wild-type mice (purple), proliferating C2C12 myoblasts (MB) (yellow) and 3-day differentiated C2C12 myotubes (MT) (green). The total number of genes for each dataset as well as the % of the single represented total number of genes are indicated. N=2 for GM, n=2 for MB and n=2 for MT.

B. Pathway analysis of the shared expressed genes among MB and GM. The pathways depicted have an FDR < = 0.05. N=2 for GM, n=2 for MB and n=2 for MT. The color code is suggested in Figure 2A.
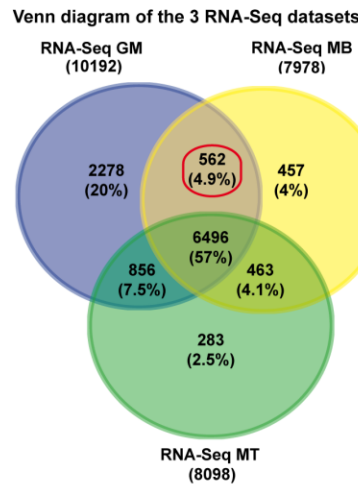
C. Heatmap of the mean-centered normalized expression of the shared expressed genes between MB and GM. The different clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

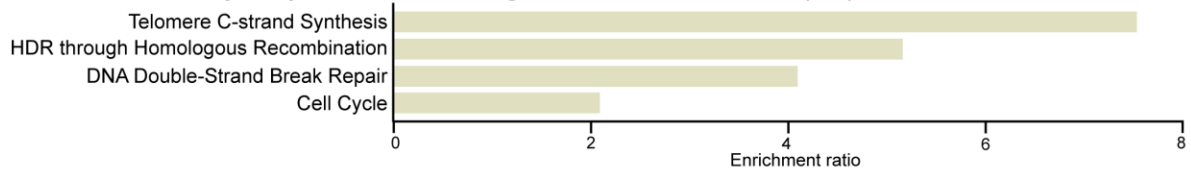D. Pathway analysis of the GM and the MB clusters. The main pathways are boxed.

E. Heatmaps of the mean-centered normalized expression of the genes of selected pathways from GM and MB clusters. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.
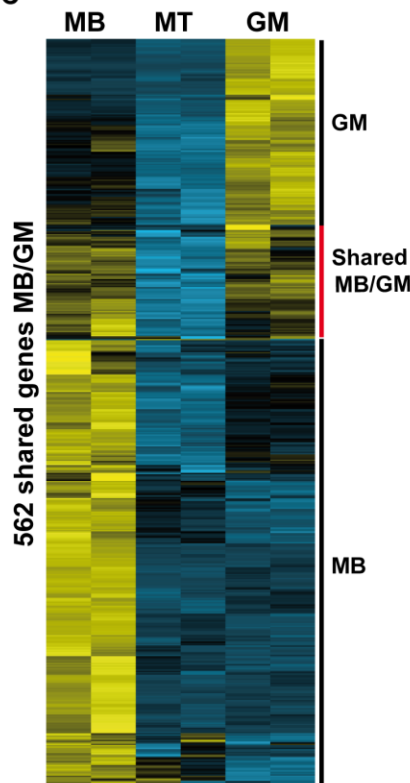
Figure 3.

A

Venn diagram of the 3 RNA-Seq datasets

B

Pathway analysis of the common genes between MB and MT (463)

C
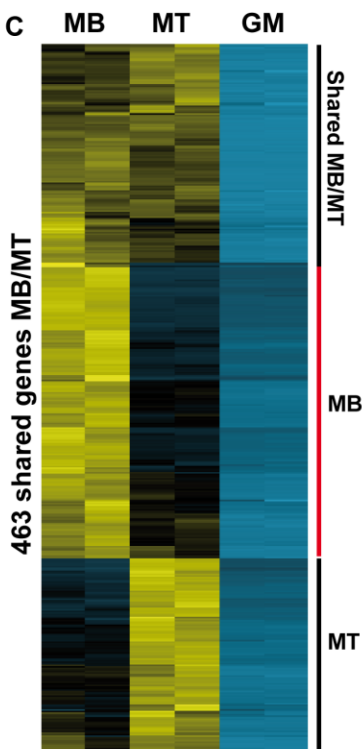
D

MB cluster: mitosis/cell cycle + axon guidance + cell shape/cytoskeleton
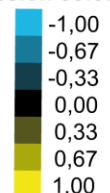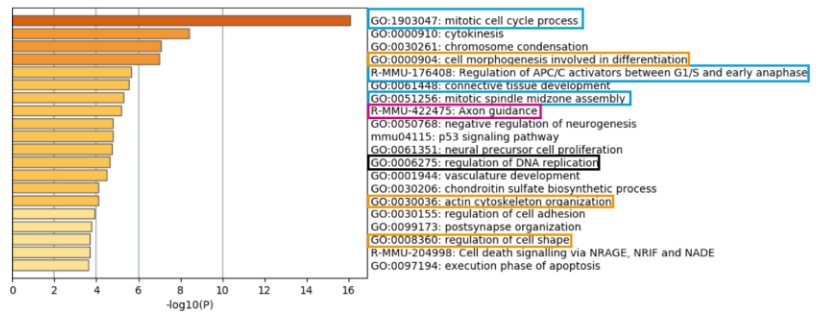

MT cluster
Axon guidance: -log10(P)= 3.52

Mean centered
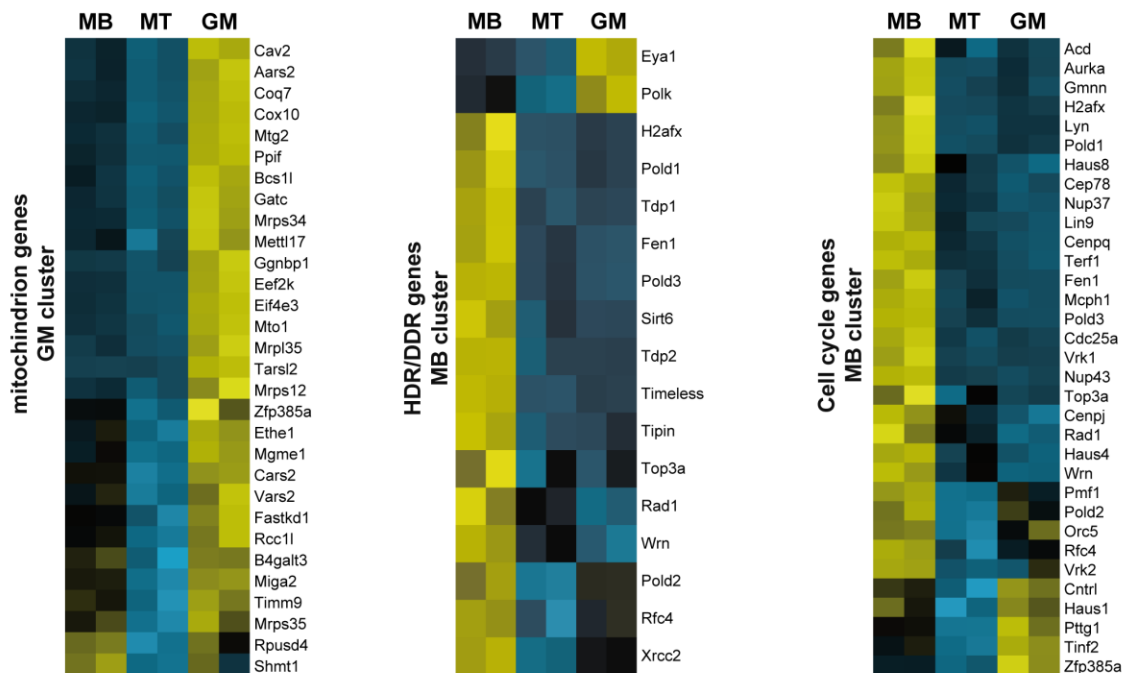normalized
expression colorbar
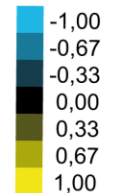
E

Figure 3. **Shared expressed genes between MB and MT.**

A. Venn diagram of expressed genes from RNA-Seq of gastrocnemius muscles (GM) of 10 week-old-wild-type mice (purple), proliferating C2C12 myoblasts (MB) (yellow) and 3-day differentiated C2C12 myotubes (MT) (green). The total number of genes for each dataset as well as the % of the single represented total number of genes are indicated. N=2 for GM, n=2 for MB and n=2 for MT.

B. Pathway analysis of the shared expressed genes among MB and MT. The pathways depicted have an FDR <= 0.05. N=2 for GM, n=2 for MB and n=2 for MT. The color code is suggested in Figure 3A.

C. Heatmap of the mean-centered normalized expression of the shared expressed genes between MB and MT. The different clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.
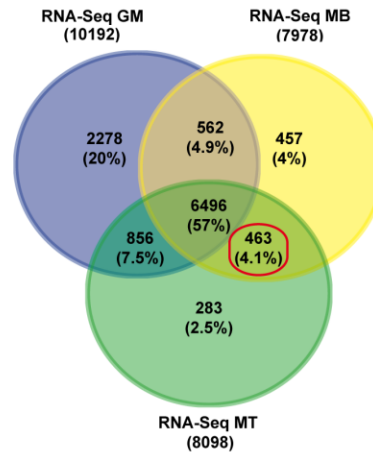
D. Pathway analysis of the MB and the MT clusters. The main pathways are boxed.

E. Heatmaps of the mean-centered normalized expression of the genes of selected pathways from MB and MT clusters. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

Figure 4.

A

Venn diagram of the 3 RNA-Seq datasets

B
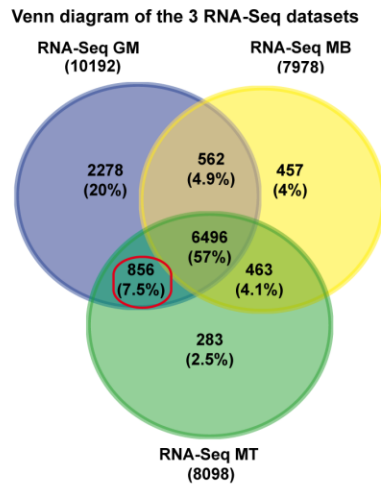
Pathway analysis of the common genes between MT and GM (856)
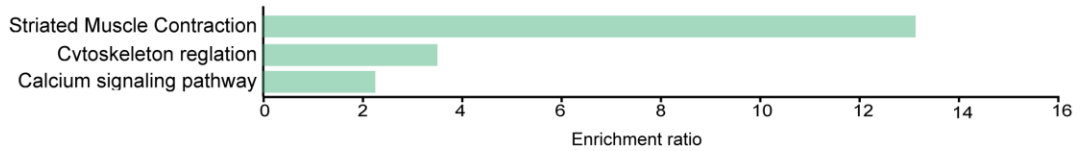
C
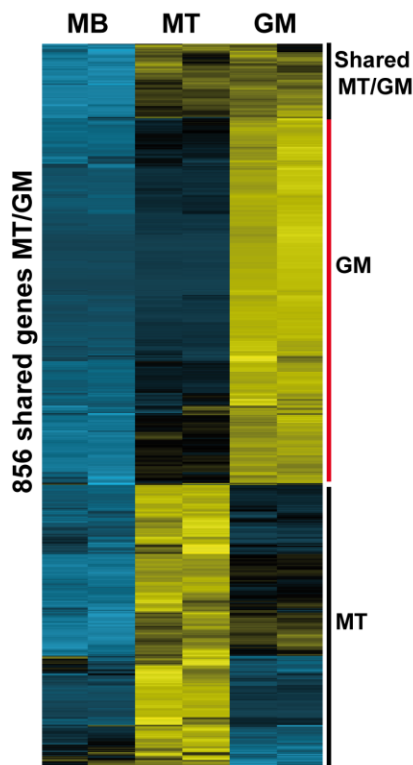
D

GM cluster: muscle contraction + ion homeostasis + mitochondrion
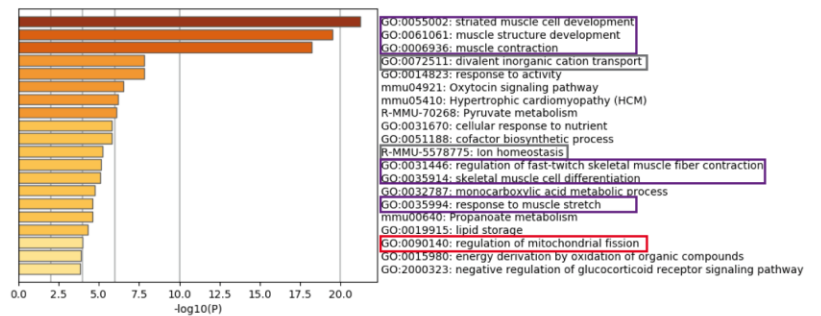
MT cluster: muscle contraction + ion homeostasis
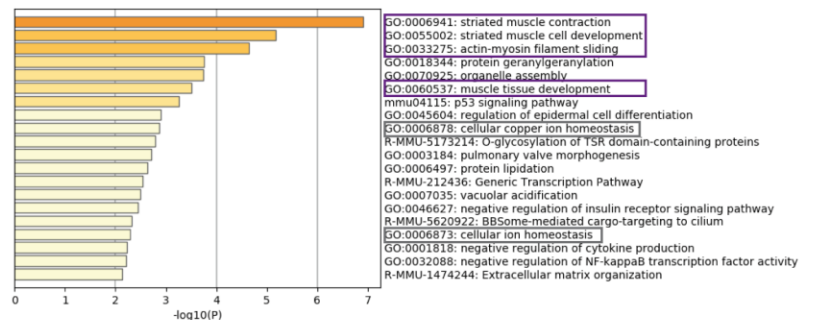
Mean centered normalized expression colorbar
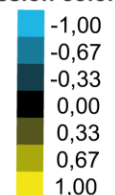
E

Figure 4. **Shared expressed genes between MT and GM.**

A. Venn diagram of expressed genes from RNA-Seq of gastrocnemius muscles (GM) of 10 week-old-wild-type mice (purple), proliferating C2C12 myoblasts (MB) (yellow) and 3-day differentiated C2C12 myotubes (MT) (green). The total number of genes for each dataset as well as the % of the single represented total number of genes are indicated. N=2 for GM, n=2 for MB and n=2 for MT.

