Thèse présentée pour obtenir le grade de
Docteur de l'Université Louis Pasteur
Strasbourg I

Discipline : Sciences du Vivant

par Thomas HUSSENET

Towards the IDENTIFICATION of ONCOGENES

by HIGH RESOLUTION MAPPING OF GENE AMPLIFICATIONS

of the 3q25-qter region

in Malignant Fibrous Histiocytoma and Squamous Cell Carcinoma

Soutenue publiquement le 12 décembre 2005 devant le jury composé de :

Docteur Stanislas du Manoir	Directeur de Thèse
Professeur Serge Potier	Rapporteur interne
Professeur Peter Lichter	Rapporteur externe
Professeur Norbert Arnold	Rapporteur externe
Docteur Bohdan Wasylyk	Examinateur
Professeur Jean Louis Mandel	Examinateur

Unité de Recherche :
Institut de Génétique et de Biologie Moléculaire et Cellulaire
INSERM U596 / CNRS - ULP UMR7104

A l'ensemble de mes Proches.

A mes Grands Pères,
ma Maman,
et toute ma famille.

A mon Papa.

« Understanding is not curing ». Tim Hunt, de passage à l'IGBMC en octobre 2005.
Oui. Mais ça peut aider. L'utopiste, novembre 2005.

REMERCIEMENTS

L'ensemble de ces travaux a été effectué depuis mon DEA dans le groupe de Cytogénétique Moléculaire, sous la direction de **Stanislas du Manoir**. Stan, ces quelques premières lignes sont pour te remercier vraiment très sincèrement d'avoir cru en moi dès le départ et pour m'avoir soutenu tout au long de cette thèse. Dès que j'essaye de prendre un peu de recul sur le chemin que j'ai parcouru depuis notre première entrevue, déjà pleine d'enthousiasme des deux côtés (un coup de foudre ? ;o), j'en prends la vraie mesure … et je trouve qu'il a pas mal évolué quand même le petit étudiant de maîtrise qui en avait marre de s'asseoir sur les bancs de la fac et qui voulait enfourcher des pipettes au galop… je pense très honnêtement que je n'en serais pas arrivé là sans toi alors un très très sincère et grand Merci pour Tout ce que tu m'as apporté et appris, ainsi que pour les très nombreuses discussions scientifiques passionnantes qu'on a pu partager. Même et peut-être surtout pour celles, les plus enflammées, où on était pas forcément d'accord et qui ont parfois pu ressembler à des débats houleux. Très simplement, merci encore pour tes qualités personnelles de grande disponibilité, de naturelle simplicité et d'Humanité. Merci finalement d'avoir su me laisser l'indépendance dont j'avais besoin par moments.

Je tiens également à remercier ici très chaleureusement **Jean Louis Mandel**. Merci Jean Louis, tout d'abord d'avoir suivi avec une attention particulière mes différents projets au cours de leur évolution, mais aussi et surtout de les avoir discutés très simplement en m'apportant ton expérience, ton savoir et tes conseils avisés lorsque tu en avais le temps et l'envie. Merci finalement pour ta simplicité et ton accessibilité.

Merci à **l'ensemble des membres du jury** d'avoir accepté d'évaluer ce travail et de le discuter, et en particulier :

Un très grand merci à **Serge Potier**, un de mes anciens professeurs de Génétique à l'ULP… et qui, certainement, devrait se sentir un peu responsable du fait que je viens de passer 5 ans dans un labo ! Au plaisir de se retrouver à l'occasion d'un match de tennis de table pour une revanche ;o).

Many thanks to **Peter Lichter** who accepted very kindly to bring his thorough knowledge to evaluate this work. I always appreciated your very well-fitted presentations, questions and/or remarks on the diverse congresses/seminars I had the opportunity to hear from you; so I want to very sincerely tell you how honoured I can feel to have you among the members of my thesis jury.

I'd like to thank **Norbert Arnold** for accepting very kindly to evaluate this work.

Merci à **Bohdan Wasylyk** d'avoir pris de son temps pour examiner ce travail.

Je voudrais également remercier ici tous les « **sponsors** » qui ont permis ce travail, qu'ils soient **institutionnels** ou **privés**.
Merci à toutes les personnes de l'école doctorale, qui après m'avoir évalué en septembre 2001, m'ont donné la chance de commencer ma thèse par un financement **MNERT**.
Merci également aux **associations**, à leurs **bénévoles** et **donateurs** :

- La Ligue Nationale Contre le Cancer, **La Ligue Régionale d'Alsace et ses 2 comités départementaux,**

- l'ARC (**Association pour la Recherche sur le Cancer**) particulièrement pour ma bourse de 4ème année de thèse.

Pour les « privés », une toute spéciale dédicace à **mes parents**, à **Anne** et enfin à **Fabrice**… et que ceux à qui je dois encore des ronds ne s'inquiètent pas trop… dès que j'ai touché l'argent du Nobel, ce qui devrait normalement plus trop tarder, j'en mettrai un peu de côté pour vous, promis !

Merci à **tous les services communs de l'IGBMC / ICS** pour leur aide et donc surtout (!) aux gens qui les animent :
merci à toutes les personnes des ateliers et des services techniques dont **Patrick** et **Serge**, des services administratifs, **Elisabeth Toussaint** en particulier, des services commerciaux et surtout **Dominique** et **Peggy**, du service informatique et en particulier à **Antoine Collet,** ainsi qu'aux « stockmens **», Serge** et **Christian**. Je remercie également tout particulièrement **Betty** et **Laure** et le service de culture cellulaire, **Serge** du séquencaçe, **Frank**, **Ingrid et Edouard** de l'ex-service de synthèse des oligos, **Pascal** pour la synthèse de peptides, **Mustapha** pour les Ac monoclonaux, **Marcel** et **Didier** de l'imagerie.
Un merci tout spécial, aux **personnes du service des Puces à ADN** avec qui j'ai particulièrement interagi: **Christelle Thibault, Doulaye Dembele**, **Tony Moutaux** et **Frederic Diemunsch**. Une attention toute particulière pour **Christine Bole-Feysot, Philippe Gerber** et **Bernard Jost** : merci pour vos compétences et implications plus que concrètes dans ces travaux, mais surtout pour votre gentillesse et disponibilité à toute épreuve !

Merci **à tous les membres passés et présents de l'équipe de Génetique Humaine**…

…et plus particulièrement tout d'abord à **Chantal**, **Christine**, **Laurence** et **Solange** pour les nombreux coups de pouce (bactéries compétentes, anticorps, sauces, Taq…). Un clin d'œil à nos belles envolées lyriques en alsacien avec toi **Christine**… dess werd ich nit morge vergesse, dü ! (je parle mais j'écris en phonétique !). Merci à toute la fine et un tout ptit peu plus jeune équipe qui pousse etarrive derrière et plus particulièrement à **Gretta** et à **Anne So** (la vraie) pour leur gentillesse et à **Myriam** pour… disons son naturel ;o). Bon courage pour 2006, l'année du sport et de la « chaise »… au plaisir de se boire une bière un de ces 4.
Ensuite, à **Yvon:** merci d'avoir été disponible à chaque fois que j'ai eu besoin de tes connaissances et tes bons conseils scientifiques et techniques. Merci enfin à **Mims** pour toutes ses qualités, mais surtout, entre autres, ses très grandes qualités de gentillesse et prévenance.

Merci à toutes **les personnes avec qui j'ai travaillé directement** et qui ont donc participé à ce travail :
Merci à l'ensemble de nos collaborateurs sans qui ce travail n'aurait pu être réalisé, **Alain Aurias** et **Elisabeth Brambilla.** Un grand merci à **Danièle Muller** et **Joseph Abecassis,** du Centre Paul Strauss, collaborateurs efficaces et sympathiques de longue durée.

Merci à **toutes les personnes qui sont passées par le labo 3045** et avec qui on a, en général, travaillé de manière courte, mais très efficace:
Un clin d'œil à **Marc** pour ses extractions de BAC et surtout ses annotations de tubes (arrrrrrggghh) ;o). Merci à **Laetitia** pour sa sympathie et sa motivation. Merci à **Gaétan** pour sa rigueur et le travail accompli (et le ptit mais important ¼ de check des analyses qPCR tout récemment !). Merci à **Fabrice**… le temps de la fac est loin mais je suis sincèrement très heureux qu'on se soit un peu retrouvés ces derniers temps. Merci à **Monika** et merci à **Julien** pour avoir plus que bien lancé l'étude sur les carcinomes du poumon. Merci à **Matthias** pour sa « fraîcheur » et à **Nedjoua** pour son travail sur les MFH. Merci à **Céline** pour sa bonne humeur et son stage sur les anticorps anti-cycline L… et surtout pour sa double compétence infirmière/porteuse de matériel alors que je me déplaçais en béquilles ! ;o). Merci à **Caroline** pour son

récent travail de DEA mais aussi pour sa gentillesse. Merci à la « toute petite dernière » arrivée, **Christelle** (de carnold gal) pour sa sympathie et simplicité et bon courage pour la plate-forme, fais nous bien tourner ça ! Merci surtout de n'avoir pas craqué avec le live de Noir Dés en boucle ! Plein de bonheur et de bons horoscopes pour la suite ! Merci à toi **Sam** pour ton amitié, ton goût pour la coinche, et ton stage même s'il fut un peu court au final mais bon… on a quand même réussi ensemble à lancer la dernière ligne droite de ma thèse, à tenir les délais et faire tout ce qui était prévu… et en plus ça t'a plu, cool! Merci **Eric** pour ton aide technique très précieuse durant ces derniers mois de thèse. Merci **Soraya** pour la courte mais très plaisante (en plus d'efficace !) période de travail que nous avons déjà pu partager. Je suis très heureux d'avoir pu ressentir ta motivation, rigueur et ton intérêt pour la science parce que ça m'a remis la pêche à un moment où j'en avais besoin et te souhaite donc plein de réussite et du bonheur pour la suite ! Votre aide à tous les trois a été vraiment capitale pour que je puisse boucler un maximum de projets presque à temps… donc, un vrai très grand merci à tous les 3 pour tout !!!

Finalement, un merci tout spécial à un **membre passé mais encore présent mais presque passé quand même de l'équipe** de génétique humaine … re-… mais surtout et avant tout mon **coloc'… Pixote**, dit Vincent ou CarnacMan à l'occase. Pixxxote !! Merci, et pas que pour le soutien ces derniers temps et surtout plein de courage pour bien redémarrer. C'est ta chance. Là bas. Il y a du thym, de la bruyère. Tout est neuf et tout est sauvage. Et tout est possible à ton âge. Tout mais pas l'indifférence. Si tu as la force et la foi. Je ne vous parlerai pas d'elle. Pas toi. Compte pas sur moi. J'accepterai la douleur. Que les vents te mènent où d'autres âmes plus belles sauront t'aimer mieux que nous puisque l'on ne peut t'aimer plus. Je crois que j'ai trop veiller tard. Fatiguéé… voilà c'était ma ptite dédicace... puisque tu pars ! Même si t'en as rien à secouer, je crois fortement en toi alors tache de pas me décevoir… j'ai parié des ronds.

Merci pour **toutes les très très belles rencontres** que cette thèse m'a permis de faire **dans le labo 3045** et qu'on a ensemble bien vite emmené au-delà du professionnel :
- et tout d'abord merci **aux deux compères de la première heure**, Krishna et Richard, exilés depuis bien trop longtemps du labo:
Krishna, je passe très rapidement sur la qualité irréprochable de ton travail en toutes circonstances mais ce pour te dire que je te remercie très sincèrement pour tout ce que tu m'as appris et as fait pour moi. Et surtout pour l'ami que tu es, tu sais combien je t'adore ! Bonne continuation à Londres ou ailleurs et qu'on arrive à se capter un peu plus ! May the force of winds and zouklove stay in you forever !
Richard, ben je crois qu'on se connaît suffisamment tous les 2 pour pouvoir t'annoncer que je vais juste te cirer les pompes vite fait bien fait, simplement parce que je sais qu'il y a d'autres personnes qui liront ce paragraphe donc j'y suis un peu obligé ☺… Je voudrai donc d'abord te remercier au niveau boulot… tout simplement parce que si tu n'étais pas « passé » avant moi, je n'aurai simplement certainement pas pu faire tout ce que j'ai fait. Loin de là. Merci aussi d'avoir été présent et à l'écoute ce fameux soir de résultats de DEA où j'ai finalement décidé de pas tout envoyer péter tout de suite et de faire cette thèse. Sans doute un tournant. Les deux années partagées ensemble au labo ont été plus que sympas. Une spéciale dédicace pour une de mes (nos même je crois) blagues préférées (sic) : merci aux puces pangénomes in France qui devaient arriver bien avant la fin de ta thèse et qui sont finalement arrivées tout juste avant la fin de la mienne (re-sic)… Tout aussi sérieusement, merci de partager fortement certaines convictions sur ce boulot. Une dédicace aux scientifiques qui mélangent à notre goût un peu trop intérêts personnels, science et politique ? No way ! Enfin, le plus important pour moi, un vrai merci très perso pour la belle amitié qu'on a pu nouer à Strasbourg et même si on est plus éloigné aujourd'hui, je voulais te dire que le ptit jaune est aussi bon sur votre gazon anglais que sur votre ex-terrasse strasbourgeoise… gageons de voir quel goût il aura sur les bord du Léman… rendez vous est pris donc ! Une très très grosse bise à **Aude** au passage.

- merci ensuite **à mes voisins directs de labo** qui ont grandement contribué à améliorer mon environnement quotidien rien que par leur présence (enfin ça, ça dépend des jours)… et qui sont également devenus de très grands amis :
Merci à toi **Fabrice**… dit Faby ou encore le Fab, pour tes extraordinaires compétences de bioch et de bio structurale… nan, ça c'était pour rire. J'ai vraiment beaucoup apprécié et continue heureusement à pouvoir le faire, nos nombreux échanges de ces dernières années et en particulier les apéros, discussions, bouffes nocturnes ou week-ends en tous genres… Longue vie aux sound-systems dans les labos, aux fraises-planètes et à tes trop nombreuses dents ;o) ... mois aussi je t'M. A bientôt pour un nouveau pèlerinage en terre sainte ?

La meilleure (ouah !!!!) des 3045 (oooooh) pour la fin (haaaaaaaannnnnnnnn lala… dis moi pas ☺) … je te remercie **Marie** pour ces années passées côte à côte, ou plutôt dos à dos (enfin ces derniers temps, je préfère ça que ventrre à ventre ;o). Merci, d'abord parce que tu m'as appris beaucoup beaucoup… sur moi, sur les autres et aussi sur moi et les autres… Merci, ensuite, pour toutes les fois où tu m'as soutenu, voire simplement tenu, moralement. Merci enfin d'être l'amie que tu es. In case you'd encounter any trouble water in the future… Bon courage pour les mois/jours qu'il te reste à être un peu grosse et surtout ne bouge pas, c'est un ordre ;o) !