B. Pathway analysis of the shared expressed genes among MT and GM. The pathways depicted have an FDR <= 0.05. N=2 for GM, n=2 for MB and n=2 for MT. The color code is suggested in Figure 4A.

C. Heatmap of the mean-centered normalized expression of the shared expressed genes between MT and GM. The different clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

D. Pathway analysis of the MT and the GM clusters. The main pathways are boxed.

E. Heatmaps of the mean-centered normalized expression of the genes of selected pathways from MT and GM clusters. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.
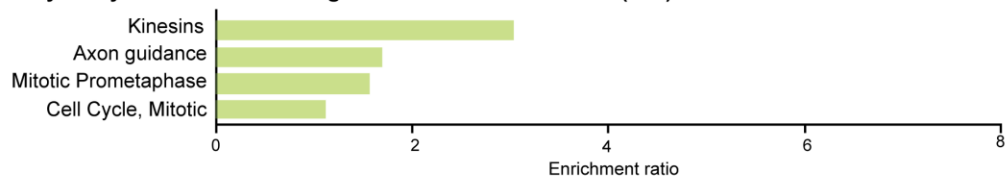
Figure 5.

A

Venn diagram of the 3 RNA-Seq datasets

RNA-Seq GM (10192) RNA-Seq MB (7978)

2278 (20%) 562 (4.9%) 457 (4%)

6496 (57%)

856 (7.5%) 463 (4.1%)

283 (2.5%)

RNA-Seq MT (8098)

B

Pathway analysis of the MB-specific genes (457)

Telomere C-strand Synthesis
HDR through Homologous Recombination
Kinesins
Mitotic Prometaphase
Cell Cycle

Enrichment ratio

C

MB MT GM

457 MB-specific genes

1st

2nd
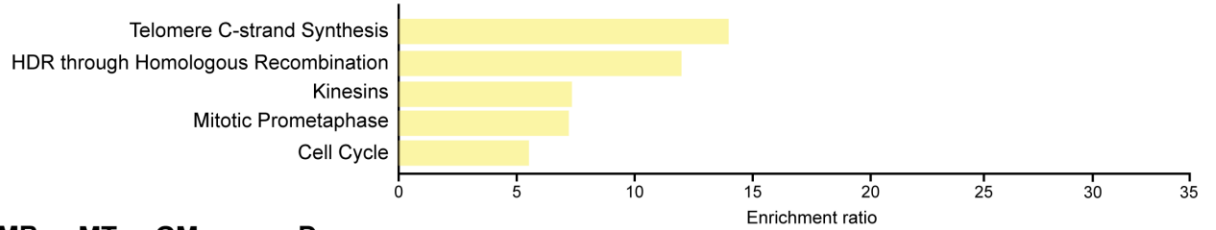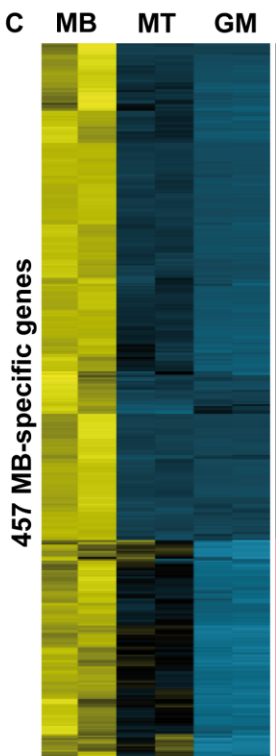
D

1st cluster: mitosis/cell cycle + DNA replication/HDR + cytoskeleton

GO:1903047: mitotic cell cycle process
GO:0006261: DNA-dependent DNA replication
GO:0010564: regulation of cell cycle process
GO:0000226: microtubule cytoskeleton organization
GO:0006974: cellular response to DNA damage stimulus
GO:0045787: positive regulation of cell cycle
GO:0140013: meiotic nuclear division
mmu04110: Cell cycle
GO:0000086: G2/M transition of mitotic cell cycle
GO:0051052: regulation of DNA metabolic process
R-MMU-5685942: HDR through Homologous Recombination (HRR)
GO:0007098: centrosome cycle
GO:0044786: cell cycle DNA replication
R-MMU-73886: Chromosome Maintenance
GO:0045005: DNA-dependent DNA replication maintenance of fidelity
GO:0000281: mitotic cytokinesis
GO:0008608: attachment of spindle microtubules to kinetochore
GO:0007143: female meiotic nuclear division
R-MMU-156711: Polo-like kinase mediated events
R-MMU-983189: Kinesins

-log10(P)

Mean centered normalized expression colorbar

-1,00
-0,67
-0,33
0,00
0,33
0,67
1,00

258

E

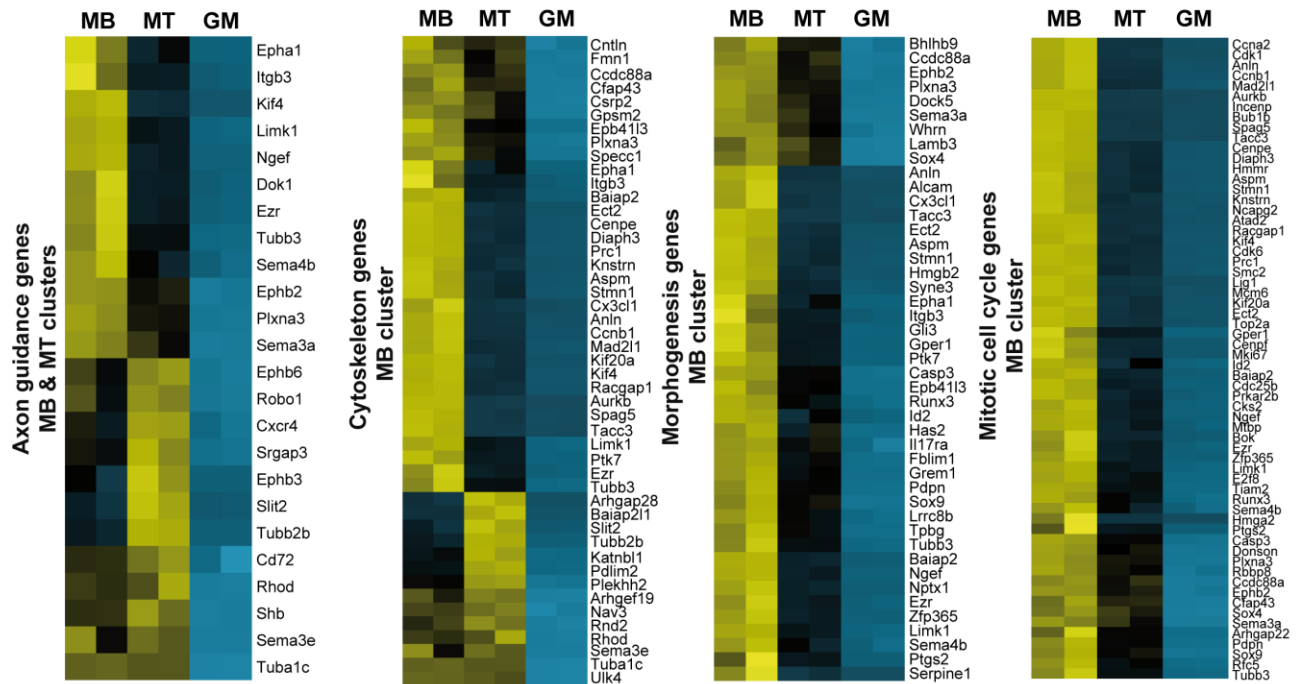**Figure 5. MB-specific expressed genes.**

A. Venn diagram of expressed genes from RNA-Seq of gastrocnemius muscles (GM) of 10 week-old-wild-type mice (purple), proliferating C2C12 myoblasts (MB) (yellow) and 3-day differentiated C2C12 myotubes (MT) (green). The total number of genes for each dataset as well as the % of the single repre-sented total number of genes are indicated. N=2 for GM, n=2 for MB and n=2 for MT.

B. Pathway analysis of the MB-specific genes. The pathways depicted have an FDR <= 0.05. N=2 for GM, n=2 for MB and n=2 for MT. The color code is suggested in Figure 5A.

C. Heatmap of the mean-centered normalized expression of the MB-specific genes. The different clus-ters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

D. Pathway analysis of the 1st cluster of the heatmap of the MB-specific genes. The main pathways are boxed.

E. Heatmaps of the mean-centered normalized expression of the genes of selected pathways from the 1st cluster of the heatmap of the MB-specific genes. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

Figure 6.

A

B

C

2nd cluster
Glycosylation: -log10(P)= 5.52

D

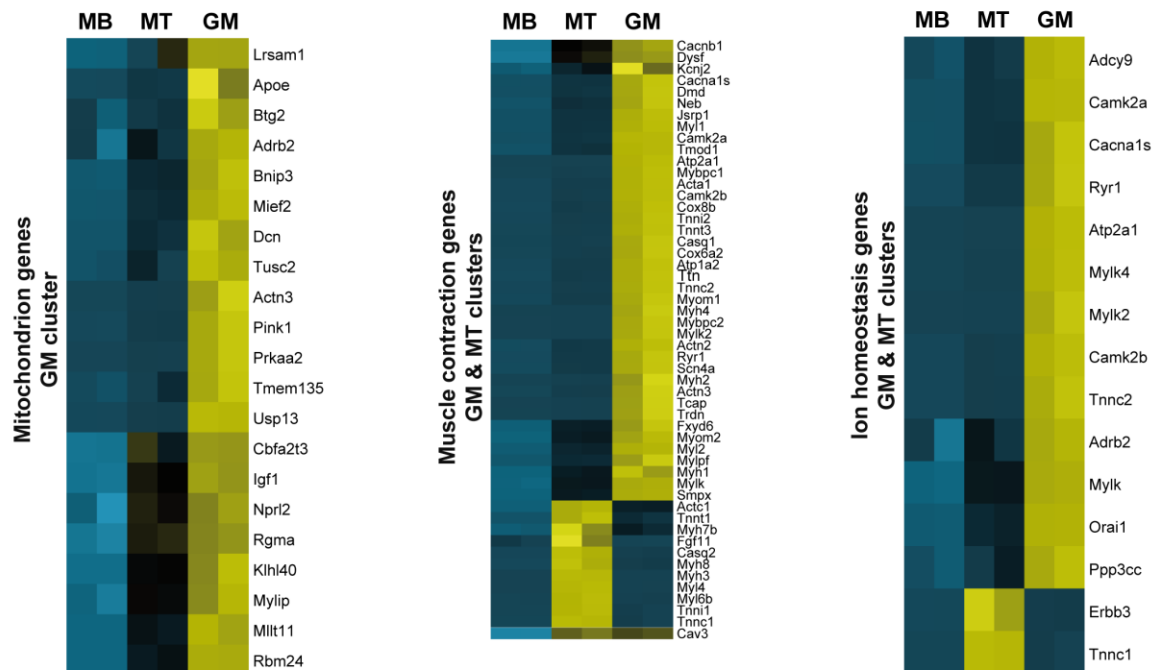**Figure 6. MT-specific expressed genes.**

A. Venn diagram of expressed genes from RNA-Seq of gastrocnemius muscles (GM) of 10 week-old-wild-type mice (purple), proliferating C2C12 myoblasts (MB) (yellow) and 3-day differentiated C2C12 myotubes (MT) (green). The total number of genes for each dataset as well as the % of the single represented total number of genes are indicated. N=2 for GM, n=2 for MB and n=2 for MT.
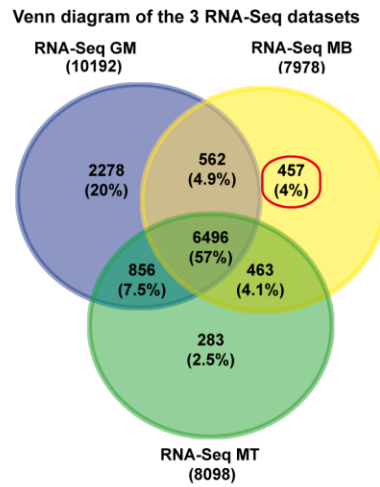
B. Heatmap of the mean-centered normalized expression of the MT-specific genes. The different clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

C. Pathway analysis of the 2nd cluster of the heatmap of the MT-specific genes.

D. Heatmap of the mean-centered normalized expression of the genes involved in glycosylation. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.
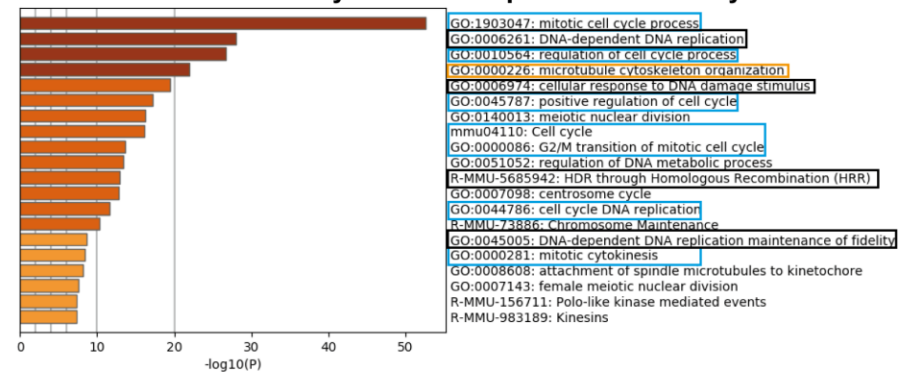
261

Figure 7.