Merci à toi **Schkoumy™** (et oui !... c'est moi qui l'a trouvé… et c'est comme ça que ça s'écrit !)… dit Schkoum', Twix, Raie-man, Café ? ou encore Stef. Même si on a mis un peu de temps au départ, je crois, et de l'avis de tout le monde, que les 2 Twix se sont bien trouvés dans une période où nos vies respectives changeaient quand même pas mal. Bref, merci pour tout ce qu'on a vécu jusqu'à présent: les apéros partagés sur les terrasses ou à domicile jusqu'à des pas d'heure, les fou-rires, les breakages, tes super bouffes, ~~les nanas~~ (ah non… ça on a pas fait encore), la piscine, les coinches, le prix de la base et autres débats acharnés…. De toutes façons, ne t'inquiètes pas, tu sais bien que « Demain…. Demain, il fera beau… sur le chemin… Manana…. Manana, es sera bien en el camino » ©Manu, Barcelone, Novembre 2004… et je rajouterais : « ok… mais con la ballelleta en el camino alors ».

Gégé, je te remercie du fond du cœur : je suis vraiment très heureux qu'on ait partagé nos vies pour faire un bon bout de chemin et grandir ensemble. On a beaucoup appris tous les deux et j'espère sincèrement qu'on ne se perdra pas trop à l'avenir.

Merci à **tous les potes** avec qui on a partagé de tellement bons moments durant ces années. Un clin d'œil spécialement amusé à toi **Lolo**… pour m'avoir mis une pression extérieure et continuelle mais salutaire pour que je « soigne le cancer »… ;o). Et à toi **François** pas que pour les quelques dépannages de ces dernières semaines.

Merci enfin à **Félix Gray** et **Didier Barbelivien…** et **à toutes les filles qu'on a aimées avant**.

Un énorme Merci **à l'ensemble de ma famille**… bonne nouvelle, ça y est presque, je vais bientôt plus être étudiant et ptêt même que je vais avoir un vrai « travail »… Plus spécialement, des grosses bises à **ma grand-mère**, **mes parents** et **mes deux ptits frangins** pour leur soutien et leur présence depuis que je suis loin. Pour l'heureux **Papa**, désolé d'avance mais je n'ai pas pu résister… comme je te l'ai déjà dit, docteur c'est sûr c'est pas ingénieur… mais c'est pas mal aussi ;o) (oh ça va c'était pour rire !). **Benj** et **Lucien**, que du bonheur et le meilleur pour tout ce qui va venir !

Merci à **Carole**… pour Tout., mais surtout d'avoir accepté de ne partager que des rares bouts de week-end ces derniers mois.

Merci à celles et ceux que j'aurai bien involontairement oubliés…

TABLE OF CONTENTS

Thèse présentée pour obtenir le grade de

Docteur de l'Université Louis Pasteur

Strasbourg I

Discipline : Sciences du Vivant

par Thomas HUSSENET

CARTOGRAPHIE HAUTE-RESOLUTION DES AMPLIFICATIONS DE LA REGION 3q25-qter pour L'IDENTIFICATION D'ONCOGENES DANS DIFFERENTS TYPES TUMORAUX : APPLICATIONS A DES FIBRO-HISTIOCYTOMES MALINS ET DES CARCINOMES EPIDERMOÏDES (POUMON ET VOIES AERO-DIGESTIVES SUPERIEURES)

Soutenue publiquement le 12 décembre 2005 devant le jury composé de :

Docteur Stanislas du Manoir	Directeur de Thèse
Professeur Serge Potier	Rapporteur interne
Professeur Peter Lichter	Rapporteur externe
Professeur Norbert Arnold	Rapporteur externe
Docteur Bohdan Wasylyk	Examinateur
Professeur Jean Louis Mandel	Examinateur

Unité de Recherche :
Institut de Génétique et de Biologie Moléculaire et Cellulaire
INSERM U596 / CNRS - ULP UMR7104

RESUME DE LA THESE EN FRANCAIS

I- INTRODUCTION

Cartographie d'amplifications géniques pour l'identification d'oncogènes

Parmi différents mécanismes, **l'amplification génique est un moyen d'activation d'oncogènes, permettant leur surexpression par effet « gène dosage »**. Ainsi, les produits (ARN et protéine) d'un oncogène amplifié sont augmentés dans les cellules contenant un nombre accru de copies du gène, si celui-ci n'est pas soumis à d'autres mécanismes de régulation.

Les amplifications géniques sont communes dans les génomes tumoraux. La mécanistique de la formation de ces amplifications, encore mal connue, repose essentiellement sur des modèles. La structure d'une amplification génique, ou « amplicon », peut être intra-chromosomique ou extra-chromosomique. Ces deux types de structures sont nommées selon l'apparence des chromosomes impliqués lors d'un examen caryotypique: dans le premier cas, les amplicons sont appelés des « régions uniformément marquées » (ou « HSR, pour Homogeneously Staining Regions ») et sont inclus dans un chromosome de la cellule; dans le second, les amplicons prennent la forme de petits chromosomes acentriques surnuméraires, appelés chromosomes « double minute » (ou « DM », leur nom provenant de leur forme et petite taille). Les HSR renferment généralement un nombre élevé de copies du locus amplifié, répétées côte à côte sur un même chromosome; à l'inverse, chaque chromosome DM renferme généralement un nombre réduit de copies du locus amplifié mais de nombreux chromosomes DM sont présents par cellule.

La cartographie fine des amplifications récurrentes dans plusieurs tumeurs, suivie de la mise en évidence de régions consensus entre ces aberrations constitue un raccourci opératoire important vers l'identification de nouveaux oncogènes. Cette démarche a été appliquée à maintes reprises avec succès et à titre d'exemples, le gène *N-myc* a ainsi été identifié il y a 22 ans, en étudiant des neuroblastomes (lignées cellulaires et tumeurs). Plus récemment, le gène *mdm2* a lui été isolé à partir d'une seule amplification génique dans une lignée cellulaire. Cette amplification, prenant la forme de chromosomes DM, avait préalablement permis la transformation spontanée de la lignée cellulaire d'origine et ainsi conféré un avantage sélectif aux cellules contenant les chromosomes DM. Les caractères oncogéniques respectifs de chacun de ces gènes n'ont été effectivement démontrés qu'après leur identification, sur la base de leurs caractérisations fonctionnelles.

L'une des thématiques du groupe de Cytogénétique Moléculaire à l'IGBMC animé par Stanislas du Manoir est d'identifier des oncogènes selon cette démarche et concerne plus particulièrement les aberrations du bras long du chromosome 3 qui sont l'un des évènements cytogénétiques les plus fréquents dans les carcinomes épidermoïdes. La région 3q25-qter est en particulier fréquemment le site d'amplifications dans ces tumeurs. Au début de ma thèse, la principale méthode de cartographie des aberrations chromosomiques utilisée dans le groupe était l'Hybridation Génomique Comparative sur chromosomes (faible résolution : une ou deux bandes cytogénétiques, soit environ 10 à 15 Mb). Dans le cadre du développement de la plate-forme des puces à ADN de l'IGBMC, notre équipe a ensuite commencé à développer une évolution de cette technique, **l'Hybridation Génomique Comparative sur puces à ADN, ou « CGH-array »** qui **permet d'effectuer un profilage à haute résolution des aberrations chromosomiques** dans un génome d'intérêt (par exemple tumoral, en comparaison à un génome normal). Elle repose sur l'utilisation de puces à ADN composées de clones d'ADN génomique, tels que les clones BACs qui ont par ailleurs servi au séquençage du génome humain. **Grâce aux bases de données de séquences génomiques, il est ensuite directement possible d'obtenir le contenu en gènes du locus concerné. La délimitation d'une région consensus d'amplification permet ensuite de restreindre la taille de la région considérée et ainsi le nombre de gènes. La discrimination entre ces gènes candidats s'effectue finalement par l'étude des conséquences transcriptionnelles des amplifications sur l'expression des oncogènes candidats** (démonstration de l'effet gène dosage) **et leur caractérisation fonctionnelle.**

La CGH-array était encore en émergence lorsque j'ai rejoint le laboratoire et j'ai participé à de nombreuses mises au point et évolutions de la technique pour pouvoir mener à bien l'analyse des aberrations chromosomiques dans les différents types tumoraux présentés dans ma thèse. Ainsi, de nombreuses versions successives de notre puce dédiée au chromosome 3 ont été construites avec des

résolutions croissantes et mes travaux au sein du groupe ont contribué aux mises au point et standardisation de la technique. Ces développements ont conduit à la mise en place récente, au sein de notre groupe, d'une plateforme nationale qui est désormais opérationnelle. Elle est employée pour des projets de cartographie par CGH-array du programme « Cartes d'Identité des Tumeurs » de La Ligue Nationale Contre le Cancer d'une part, et impliquée dans des études émanant de la Cancéropôle Grand-Est d'autre part.

Les différents projets que j'ai mené au sein de notre groupe de Cytogénétique Moléculaire ont principalement consisté à analyser les aberrations du chromosome 3 pour des carcinomes épidermoïdes, (du poumon et des voies aéro-digestives supérieures), dans le but d'identifier des oncogènes. Un projet secondaire a été consacré à l'étude des aberrations du chromosome 3 dans des sarcomes des tissus mous (Fibro-Histiocytomes Malins ou MFH).

II- Aberrations du chromosome 3 dans les fibro-histiocytomes malins

<u>Le gène *hsa-miR-28*, codant pour un microARN, comme oncogène candidat.</u>

Travaux effectués en **collaboration avec Alain Aurias, Institut Curie, Paris.**

Nous avons au départ entrepris d'analyser une lignée cellulaire, COMA, établie par nos collaborateurs. La caractérisation cytogénétique initiale de la tumeur, un sarcome de tissu mou diagnostiqué comme Fibro-Histiocytome Malin (MFH), et de la lignée cellulaire dérivée, COMA, a montré la présence de deux amplifications de très haut niveau, sous forme de chromosomes DM, impliquant les régions chromosomiques 3q26-qter et 8q24, locus de l'oncogène *c-myc*. Par CGH-array, nous avons cartographié l'amplicon du locus 3q28. Cette amplification a été confirmée par FISH (Hybridation In Situ en Fluorescence) sur des noyaux et des étalements de chromosomes métaphasiques de la lignée: deux populations exclusives de chromosomes DM coexistent dans la lignée, l'une contenant le gène *c-myc*, l'autre le locus 3q28. Une cartographie plus précise de l'amplicon du locus 3q28 dans la lignée cellulaire COMA a été ensuite réalisée par PCR semi-quantitative permettant de réduire sa taille à environ 1Mb et ses bordures à des intervalles de 62 et 150 kb. Une analyse bioinformatique de la bordure télomérique révèle un enrichissement significatif en séquences répétées dans le génome, comme précédemment décrit pour d'autres amplicons et qui pourraient être impliquée dans la genèse de cet amplicon.

Notre étude a ensuite été étendue à 8 sarcomes primaires indifférenciés avec des gains ou amplifications de la région 3q26-qter (pré-criblage réalisé par nos collaborateurs). Par CGH-array, nous avons délimité une **région consensus d'amplifications entre deux de ces tumeurs primaires et la lignée COMA**. L'analyse bioinformatique de cette région révèle la **présence d'au moins trois gènes** (*LPP*, *FLJ42393*, un gène codant pour un microARN, *hsa-miR-28*), ainsi que plusieurs ESTs ou « morceaux de séquences exprimées ».

Pour étudier les conséquences transcriptionelles de l'amplification du locus 3q28, nous avons comparé le niveau d'expression de ces gènes dans la lignée COMA et dans une autre lignée tumorale dérivée d'un MFH mais ne contenant pas d'altération sur le bras long du chromosome 3 (l'origine histologique des MFH étant inconnue, aucun tissu normal correspondant n'est disponible). Par RT-PCR, nous avons montré que les gènes *LPP* et *FLJ42393* ne sont pas ou très peu surexprimés alors que **deux ESTs**, (AI338598 et BX118304), **sont fortement surexprimées mais sont toutefois très peu caractérisées *in silico***. Finalement, le gène *hsa-miR-28*, est exprimé à des niveaux similaires dans les deux lignées. Nous n'avons ainsi pas pu formellement identifier le gène cible de l'amplification du locus 3q28 présente dans la lignée cellulaire COMA ainsi que dans deux tumeurs primaires.

Un manuscrit a été rédigé (en cours de soumission à la revue Cancer Genetics & Cytogenetics, publication 1).

Toutefois, nous ne pouvons pas exclure que l'expression d'un des gènes candidats étudiés ci-dessus puisse être régulée par un mécanisme alternatif à l'amplification génique dans la lignée contrôle utilisée. De plus, nous avons observé que **l'expression de miR-28 est nettement plus élevée dans ces deux lignées cellulaires de MFH en comparaison avec d'autres lignées cellulaires tumorales d'origine histologique différente** (type épithélial). L'importance physiopathologique de cette nouvelle classe de gènes codant pour les microARNs commence seulement à être caractérisée.

Pour la première fois, le statut oncogénique d'un gène polycistronique codant pour plusieurs microARNs, amplifié et surexprimé de manière récurrente dans des leucémies (avec un mauvais pronostic associé pour les patients), a été proposé dans une publication très récente parue dans la revue Nature en Juin 2005. Au vu de l'ensemble des données récentes suggestives de l'importance de cette nouvelle classe de gènes codant pour les microARNs, nous avons été hautement intéressés par la caractérisation fondamentale et fonctionnelle de miR-28 et son lien potentiel au cancer.

Ainsi, **en inhibant spécifiquement l'expression de miR-28, la prolifération des lignées cellulaires tumorales de type MFH est ralentie** alors que celle des lignées tumorales de type épithélial est non affectée ou légèrement stimulée. Ce ralentissement de la prolifération pourrait résulter d'apoptose. Les microARNs jouent un rôle physiologique pléiotropique: chacun régule le niveau d'expression de plusieurs protéines en s'appariant aux ARNm correspondants et en contrôlant leur traduction (inhibition de la traduction ou dégradation de ces ARNm). Nous avons donc recherché des cibles potentielles de miR-28 à l'aide d'outils bioinformatiques récemment développés et dédiés à la prédiction de cibles des microARNs. Nous avons sélectionné deux protéines, qui, de manière intéressante, appartiennent à la même famille des facteurs de transcription E2Fs/DP. Ces facteurs sont essentiels au contrôle du cycle cellulaire, ont un rôle dans l'apoptose et sont altérés dans plusieurs cancers. Nous venons de montrer que **le niveau d'expression de la protéine DP2 est effectivement régulé par miR-28**. Ce résultat, en soi, est d'importance étant donné le peu de cibles identifiées jusqu'à présent pour les microARNs. Nous testons d'autres protéines de la même famille E2Fs/DP. Il est désormais acquis que ***hsa-miR-28*** **est un oncogène candidat dans les MFH**. Des études complémentaires de son expression dans une série de sarcomes détermineront si cet oncogène candidat peut être utilisé comme un marqueur diagnostique spécifique d'un sous-groupe de sarcomes.
Cette étude consacrée au microARN miR-28 sera très prochainement finalisée par la rédaction et la soumission d'un manuscrit.