A

Venn diagram of the 3 RNA-Seq datasets

B

Pathway analysis of the GM-specific genes (2278)

C

MB   MT   GM

2278 GM-specific genes

1st

2nd

D

1st cluster: complement + peroxisome

2nd cluster: break repair + inflammation

GO:0006399: tRNA metabolic process
GO:0006302: double-strand break repair
GO:0019886: antigen processing and presentation of exogenous peptide antigen via MHC class II
GO:1903799: negative regulation of production of miRNAs involved in gene silencing by miRNA
GO:0031848: protection from non-homologous end joining at telomere
GO:0044782: cilium organization
GO:0060122: inner ear receptor cell stereocilium organization
GO:0007219: Notch signaling pathway
GO:0042035: regulation of cytokine biosynthetic process
GO:0003382: epithelial cell morphogenesis
R-MMU-5693532: DNA Double-Strand Break Repair
mmu04142: Lysosome
GO:0032642: regulation of chemokine production

Mean centered normalized expression colorbar

	-1,00
	-0,67
	-0,33
	0,00
	0,33
	0,67
	1,00

E

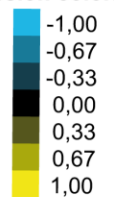**Figure 7.** **GM-specific expressed genes.**

A. Venn diagram of expressed genes from RNA-Seq of gastrocnemius muscles (GM) of 10 week-old-wild-type mice (purple), proliferating C2C12 myoblasts (MB) (yellow) and 3-day differentiated C2C12 myotubes (MT) (green). The total number of genes for each dataset as well as the % of the single repre-sented total number of genes are indicated. N=2 for GM, n=2 for MB and n=2 for MT.

B. Pathway analysis of the GM-specific genes. The pathways depicted have an FDR < = 0.05. N=2 for GM, n=2 for MB and n=2 for MT. The color code is suggested in Figure 7A.

C. Heatmap of the mean-centered normalized expression of the GM-specific genes. The different clus-ters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

D. Pathway analysis of the 1st and 2nd clusters of the heatmap of the GM-specific genes. The main pathways are boxed.

E. Heatmaps of the mean-centered normalized expression of the genes of selected pathways from the 1st and 2nd clusters of the heatmap of the GM-specific genes. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.
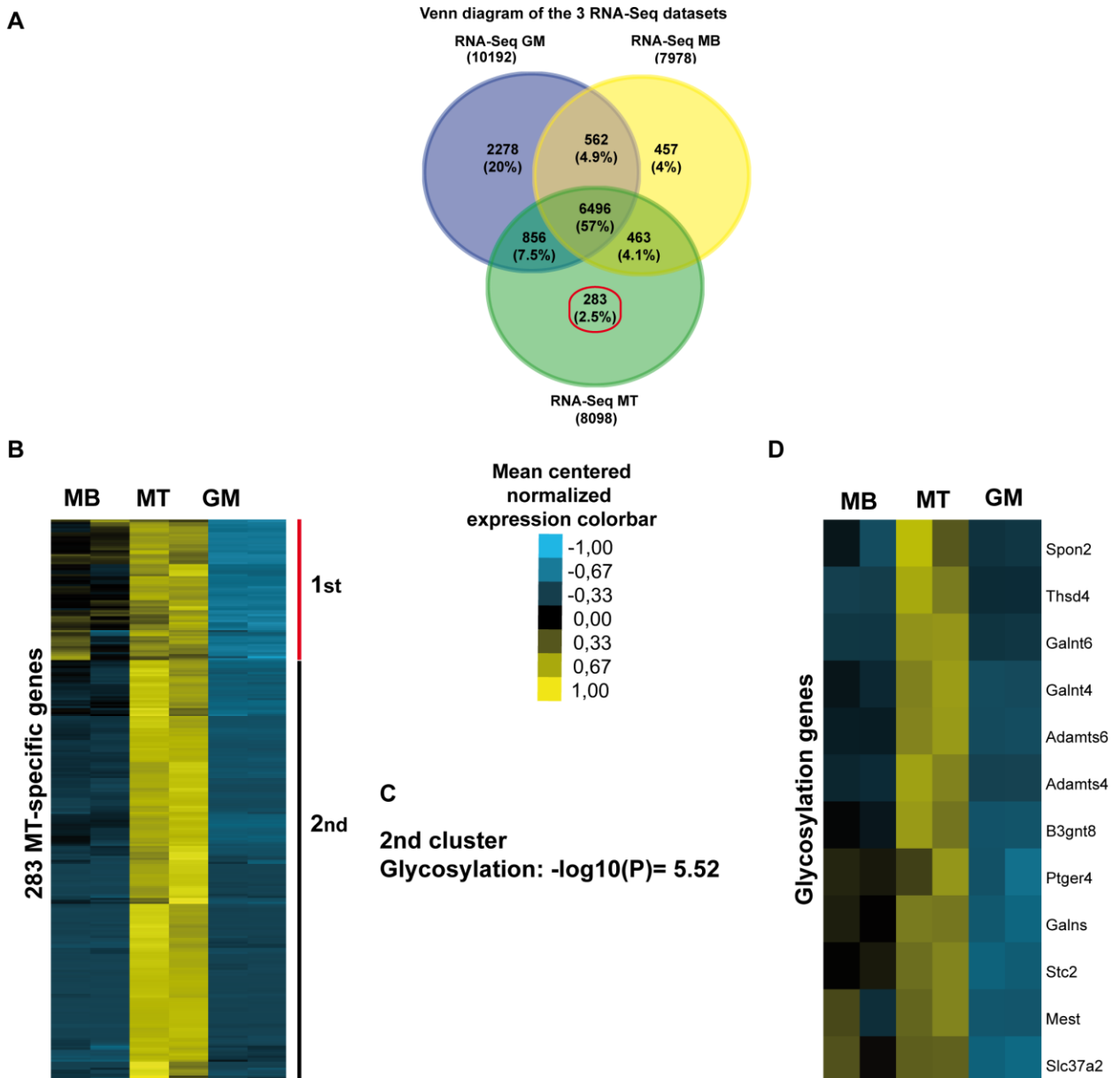
Figure 8.

A
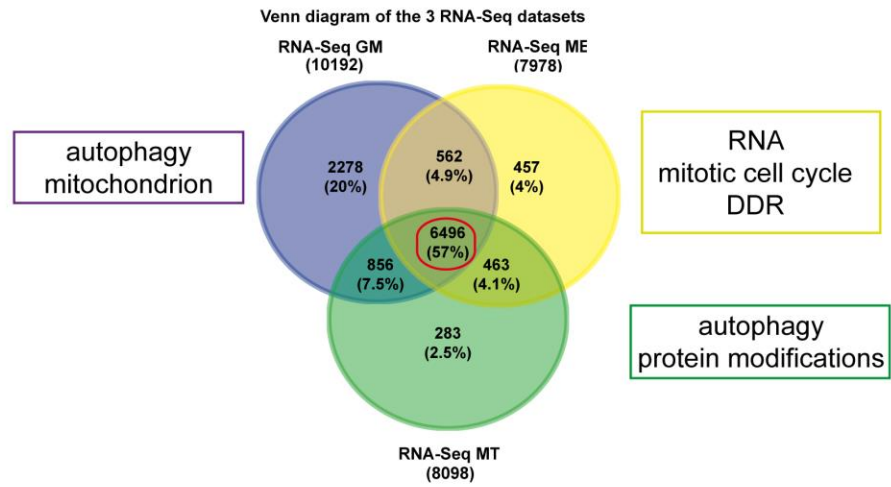
B
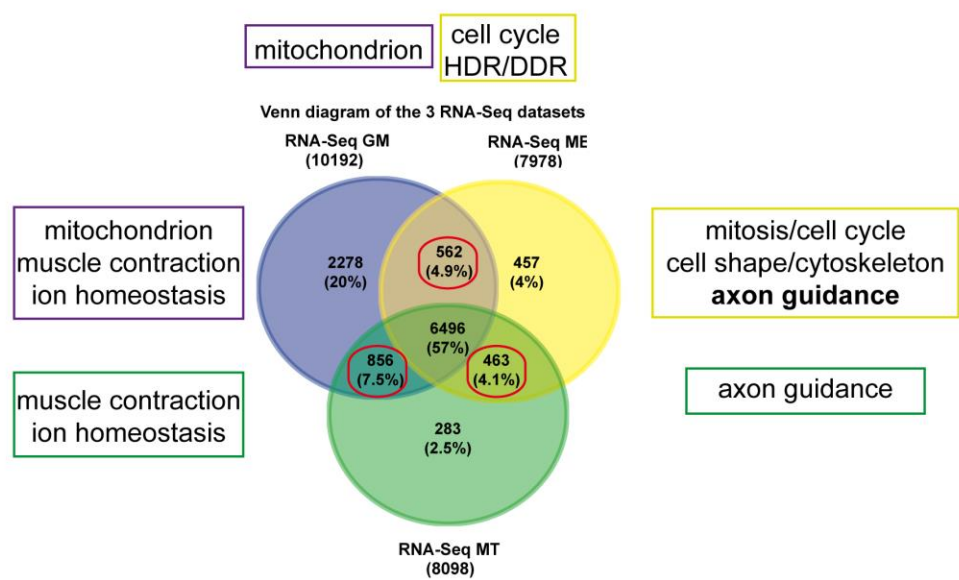
C
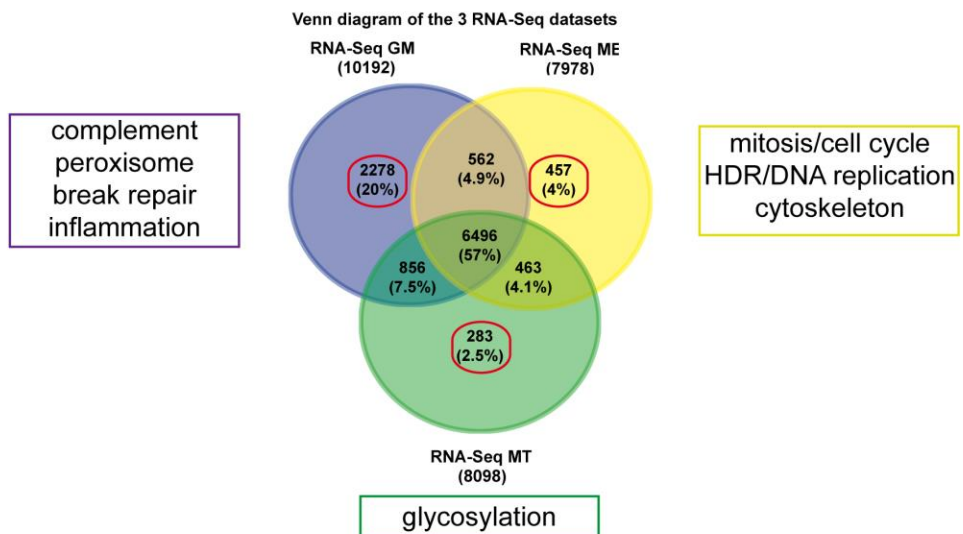
Figure 8. Shared and specific enriched pathways.

A. Summary scheme of the pathways enriched among the common expressed genes in MB, MT and GM. Boxes in purple for GM, in yellow for MB and in green for MT.

B. Summary scheme of the pathways enriched among the shared expressed genes between MB and GM, MB and MT and MT and GM. Boxes in purple for GM, in yellow for MB and in green for MT.

C. Summary scheme of the pathways enriched in the MB-, MT- and GM-specific expressed genes. Boxes in purple for GM, in yellow for MB and in green for MT.

Figure 9.

A

Venn diagram of the 3 RNA-Seq datasets

B

Venn diagram of the 3 RNA-Seq datasets

C

Venn diagram of the 3 RNA-Seq datasets

Figure 9. **Shared and specific enriched pathways.**

A. Summary scheme of the enriched TFs families among the common expressed genes in MB, MT and GM. With **bold** are indicated the TFs families that are enriched throughout the differentiation. The p-value to identify the enriched pathways was less than 0.05. The color code of the box is suggested from the venn diagram.

B. Summary scheme of the enriched TFs families among the shared expressed genes between MB and GM, MB and MT and MT and GM. With **bold** are indicated the TFs families that are enriched throughout the differentiation. With red are indicated the TFs families that are common between the shared gene lists. The p-value to identify the enriched pathways was less than 0.05. The color code of the boxes is suggested from the venn diagram.

C. Summary scheme of the enriched TFs families in the MB-, MT- and GM-specific expressed genes. With **bold** are indicated the TFs families that are enriched throughout the differentiation. With red are indicated the TFs families that are common between the specific gene lists. The p-value to identify the enriched pathways was less than 0.05. The color code of the boxes is suggested from the venn diagram.

Figure 10.