III- Aberrations du chromosome 3 dans les carcinomes épidermoïdes

a. <u>Carcinomes épidermoïdes du poumon</u>

Travaux effectués en **collaboration avec Elisabeth Brambilla, Institut A. Bonniot, Grenoble.**

Dans ces tumeurs, de nombreux travaux ont montré la présence de gains et d'amplifications dans la région terminale du bras long du chromosome 3 (3q25-qter) mais les méthodes employées n'étaient pas suffisament résolutive pour établir une cartographie fine de ces évènements. En utilisant la CGH-array, nous avons criblé les aberrations du chromosome 3 dans 30 tumeurs primaires de stades avancés (tumeurs T3 et T4). Six tumeurs présentent des amplifications de haut niveau du locus 3q26.3. Cette haute **fréquence d'amplifications (20%),** parmi les plus élevées de celles décrites pour d'autres oncogènes connus, indique l'importance fonctionnelle de cette aberration pour le développement ou la progression de ce type tumoral.

Nous avons alors produit une nouvelle version de notre puce avec une couverture globale du chromosome 3 nettement augmentée (passage de 240 à 350 clones), et une représentation exhaustive du locus d'intérêt 3q26.3 (un contig représentant plus de 5 Mb). Cet outil a permis de réaliser une cartographie fine des amplicons et de leurs bordures dans les 6 tumeurs concernées. Finalement, **la région consensus d'amplifications dans ces tumeurs couvre environ 2 Mb et ne contient qu'un nombre limité de gènes** (dix gènes appartenant aux catégories Refseq et ARNm Genbank).

Nous avons analysé leurs niveaux d'expression dans les 6 tumeurs avec amplification en comparaison avec des tumeurs sans amplification du locus 3q26.3. **Un effet gène dosage est observé pour la plupart des gènes** (9/10), ne permettant pas d'isoler parmi ceux-ci un gène « cible » préférentiellement activé. Certains de ces gènes pourraient être amplifiés et surexprimés de par leur position sur le génome mais sans être pour autant liés au phénotype tumoral, comme l'ont déjà montré plusieurs études. Dès lors, une analyse fonctionnelle dans le contexte de la carcinogenèse s'impose.

Pour effectuer une **discrimination entre ces dix gènes candidats**, nous avons donc dû analyser les conséquences fonctionnelles de ces aberrations dans les cellules tumorales à l'aide d'un modèle cellulaire. Les cellules NIH-3T3, classiquement utilisées pour des tests fonctionnels de

détermination de rôle oncogénique *in vitro* ont été retenues pour **étudier les effets de la surexpression de ces gènes sur des traits du phénotype tumoral (prolifération, invasion et résistance à l'apoptose)**. Ainsi, nous avons sous-cloné les ADNc codés par ces dix gènes dans un même vecteur d'expression et effectué des travaux de mises au point des différents tests. Nos résultats préliminaires indiquent que plusieurs gènes semblent importants. Ces tests nous permettront de déterminer effectivement le ou les oncogènes du locus 3q26.3, activé de manière récurrente par amplification du locus 3q26.3 dans les carcinomes épidermoïdes du poumon et contribuant au phénotype tumoral.

Cette étude originale constitue une avancée importante: pour la première fois, sur la base de preuves expérimentales formelles obtenues en suivant une démarche exhaustive, nous sommes en mesure d'identifier un ou plusieurs oncogènes cible(s) des amplifications du locus 3q26.3 dans des cancers épidermoïdes.

L'ensemble de ce travail sur les cancers du poumon est actuellement en cours de finalisation et un manuscrit, d'ores et déjà avancé, devrait être soumis pour publication dès les tests fonctionnels achevés.

.

b. Carcinomes épidermoïdes des Voies Aéro-Digestives Supérieures (VADS)

Travaux effectués en **collaboration avec le groupe de Joseph Abecassis, Centre Paul Strauss, et B. Wasylyk, IGBMC, Strasbourg.**

i- Identification de la cycline L1 comme oncogène potentiel dans les cancers des VADS

Nous avons initialement montré **l'amplification et la surexpression de la cycline L1, gène localisé en 3q25.3, dans une lignée cellulaire VADS, ainsi que sa surexpression dans une série de 20 tumeurs**. Au cours de cette étude, j'ai d'abord eu la responsabilité de l'analyse bioinformatique exhaustive de la région d'intérêt et de son annotation. Ceci a alors notamment permis d'identifier le gène de la cycline L1 (*CCNL1*) comme gène localisé dans la région à partir de la séquence du génome et son assemblage par Celera (la séquence du génome humain du consortium public, habituellement utilisée, était alors incomplète dans cette région). J'ai eu également la charge de l'analyse de l'expression des gènes par puces à ADN et par RT-PCR quantitative.

Ces travaux ont été publiés dans Cancer Research fin 2002 (publication n°2).

Nous avons ensuite étendu cette étude *in vivo* dans une large série de tumeurs des VADS, en déterminant par PCR quantitative le nombre de copies du gène (pour 35 tumeurs) et son niveau d'expression (pour 96 tumeurs). **Le nombre de copies du gène *CCNL1* est augmenté dans environ 30% des tumeurs; de plus, celui-ci est surexprimé dans 57% des tumeurs**. Aucune corrélation avec les paramètres cliniques n'a toutefois été observée. Nous avons développé à l'IGBMC un anticorps monoclonal spécifique de la cycline L1 (isoforme α) et avons validé son utilisation pour la détection de la protéine sur des coupes de tumeurs des VADS par immunohistochimie.

Ces travaux ont été soumis récemment pour publication à British Journal of Cancer (publication n°3).

Cet anticorps devrait permettre d'explorer l'expression de la cycline L1 pour un grand nombre de tumeurs par immunohistochimie.

Par ailleurs, nous avons récemment établi des **modèles cellulaires *in vitro* d'expression inductible pour les deux principaux transcrits alternatifs (α et β) de la cycline L1**. Ces clones, transfectés de manière stable, ont été sélectionnés, sont en cours d'amplification et vont permettre de caractériser le rôle respectif de chacune des isoformes protéiques de la cycline L1 dans le cycle cellulaire, mais aussi leurs potentiels oncogéniques.

ii- Aberrations du chromosome 3 dans les carcinomes épidermoïdes des VADS

Les carcinomes épidermoïdes des Voies Aéro-Digestives Supérieures et du poumon ont en commun leur origine histologique, tous deux étant dérivés d'un épithélium desquamant. Pour définir si les régions génomiques du chromosome 3 amplifiées sont identiques dans ces deux types tumoraux et impliqueraient les mêmes oncogènes, un criblage de 40 tumeurs VADS de stade avancé (tumeurs T3 et T4) a été réalisé. Cette étude, réalisée par CGH-array, fait apparaître **dans les tumeurs des VADS, 2 locus de la région 3q25-qter** qui **sont fréquemment altérés** : le premier, en **3q26. 3**, est **amplifié dans 12,5 % des tumeurs et recouvre la région consensus définie pour les carcinomes épidermoïdes du poumon** mais de manière plus large; de manière intéressante, le second site fréquemment altéré est la région **3q25**, locus du gène *CCNL1* que nous avons précédemment identifié comme oncogène potentiel dans les tumeurs des VADS.

Ce travail devrait donner lieu également à la rédaction d'un manuscrit.

IV- CONCLUSIONS

Ce travail de thèse illustre la puissance de la stratégie de la cartographie fine des amplifications géniques pour l'identification de nouveaux oncogènes, ayant déjà permis d'identifier deux oncogènes candidats, les gènes *hsa-miR28* et *CCNL1*, respectivement dans des fibro-histiocytomes malins et des carcinomes épidermoïdes des VADS. De plus, nous sommes sur le point d'isoler et de caractériser fonctionnellement un ou plusieurs oncogènes majeurs des carcinomes épidermoïdes (du poumon et des VADS), localisés dans la région 3q26.3, ce locus étant très fréquemment amplifié dans ces tumeurs.

ABSTRACT

Cancer cells are characterized by a certain number of hallmarks. Among these, genetic instability is admitted to be a near universal one. Consequently, most of tumour cells presents genetic alterations and particularly at the chromosomal level. Actually, interstitial deletions, low-level copy number gains and high-level copy number amplifications of genomic regions are recurrently observed in tumour genomes, and produce deregulations in the expression of tumour suppressor genes and oncogenes. We used a positional cloning strategy by mapping of chromosomal aberrations to define common genomic intervals that are altered in different tumours and identify candidate oncogenes. This thesis work specifically applied this strategy to chromosome 3 aberrations in different tumour types and particularly to identify new candidate oncogenes located at 3q, amplified and consequently over-expressed.

A first part was dedicated to the study of some soft-tissue sarcomas, **Malignant Fibrous Histiocytomas** (MFH). In a first approach, we characterized an amplicon at 3q28 in a MFH cell line. This region was also found to be amplified in two other MFH primary tumours but no obvious candidate oncogene was isolated (Hussenet *et al*., submitted). From this region, we finally identified a microRNA encoding gene, *hsa-miR-28* representing an oncogene candidate: its inhibition indeed slows down the proliferation of 2 MFH cell lines. We further identified three targets of miR-28 (DP2, E2F6 and CDC14A): all play roles in the regulation of the cell cycle and interestingly the *hsa-miR-28* gene expression itself seems to be cell cycle regulated (Hussenet *et al.*, in preparation).

A second part was dedicated to **Squamous Cell Carcinoma** (SCC) of two localizations (lung & head and neck).

We analyzed chromosome 3 aberrations in a series of 25 **lung SCC** using array CGH. This work allowed us to delineate three regions of deletions at 3p, low-level copy number gain of the 3q26-qter region in a majority of tumours, as well as a common region of high-level amplifications at 3q26.3 for 20% of the tumours. Further high-resolution mapping of these amplifications using array CGH with an array providing tiling coverage of the 3q26.3 locus pinpoint a consensus region that is approximately 2 Mb in size. Analyses of the transcriptional consequences of these high-level amplifications were carried out for 9 genes of the region. Most of them are recurrently over-expressed but two, *SOX2* and *SOX2OT*, likely represent the 3q26.3 amplification driver genes in these tumours. Ongoing *in vitro* functional assays will assess the effect of over-expression of candidate oncogenes located in the consensus region of high-level amplifications at 3q26.3 (Hussenet *et al.*, in preparation).

Our positional cloning strategy also enabled us to isolate the *cyclin L1* (*CCNL1*) gene at 3q25.3 as a candidate oncogene, amplified and over-expressed in a **head and neck SCC** (HNSCC) cell line, as well as over-expressed in HNSCC tumours (Redon *et al.*, 2002). Further investigations of *CCNL1* gene alterations in a larger series of HNSCC revealed consistent low-level copy number gains and over-expression in these tumours (Muller *et al.*, submitted). Finally we also screened a series of 25 HNSCC for chromosome 3 aberrations using array CGH. We were able to delineate several recurrent sites of deletions at 3p, low-level gains and high-level amplifications at 3q. A common region of high-level amplifications in HNSCC is overlapping with the consensus region we have previously defined for lung SCC and suggests that similar genes at 3q26.3 may be involved in SCC pathogenesis independently of the localization.

In conclusion, this work illustrates the power of the positional cloning strategy applied to tumour genomes through the use of high resolution methods and tools to map the recurrent sites of genomic alterations. We indeed were able to isolate several oncogene candidate located at 3q that are amplified and consequently over-expressed in different tumour types. Further investigations are still needed to assess their oncogenic status as well as putative relationships with clinical parameters.

AUTHOR's list and personal contribution from Thomas Hussenet
for the not yet submitted works.
(written by Stanislas du Manoir, PhD supervisor)

Thomas was instrumental in these studies, others were short time students or technical people.

Lung SCC
Authors
Thomas Hussenet, Julien Exinger, Richard Redon, Bernard Jost, Beatrice Orsetti, Charles Theillet, Elisabeth Brambilla and Stanislas du Manoir.

Relative contributions :
Julien Exinger during its one year master's degree did the initial array CGH screen using the arrays we had prepared in a collective effort (Thomas Hussenet, Julien Exinger, Richard Redon and Bernard Jost for preparation and spotting of the arrays). Thomas prepared (with help from Beatrice Orsetti in Charles Theillet's lab, Monptellier, for the extraction of some BAC clones) and used the contig array for the high resolution mapping of 3q26.3 high level amplifications found in the first screen. Thomas had the full charge of the qPCR experiments (primers design and validation; RNA extractions, quantitative RT-PCR) Thomas initiated and managed the cloning/subcloning experiments of the cDNAs in a common mammalian expression vector, for which he received technical support from Eric Flatter.

Head and Neck SCC
Authors
Thomas Hussenet, Richard Redon, Bernard Jost, Gaetan Bour, Danièle Muller, Bohdan Wasylyk, Joseph Abecassis and Stanislas du Manoir.

Relative contributions :
Thomas realized the array CGH experiments as well as DNA chip preparation (see above).

miR-28 microRNA as an "onco-miR" candidate in MFH
Authors
Thomas Hussenet, Soraya Dali, Alain Aurias and Stanislas du Manoir.

Relative contributions :
This a project from Thomas, he really pushed it and was a driving force. Technical support from Soraya Dali with cell culture and western blots experiments.

Abbreviations

BAC *Bacterial Artificial Chromosome*

Bp *Base pairs*

CIN *Chromosomal INstability*

CGH *Comparative Genomic Hybridization*

DNA *DesoxyriboNucleic Acid*

dsDNA double stranded DNA

ESP *End Sequence Profiling*

HNSCC *Head and Neck Squamous Cell Carcinoma*

Kb *Kilobases*

MIN *Microsatellite INstability*

Mb *Megabases*

NIN *Nucleotide excision repair INstability*

PCR *Polymerase Chain Reaction*

qPCR *quantitative PCR*

RNA *Ribonucleic Acid*

RT-PCR *Reverse-Transcription PCR*

SCC *Squamous Cell Carcinoma*

YAC *Yeast Artificial Chromosome*

Part 1

A general introduction

on some aspects of

CARCINOGENESIS

-I- Cancer is a GENETIC disease: first evidences and concepts

In addition to the cited monographies, this part is mainly inspired from different sources (reviews and books): Nowell (1976); Nowell, Rowley and Knudson (1998); Cahill *et al.* (1999); Knudson (2001); Balmain (2001); Heim and Mitelman (Cancer Cytogenetics, 2nd edition, ©1995); The Metabolic & Molecular Bases of Inherited Disease, 8th edition (©2001).

a. A historical perspective:

a-1 First evidences from chromosomes observations

i- The German Precursors

Initial evidences accounting for the genetic origin of cancer came from early cytogenetic studies that were carried out more than a century ago. These first important observations and descriptions of phenotypic traits in neoplasms, in accordance with the technical means available at that time, have provided not only with the conceptual bases for the understanding of cancer, but also with the practical approaches needed to dissect cancer.