A

B

Homeobox

MB MT GM

Hoxa11
Hoxa7
Hoxa1
Hoxb1
Hoxa9
Hoxa6
Hoxa5
Hoxc11
Hoxc5
Hoxc8
Hoxc6
Hoxc9
Hoxd3
Hoxc12
Hoxc13
Hoxa2
Hoxa3
Hoxc4
Hoxa10
Hoxd11
Hoxb4
Hoxb2
Hoxb8
Hoxb3
Hoxb9
Hoxa4
Hoxc10
Hoxa13
Hoxd8
Hoxd9
Hoxb5
Hoxd4
Hoxb6
Hoxb7
Hoxd10

MB

GM

**Mean centered
normalized
expression colorbar**

-1,00
-0,67
-0,33
0,00
0,33
0,67
1,00

SIX homeobox

MB MT GM

Six1 MT/GM
Six2 GM
Six5
Six3 MB
Six4

AP-1 subunits

MB MT GM

Fos GM
Jun
Fosb MB
Fosl1
Junb
Fosl2 MT

C Notch signaling

MB MT GM

Notch1 — MB/MT
Notch2 — MB/MT
Hes1 — MT
Notch3 — MT
Heyl — MT
Notch4 — GM
Hey1 — GM

Signal Transducer and activator of transcription

MB MT GM

Stat2 — MT
Stat3 — MT
Stat5a — MT
Stat4 — GM
Stat1 — GM
Stat5b — GM
Stat6 — GM

Myocyte enhancer factors 2

MB MT GM

Mef2a — MT
Mef2c — GM
Mef2d — GM

D Kruppel like factor

MB MT GM

Klf15 — GM
Klf12 — GM
Klf2 — GM
Klf1 — MB
Klf16 — MB
Klf13 — MB
Klf7 — MB
Klf4 — MT
Klf3 — MT
Klf5 — MT

E2f transcription factors

MB MT GM

E2f1 — MB
E2f5 — MB
E2f3 — MB
E2f4 — MB
E2f2 — MT
E2f6 — GM

Mean centered normalized expression colorbar

-1,00
-0,67
-0,33
0,00
0,33
0,67
1,00

H

Myosin heavy chains

	MB	MT	GM	
Myh15				MB
Myh10				
Myh9				
Myh13				GM
Myh14				
Myh11				
Myh4				
Myh2				
Myh1				
Myh7				
Myh6				MT
Myh3				
Myh8				
Myh7b				

Myosin light chains

	MB	MT	GM	
Myl12b				MT
Myl12a				
Myl4				
Myl6b				
Mylk3				
Myl6				
Myl1				GM
Mylpf				
Myl2				
Mylk				
Myl10				
Myl3				
Mylk2				
Mylk4				
Mylip				
Myl7				MB
Myl9				

Myosins

	MB	MT	GM	
Myo16				
Myo15b				
Myo15				
Myo3b				
Myo1a				MB
Myo19				
Myo1h				
Myo5a				
Myo7b				
Myo1d				
Myo9b				
Myo10				MT
Myo9a				
Myo1c				MB/MT
Myo1e				
Myo1g				GM
Myo3a				
Myocd				
Myo1b				
Myo1f				
Myo5b				
Myo6				
Myo18b				
Myo7a				
Myo5c				
Myo18a				

Mean centered normalized expression colorbar

-1,00
-0,67
-0,33
0,00
0,33
0,67
1,00

Figure 10. Heatmaps of mean-centered normalized expression of TFs families.

A. Heatmaps of the mean-centered normalized expression of the **Zinc finger proteins factors**, the **NF-kappa B subunits** and the TFs and families involved in **mitochondrial function** in MB, MT and GM. The specific and shared clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

B. Heatmaps of the mean-centered normalized expression of the **AP-1 subunits**, the **Homeobox factors**, the **SIX homeobox factors**, the **SRY-box factors**, the **paired box factors** and the **ETS-related factors** in MB, MT and GM. The specific and shared clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

C. Heatmaps of the mean-centered normalized expression of the **Notch signaling factors**, the **Signal transducer and activator of transcription factors** and the **Myocyte enhancer 2 factors** in MB, MT and GM. The specific and shared clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

D. Heatmaps of the mean-centered normalized expression of factors involved in **cell cycle regulation** and especially the **Kruppel-like family** and the **E2F factors** in MB, MT and GM. The specific and shared clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

E. Heatmaps of the mean-centered normalized expression of factors involved in **DNA replication**, of the **Transcription Factors family** and the **Forkhead box family** in MB, MT and GM. The specific and shared clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

F. Heatmaps of the mean-centered normalized expression of the **TEA-domain factors** and the **myogenic regulatory factors** in MB, MT and GM. The specific and shared clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

G. Heatmaps of the mean-centered normalized expression of the **POU domain factors** and the **Nuclear receptors** in MB, MT and GM. The specific and shared clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

H. Heatmaps of the mean-centered normalized expression of the **myosin heavy chain factors**, the **myosin light chain factors** and the other **myosins** in MB, MT and GM. The specific and shared clusters are indicated. The levels of expression are indicated by the color bar. N=2 for GM, n=2 for MB and n=2 for MT.

Figure 11.

A

H3K27ac

MB   MT   GM

32518 MB H3K27ac peaks

MB/MT cluster
6638 peaks
4252 genes

MB cluster
13802 peaks
8548 genes

-5 kb +5 kb
peak center

Examples

under expressed in GM
overexpressed in GM

Myod1	Pax7	Mef2a
Mef2a	Tcf4	Sta3
Myl9	Smad7	Fos
Stat3	Smad6	Mylip
Etv4	Myo1e	Smad6
Sox8	Myh9	Smad3
Sox11	Elk3	Foxo3
Notch3	Notch1	Myl10
Myl12b	Musk	Mybpc1
Nfkb1	Pparg	Myf6
Smad1	Nfkbiz	Musk
Mylk3	Foxo6	Foxo6

Examples

under expressed in GM
overexpressed in GM

Pax7	
E2f8	
Smad1	
Foxl1	Fos
Foxq1	Sox13
Elk3	Foxo3
Hoxc13	Myl10
Nr2f1	Nrbf2
Nr4a2	Myl3
Nfkb1	Mylk

B

H3K27ac

MB   MT   GM

37021 MT H3K27ac peaks

MB/MT cluster
4957 peaks
3536 genes

MT cluster
4445 peaks
2916 genes

MT cluster
18869 peaks
9067 genes

-5 kb +5 kb
peak center

under expressed in MT
overexpressed in MT

Pax7	Pax7	Myog
Mef2c	Mef2c	Mef2a
Esr1	Foxo6	Myo10
Elk4	Esr1	Myh3
Myl2	Myh7	Myh2
Myh9	Foxo3	Myh1
Nfkbib	Myl2	Nfel1
Smad3	Myh9	Myh7b
Myh2	Musk	Stat3
Myh1	Nfkbiz	Notch3
Nfe2l1	Smad3	Sox11
Nfkbia	Pou2af1	Myoz2
Myoz2	Fos	Mylk3
Mylip		Mylip

C

H3K27ac

MB | MT | GM

32480 GM H3K27ac peaks

GM cluster
4829 peaks
2544 genes

GM cluster
13862 peaks
8478 genes

-5 kb | +5 kb
peak center

Examples

under expressed in GM
overexpressed in GM

Myog   Myoz1   Myog
Mylk2   Mustn1   Nr2f2
Nr2f2   Myo5c   Ppard
Myh1   Mlip   Musk
Myh4   Foxk1   Fosl2
Myo18a   Rxra   Mustn1
Stat5a   Ppara
Ppard   Nr4a1
Musk   Foxo3
Myoz3   Hoxc8
Fos   Nrbf2
Mylk4   Sox13
Stat1   Elk4
Myl1   Myo18b
Mef2c   Mybpc1

D

Myoblasts (MB) → Myotubes (MT) → Gastrocnemius muscle (GM)

Myoblasts (MB)	Myotubes (MT)	Gastrocnemius muscle (GM)
bHLH	bHLH	bHLH
bZIP	bZIP	bZIP
C2H2 zinc finger	C2H2 zinc finger	C2H2 zinc finger
Forkhead box	Forkhead box	Forkhead box
Homeo domain	Homeo domain	Homeo domain
TEA domain	TEA domain	
Runt domain	Runt domain	
ETS-related	ETS -related	
Sox- related		
	Steroid hormone receptors	Steroid hormone receptors
	ARID domain	
	MADS box	
	SMAD/NF1 DNA binding domain	
	Paired box	
	POU domain	
		FTZ-F1-related receptors
		NGFI-B-related receptors
		RXR-related ceptors
		Thyroid hormone related receptor
		Rel homology region

Figure 11. **Comparative epigenetic profiling between murire skm and C2C12 cells.**

A. Tag density map of H3K27ac ChIP-Seq reads of MB, MT and GM, using as a reference the H3K27ac MB peaks, in a window of +/- 5 kb around the center of the peaks. The shared and the unique clusters are indicated as well as the number of peaks and genes for these clusters. For the MB/MT cluster a selection of genes is indicated, in green the under expressed in GM and in red the over expressed in GM.

B. Tag density map of H3K27ac ChIP-Seq reads of MB, MT and GM, using as a reference the H3K27ac MT peaks, in a window of +/- 5 kb around the center of the peaks. The shared and the unique clusters are indicated as well as the number of peaks and genes for these clusters. For the MB/MT cluster a selection of genes is indicated, in green the under expressed in GM and in red the over expressed in GM. For the MT-specific cluster a selection of genes is indicated, in green the under expressed in MT and in red the over expressed in MT.

C. Tag density map of H3K27ac ChIP-Seq reads of MB, MT and GM, using as a reference the H3K27ac GM peaks, in a window of +/- 5 kb around the center of the peaks. The shared and the unique clusters are indicated as well as the number of peaks and genes for these clusters. For the GM-specific cluster a selection of genes is indicated, in green under expressed in GM and in red the over expressed in GM.

D. Summary scheme of the *de novo* motifs of TFs families enriched in the shared and specific clusters. With **bold** are indicated the TFs family motifs that are enriched throughout the differentiation. E-value used less or equal to 0.05.

278

Figure 12.

A

MB-specific gene: Foxc2

B

MT-specific gene: Myog

C

GM-specific gene: Mef2d

chr3 (mm10)

D

common gene MB, MT, GM: Notch3

chr17 (mm10)
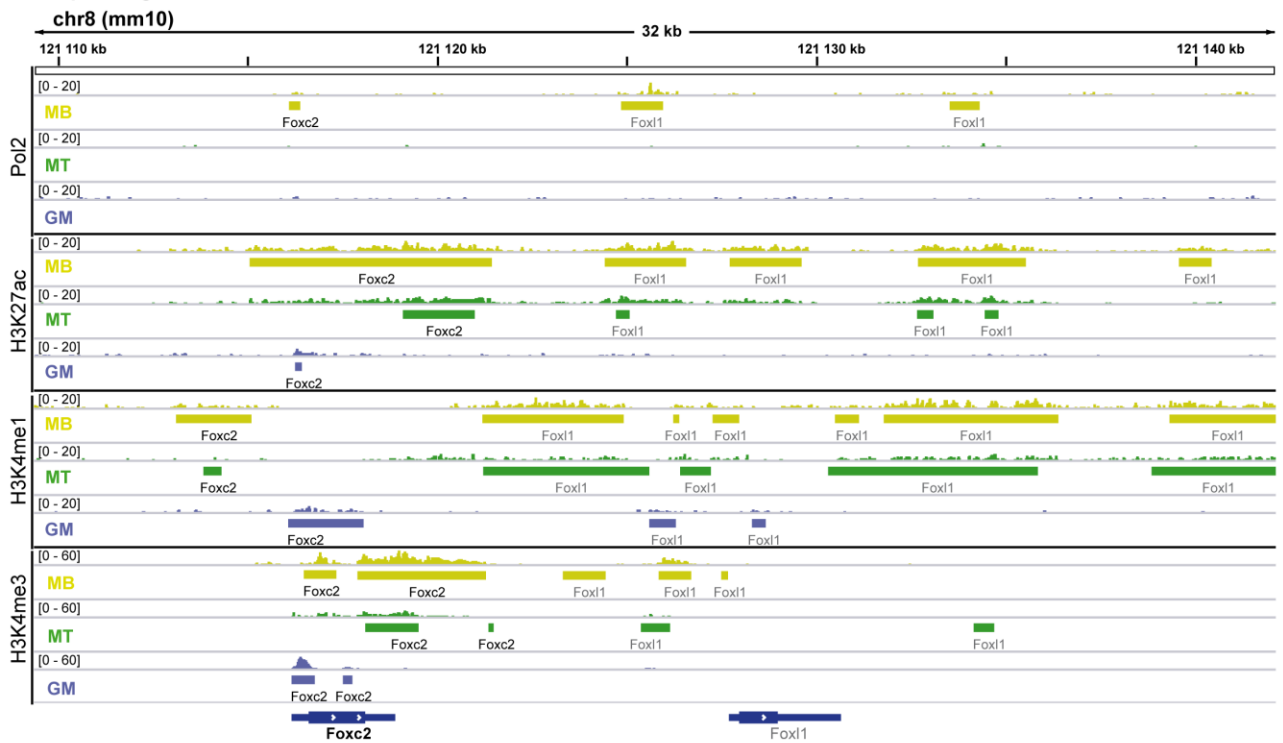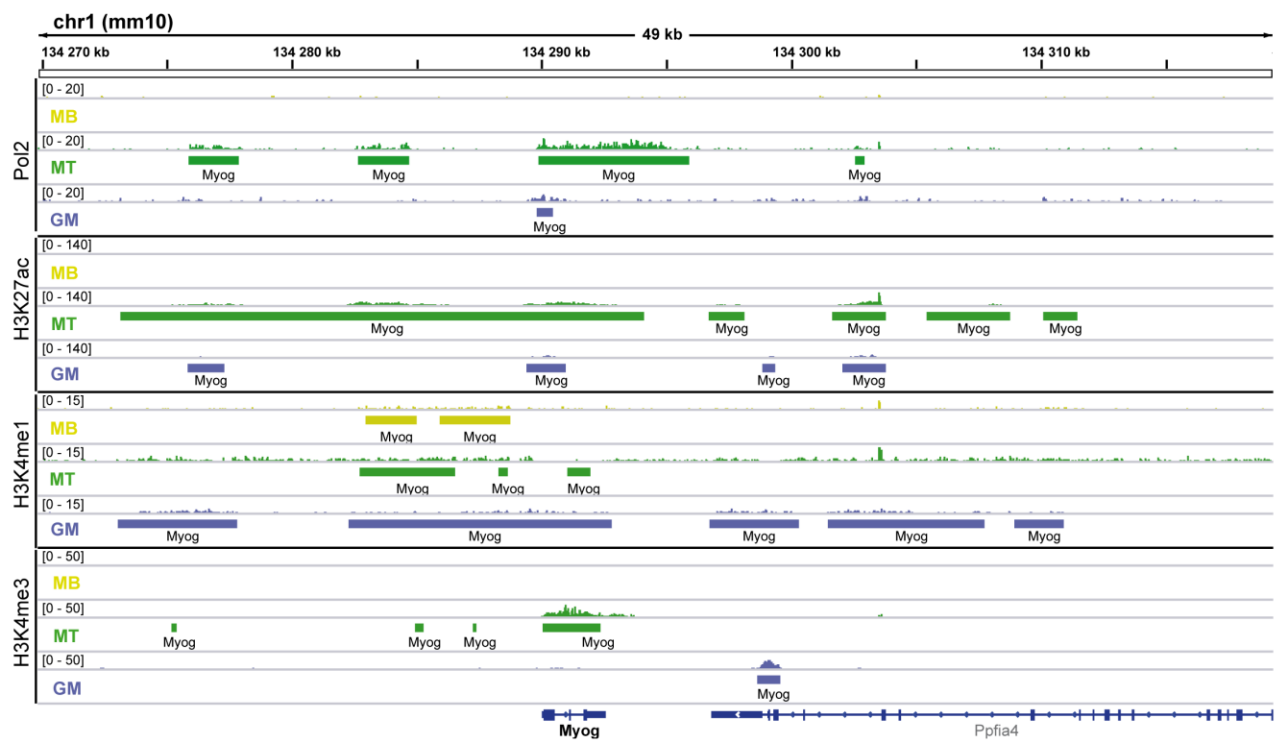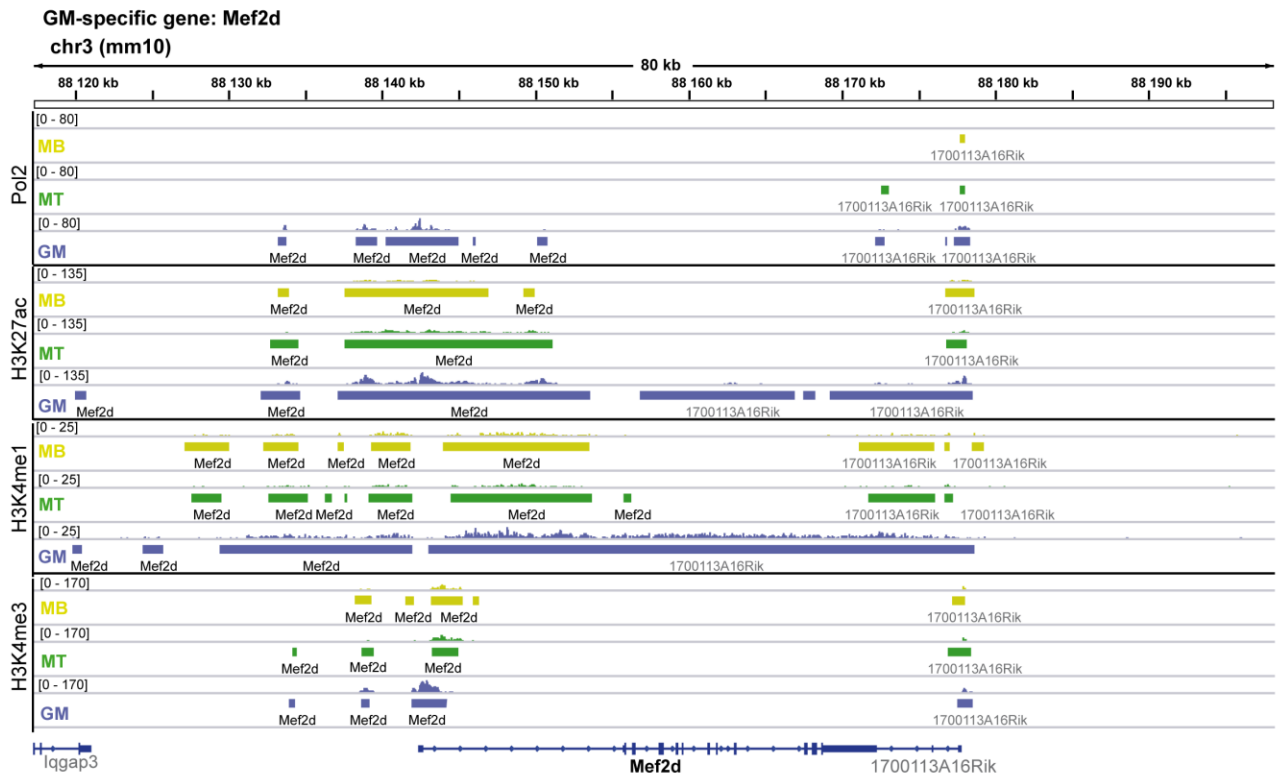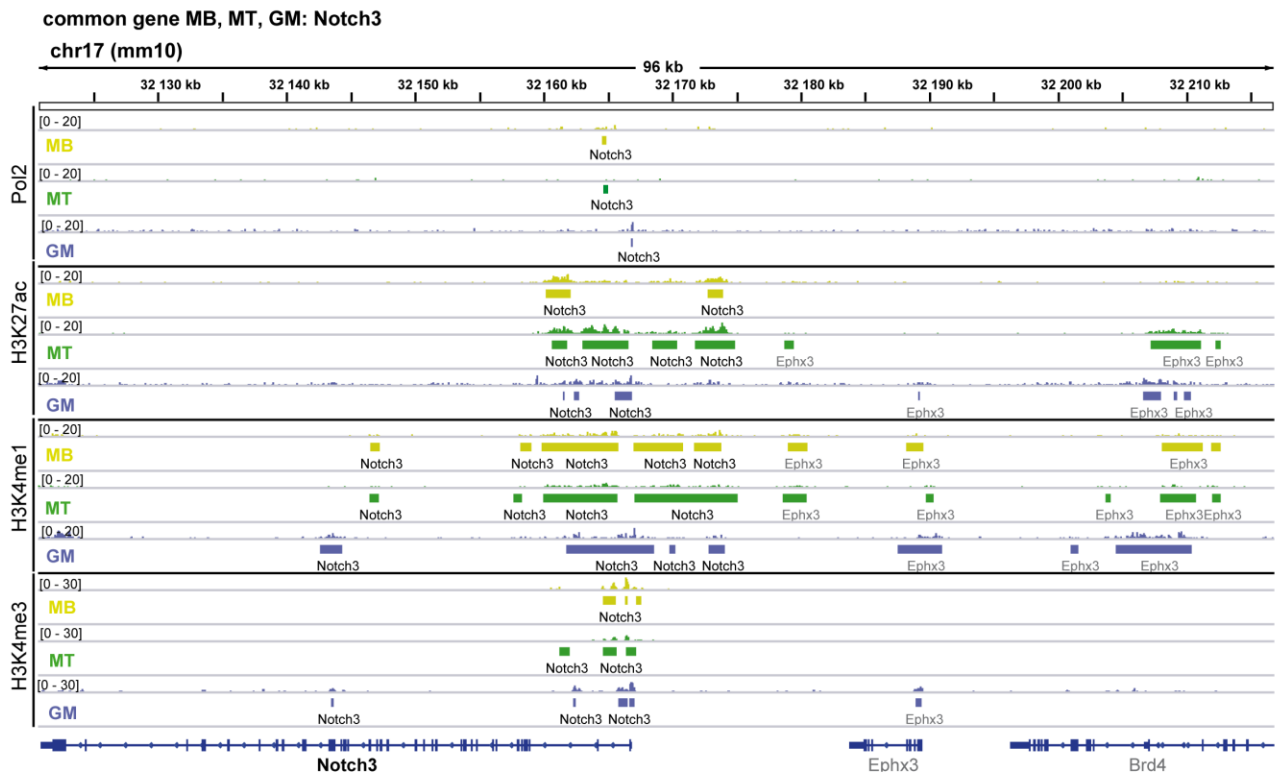
Figure 12. **Genome-wide chromatin state maps of murine skm and C2C12 cells.**

A. Overview of the genomic region surrounding the **Foxc2** gene in the IGV browser featuring panels of chromatin marks like Pol2, H3K27ac, H3K4me1, H3K4me3 showing ChIP-Seq data in MB, MT, and GM. The scale for each track is represented and possibly adjusted to account for lower total ChIP-Seq signals. The color of the tracks represents the different stages, MB (yellow), MT (green) and GM (purple).

B. Overview of the genomic region surrounding the **Myog** gene in the IGV browser featuring panels of chromatin marks like Pol2, H3K27ac, H3K4me1, H3K4me3 showing ChIP-Seq data in MB, MT and GM. The scale for each track is represented and possibly adjusted to account for lower total ChIP-Seq signals. The color of the tracks represents the different stages, MB (yellow), MT (green) and GM (purple).

C. Overview of the genomic region surrounding the **Mef2d** gene in the IGV browser featuring panels of chromatin marks like Pol2, H3K27ac, H3K4me1, H3K4me3 showing ChIP-Seq data in MB, MT and GM. The scale for each track is represented and possibly adjusted to account for lower total ChIP-Seq signals. The color of the tracks represents the different stages, MB (yellow), MT (green) and GM (purple).

D. Overview of the genomic region surrounding the **Notch3** gene in the IGV browser featuring panels of chromatin marks like Pol2, H3K27ac, H3K4me1, H3K4me3 showing ChIP-Seq data in MB, MT and GM. The scale for each track is represented and possibly adjusted to account for lower total ChIP-Seq signals. The color of the tracks represents the different stages, MB (yellow), MT (green) and GM (purple).

Supplementary Figure 1.

A **GM vs MB**

6587 — overexpressed in MB
6631 — overexpressed in GM

B **GM vs MT**

5957 — overexpressed in MT
6453 — overexpressed in GM

Overexpressed in MB	LFC	Function
Fosl1	-9,81	cell prolif/diff/transform
Ascl4	-9,54	cell fate/diff/develop
Hoxc13	-9,48	morphogenesis
Ttk	-9,30	mitosis/cell cycle
Bub1	-9,01	mitosis/cell cycle
Kif2c	-8,63	mitosis/cell cycle
Aurkb	-8,02	mitosis/cell cycle
Prc1	-7,22	mitosis/cell cycle
Cks2	-6,46	mitosis/cell cycle
Eme1	-11,76	DDR/HDR
Exo1	-7,92	DDR/HDR
Bard1	-7,67	DDR/HDR
Brca1	-6,27	DDR/HDR
Chek1	-5,41	DDR/HDR
Brip1	-5,07	DDR/HDR

Overexpressed in MT	LFC	Function
Slc24a5	-11,75	sodium/calcium channels
Sc210a	-10,04	sodium/calcium channels
Slc44a4	-9,10	sodium/calcium channels
Sox11	-11,30	embryonic develop
Isx	-10,04	embryonic develop
Foxq1	-8,90	embryonic develop
Wnt10a	-7,90	embryonic develop
Myh3	-7,89	embryonic develop
Ascl4	-9,79	cell fate/diff/develop
Hoxc13	-9,44	morphogenesis
Chrng	-10,84	muscle contraction/ca2+
Tnnt2	-8,32	muscle contraction/ca2+
Myl4	-7,20	muscle contraction/ca2+ early develop
Chrna1	-5,50	muscle contraction/ca2+
Casq2	-4,47	muscle contraction/ca2+
Tnnc1	-4,24	muscle contraction/ca2+
Tnni1	-4,32	muscle contraction/ca2+ early develop
Myh8	-4,44	muscle contraction/ca2+ early develop

Overexpressed in GM	LFC	Function
Ckm	19,62	muscle related
Mybpc2	19,11	muscle related
Myh4	22,18	muscle related
Mb	17,37	muscle related
Myl1	17,16	muscle related
Pvalb	17,07	muscle contraction/ca2+
Tnnt3	16,87	muscle contraction/ca2+
Trdn	16,78	muscle contraction/ca2+
Atp2a1	16,39	muscle contraction/ca2+
Myot	16,27	muscle contraction/ca2+
Myh2	16,03	muscle contraction/ca2+
Tnnt3	16,87	muscle contraction/ca2+
Tnni2	15,67	muscle contraction/ca2+
Mylpf	14,65	muscle contraction/ca2+
Actc1	13,15	muscle contraction/ca2+
Acta1	11,91	muscle contraction/ca2+
Ckmt2	16,50	mitochondrion
Cspr3	10,72	cytoskeleton

Overexpressed in GM	LFC	Function
Scn4b	15,26	sodium channels
Eef1a2	17,03	muscle related
Myh4	9,31	muscle related
Mybpc2	8,12	muscle related
Ckm	7,13	muscle related
Tcap	6,28	muscle related
Sox18	7,97	embryonic develop
Sox10	7,63	embryonic develop
Notch4	7,48	Notch signaling
Mylk2	6,79	muscle contraction/ca2+
Atp2a1	6,42	muscle contraction/ca2+
Trdn	6,31	muscle contraction/ca2+
Myh2	5,72	muscle contraction/ca2+
Mylk4	5,62	muscle contraction/ca2+
Mybpc1	5,59	muscle contraction/ca2+
Actn3	4,45	muscle contraction/ca2+
Tnni2	4,35	muscle contraction/ca2+
Tnnt3	4,55	muscle contraction/ca2+
Acta1	4,48	muscle contraction/ca2+
Casq1	5,10	mitochondrion