Chromosomal abnormalities in cancer cells were first reported at the end of the 19th century by David von Hansemann, a German pathologist who microscopically examined many epithelial tumour sections and noted frequent aberrant mitotic figures (von Hansemann, 1890).

Theodor Boveri, a German Biologist, originally bought new insights into the important roles of chromosomes for different physiopathological processes. Even if he was already well recognized by the scientific community (on the contrary to Mendel), Theodor Boveri has to be considered as the founder of cancer genetics.

His first studies were dedicated to microscopic observations in a horse nematode and led him to identify chromosomal individuality and to be among the precursors to the discovery of meiosis.

Still addressing the role of chromosomes, he then went on answering scientific questions using an experimental model, the fertilization of the sea urchin and was able to observe developmental defects associated with mitotic abnormalities. After Walter Sutton's observations that each chromosome is present twice in grasshoppers cells, and linking that to Mendel's laws of inheritance (at that time recently and independently rediscovered by DeVries, Correns and Von Tschernak), the chromosomal theory of genetic inheritance was established. Later on at the end of his life (Boveri, 1914), Boveri was able to translate all the recently accumulated knowledge about chromosomes and underlying genetic concepts into new fundamental concepts for cancer. In fact, by linking genetic changes and tumour growth, he suggested the somatic mutation theory of cancer. The "somatic mutation" term was only applied two years later (Tyzzer, 1916).

Most importantly, Boveri first suggested the genetic basis underlying cancer in that tumorigenesis could be due to aneuploïdy and secondly, in that tumors could clonally originate from a single cell. He implied thirdly that genetic instability could be an essential determinant in tumor cells and finally, that specific chromosomal or submicroscopic changes could be relevant for cancer cells. He already intuited, at the chromosomal level, the more

recently proven concepts of oncogenes and tumour suppressor genes, cell cycle checkpoints and cancer predispositions.

Boveri's global work made of him a real pioneer in the field of genetics, both in terms of concepts and practical issues. Very remarkably, all the intuitive and deductive hypotheses he posited with respect to cancer genetics were in turn verified.

From Boveri's work to the emergence of the modern cytogenetics in the 50s, only little advances have been made towards the understanding of the genetic contributions to cancer. The most influential works were from O. Winge who defined the stemline concept predicted by Boveri (Winge, 1930) and from tumour transplantation studies in rodents, as well as from the study of radiation effects (Heim and Mitelman, 1995).

ii- The modern Cytogenetics era

a) The "pre-banding" era

From the beginning of the 1950s, technical improvements in both cell culture techniques and slide preparations paved the way to improved karyotype examinations of tumour cells. At this point, studied tumours were highly advanced cases and mostly malignant. These early studies already uncovered three specific traits of chromosomal abnormalities in tumours: a high level of genetic heterogeneity for those tumour cell populations, the proof of the formerly proposed stemline concept (most frequent chromosome constitution in a tumour cell population besides the sidelines that have deviating compositions), and the stemlines competition in response to environmental alterations resulting in labile chromosome equilibria. Notably, this period also led to the description of the correct chromosome number in man (Tijo and Levan, 1956).

A major breakthrough in cancer cytogenetics was then achieved by a collaborative effort of Nowell and Hungerford: they discovered a karyotypic marker in chronic myeloid leukaemia (CML) that was present in all cancer cells of nearly every typical CML they had examined. As the specific chromosome involved could not be identified, they followed the newly established nomenclature rules (from the First International Conference on Cytogenetics, 1960) and called it the "Philadelphia" (Ph) chromosome after the name of the city where this discovery had been done (Nowell and Hungerford, 1960).

b) The "banding revolution"

In the early 1970s, the chromosome banding staining were introduced and can be considered as a technical revolution in the field, allowing cancer cytogeneticists to look to tumour cells at sub-chromosomal levels. This mostly and importantly helped to convince numerous sceptics that chromosomal gains and losses observed in tumour cells were not only random events.

J. Rowley confirmed that the Ph chromosome was composed of chromosome 22 material in CML, but also noticed the presence of an additional band to the chromosome 9 in each examined karyotype. Thus she was able to identify the Ph chromosome as resulting from the translocation of a chromosome 22 segment to the chromosome 9 (Rowley, 1973). This translocation was finally characterized only 11 years later with cloning techniques (Groffen *et al.*, 1984) and shown to involve the *ABL* and *BCR* genes. Other translocations specifically associated with other tumour types were then described and characterized in leukemias,

lymphomas and sarcomas (Rowley, 1984). Most of the described translocations create chimeric transcription factors. Since then, the overall most common mutation found in cancer genes is translocation that creates chimeric genes (Futreal *et al.*, 2004).

By studying sarcomas induced by the Rous Sarcoma Virus (RSV) in rats, F. Mitelman showed a stepwise increase in chromosomes number (sequential gains of chromosomes 7, 13 and 12) in cancer cells that was typically associated with progressive loss of differentiation for the sarcoma cells, observed histologically and biochemically by decreased collagen production (Mitelman, 1974).

In less than a century, great progresses on the understanding of the genetic basis of cancer were made through cytogenetics. These first numerous and important observations led to elaborate new concepts for the understanding of how a tumour could initiate and progress.

a-2 Place of chromosomal aberrations in the Darwinian evolution of the cancer cell and clonal tumour cell populations.

As noted in the previous part, Boveri already predicted that tumours could arise from a single cell that would clonally expand. This was further defined as the stemline concept (Winge, 1930).

Shortly after the identification of the Ph chromosome in CML, Cole and Nowell suggested (1965) that tumour progression could be due to a sequential acquisition of multiple genetic changes to the initiated neoplastic clone. Of importance, they also suggested that genetic instability could facilitate this process. But it's only eleven years later, based in particular on accumulated cytogenetic observations in tumours and on the Darwinian selection principles he applied to cell populations, that P.C. Nowell was able to propose a detailed and argued model explaining the clonal evolution of tumour cell populations: the neoplastic progression results from stepwise selections of malignant cells presenting genetic instability (Nowell, 1976).

In this model, cancer initiates by genetic alterations in a single normal cell of origin which will then give rise to a progeny with genetic instability. From that instability, the daughter cells population will present acquired genetic variability along successive cell divisions. Among the progeny, a cell will further be sequentially selected and expand as its genetic variability has provided it with a selective advantage (i.e. growth advantage such as increased proliferation rate) allowing this neoplastic clone to overcome others. And this sequence of events will further be repeated: increase of genetic variability in the progeny due to genetic instability, which will facilitate the emergence of a new clone with a selective advantage possibly under new environmental pressures and that will clonally expand to the detriment of other cells.

In conclusion, tumour progression can be explained after initiation of a neoplastic cell by several waves of clonal expansions, each of these conferring cells with a new selective advantage towards a best fitted malignant phenotype (figure 1). Experimental confirmations of this pioneer conceptual thesis of Nowell had to wait until the emergence of molecular cancer genetics. This conceptual frame is now accepted by the scientific community.

In any case, it has now clearly been demonstrated and accepted that chromosomal changes in almost all human cancer types are non-random reflections of clonal transmission. Numerous specific chromosomes alterations have been described for a wide majority of tumour types. With respect to the model established by Nowell, chromosomal aberrations should be divided into two categories, depending on their phenotypic contribution to malignancy (Heim and Mitelman, 1995).

Primary chromosomal aberrations refer to the aberrations which are involved in the earliest stages of carcinogenesis and thus represent specific markers of given neoplasms. Their primary involvement doesn't preclude any previous somatic mutation not detectable cytogenetically and putatively causative of initiation.

On the contrary, secondary chromosomal aberrations refer to the abnormalities that even if non-random and clonal, are less specific, rarely found alone and as such are numerous in cells of advanced tumours.

Importantly, as genetic instability is a main contributor to carcinogenesis and cancer progression, tumour cells genomes have been described as highly variable with a lot of

apparently random and non clonal genetic aberrations. Such chromosomal aberrations have been referred to as cytogenetic noise.

Concerning tumour initiation, it is still unclear whether cancer initiation is a mono or polyclonal event. In other words, does cancer initiates in a single cell or begins by simultaneous transformation of multiple cells ? With respect to Nowell's model, this question is clearly open for both answers. There is no doubt that both hypotheses should turn out to be verified in different cases. It seems that for some cancers at least, initiation is polyclonal, but later stages of cancer progression may concern selection of oligo- or monoclonal cell populations.

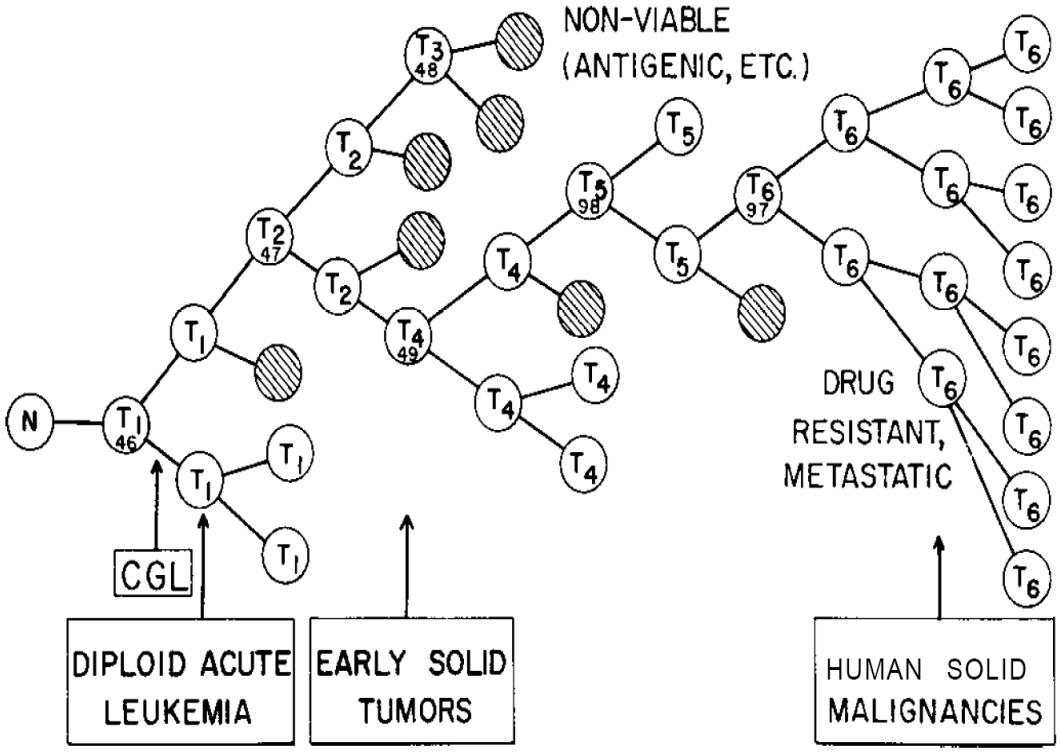


Figure 1. Model of tumour cell clonal evolution in cancers.
Taken from Nowell, Science 1976.

After acquisition of the malignant phenotype, a normal cell (N) is converted to a tumoural cell (T1). This cell's karyotype can be eventually composed of 46 chromosomes (as noted) or could already present variation of its chromosome numbers. Genetic instability in the progeny produces variants (T2 to T6) with different chromosomes numbers. Most of cells die (hatched circles) due to any disadvantage, as for example a too high genetic instability. In some cases, a cell variant will acquire an advantage through an additional chromosome for example as for T2, noted with 47 chromosomes. T2 will accordingly produce a progeny that will overcome the other cells. Tumour progression may thus be explained by such step-wise selections of cancer cells towards a best fitted malignant phenotype, partially determined by environnmental selective pressures.

a-3 Getting further insight from genes: discovery of Oncogenes and Tumour Suppressors

i- Oncogenes and proto-oncogenes

The oncogene term (onco = bulk or mass in Greek) was first introduced to describe viral genes with transforming activity among the viral genome (called "virogene"; Huebner and Todaro, 1969). The discovery of oncogenes came from studies of transforming retroviruses that are able to induce tumourigenesis in animal models. At that time, the hypothesis of tumours being induced by those viruses was one of the prominent hypotheses to explain carcinogenesis.

In particular, the Rous Sarcoma Virus, a Retrovirus capable of infecting and transforming avian cells, was extensively studied in the 60s-70s period. Several genetic studies showed that its *v-src* gene was responsible of its transforming activity in chicken but not necessary for viral replication (Martin, 2004). Trying to get a specific nucleic probe for a RSV variant, a cellular homologue of the *v-src* gene was isolated in the chicken genome and its sequence shown to be conserved in several other avian species (Stehelin *et al.*, 1976). The authors suggested the cellular origin of the viral oncogene. Notably, they anticipated the potential normal role in cell growth and development as well as the putative transforming activity upon alteration of the cellular homolog. It finally turned out that the *v-src* gene was indeed of cellular origin and had been captured by the retrovirus. It was then shown later that the *c-src* gene is indeed a cellular functional protein encoding gene, coding for a tyrosine kinase protein, and able to induce cellular transformation when altered by activating mutations or overexpression (Martin, 2004).

The term "proto-oncogenes" was applied to describe those cellular genes with transforming capacity upon alteration to distinguish them from the putative ones of viral origin (previously defined as "oncogenes"). It turned out that these retroviral oncogenes are in fact of cellular origin, and notably, this global discovery was rewarded by the Nobel Prize to H.E. Varmus and J.M. Bishop (Bishop, 1990; Varmus, 1990).

Following the same strategy as for *c-src* identification, several other proto-oncogenes were then characterized, such as for memory among now famous ones, the *c-erb* and *c-myb* (Roussel *et al*., 1979; Vennstrom *et al.*, 1982a), and the *c-myc* (Sheiness *et al.*, 1980; Vennstrom *et al.*, 1982b) proto-oncogenes. Since the cellular origin of the formerly defined oncogenes has been clearly established, proto-oncogenes are now commonly referred as to oncogenes.

Finally, one can define an oncogene as a gene that upon alteration of its normal function by any activating mechanism enhances the malignant phenotype.

Activation of oncogenes may be achieved through increase in their dosage (due to translocation, gene copy number amplification or any other mechanism), activating mutations or translocations.

ii- Tumour suppressor genes

Precursor work came from A. Knudson, trying to understand which relation could exist between hereditary and non hereditary forms of retinoblastoma. Through an epidemiological study of these childhood cancers he established a mathematical model of both forms of retinoblastoma cancer incidence (Knudson, 1971). Accordingly, he was able to postulate the later on so-called "two-hit" hypothesis: two events of mutations are needed to inactivate a tumour suppressor gene, or TSG.

In the hereditary form of retinoblastoma, the first mutation is inherited from the parent (as a carrier, or through a de novo mutation which is from far the most frequent) and the second is acquired early during development. In this case, most affected children develop bilateral affections. In the non hereditary form, the two mutations are somatically acquired, and most cases are unilateral.