283

C

MT vs MB

3545 — overexpressed in MB
3669 — overexpressed in MT
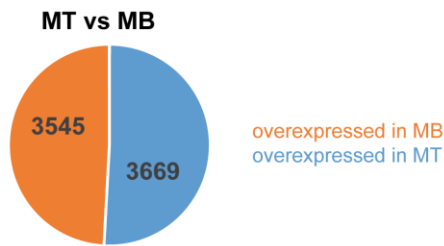
Overexpressed in MB	LFC	Function
Rmrp	-9,92	mitochondrial RNA
Lars2	-6,86	mitochondrial RNA
Rpph1	-9,16	RNA component
Hmx2	-6,50	embryonic develop
Fgf15	-5,94	embryonic develop
Hhip	-5,56	embryonic develop
Cacng7	-5,66	calcium channels
Id3	-4,77	typical MB gene
Rmi2	-5,36	HDR
Fosl1	-5,29	cell prolif/diff/transform
Prokr2	-5,25	muscle contraction
Actr3b	-5,01	cytoskeleton/cell shape/motility
Shroom4	-4,24	cytoskeleton /cell shape/motility
Tircc	-4,55	DNA replication
Cdt1	-4,25	DNA replication
Mcm5	-4,19	DNA replication
Mms22l	-4,05	DNA repair
Psrc1	-4,24	cell division/mitosis
Cdc25c	-4,49	cell division/mitosis
Ccno	-4,45	cell cycle
Ccnd1	-4,45	cell cycle
Mybl2	-4,55	cell cycle
Bub1b	-3,23	mitotic cell cycle
Cdk1	-3,15	mitotic cell cycle
Incenp	-3,27	mitotic cell cycle
Ccna2	-3,00	mitotic cell cycle
Ptprn	-6,73	mitotic cell cycle
Cdc25c	-4,49	mitotic cell cycle

Overexpressed in MT	LFC	Function
Myh3	15,37	embryonic develop
Mb	14,82	muscle related
Myl1	14,26	muscle related
Myh4	12,84	muscle related
Myh7	12,49	muscle related
Mylpf	12,40	muscle related
Myh8	13,99	muscle contraction/ca2+ early develop
Actc1	13,30	muscle contraction/ca2+
Mylpf	12,40	muscle contraction/ca2+
Tnnt3	12,33	muscle contraction/ca2+
Casq2	12,26	muscle contraction/ca2+
Tnni1	12,14	muscle contraction/ca2+ early develop
Tnni2	11,30	muscle contraction/ca2+
Tnnc2	11,26	muscle contraction/ca2+
Trdn	10,45	muscle contraction/ca2+
Scn4a	10,33	muscle contraction/ca2+
Myh2	10,29	muscle contraction/ca2+
Casq1	10,22	muscle contraction/ca2+
Myl4	10,12	muscle contraction/ca2+ early develop
Myoz2	12,88	calcineurin
Ano5	12,74	calcium
Cacna1s	10,18	calcium
Ckm	12,47	muscle related
Mybpc2	10,97	muscle related
Neb	11,75	cytoskeleton
Mypn	10,82	cytoskeleton
Csrp3	10,73	cytoskeleton
Mef2c	10,95	myogenesis
Myog	10,88	myogenesis
Art1	10,88	PTM
Cav3	10,98	mature MT

284

Supplementary Figure 1. **Differentially expressed genes**

A. Differentially expressed genes between GM and MB and functional annotation of the most over expressed in MB (orange) and in GM (blue). The LFC for each gene is indicated. N=2 for GM and n=2 for MB.

B. Differentially expressed genes between GM and MT and functional annotation of the most over expressed in MT (orange) and in GM (blue). The LFC for each gene is indicated. N=2 for GM and n=2 for MT.

C. Differentially expressed genes between MT and MB and functional annotation of the most over expressed in MB (orange) and in MT (blue). The LFC for each gene is indicated. N=2 for MB and n=2 for MT.

General discussion

General discussion

My thesis work focuses on investigating genome-wide signaling pathways controlled by steroid receptors, in tissues in which they have opposite effects, namely skeletal muscle and prostate. In addition, we compared the genomic landscape of skeletal muscle tissue to C2C12 myoblasts and myotubes.

In skeletal muscle, glucocorticoids regulate glucose, lipid and protein metabolism and thus contribute to energy homeostasis (Tanaka et al., 2017). The objective of the first part of the thesis was to provide insights into the regulation of anabolic and catabolic pathways in skeletal muscles controlled by myofiber GR at physiological glucocorticoid levels. To delineate the physiological and molecular function of GR in mature skeletal muscles, the lab generated GR[(i)skm-/-] mice in which GR is selectively ablated in myofibers at adulthood. The study of the model showed that myofiber GR down-regulates muscle mass and strength.

Using transcriptomic analysis in control and GR[(i)skm-/-] mice, we identified GR-regulated genes in muscle. Our results demonstrate that physiological glucocorticoids, via myofiber GR, reduce the expression of several anabolic factors (e.g. Akt3, Rps6kb1, Pik3cg) and induce that of anti-anabolic factors (e.g. Eif2ak1, Eif4ebp2, Ddit4, Pik3r1), leading to a decreased anabolic pathway, thereby limiting muscle fiber size and mass, without stimulating catabolic pathways.

Moreover, we determined genome-wide GR chromatin occupancy under physiological glucocorticoid levels in muscle by chromatin immunoprecipitation, followed by massive parallel sequencing (ChIP-seq), in order to delineate GR target genes in skeletal muscle. We identified binding sites located in many genes, including several encoding anti-anabolic and anabolic factors. The binding sites were distributed mostly at promoter-TSS regions (-1000 bp; + 100 bp).

We also performed ChIP-seq analysis for various histone marks in order to characterize the genomic landscape of GR binding sites (GRBS), and we demonstrate that GR is bound at active promoters, defined by the presence of H3K27ac, H3K4me3 and Pol2, and low H3K4me1 levels, and active enhancer regions, defined by the presence of H3K27ac, H3K4me1 and Pol2, and low H3K4me3 levels. Our data show that, whereas GR was bound at genomic regions encompassing response elements corresponding to ARE and GRE (termed ARE/GRE) in more than 30 % of the enhancer sites, promoter occupancy of GR correlated with less than 2 % of AREs/GREs, and Nrf1 appeared as one the most frequent motifs by *de novo* motif search (Table 1, 2). Thus, GR bound at enhancers might interact with Nrf1 at the promoter region of GR-regulated genes to initiate transcription.

To provide insights into the molecular determinants of GR transcriptional regulation in skeletal muscle, we performed ChIP-seq known motif search at ARE/GRE-containing enhancers. Myod1, Ctcf, Stat1 and Stat3 were the most enriched motifs identified (Table 1). Myod1 is recruited at half of GR target genes and half of the genes down-regulated in GR$^{(i)skm-/-}$ mice were bound by GR and Myod1. In addition, a mass spectrometry assay revealed interacting partners of GR and among them we identified Stat1 and Stat3 as well as Foxf2. The interactions of GR with Stat3 and Foxf2 were confirmed experimentally and we have shown that Foxf2 putative binding sites surround enhancer localized GR peaks (Table 1).

To determine the impact of GR, Myod1 and Foxf2 on gene regulation of the anti-anabolic factors *Eif4ebp2* and *Pik3r1*, small interfering RNA (siRNA)-mediated knock-down of each of these factors was performed in C2C12 myotubes. Myod1 and Foxf2 are required for efficient GR binding at enhancer regions, and Foxf2 is recruited in a GR-dependent manner. Taken together, our data showed that Myod1 and Foxf2 bind DNA at the vicinity of AREs/GREs of genes encoding anti-anabolic factors in a GR-dependent manner, and that these two co-factors facilitate GR binding at its response elements.

It would be interesting to determine the role of myofiber GR under pharmacological glucocorticoid levels and to investigate the molecular determinants controlling muscle mass and strength under these conditions, using a combination of phenotypic, transcriptomic and cistomic analysis after dexamethasone (DEX) treatment. Moreover, it was previously shown that Myod1 expression is glucocorticoid-sensitive, since a DEX treatment decreased Myod1 expression within few hours (Sun et al., 2008). This suggests that upon DEX treatment, GR might be released from pre-existing complexes through Myod1 degradation, and associates with other partners like Foxo1 (Waddell et al., 2008), to promote muscle atrophy.

Moreover, since GR is bound at AREs/GREs at enhancers and might interact with Nrf1 at the promoter region of GR-regulated genes, an interaction not demonstrated before in any tissue, and Ctcf is recruited at genomic regions encompassing GR response elements, it would be interesting to determine whether chromosomal conformation changes are taking place to coordinate transcription. To elucidate whether the chromosomal interactions involve the looping of the enhancer region with the promoter region, a conformation capture on chip (4C) should be performed. In addition, given GR's preference for binding at accessible chromatin, we could analyze this accessibility, identify differences that explain the locus-specific binding of GR and determine alternations in GR genomic binding, caused by other co-factors, by ATAC-seq (assay for transposase-accessible chromatin) under physiological and pharmacological conditions. We could also correlate the GR-binding, the acetylated histone mark and the binding of Ctcf with the chromatin accessibility.

Given the fact that GR is also expressed in prostate, we investigated GR signaling in this tissue. We determined the GR cistrome and transcriptome and observed that prostate GRBS are mostly located in intergenic and intronic regions. The most enriched *de novo* motifs identified under these binding sites were canonical AREs/GREs as well as TP53/63/73 motifs (Table 1, 2). We also identified more than 10000 expressed genes in WT prostates by setting a threshold of raw read counts greater than 100.

Moreover, to provide insights into GR tissue-specificity under physiological conditions, we compared the cistromes and transcriptomes of GR in murine prostate and murine skeletal muscle. The data overlap of the two GR ChIP-seqs showed that more than 3500 genes were bound by GR in both tissues. *De novo* motif analysis of the sites bound by GR in both tissues, in intergenic and intronic regions, on the one hand, using the skeletal muscle coordinates, revealed AREs/GREs and CTCF motifs, and on the other hand, using prostate coordinates, revealed AREs/GREs and TP53/63/73 motifs (Table 1, 2). Thus, GR binds to AREs/GREs in intergenic and intron regions, and not in promoter regions, where it probably interacts or cooperates with distinct factors depending on the tissue. Of note, the CTCF motifs were found in muscle, not in prostate, and were located either very close to GREs (0 to 1 bp between the two motifs) or more than 145 kb from each other.

The cistromic intersection of GR in muscle and in prostate revealed that 56% of the GRBS are shared by the two tissues. However, we also identified unique binding sites in muscle. *De novo* motif analysis of the shared binding sites revealed canonical and half-site AREs/GREs as well as Nrf1 and Stat3 motifs, whereas similar analysis of the selective binding sites revealed canonical and half-site AREs/GREs as well as Myod1, MEF2A/B/C/D and TEAD2 motifs (Table 1). In the manuscript in preparation (Results part I), we showed that GR interacts with Stat3 and Myod1 in muscle and that their cognate response elements are in the vicinity of enhancer-GRE containing regions. In addition, we provide evidence that enhancer GRE-bound GR might interact and communicate with Nrf1, bound to its response element located in promoter regions, to regulate gene transcription.

Detailed analysis of the response elements in the different tissues and genomic repartitions revealed that the identified AREs/GREs bound by GR do not exhibit any selectivity between the two tissues (Table 2) and that the tissue-specificity of GR is defined by the distinct surrounding co-factors (Table 1).

It would be interesting to investigate GRBS in prostate at pharmacological glucocorticoid levels. Moreover, we did not investigate the genomic landscape of GR binding in prostate using histone marks and Pol2. In addition, we could compare the GRBS we identified in muscle and prostate with those of other studies in adipose tissue or liver. A preliminary analysis of GRBS in prostate, muscle and liver (Lim et al., 2015) showed that GR binds to almost 2400/15263 genes (15.7%) in the three tissues. *De novo* motif analysis of the binding

sites located in these genes revealed AREs/GREs and additional factors that might impose the tissue-specificity, e.g. bHLH factors (Myod1) for muscle, p53 domain factors (TP53/63/73) for prostate, and Forkhead box factors (FOX) and nuclear receptors with C4 zinc fingers (Rxra) for liver.

From a clinical point of view, AR and GR are therapeutic targets for several diseases including asthma, sarcopenia, allergies and cancers. The AR ligands, androgens, induce the growth and proliferation of prostatic epithelial cells and therefore increase the risk of prostate cancer (Banerjee et al., 2018). On the other hand, glucocorticoids are frequently applied to patients with prostate cancer combined with chemotherapies (Montgomery et al., 2014). It is also supported that the signaling pathways of both receptors are interconnected and there is interplay between them in prostate cancer (Arora et al., 2013; Isikbay et al., 2014). However, the molecular determinants of this interplay are not well characterized for the moment and this is a major issue for the pharmaceutical companies.

We thus characterized AR signaling in murine prostate by determining the AR cistrome and transcriptome. We identified that the prostate AR binding sites (ARBS) are mostly located in intergenic and intronic regions and the most enriched *de novo* motifs under these binding sites were also AREs/GREs (Table 1, 2).

Furthermore, to investigate what confers to AR and GR their selectivity under physiological conditions, we performed a genome-wide comparative analysis between AR and GR cistromes and transcriptomes in prostate. The overlap of the AR and GR ChIP-seqs revealed that almost 2000 genes were bound by both AR and GR in prostate. *De novo* motif analysis of the ARBS and GRBS located in intergenic and intronic regions of these genes, identified AREs/GREs using AR coordinates and AREs/GREs and TP53/63/73 motifs using GR coordinates (Table 1, 2). Thus, AR and GR bind to AREs/GREs in intergenic and intron regions and not in promoter regions, where they probably interact or cooperate with distinct factors depending on the receptor. Interestingly, we observed that the response elements bound by both AR and GR in intergenic and intronic regions are composed elements consisting of the 1st half site of a canonical 15-bp ARE, as described by the Jaspar motif database, with a very conserved **G** at the position 2 or 3 and the same probability for the **A** and **C** in positions 4 or 5 and 5 or 6, respectively, and the 2nd half site of a canonical 15-bp GRE, as described by the Jaspar database, with a very conserved **G** at the position 11, a conserved **C** at the position 14 and two **T**, the first less conserved that the other, at the positions 10 and 12, respectively (Table 2) (for detailed representation see figure 8A of Results part II "II. Genome-wide comparative analysis of androgen & glucocorticoid receptors' cistomes and transcriptomes in prostate").