Knudson predictions were verified: the involved locus, on chromosome 13q14 was shown to be a deleted region in retinoblastoma, (Francke and Kung, 1976; Knudson *et al.*, 1976) which was then further mapped by Loss Of Heterozygosity (LOH, Cavenee *et al.*, 1983). The involved first tumour suppressor gene, ***RB***, was finally cloned (Friend *et al.*, 1986). Interestingly, we know now that the encoded pRb protein belongs to a cellular central proliferation pathway which is altered at several levels in a wide majority of tumours.

In the meantime, the p53 protein was identified in a complex with the SV40 large T antigen (Linzer and Levine, 1979). From that and other studies, the *TP53* gene was later proposed as an oncogene (Parada *et al.*, 1984). Further studies then proposed that it could on the contrary be a tumour suppressor gene (Wolff and Rotter, 1984 and 1985; Mowat *et al.*, 1985). This controversial subject was highly debated. The ultimate proofs of p53 role as a TSG came from studies showing that the gene locus is very frequently deleted in colorectal carcinomas and other tumour types (Baker *et al.*, 1989; Nigro *et al.*, 1989), and that the wild type *p53* gene is able to suppress colon carcinoma cell growth (Baker *et al.*, 1990). Fitting with this tumour suppressor role, it turned out that the p53 protein functions as a "guardian of the genome" and is virtually the TSG most frequently altered in cancer (Levine *et al.*, 1997).

[Please see part A-2-a "Cellular signalling pathways and functions deregulated in cancer cells" for more details on the pRb and p53 pathways and their involvement in cancer cells].

A tumour suppressor gene can be defined as a gene that upon loss of its physiological function by any inactivating mechanism contributes to the malignant phenotype. Loss of TSG activity in tumour cells can be achieved through inactivating mutations in their sequence, loss of expression by gene deletion or silencing, or inhibition of the activity of the encoded protein by any means.

b. Genetic Instability of the cancer cell

b-1 Introduction

Cancer has been largely described as a multi-step, and somehow stochastic process resulting in the cellular physiology to be highly perturbed in tumour compared to normal cells. Advances in our understanding of this pathological physiology has enabled however to define common traits of cancer cells and even to link sometimes these traits to tumour initiation and progression.

Among the hallmarks that can be defined for tumour cells, it seems quite clear now that a wide majority, if not all, of these cells present a high genetic instability (Hanahan and Weinberg, 2000).

[Others hallmarks that have been defined for cancer cells are presented and further slightly detailed in the part A-2 "Towards a comprehensive understanding of cellular transformation and cancer progression"]

This genetic instability can thus be seen as an intrinsic property of cancer cells and, as such, has been proposed to participate to both cancer initiation and progression and even to have causative roles of these different steps in certain circumstances. Once initiated, a cell will have to accumulate several types of genetic aberrations to become a malignant cell. These genetic aberrations, or mutations, can be observed at both the DNA sequence and the chromosomal levels. Which of these could be real driving forces of tumourigenesis and tumour progression is still debated and controversial (Sieber *et al.*, 2003).

The following part is an overview of current knowledge about genetic instability, principally resuming its different types and presumed respective contributions to the initiation and progression of the malignant phenotype, in general and in selected examples.

b-2 Different types of genetic instabilities putatively underlying cancer initiation

i- Types and origins of genetic instabilities: lessons from colorectal cancers

The study of colorectal carcinogenesis has provided great insights to the understanding of genetic instability and its contribution to cancer. Importantly, several works have shed new lights on colorectal carcinogenesis, both in hereditary (FAP, for Familial Polyposis, and HNPCC, for Hereditary Non Polyposis Colon Cancer) and sporadic tumours (Lengauer *et al.*, 1998). Main contributors to these remarkable studies were conducted by C. Lengauer, K. W. Winzler, B. Vogelstein and coworkers.

Three main categories of genetic alterations are believed to play a role in cancer initiation: the first concerns point mutations affecting the DNA sequence; the second, aneuploïdy refers to an abnormal number of chromosomes; and the third, aneusomy, to an abnormal structure of chromosomes. Point mutations, as well as numerical and structural rearrangements of chromosomes are indeed recurrently observed in cancer cells, and as such have been postulated to participate to cancer initiation. These three types of alterations can be observed at two different levels and consequently, two main categories of genetic instability underlying them have been defined: the first is at the nucleotide level (that is the sequence) and the second at the chromosomal level of DNA.

a) At the sequence level

Two types of genetic instability have been described and are due to deficiencies in various DNA repair processes. The cellular machineries responsible of these are indeed normally controlling the genome sequence to avoid mutagens effects and to ensure DNA replication fidelity.

A first type of genetic instability at the sequence level is due to Nucleotide Excision Repair (NER) deficiency, and known as NIN, for NER-associated INstability.

The NER process is responsible of genome surveillance to avoid exogens mutagens effects on DNA. Definition of NIN came from xeroderma pigmentosum (and others related disorders such as Cockayne syndrome) study (Cleaver, 1968; De Weerd-Kastelein *et al.*, 1972), an hereditary syndrome mostly predisposing to skin tumours. In these cases, cutaneous cancers are mainly initiated through inherited mutations in NER genes such as XPA, XPB or XPC (and others) of which the encoded proteins are consequently unable to correct for DNA adducts induced by ultraviolet light, a major environmental mutagen (Cleaver, 2005).

A second type of genetic instability at the sequence level is Microsatellite Instability, called MIN (or MSI).

Its definition came from observations that microsatellites, small nucleotides repeated tracts present throughout the whole genome, are highly variable in colon tumour cells (Ionov *et al.*, 1993; Thibodeau *et al.*, 1993). Like approximately 15% of sporadic colorectal cancers, some colorectal cancer cell lines do also present MIN and have a stable near diploid karyotype (Lengauer *et al.*, 1997).

MIN is due to DNA MisMatch Repair (MMR) deficiency and mutations affecting genes involved in the MMR process have been first identified in Hereditary Non Polyposis Colon Cancer, or HNPCC (Fishel *et al.*, 1993; Strand *et al.*, 1993; Parsons *et al.*, 1993; Leach

et al., 1993): HNPCC is due to a heritable predisposition to colon cancer through transmission of germline mutations in MMR genes such as *MSH2*, *MLH1*, *PMS1*, *PMS2*, or *MSH6*. Mutations in these genes were further also found in sporadic colorectal tumours but rarely (Narayan and Roy, 2003).

b) At the chromosomal level

On the contrary, other colorectal cancer cell lines, like a wide majority of colorectal tumours present abnormal karyotypes with whole chromosomes that are gained and others lost. Aneuploïdy is a common trait of tumour cells but the question of whether these cells present constant chromosomal segregation defects or if aneuploïdy results from one or a few isolated crisis remained unanswered for decades. First clear experimental proofs were obtained by Lengauer and coworkers showing that the rate of gains or losses of whole or large portions of chromosomes by cellular generation is increased in colorectal cancer cell lines without MIN phenotype. This property was named Chromosomal INstability (CIN, Lengauer *et al.*, 1997). Chromosomal INstability, or CIN, may be responsible of the consistent aneuploïdy state observed in tumour cells, and importantly not the reverse. Aneuploïdy is a state which *per se* does not reflect an obligatory CIN phenotype. Moreover, cell fusions experiments clearly indicate that CIN is a dominant heritable trait over MIN. CIN can thus be defined as an accelerated rate of gains or losses of whole or large portions of chromosomes (Lengauer *et al*., 1997) which is reflected at the molecular level, by a Loss Of Heterozygosity of the involved loci (Rajagopalan *et al.*, 2003).

Further works have shown that mutations in genes involved in mitotic checkpoints and chromosome segregation can be at the origin of CIN in colorectal cancers (*BUB1* and *BUBR1* genes, Cahill *et al.*, 1998; *CDC4,* Rajagopalan *et al.*, 2004; *MRE11*, *hZw10*, *hZwilch*, *hRod*, and *Ding* genes, Wang *et al.*, 2004a). Depending on the mutations nature and number they require to produce CIN, those genes can be classified into three categories (one dominant, inactivating or activating mutation; or two inactivating recessive mutations needed). Thus it seems that alterations in the handling of chromosomes during mitosis can be at the origin of CIN.

Alternative mechanisms to these specific gene mutations have also been proposed as able to induce CIN: telomere dysfunction and centrosomes aberrations could be initial events leading to genomic instability (Feldser *et al.*, 2003; Nigg, 2002). On the other hand, aneuploïdy by itself has been proposed as causative of CIN (Duesberg *et al.*, 2004).

[Please see in next sections, "genetic instability at the origin of cancer ?" and "chromosomal aberrations and cancer" of this general introduction part, for more details on these putative contributions to CIN.]

Finally, the two main types of genetic instability, MIN and CIN, are mostly exclusive in colorectal cancer cell lines suggesting a somewhat equivalent contribution to tumour progression even if in some rare instances they have been found to occur.

c) Caretakers and Gatekeepers genes

Since they are in charge of both the surveillance and maintenance of genome integrity throughout cell life, the genes whose normal function inactivation leads to genomic instability in pre-malignant cells should be regarded as "caretakers" (Kinzler and Vogelstein, 1997). Their inactivation doesn't conduct directly to tumour formation but rather facilitates the likelihood of malignant conversion by means of secondary mutations accumulation in the genome.

Among these, some will affect other crucial genes such as the "gatekeepers" which are supposed to regulate directly tumour growth, either by inhibiting cell proliferation or by promoting cell death. (figure 2).

It should be here pointed out that the genes responsible for NIN, MIN and CIN phenotypes that have been identified so far could be regarded as "caretakers". However, the inactivation events required to produce the respective genetic instabilities is different: two mutations and an additional mutagen (UV light) may lead to NIN, whereas two mutations only may be required to produce MIN. On the contrary, CIN may result from a single dominant mutation, as exemplified by *BUB1* mutations (Lengauer *et al.*, 1998).

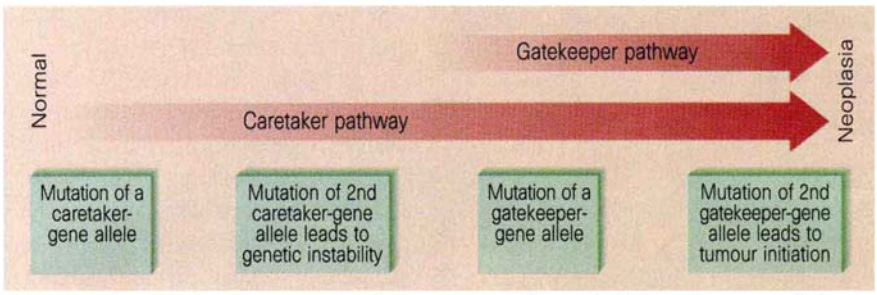


Figure 2. Caretakers and Gatekeepers genes.
Taken from Kinzler and Vogelstein, Nature 1997.

Two main ways may lead to cancer. **Caretakers genes** are responsible of the integrity of the genome. Their mutational inactivation leads to genomic instability in tumour cells. Genomic instability may produce random mutations that will also affect **gatekeepers genes** which directly regulate tumour growth either by inhibiting cell proliferation or by promoting cell death. Inactivation of gatekeepers genes finally allows tumour formation. Inactivation of gatekeepers genes by mutations may also be sufficient to lead to cancer as frequently observed for several familial forms of cancer.

ii- Genetic instability at the origin of cancer ?

Even if genetic instability can be observed in most cancer cells, it should be underlined that **its contribution to initiation and progression of the malignant state is not an obligatory mechanism** (Sieber *et al.*, 2003). Consistent with that, it has been shown for example that a non-negligible subset of colorectal tumours arise without MIN nor CIN (Tang *et al.*, 2004; and other references herein).

On the road to become malignant, a cell can for example have its two different alleles of a given gatekeeper gene that will be inactivated and as such will be more likely to become a malignant cell and produce a tumour. Such events are classically observed in several inherited predispositions to cancer where specific genes have been identified: the first mutant allele is inherited from one of the parents, and the second mutation is then somatically acquired, leading to tumour formation. Examples concern retinoblastoma (*Rb* gene), von Hippel Lindau (*VHL*), neurofibromatosis type 1 (*NF1*) or Adenomatous polyposis coli (*APC*) cancer predisposing syndromes.

However, several lines of evidences suggest that **the different types of genetic instabilities may play crucial roles in the very first steps of cancer initiation and progression**.

Colorectal carcinogenesis is thought to be initiated by inactivation of the APC (Adenoamtous Polyposis Coli protein) / beta-Catenin pathway. Mutations in *APC* gene are found in more than 85% of colon tumours whereas most of remaining harbours *beta-Catenin* gene mutations (Rajagopalan *et al.*, 2003).

Histologically, this initial event leads to cell dysplasia and slowly allow further formation of an adenoma (or polyp), which will possibly evolve to a larger form through acquisition of other mutations, in *K-RAS* or *B-RAF* oncogenes for example. Sequential acquisition of additional mutations in important cancer related genes (belonging, among others, to the TGF-Beta or p53 pathways) can then allow the formation of a carcinoma (figure 3).

In such a sequential model of carcinogenesis and progression, evidences that MIN and CIN / aneuploïdy do occur at very early stages and may actually represent initiating events of carcinogenesis have been bought.

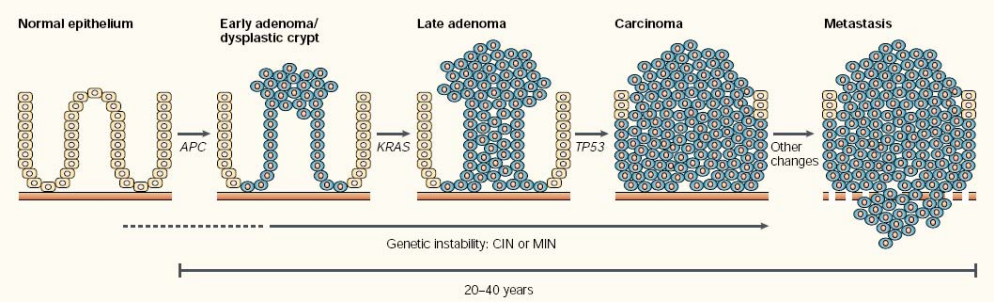


Figure 3. A progression model for colorectal carcinogenesis.
Taken from Rajagopolan *et al.*, Nature Reviews Cancer 2003.

Inactivation of the APC (Adenomatous Polyposis Coli) / beta-Catenine pathway is thought to initiate colorectal carcinogenesis in a colonic crypt. This is believed to arise trough bi-allelic inactivation of APC in most cases and leads to the formation of an adenoma. Upon acquisition of others mutations, a late adenoma will possibly form and evolve to give rise to an *in situ* carcinoma. Further acquisitions of other mutations can allow tumour progression up to the formation of an invasive carcinoma which can lead to metastasis formation. The whole tumour progression course is believed to take from 20 to 40 years.

Genetic instability, either MIN or CIN, may occur very early and represent a driving force of tumourigenesis. APC inactivation by itself indeed leads to both chromosomal instability and sustained proliferation (through activation of the Wnt/beta-Catenine pathway) in colon cancer cells.

a) MIN as a contributor to carcinogenesis

As a consequence of a first acquired MIN phenotype, through loss of a MMR gene, an initiated malignant cell will accumulate point mutations at higher rates, but not aneuploïdy. Accordingly, mutations rates around 2 to 3 fold higher are observed in MIN colon cancer cells when compared to normal cells (Rajagopolan *et al.*, 2003).

It is therefore reasonable to conceive that such MIN cells are likely to accumulate randomly mutations at higher rates in various genes. Some of them will be critical for the cell to progress towards a malignant state. These cells will then be prone to randomly acquire over time other mutations needed for malignant progression that will ultimately get fixed in a clonal cancer cell population. It is hence commonly admitted that the MIN phenotype represent a driving force for tumour progression but its participation to carcinogenesis is however not an obligatory mechanism (Sieber *et al.*, 2003).

b) CIN and / or Aneuploïdy as a tumourigenesis driver ?

On the contrary, the questions of whether CIN and / or aneuploïdy could represent very early events and driving forces of both tumour initiation and progression has been even more long-standing debated (for example Sieber *et al.*, 2003).

Mutations in CIN causing genes

The genes that have been identified originally as mutated in colon cancer cell lines presenting CIN and others identified since then have been found mutated in only a small proportion of human sporadic tumours (Wang *et al.*, 2004a).

On the contrary, mutations in ***APC*** are present in a wide majority of these tumour cells Inactivation of the APC pathway is thought to be an initiating event of colorectal tumourigenesis (Rajagopolan *et al.*, 2003), and even exclusively by bi-allelic inactivation of the *APC* gene (Sieber *et al.*, 2003).

Apart of their role in activating the beta-Catenin pathway to sustain cell proliferation in colon carcinoma cells, *APC* mutations have been recently also demonstrated to induce chromosomal instability *in vitro* (Kaplan *et al.*, 2001; Fodde *et al.*, 2001a).

These findings were confirmed by showing that the mitotic abnormalities observed in CIN colorectal cancer cell lines correlate with APC status in these, and that an APC truncating mutation, reflecting the situation that occurs in tumours *in vivo* interfere in a dominant manner with microtubule attachment of chromosomes to the spindle (Green and Kaplan, 2003). Importantly, these results were independently confirmed (Tighe *et al.*, 2004). Moreover, aneuploïdy is significantly more frequent in colon adenomas with mutations in APC, concerning 47% of these tumours (*versus* 26% for those without) and finally, specific point mutations of APC, affecting the central part (known as the MCR, for Mutated Cluster Region) of the protein or the more C-terminal part seem to be associated with CIN occurrence *in vivo* in these colon adenomas (Giaretti *et al.*, 2004).

It thus seems that the loss of functional C-terminal part of the APC protein could be a critical initiating event leading to CIN in a dominant manner. This would then further increase the likelihood of loosing the remaining WT allele of *APC*, which would fully activate the beta-Catenin pathway: this cell will have both high proliferation capacity and CIN, outstanding features of malignancy (Fodde *et al.*, 2001b; Fodde, 2002). APC loss of function

per se could also act further on other tumoural traits such as the coordinated regulation of cell adhesion and motility (Fodde, 2003).

Nevertheless, it is likely that other not yet identified CIN causing genes could be also found mutated and responsible for the CIN initiating phenotype. Consistent with that hypothesis, CIN can result from the inactivation of more than 50 different genes in yeast for example (Kolodner *et al.*, 2002), suggesting that more mutated genes are to be identified as responsible for CIN in human cancers (Wang *et al.*, 2004a).

Consequences of CIN

The first evident advantage towards malignancy that CIN could confer to a cell is an improved likelihood of inactivating TSGs: the loss of one of the chromosome harbouring a TSG can provide either the first or the second hit needed to inactivate such a gene.

CIN can furthermore cause changes in the expression of the numerous genes concerned by a chromosome that is gained or lost. In both cases, such dosage changes are however susceptible to be countered for most of the genes by either feedforward or feedback loops that usually regulate a majority of cellular pathways (Rajagopalan *et al.*, 2003). Moreover, it has been described that when a chromosome is lost, the normal conserved chromosome can undergo reduplication, surely to avoid haploinsufficiency that could concern a lot of genes (Brat *et al.*, 1997; Thiagalingam *et al.*, 2001).

Even though it is likely that expression changes may be sufficiently induced or diminished for some genes and even if transitory, this could provide the cell in which an entire chromosome was gained or lost with an advantage towards malignancy.

Aneuploïdy in very early pre-neoplastic and neoplastic lesions

Consistent with CIN / aneuploïdy putatively underlying cancer initiation, it has been shown in several cases that neoplastic or still pre-neoplastic cells in very early lesions from various tumour types do already present aneuploïdy.

Analysis of colorectal early cancerous lesions (tiny adenomas of very small sizes, between 1 to 3 mm) uncovered allelic imbalances in a wide majority of them (at least one chromosome affected in 90% of the samples; Shih *et al.*, 2001). Furthermore, several recent mathematical modelling studies of colorectal cancer initiation point out that early occurrence of CIN is a very likely phenomenon (Michor *et al.*, 2004; Komarova and Wodarz, 2004).

A second and even more strinking example concerns retrospective studies of premalignant dysplastic lesions from the oral cavity. These early lesions can be of two different types, erythroplakia and leukoplakia (Hunter *et al.*, 2005). Erythroplakia are very rare, generally show features of dysplasia and are likely to progress to HNSCC (50% over 10 years; Mashberg, 1977). Leukoplakia are believed also to be very early and still non malignant lesions, but are rarely dysplastic (5-15% of them; Sudbo *et al.*, 2001). On the contrary to erythroplakia, they more rarely further evolve and lead to HNSCC (2-5% over 10 years; Mashberg, 1977). Recent studies strikingly uncovered that the presence of aneuploïdy in dysplastic leukoplakia is a consistent predictor of tumour progression and even of patient's poor prognosis (Sudbo *et al.*, 2001; Sudbo *et al.*, 2004).

Aneuploïdy by itself as a tumourigenesis driver and at the origin of CIN ?

Long standing efforts to understand the **contribution of aneuploïdy to cancer initiation** and progression have also been made by P. Duesberg and coworkers.

[I would strongly emphasize that their scientific positions and works related to cancer are the sole that have to be rationally and objectively considered here. Reporting these data/concepts doesn't imply any further adhesion to their thoughts on any other scientific subject such as questioning the origin of AIDS.]

Several convincing arguments have been advanced supporting the view that random "aneuploïdization", either occurring spontaneously or induced by a non mutagenic carcinogen could be an initiating event in carcinogenesis, and even preceding CIN (Duesberg and Rasnick, 2000; Duesberg *et al.*, 2004). They established a model, in which the initial carcinogenic event in a normal cell is the generation of an aneuploïd karyotype, which will then auto-catalyze further karytotypic evolutions. Most of these will be consequently lethal, but on rare occasions could allow the emergence of a new chromosomal combination that is more viable for the cell in its habitat than the one of a normal cell. This cell with acquired malignant capabilities and genomic instability will expand clonally and give rise to other new karyotypic evolutions in the progeny (figure 4). In this view, aneuploïdy destabilizes the karyotype by corrupting the proteins involved in DNA synthesis and repair as well as chromosomes segregation. Consequently, aneuploïdy is proposed as the primary cause of the genomic instability observed in cancer cells. They have proposed that most of the spontaneous aneuploïdizations could be initiated through chromosome fragmentation or direct alterations of the mitotic spindle.

Conclusion

All these studies even if not direct proofs that CIN can be responsible for cancer initiation indicate at least that acquisition of aneuploïdy, very likely through CIN, can be a very early event in tumour progression and even at early stages related to patient clinical outcome.

This is in support for the hypothesis that CIN induced aneuploïdy could be an initiating event for cancer and at least a strong driving force for tumour progression even in its early stages in some cases. It is thus more conceivable and accepted that aneuploïdy may represent "obligatory second hits" that could allow irreversibility of transfomation progress (Ried *et al.*, 1999).

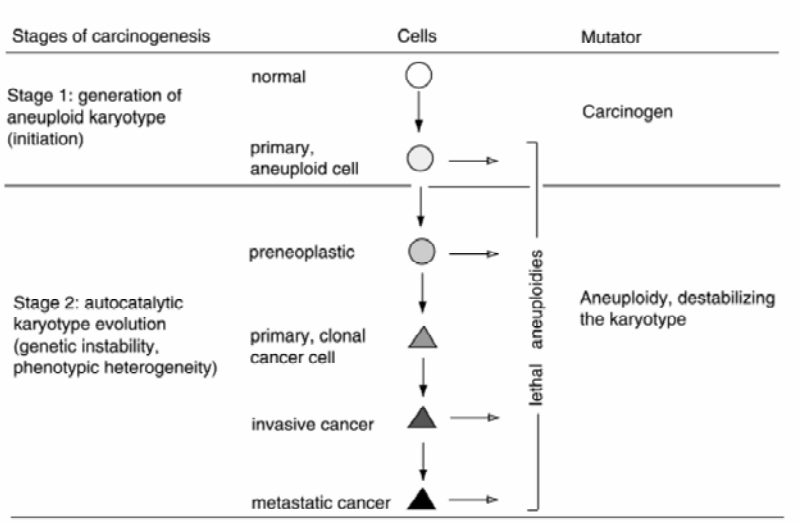


Figure 4. Aneuploidy as an autocatalyzer of chromosomal instability.
Taken from Duesberg and Rasnick, Cell Motility and Cytoskeleton 2000.

In this model, aneuploidy is first induced by a non mutagenic carcinogen in a non cancerous cell (circle). This aneuploidy state *per se* will further lead to the generation of cells with other karyotypes. Most of these will be lethal but a few could create a normal preneoplastic cell (circle) with a higher degree of aneuploidy. This autocatalytic mechanism could then lead to the emergence of a cancer cell (triangle) that could further evolve. Increasing densities of black color inside circles or triangles depict increased degrees of aneuploidy.

-II- Towards a comprehensive understanding of cellular transformation and cancer progression

This part is a rapid overview of some important works that have contributed to a better understanding of cancer until now.

In accordance with the work and results that are presented in the second part of this thesis, the following sections of this introduction part especially try to describe in a more comprehensive and precise manner with respect to cancer: first, some insights that have been bought by genome-wide gene expression profiling studies; secondly, the cell cycle with a more detailed view on some crucial players and the exit of mitosis; thirdly, the microRNAs as newly discovered post-transcriptional regulators of gene expression; and last, the chromosomal aberrations and their formation, with a specific focus on large chromosomal gains, gene amplifications, their transcriptional consequences and biological relevances.

This part does not pretend to represent an exhaustive review of the state of the art in our understanding of cancer and accordingly some other important points are also mentioned but more rapidly overviewed.

a. Cellular signalling pathways and functions deregulated in cancer cells

a-1 Hallmarks of cancer cells: Cancer Genes and the pathways they control

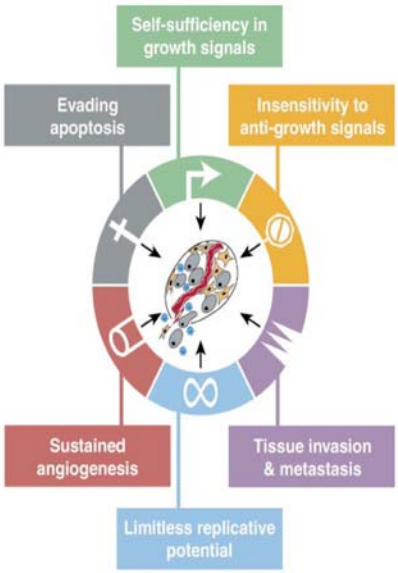
i- Common pathways and functions deregulated

An attempt has been made towards the definition of phenotypic hallmarks of cancer cells: common tumoral traits indeed can be aggregated in six acquired capabilities a cell may have to acquire to become fully malignant (Hanahan and Weinberg, 2000). Namely, these capabilities are the self-sufficiency in growth signals and the insensitivity to anti-growth signals, the possibility to evade apoptosis, the acquisition of immortality through limitless replicative potential, the angiogenesis and invasive potentials (figure 5A). The order of acquisition of these capabilities by cancer cells may vary from tumour to tumour both in terms of the way and of the chronological order they are acquired (figure 5 B and C). Furthermore, some genes altered in cancers may participate to the regulation of more than one of these acquired capabilities.

Efforts to integrate the molecular alterations observed in cancer cells in terms of cellular pathways that are deregulated and interconnected may in turn lead to the establishment of a complete picture of the molecular circuits underlying cancer (Hahn and Weinberg, 2002a). Examples of such circuits concerning the crucial pathways that are recurrently altered in cancer cells are given in figure 6.

Accordingly, this part is a very brief overview of these fundamental pathways of which the deregulation seem to be quite obligatory for the cell to acquire malignancy. For more extensive reviews (from which this part is largely inspired) of these pathways, please see: Hanahan and Weinberg, 2000; Hahn and Weinberg, 2002b; Vogelstein and Kinzler, 2004.

A.



B.

Component	Acquired Capability	Example of Mechanism
	Self-sufficiency in growth signals	Activate H-Ras oncogene
	Insensitivity to anti-growth signals	Lose retinoblastoma suppressor
	Evading apoptosis	Produce IGF survival factors
	Limitless replicative potential	Turn on telomerase
	Sustained angiogenesis	Produce VEGF inducer
	Tissue invasion & metastasis	Inactivate E-cadherin

C.

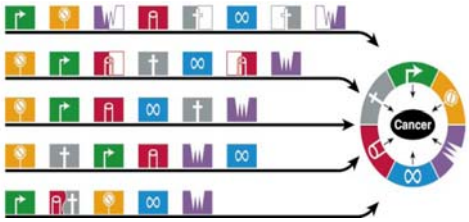
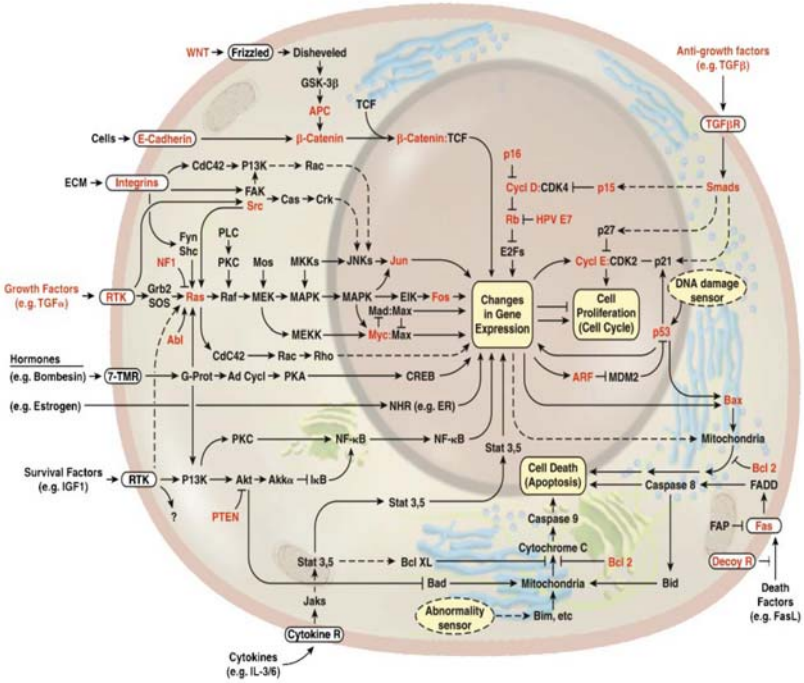


Figure 5. Phenotypic hallmarks of cancer cells.
Taken from Hanahan and Weinberg, Cell 2000.