Finally, the cistromic intersection of AR and GR in prostate showed that 66% of the GRBS are not bound by AR. *De novo* motif analysis of the shared binding sites revealed AREs/GREs and

motifs of C2H2 zinc finger factors, whereas similar analysis of the unique sites revealed TP53/63/73 motifs and motifs of Forkhead box, TEA-domain and Homeo domain factors (Table 1).

Importantly, our results indicate that the specificity of the responses is not only based on the tissue-specific expression of the receptors, but also on the cooperation with distinct surrounding transcription factors, rather than the sequence of the response elements.

It would be interesting to investigate the ARBS in prostate also under pharmacological conditions, e.g. using DEX for GR and dihydrotestosterone (DHT) for AR. Moreover, we did not investigate the genomic landscape of AR binding in prostate using histone marks and Pol2. A better characterization of the response elements and the identification of additional binding sites for co-partners that surround the AREs/GREs might provide important insights into AR/GR specificity in a given tissue. Furthermore, as the quality of the AR ChIP-seq was not optimal, due to the used antibody, to experimental challenges of the assay or to the sequence depth, we could not perform a detailed comparison and this assay should be repeated.
Of note, a complementary study of AR binding in skeletal muscle in combination with the one in prostate will give useful insights into the tissue-specificity of AR.

All these observations strongly support a cross talk between AR and GR in gene regulation that needs to be characterized thoroughly, genome-wide, in terms of common and selective binding sites, DNA sequences, co-factors, transcriptional activity, genes controlled by ligands and their corresponding responses. The investigation of the cell-specificity, using high-throughput single cell RNA-seq, will provide additional information on this cross talk. Many of our results are descriptive and require experimental validation in order to understand the mechanisms responsible for the different effects and identify discriminative response elements that could be used to identify analogs inducing specific genes, and thus with reduced side effects, with the ultimate goal to ameliorate existing therapies.

The objective of the third part of my thesis work was to compare the expression profile and genomic landscape C2C12 myoblasts and myotubes, and skeletal muscle. C2C12 cells derived from murine skeletal muscle cells and are considered as a well-established model which mimics the development of skeletal muscle in vivo, appropriate to study muscle regeneration and differentiation as well as myogenic regulation (Burattini et al., 2004; Cheema et al., 2003; Manabe et al., 2012; Moran et al., 2002; Nedachi et al., 2008; Schoneich et al., 2014).

To provide insights into the molecular mechanisms and transcriptional programs underlying muscle differentiation, we performed a genome-wide comparative transcriptomic and cistromic analysis between undifferentiated C2C12 myoblasts (MB) and differentiated C2C12

myotubes (MT) and murine gastrocnemius muscles (GM) in terms of gene expression, levels of expression, functional annotation, families of transcription factors, binding motifs and histone modifications.

Our analysis revealed that 57% of the expressed genes were common among the three datasets. Moreover, 20% of the genes are selectively expressed in GM-specific genes. Among the expressed genes of the shared stages, the MT/GM showed 7.5% overlap of genes. There are more than 6400 common expressed genes among MB, MT and GM, but at various levels. In addition, there are 13900 differentially expressed genes between GM and MB reporting the differences between the two stages. Almost ⅓ of the genes are specifically expressed in either MB, MT or GM.

Pathway analysis revealed that the three stages exhibit specific signatures (Figure 1A):
   i.   *MB-signature*: mitosis, cell cycle, DNA replication, cytoskeleton regulation, DDR, HDR, axon guidance,
   ii. *MT-signature*: protein modifications, autophagy, muscle contraction, ion homeostasis, axon guidance and
   iii. *GM-signature*: mitochondrial functions, autophagy, muscle contraction, ion homeostasis, complement activation, peroxisome, break repair, inflammation.

The differentiated C2C12 cells share more common characteristics with skeletal muscle, than the undifferentiated ones.

During the muscle differentiation process specific and shared classes of transcription factors are enriched (Figure 1B). The factors of the classes of bHLH, bZIP, C2H2 zinc finger and paired box are enriched throughout this process. In addition, the shared stages are additionally enriched for factors of the class of the bHSH and the family of E2F-related factors. The specific stages were also enriched for factors of the class of Rel homology region.

In addition, *de novo* motif analysis of the MB/MT, MB-specific, MT-specific and GM-specific binding sites identified various motifs (Figure 1C). Those corresponding to the classes of bHLH, bZIP, C2H2 zinc finger, Forkhead and Homeo domain factors are enriched throughout the differentiation. MB and MT are additionally enriched for motifs of the classes of TEA domain, Runt domain and the family of ETS-related factors. MT and GM are also enriched for motifs of the family of steroid hormone receptors. The motifs of the family of SOX-related factors was selectively enriched in MB stage, the motifs of the families ARID domain, MADS box, SMAD/NF1 DNA binding domain, POU domain factors and the class of paired box factors were uniquely enriched in MT stage and motifs of most of the factors of the class of nuclear receptors with C4 zinc fingers and the class of the Rel homology region were specifically enriched in GM.

Importantly, our study highlighted the expression selectivity of transcription factors and transcription factors' families involved in myogenesis and in other biological processes (Figure 1D, E). Of note, members of the same subfamilies are not over expressed at the same myogenic stage, as for example the members of the family of nuclear receptors and especially the members of the Nerve Growth Factor IB-like subfamily that are highly expressed in GM (Nr4a1, Nr4a3) but also in MB (Nr4a2).

Moreover, by correlating the epigenetic marks with gene expression, we found that the over or under expression of genes in a specific cluster correlate with the absence or presence of H3K27ac, an open chromatin mark, in the epigenetic locus. Finally, the chromatin state maps across genome (genome-wide changes in the epigenetic landscape) unraveled specific characteristics for each stage and revealed more differences between murine skeletal muscle and C2C12 cells at the enhancer level than at promoter regions, indicating that distant enhancers provide selectivity in the differentiation process. Nevertheless, as the publically available datasets used in this study originate from different labs, we cannot exclude the presence of biases in the analysis.

Overall, this study provided a detailed description of genes, signaling pathways and transcription factors that are differentially expressed during myogenic proliferation and differentiation. The similarities of differentiated C2C12 with skeletal muscle indicate that the former can be used to some extend as an *in vitro* model of skeletal muscle tissue that recapitulates myogenic differentiation.

It would be of interest to investigate the epigenetic status of additional histone marks and Pol2 during the differentiation process and in correlation with gene expression. These results should be validated experimentally in vitro and in vivo by qPCR, Westen Blots, siRNA treatments, fluorescence microscopy and Hematoxilin Eosin staining. Moreover, a Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) of the over and under expressed genes, in order to identify molecular signatures, would complete this study.

Taken together, our results provide a characterization of genes controlled by androgens and glucocorticoids and the corresponding response in skeletal muscle and prostate. These results provide the basis of a molecular understanding of tissue- and/or promoter-specific activity of androgens and glucocorticoids, and thus open new avenues to design screens for analogs inducing genes selectively, using cell-based assays.

Recapitulative tables of the part I and II of the thesis

Table 1. Enriched response elements among the genomic repartitions

	Enhancer region	**Promoter region**	**Undetermined region**
GRBS skm	GRE (> 30%) Ctcf Myod1 Stat3/1 Foxf2	GRE (< 2%) Nrf1 Sp1 Yy1 Nfya/b ETS-related	Half-site GRE Mef2a/b/c/d Tead2
GRBS prostate	GRE Tp53/63/73	Znf263 E2f6/4 Sp2	Half-site GRE Nrf1 Forkhead box Stat3/1 Nfya/b ETS-related C2H2 zinc finger Tead2/3 Homeo domain SMAD/NF-1 DNA binding domain
ARBS prostate	ARE	-	Half-site ARE C2H2 zinc finger Tead2/3 Homeo domain

Table 2. Enriched consensus AREs & GREs among the genomic repartitions

	Intergenic region	**Intronic region**	**Promoter region**
ARBS prostate			-
ARBS/GRBS prostate			-
selective ARBS prostate	-	-	-
selective GRBS prostate			-
GRBS prostate			-
GRBS pro/skm			-
selective GRBS skm			-
selective GRBS prostate			-
GRBS skm			

References

Arora, V.K., Schenkein, E., Murali, R., Subudhi, S.K., Wongvipat, J., Balbas, M.D., Shah, N., Cai, L., Efstathiou, E., Logothetis, C., *et al.* (2013). Glucocorticoid receptor confers resistance to antiandrogens by bypassing androgen receptor blockade. Cell *155*, 1309-1322.

Banerjee, P.P., Banerjee, S., Brown, T.R., and Zirkin, B.R. (2018). Androgen action in prostate function and disease. Am J Clin Exp Urol *6*, 62-77.

Burattini, S., Ferri, P., Battistelli, M., Curci, R., Luchetti, F., and Falcieri, E. (2004). C2C12 murine myoblasts as a model of skeletal muscle development: morpho-functional characterization. Eur J Histochem *48*, 223-233.

Cheema, U., Yang, S.Y., Mudera, V., Goldspink, G.G., and Brown, R.A. (2003). 3-D in vitro model of early skeletal muscle development. Cell Motil Cytoskeleton *54*, 226-236.

Isikbay, M., Otto, K., Kregel, S., Kach, J., Cai, Y., Vander Griend, D.J., Conzen, S.D., and Szmulewitz, R.Z. (2014). Glucocorticoid receptor activity contributes to resistance to androgen-targeted therapy in prostate cancer. Horm Cancer *5*, 72-89.

Lim, H.W., Uhlenhaut, N.H., Rauch, A., Weiner, J., Hubner, S., Hubner, N., Won, K.J., Lazar, M.A., Tuckermann, J., and Steger, D.J. (2015). Genomic redistribution of GR monomers and dimers mediates transcriptional response to exogenous glucocorticoid in vivo. Genome Res *25*, 836-844.

Manabe, Y., Miyatake, S., Takagi, M., Nakamura, M., Okeda, A., Nakano, T., Hirshman, M.F., Goodyear, L.J., and Fujii, N.L. (2012). Characterization of an acute muscle contraction model using cultured C2C12 myotubes. PLoS One *7*, e52592.

Montgomery, B., Cheng, H.H., Drechsler, J., and Mostaghel, E.A. (2014). Glucocorticoids and prostate cancer treatment: friend or foe? Asian J Androl *16*, 354-358.

Moran, J.L., Li, Y., Hill, A.A., Mounts, W.M., and Miller, C.P. (2002). Gene expression changes during mouse skeletal myoblast differentiation revealed by transcriptional profiling. Physiol Genomics *10*, 103-111.

Nedachi, T., Fujita, H., and Kanzaki, M. (2008). Contractile C2C12 myotube model for studying exercise-inducible responses in skeletal muscle. Am J Physiol Endocrinol Metab *295*, E1191-1204.

Schoneich, C., Dremina, E., Galeva, N., and Sharov, V. (2014). Apoptosis in differentiating C2C12 muscle cells selectively targets Bcl-2-deficient myotubes. Apoptosis *19*, 42-57.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Sun, L., Trausch-Azar, J.S., Muglia, L.J., and Schwartz, A.L. (2008). Glucocorticoids differentially regulate degradation of MyoD and Id1 by N-terminal ubiquitination to promote muscle protein catabolism. Proc Natl Acad Sci U S A *105*, 3339-3344.

Tanaka, H., Shimizu, N., and Yoshikawa, N. (2017). Role of skeletal muscle glucocorticoid receptor in systemic energy homeostasis. Exp Cell Res *360*, 24-26.

Waddell, D.S., Baehr, L.M., van den Brandt, J., Johnsen, S.A., Reichardt, H.M., Furlow, J.D., and Bodine, S.C. (2008). The glucocorticoid receptor and FOXO1 synergistically activate the skeletal muscle atrophy-associated MuRF1 gene. Am J Physiol Endocrinol Metab *295*, E785-797.
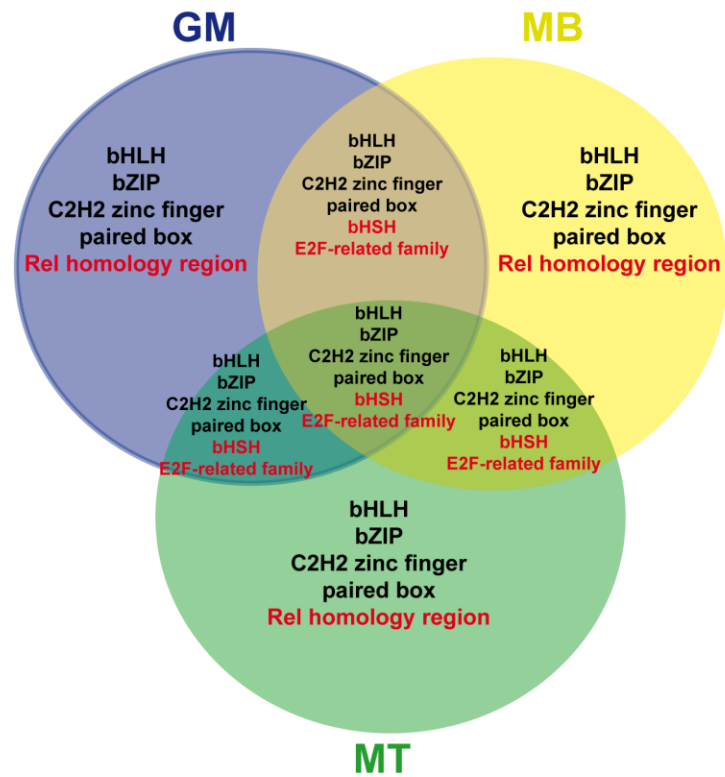
Recapitulative schemes of the part III of the thesis
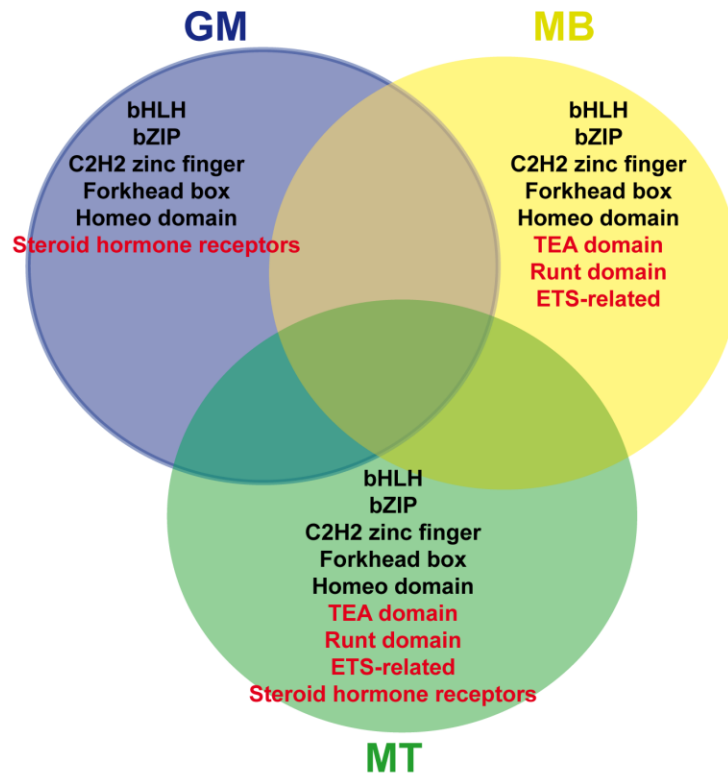
Figure 1.