A. Apart of genetic instability, cancer cells may have to acquire the six presented capabilities to become fully malignant.

B. Cancer cells may acquire these capabilities through different molecular alterations. Examples of molecular mechanisms to acquire these are given.

C. Cancer cells may acquire these capabilities in different orders (symbols equivalence given in A and B) suggesting the existence of parallel pathways that can lead to carcinogenesis.



<u>Figure 6.</u> Attempts towards the modelling of integrated circuits of the cancer cell.
Taken from Hanahan and Weinberg, Cell 2000.

An emergent integrated circuit of the cell: intracellular signalling cascades allows to transmit various informations in a normal cell through the interplay of numerous molecular actors and interconnected ways. Some of the players play crucial roles in integrating different informations by their cross-roads position in this scheme.
Those depicted in red are known to be deregulated in cancer cells.

a) The pRB and p53 pathways

The two first TSG historically identified, *pRB* and *p53* turned out to encode proteins crucial for normal cell physiology, at the crossroads of different signalling pathways regulating both normal cell proliferation and death. These pathways are inactivated in a majority of cancer cells, but globally, alterations can be found at different levels in the pathway (figure 7).

The pRB protein plays a crucial role in regulating cell cycle progression during the G1 phase, with other members of the pocket protein family it belongs to. Depending on tumour types, pRB can be inactivated directly by gene mutations or by oncogenic proteins from carcinogenic viruses (as exemplified by human papillomaviruses in cervical cancers). Its tumour suppressor role can also be overcome by activation of other proteins, such as cyclins D or E in other tumours types.

The p53 protein is a pivotal protein that can stop cell growth and / or trigger apoptosis in response of various stresses such as DNA damage. The p53 gene is mutated in at least half of human tumours, whereas some others overexpress the *hdm2* oncogene (human homolog of the mouse *mdm-2*) which targets p53 for degradation. Finally, some tumours inactivate the p53 pathway by suppressing $p19^{ARF}$ expression (by gene deletion or promoter methylation), which encodes a HDM2 antagonist.

b) Telomeres and telomerase

Telomeres status is believed to play a critical role in tumourigenesis since the telomerase enzyme is reactivated in most of cancer cells. This seems even to be a pre-requisite to cellular immortalization by increasing the replicative lifespan of the cell.

A telomere crisis do occur when telomeres get too short and provokes the formation of numerous chromosomal aberrations in breast cancer cells (Chin *et al.*, 2004). Telomerase is reactivated during this crisis and could participate to the fixation of the aberrations and even mark the transition from hyperplasia (hyperproliferative epithelium) to the formation of an *in situ* carcinoma in breast cancer (Chin *et al.*, 2004).

Alternative mechanisms to telomerase re-activation seem to account for telomere maintenance in some tumours (Bryan *et al.*, 1997). Some telomerase activity seems to be required for human stem cells and early progenitors cell divisions and homeostasis in a broad range of human tissues (Forsyth *et al.*, 2002). These stem and stem-like cells could thus represent targets of choice for cancer initiation.

c) Others important pathways

Cancer cells are believed to acquire self sufficiency in growth signals (with respect to these coming from their environment) by constitutively activating the signalling pathways that these external growth factors usually use. Many growth factor receptors and their intracellular downstream signalling cascades are important with respect to cancer. With this respect the Ras / MAP kinases pathways is archetypal. Numerous other signalling pathways and cellular functions are altered in cancer cells and affect general metabolism (such as glucogenesis), DNA repair and apoptosis for selected examples.

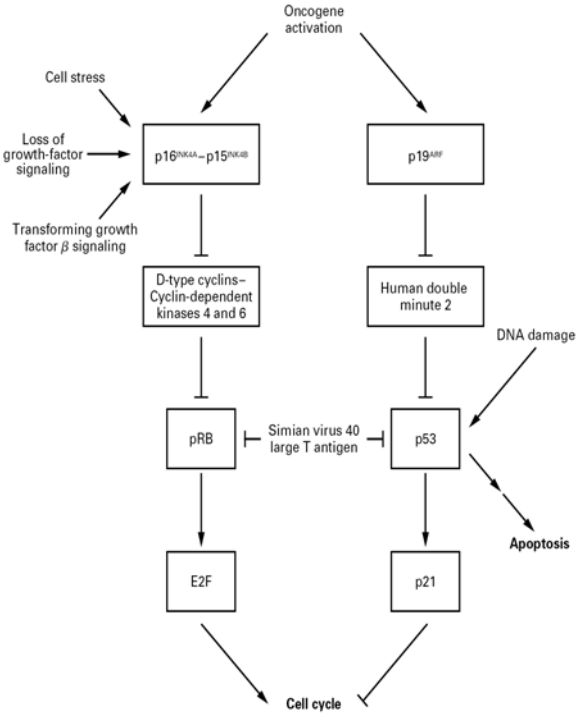


Figure 7. pRB and p53 pathways
Taken from Hahn and Weinberg, New England Journal of Medicine 2002.

The pRB and p53 proteins are pivotal players of two major cellular pathways involved in the control of normal cell cycle and the response to DNA damage. As such these pathways have been found recurrently altered in cancer cells: *p53* and *pRB* genes may be directly inactivated by mutation, whereas the overexpression of their respective inhibitors, HDM2 and cyclinD/CDK-4/6, can also overcome their normal function.
Arrows denote activating interactions and barred lines inhibitory ones.

ii- New insights from large scale gene expression profiling

Since their introduction (Schena *et al.*, 1995), DNA microarrays turned out to be extremely powerful tools, virtually allowing profiling of expression levels of all genes in a single experiment. Microarrays have been widely used to address very different fundamental questions such as for example, the transcriptional profiling of human cells during their cell cycle (Iyer *et al.*, 1999; Whitfield *et al.*, 2002).

This genome-wide approach gives a fresh global view starting from genes expression towards the identification of cellular pathways deregulated in cancer cells. More objective description of the relative weight of the different cellular pathways can be obtained if unsupervised analysis with statistical validation of the pathways involved is carried out. These studies have fuelled transversal comparisons which have the potential to unveil common pathways shared by subgroups of different tumour types. In the long term they may have major consequences for the individual tailoring of therapy for a given patient. Most large scale expression profiling data are publicly available and an integrated view can be extracted, giving further insight on the genes and regulatory modules / pathways that are recurrently altered in cancer cells (Segal *et al.*, 2005).

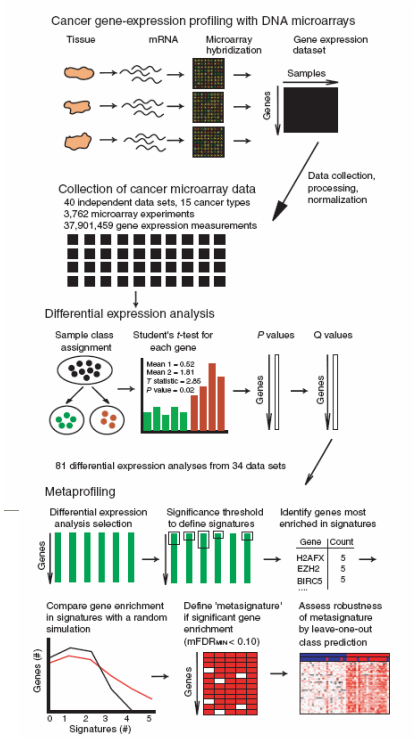
Global integration of gene expression profiling in cancer is clearly challenging (Hanash, 2004). Methodological issues are indeed crucial towards the compilation of heterologous results and informations from different sources. Furthermore, integration of data can be made at different levels, such as for examples “simple” data meta-analysis, functional enrichment analysis or interactome networks (Rhodes and Chinnaiyan, 2005; Segal *et al.*, 2005).

Such ongoing attempts rely on the establishment of integrated databases and dedicated computational tools. Examples of databases collecting large scale expression profiling data and especially from microarray experiments include Gene Expression Omnibus at NCBI (or GEO, http://www.ncbi.nlm.nih.gov/geo/) and Source held at Stanford University (http://source.stanford.edu/cgi-bin/source/sourceSearch). Oncomine is a more specialized databank for cancer related microarrays data held by Michigan University (http://www.oncomine.org/main/index.jsp). This database is moreover implemented which efficient analysis tools available online and allows for objective comparisons of data.

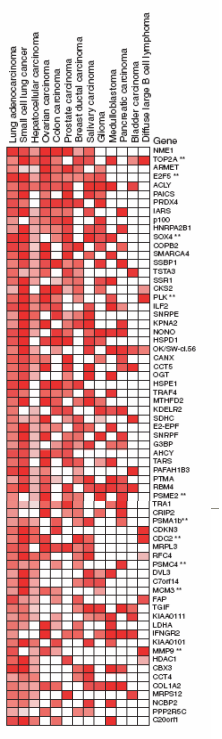
A prototype example has been recently published aiming at the meta-analysis of genes expression profiles from different cancer studies (Rhodes *et al.*, 2004) This study was indeed able to define two different meta-signatures compiled from 40 datasets rerpresenting more than 3500 individual tumours profiled. A first meta-signature associated with neoplastic transformation comprises 67 genes commonly over-expressed in cancers (comparison *vs* normal). The second meta-signature associated with cancer undifferentiation is constituted of 69 genes commonly over-expressed in undifferentiated tumours (compared *vs* well differentiated tumours). As an example of this impressive work, the approach and the meta-signature of neoplastic transformation is illustrated by figure 8.

Large scale deregulated genes expression in cancer cells can also be integrated according to the function of genes or their participation in a given cellular pathway. Recent advances in determining Gene Ontologies (Gene Ontology Consortium, 2001; Harris *et al.*, 2004; http://www.geneontology.org/) have greatly contributed to the meta-analysis of gene expression profiling (Rhodes and Chinnaiyan, 2004) and virtually all expression profiling studies are now looked up in terms of putative gene ontologies that are over- or under-represented in the gene expression patterns delineated.

A.



B.



<u>Figure 8.</u> Towards meta-analysis of genowe-wide expression profiles.
Taken from Rhodes and Chinnaiyan, Nature Genetics 2005.

A. The meta-analysis process. Methodological issues (collection and standardization of data for their analysis) are crucial to meta-analyze large scale data. Computational tools are also needed particularly to ensure robust statistical validations of the analyses.

B. Example of the « cancer meta-signature » obtained by meta-analysis. This represents the list of the 67 genes that are commonly over-expressed in common human cancer types when compared to their normal counterparts. Level of over-expresssion in respective tumours is represented by the red color strength.

A gene function based approach has been realized by an integrative meta-analysis for 22 tumour types that aimed at defining "modules" (Segal *et al.*, 2005). These modules represent sets of genes that are co-acting in a specific function. Their definition enables to draw a module map of functions that are commonly altered in several cancers or others that are more specific to a subtype which may have direct consequences for our understanding of tumour progression.

Another recent large scale integrative approach tried to establish links between the commonly deregulated genes in cancer cells and the regulatory circuits involved. The goal was to decipher the transcription factors that may explain consistent deregulations of genes in various cancers through analysis of the promoters of these genes and their transcription factor binding sites present in these promoters (Rhodes *et al.*, 2005). This study uncovered that the responsive genes regulated by the E2F transcription factors family, implicated in cell cycle control, are consistently over-expressed in a wide variety of tumours and thus reinforced that activation of the E2F pathway is a prevalent event in human cancers as already previously proposed (Muller *et al.*, 2001). Others important regulatory circuits were underlined but for specific tumour types. Among these, the c-myc regularory program was shown to be activated in different carcinomas types.

[For more details on the E2F pathway, its involvement in cell cycle control and deregulations in cancer cells, please see the next section "Focus on cell cycle" hereafter.]

a-2 Focus on Cell Cycle

In addition to cited articles, this part is mainly compiled from Clurman and Roberts (Cell cycle: an overview, in The Metabolic & Molecular Bases of Inherited Disease, 8th edition). With respect to this thesis work, special attention is given to some parts.

i- Normal cell cycle, key players and checkpoints

a) An introduction to the cell cycle

The "cell cycle" term refers to the process of cell reproduction that is the different steps that a cell has to go through to divide into two daughter cells. Depending on its environment and fate, a somatic cell can indeed either proliferate, or exit the cell cycle to stay quiescent or differentiate (this quiescent state being symbolised by G0 for *in vitro* cultured cells).

Basically, eukaryotic cell cycle necessarily requires genome replication, doubling of cellular mass and perfect segregation of the chromosomes accompanied with a more or less equal distribution of other cellular components. It comprises 4 different phases (G1, S, G2 and M).

The G1 and G2 phases are "Gap" or "Growth" phases, and so called because of no obvious activity being microscopically observable in those cells apart of cell size growth.

The S phase is the DNA replication period (S standing for Synthesis) during which each chromosome is replicated to give rise to a bi-chromatidian chromosome.

In the M phase, the cell undergoes mitosis which is separated into four phases: prophase, metaphase, anaphase and telophase. Mitosis allows the segregation of the bi-chromatidian chromosomes in two equivalents sets of mono-chromatidian chromosomes during anaphase, to give birth to two genetically identical daughter cells at the end of telophase after cytokinesis completion (figure 9)

Several key players involved in controlling the cell cycle and its progression have been identified. These are mainly the Cyclins / Cyclin Dependent Kinases (or CDKs) protein complexes and their regulatory proteins the CDK inhibitors (CDKIs), as well as the E2Fs family of transcription factors. The functions of these molecular players are intimately linked and integrated to provide a tighten control of the cell cycle. Accordingly, these key players have been recurrently found to be deregulated in cancer cells.