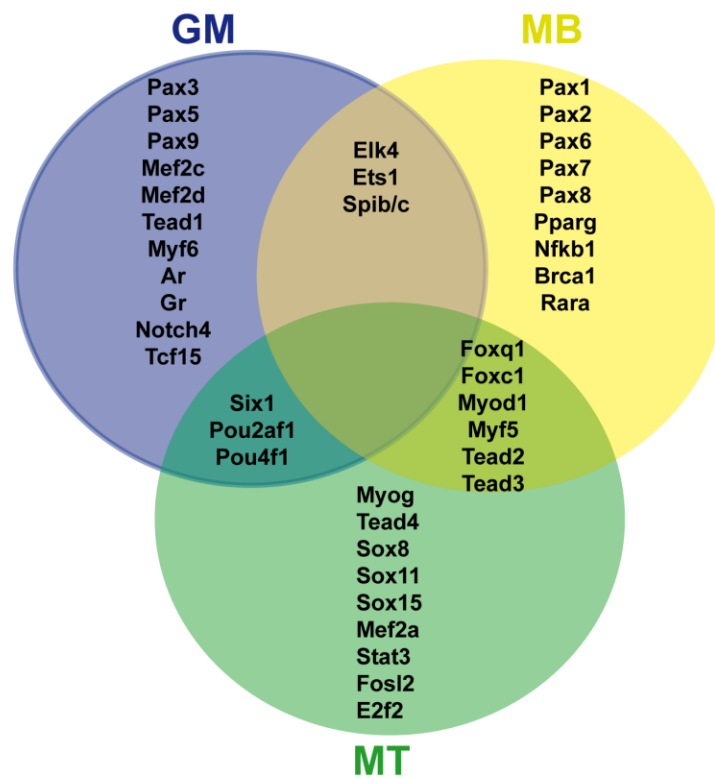 **A   Shared and specific signatures in MB, MT and GM**
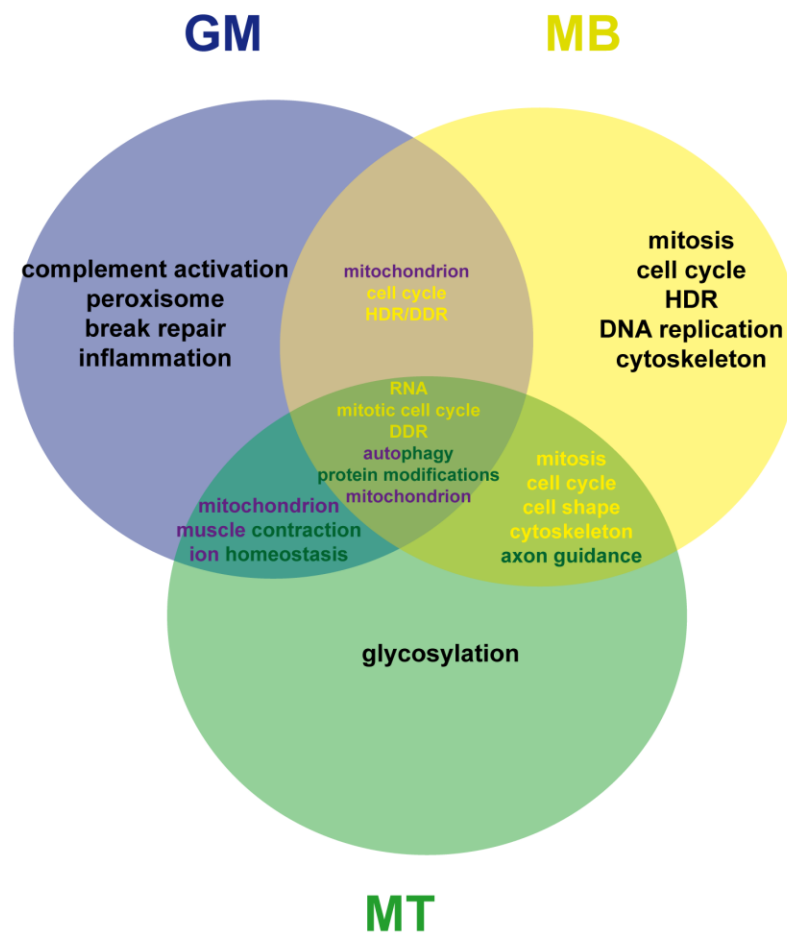
B   Enriched families and classes of transcription factors in MB, MT and GM
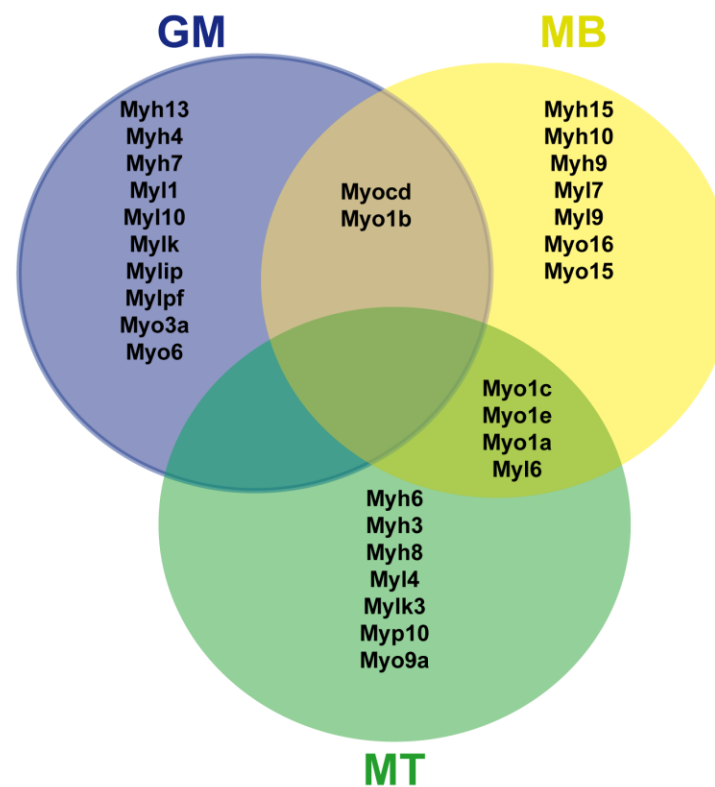

C   Enriched *de novo* motifs of families and classes of transcription factors in MB, MT and GM

D   Over expressed transcription factors in MB, MT and GM

GM **MB**

Pax3
Pax5
Pax9
Mef2c
Mef2d
Tead1
Myf6
Ar
Gr
Notch4
Tcf15

Elk4
Ets1
Spib/c

Pax1
Pax2
Pax6
Pax7
Pax8
Pparg
Nfkb1
Brca1
Rara

Six1
Pou2af1
Pou4f1

Foxq1
Foxc1
Myod1
Myf5
Tead2
Tead3

Myog
Tead4
Sox8
Sox11
Sox15
Mef2a
Stat3
Fosl2
E2f2

MT

E   Over expressed myosins in MB, MT and GM

GM **MB**

Myh13
Myh4
Myh7
Myl1
Myl10
Mylk
Mylip
Mylpf
Myo3a
Myo6

Myocd
Myo1b

Myh15
Myh10
Myh9
Myl7
Myl9
Myo16
Myo15

Myo1c
Myo1e
Myo1a
Myl6

Myh6
Myh3
Myh8
Myl4
Mylk3
Myp10
Myo9a

MT

UNIVERSITÉ DE STRASBOURG

Anna-Isavella RERRA

Genome-wide analyses of signaling pathways controlled by steroid receptors

École Doctorale
des Sciences de la Vie
et de la Santé
STRASBOURG

Résumé

Les androgens(ADs) et les glucocorticoïdes (GCs) sont des hormones stéroïdiennes qui exercent des effets pléiotropes chez les mammifères. Leurs effets sont relayés par deux récepteurs nucléaires, le récepteur des androgènes (AR) et le récepteur des glucocorticoïdes (GR), respectivement. Même si les GCs sont fréquemment utilisés pour traiter les maladies inflammatoires et les antiandrogènes pour le cancer de la prostate, les traitements à long terme induisent des effets secondaires majeurs, notamment l'atrophie musculaire.

Afin de préciser les mécanismes d'action de ces hormones, nous avons réalisé des analyses phénotypiques, transcriptomiques et cistromiques. La première partie de ce travail démontre que GR des myofibres contrôle négativement la masse et la force musculaire aux niveaux physiologiques de GCs. La perte de GR dans les muscles squelettiques n'affecte pas les voies cataboliques, mais augmente l'expression de facteurs anaboliques et réduit celle de facteurs anti-anaboliques. Nous avons également montré que GR se lie à des éléments de réponse du GR (GREs) situés aux enhancers, en association avec Myod1 et Foxf2, et interagit avec des facteurs liés aux promoteurs, tels que Nrf1, pour favoriser la transcription des gènes.

Dans la deuxième partie de ce travail, nous avons comparé le cistrome et le transcriptome du GR dans la prostate et le muscle squelettique, et identifié des sites de liaison pour d'autres facteurs de transcription proche des GREs, indiquant que ces facteurs contribuent à la spécificité tissulaire. De plus, en comparant les cistromes et transcriptomes d'AR et de GR dans la prostate, nous montrons que les éléments de réponse liés par les deux récepteurs sont distincts de ceux liés uniquement par AR ou GR, et que la sélectivité du récepteur dépend de la liaison d'autres facteurs de transcription.

Enfin, nous avons comparé les données transcriptomiques et épigénétiques du tissu musculaire squelettique et de myoblastes et myotubes C2C12, et nous fournissons une description détaillée de gènes, voies de signalisation et facteurs de transcription exprimés de façon différentielle pendant la différenciation myogénique.

En conclusion, nos travaux ont permis de clarifier les mécanismes moléculaires régulant l'homéostasie musculaire et ont établi la base d'une compréhension moléculaire des effets spécifiques des ADs et des GCs dans divers types cellulaires.

Mots clés: glucocorticoïdes, androgènes, GR, AR, GRE, ARE, muscle squelettique, myofibres, masse/force musculaire, voies anaboliques/cataboliques, prostate, spécificité tissulaire, sélectivité des récepteurs, enhancer, promoteur, modifications des histones, cellules C2C12, facteurs de transcription, myogenèse, épigénétique

Abstract

Androgens (ADs) and glucocorticoids (GCs) are steroid hormones exerting pleiotropic effects in mammals. Their effects are mediated by two nuclear receptors, the androgen (AR) and the glucocorticoid (GR) receptor, respectively. Although GCs are extensively used to treat inflammatory diseases and antiandrogens for prostate cancer, long-term treatments induce major side effects such as muscle atrophy.

To determine the mechanisms underlying their effects in muscle, we performed phenotypic, transcriptomic and cistromic analyses. The first part of this work demonstrates that myofiber GR negatively controls muscle mass and strength under physiological GCs levels. GR loss in skeletal muscle did not affect catabolic pathways, but enhanced the expression of anabolic factors and reduced that of anti-anabolic ones. We also showed that myofiber GR binds DNA to GR response elements (GREs) located at enhancers, in association with Myod1 and Foxf2, and interact with promoter-bound factors such as Nrf1 to promote gene transcription.

In the second part of this work, we compared GR cistromes and transcriptomes in prostate and skeletal muscle, and identified binding sites for additional transcription factors in the vicinity of GREs, indicating that they contribute to the tissue specificity. In addition, by comparing the AR and GR cistromes and transcriptomes in prostate, we show that the response elements bound by both receptors are distinct from those bound by either AR or GR, and that the receptor-selectivity depends mostly on the surrounding factors.

Finally, we compared transcriptomic and epigenetic data of skeletal muscle tissue and C2C12 myoblasts and myotubes and provide a detailed description of genes, signaling pathways and transcription factors that are differentially expressed during myogenic differentiation.

In conclusion, our work allowed to clarify the molecular mechanisms regulating muscle homeostasis and provides the basis of a molecular understanding of tissue- and/or promoter-specific activity of ADs and GCs.

Keywords: glucocorticoids, androgens, GR, AR, GRE, ARE, skeletal muscle, myofibers, muscle mass/strength, anabolic/catabolic pathways, prostate, tissue-specificity, receptor-selectivity, enhancer, promoter, histone modifications, C2C12 cells, transcription factors, myogenesis, epigenetics