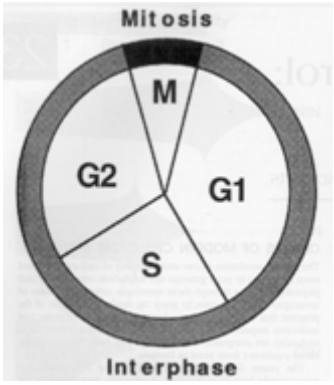
b) Key players in cell cycle regulation and control

The E2Fs/DP transcription factors

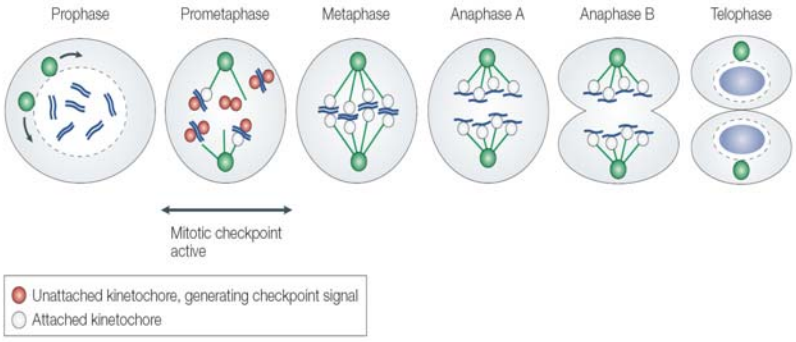
E2Fs family members

The E2Fs transcription factor family actually refers to the E2Fs members but also to their heterodimeric protein partners, the DP proteins. In Human, eight E2Fs (E2F1 to 8) and two DPs (DP1 and DP2) have been identified to date. The E2Fs family indeed enlarged recently with the discovery of two new members, E2F7 and E2F8 (de Bruin *et al.*, 2003; Maiti *et al.*, 2005).

A.



B.



<u>Figure 9.</u> The different phases of the cell cycle
Taken from Clurman and Roberts (Cell cycle: an overview, in The Metabolic & Molecular Bases of Inherited Disease, 8th edition) and Kops *et al*., Nature Reviews Cancer 2005.

A. The whole cell cycle is constituted by 4 different phases: G1, S, G2 and M.
The cell cycle can be subdivided to the mitosis phase and another predominant phase, the interphase which is composed of the S phase (DNA replication or Synthesis period) and the G1 and G2 phases (G standing for Gap or Growth).

B. The different phases of mitosis.
MItosis is the final key cell cycle step towards the creation of two genetically identical daughter cells by equally segregating the chromosomes just before cytokinesis. Four phases compose mitosis : prophase, metaphase, anaphase and telophase. Some of them may be further subdivided as mentioned on the panel. These phases are defined on the basis of the events taking place towards achieving the segregation of chromosomes into two perfectly identical sets. In accordance, these processes are highly regulated and involve for example a « checkpoint signal » in prometaphase as mentioned. After mitosis completion, cells can then either re-enter a new cell cycle (G1) or differentiate (« G0 », not figured on this cell cyle schematic view).

Chromosomes are figured in blue, the centrosomes and the mitotic spindle in green. Kinetochores are the fundamental molecular links between chromosomes and microtubules from the mitotic spindle. They are figured in red when unattached to the spindle, in white otherwise.

Depending on their transcriptional activity, two classes can be distinguished: the activators E2Fs, (E2F1, E2F2, and E2F3a) and the repressors E2Fs (E2F3b, E2F4, E2F5, E2F6, E2F7 and E2F8). Trials to decipher respective roles of E2Fs have been clearly hampered by the fact that they share relatively common binding sites in the promoters of their target genes (Tao *et al.*, 1997). In general, it is believed that activators E2Fs promote cell cycle progression whereas the repressors may be required for cell cycle exit and differentiation (Dimova and Dyson, 2005).

The protein domains of E2Fs with a global summary of their interactome (network of protein protein interactions), a simple model of the control of cell cycle by activators *vs* repressors E2Fs, as well as a non-exhaustive list of some of their important targets genes are presented in figure 10. For more details on E2Fs functions, please refer to the following excellent and more exhaustive reviews that were used for this part: Degregori, 2002; Stevaux and Dyson, 2002; Dimova and Dyson, 2005.

Activators E2Fs

Several genetic studies about E2F1-3 have pointed out a putative functional redundancy between these proteins (Degregori, 2002). E2F1 to 3 are essential for normal cellular proliferation since triple knockout cells are unable to progress to S phase (Wu *et al.*, 2001).

The best characterized E2Fs family member, E2F1 was originally identified as a factor inducible by the E1A adenoviral protein and able to bind E2 adenoviral promoter to induce its expression (Reichel *et al.*, 1987). It was shown to be a transcription factor with transcription transactivation activity in cooperation with the DP1 protein (Girling *et al.*, 1993; Helin *et al.*, 1993). Since then, E2F1 physiological roles have been widely investigated, through identification of numerous target genes: E2F1 not only controls cell proliferation but is also able to induce apoptosis (Dimova and Dyson, 2005).

The *E2F2* gene was cloned 12 years ago (Ivey-Hoyle *et al.*, 1993) but eventhough its role remained less investigated than E2F1.

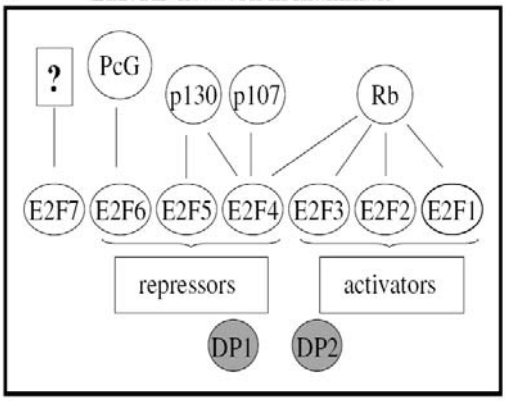
E2F3 is essential for proper cell cycle progression, and among the numerous E2Fs responsive genes, some have been shown to be specifically induced by E2F3 who has been accordingly shown to be critical for global cell proliferation and timing of the G1/S transition (Humbert *et al.*, 2000). Its role in repressing the ARF tumour suppressor, who belongs to the p53 pathway, has been shown recently to be crucial for normal cell cycle kinetics (Aslanian *et al.*, 2004).

A recent study analyzed large scale expression signatures to distinguish genes profiles that are dependent of E2F1 or E2F3 (Black *et al.*, 2005). It seems as previously known that both are inducing genes necessary for cell cycle entry and progression up to S phase and that E2F1 is also specifically regulating genes involved in apoptosis. Of importance, this study highlights that E2F3 seems to specifically control the G2/M by regulating the expression of several mitotic genes.

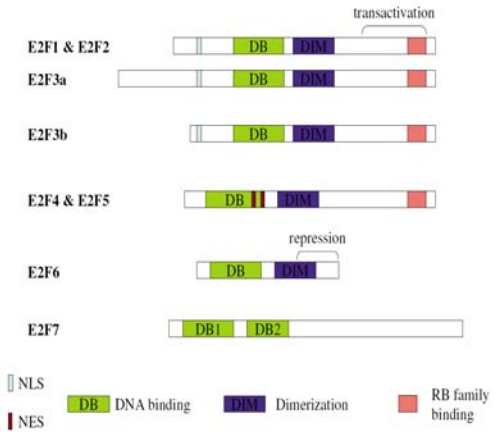
Repressors E2Fs

E2F4 and E2F5 are two E2Fs repressors of E2Fs responsive genes through binding to their promoter with pocket proteins (pRB, p107, p130 for E2F4; p130 fro E2F5).

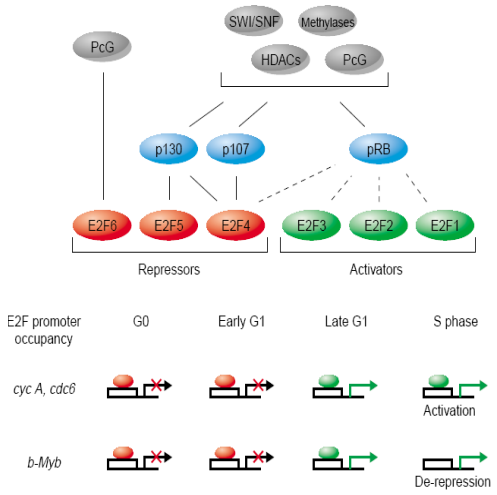
A.



B.



C.



D.

G1/S cell cycle regulators

Cyclin A
Cyclin E
B-myb
c-myb
Cdk2
E2F1,2, 3a
DP-1
Cdc25a
Cdk7

Inhibitors
p18Ink4c
p19Ink4d
Rb
p107
p21
p27

Nucleotide synthesis

thymidine kinase (tk)
thymidylate synthase (ts)
ribonucleotide reductatase
RNR1, 2
deoxycytidine kinase
dihydrofolate reductase (DHFR)

DNA replication

PCNA **histoneH2A**
DNA polα and polδ
DNA primase
DNA ligase
replication protein A2, A1
replication factor C4
topoisomerase IIα
Cdc6, orc1
Mcm2, 3, 4, **5**, **6**, 7
Dbf4
Flap endonuclease I

G2/M cell cycle regulators

Cdc2
Cyclin A
Cyclin B1, B2
Ki-67
Stathmin
importin 2

checkpoint regulators
Bub1
Cdc20
RanBP1

DNA repair

Brca1
Rad51
uracil-DNA-glycosylase 2

Development

Frizzled homologs 1,2,3
Enhancer of zeste
Homeobox A10,A7,A9, A13
TGF (& homologs & receptors)

Apoptosis

p73 ARF
Apaf-1
caspase 3,7
MAP3 kinase 5

Other

ID4
c-fos
HMG2,4
NFATc1,c3

Repressed by E2F independent of Rb

uPA
PAI-1

Figure 10. The E2Fs transcription factors family
Taken from Dimova and Dyson, Oncogene 2005; Stevaux and Dyson, Current Opinion in Cell Biology 2002; Degregori, Biochimica et Biophysica Acta 2002.

A. The E2FSs interactome in mammals. The E2Fs transcription factor comprises 8 family members: E2F-1 to -8 (E2F8 has just been recently identified and is not represented). E2F-1 to -6 heterodimerize with DP1 or DP2 to become active transcriptional activators or repressors. E2F-1 to -5 can selectively bind pocket proteins (p107, p130 and pRB). E2F6 can bind several Polycomb group proteins (PcG, transcriptional repressors). E2F-7 and -8 are repressors but with no known binding partner.

B. The different E2Fs and their proteins domains structures. E2F-1 to -6 contain one DNA binding domain (DB) and one DP dimerization domain (DIM). E2F-1 to -5 contain one transactivation domain and either a nuclear localisation signal (NLS) or nuclear export signals (NES). E2F7 (and E2F8, not represented) only contain two DNA binding domains.

C. Simplistic view of E2Fs interplay for the control of cell cycle. The repressors E2Fs (in red) occupy E2Fs target genes promoters in G0 and early G1. In late G1, they are replaced by activators E2Fs which bind either transiently in G1 (as for the *b-Myb* promoter) or from G1 to late S phase (as for the *cyclin A* and *cdc6* promoters).

D. Some E2Fs responsive genes. This list represents some well-known E2Fs responsive genes classified according to their participation to various cellular processes.

E2F6 was originally identified as a new E2F family member able to bind both DP proteins and acting as an E2Fs repressor, that when ectopically over-expressed leads to the accumulation of cells in S phase (Cartwright *et al.*, 1998).

Since then, E2F6 repressive role has been extended since it takes part of three different multimeric chromatin modifier complexes.

The first identified is a Polycomb (PcG) protein complex, comprising RYBP, BMI1 and other proteins (Trimarchi *et al.*, 2001).

The second was shown to comprise DP1 and other DNA binding proteins: Mga, a c-myc antagonist, and Max, a partner of c-myc. Polycomb group proteins and HP1gamma were also found and are associated with gene silencing activities. This "E2F6.com1" complex is binding to E2Fs and c-myc responsive genes, including *c-myc*, and *E2F1* themselves, preferentially in G0 fibroblasts (Ogawa *et al.*, 2002). The authors suggested that this complex could provide the silencing of *E2Fs* and *c-myc* co-regulated genes during quiescence stage through chromatin modifications.

Finally, a recent study enlarged the repressive role of E2F6 through its belonging to a third Polycomb modifier complex present in cycling cells (Attwooll *et al.*, 2005). E2F6 was indeed shown to be present with DP1 and the PcG protein EPC1 in this complex, which interacts with EZH2 (through EPC1) only in proliferating cells and apparently starting from late G1 phase. EZH2 is a PcG transcriptional repressor required for cell proliferation and accordingly, this complex has been proposed to regulate genes required for cell cycle progression (Attwooll *et al.*, 2005).

In addition to these roles, an alternative E2F6 mediated repression mechanism has also been demonstrated. It basicly relies on a competition between the E2F1 activator and the E2F6 repressor for E2F binding site occupancy on some target gene promoters but not all as for example it doesn't seem to be the case for the c-myc one (Oberley *et al.*, 2003). In any case, this is of interest in that it provides a simple mechanism, independent of chromatin modifications, for direct E2F6 mediated repression of some E2Fs responsive genes at least.

Concerning its **target genes**, E2F6 was shown to be bound to the *cyclin A* (G1/S cyclin) and *Cdc2* (G2/M) genes promoters in NIH-3T3 fibroblasts during G0, and to the *cyclin B2* (G2/M Cyclin) promoter later in G1 phase (Caretti *et al.*, 2003). With the "E2F6.com1" complex it has been shown among others, to bind to *c-myc* and *E2F1* genes promoters (Ogawa *et al.*, 2002). E2F6 was also found, among others, to bind to *BRCA1*, *HP1alpha* and *art-27* gene promoters (Oberley *et al.*, 2003). Finally, one of the role of E2F6 is to repress specific genes in somatic cells, such as the genes encoding meiotic cohesins, *SMC1beta* and *STAG3* (Storre *et al.*, 2005), which are involved in sisters chromatids cohesion during meiosis, as well as some male germ-cell restricted genes, such as specific tubulins encoding genes (Pohlers *et al.*, 2005).

A recent study pinpointed the role of the activating E2Fs in the transcriptional activation at G1/S of the E2F6 repressor (Lyons ***et al.***, 2005). Moreover, a role for E2F6 in distinguishing the G1/S from the G2/M E2Fs transcriptional program has been shown: by its specific association starting from S phase to the promoters of G1/S E2Fs responsive genes but not G2/M ones, E2F6 allows the specific repression of G1/S transcription during G2 and M phases of these genes (Giangrande *et al.*, 2004).

Both **E2F7** and **E2F8** have been isolated only very recently (de Bruin *et al.*, 2003; Maiti *et al.*, 2005; Logan *et al.*, 2004). They form a distinct subgroup in the E2F family, since they only present two DNA Binding domains (DB1 and 2) but no pocket protein binding

domain and no heterodimerization domain. They more look like the E2F-Like, or E2F-L proteins recently identified in *Arabidopsis thaliana* (Mariconti *et al.*, 2002; Kosugi and Ohashi, 2002) and bind consensus E2Fs sites *in vivo* but independently of DP proteins. Their ectopic expression delays cell cycle progression (Logan *et al.*, 2004) which suggests a counteracting role for E2F-7 and E2F-8 repressors to the cell cycle promoting roles of the activators E2Fs.

The Dimerization Partners of E2Fs: DP1 and DP2

Little has been investigated to address the respective roles of DP1 and DP2 proteins with respect to the E2Fs they can bind to. DP proteins have been both shown to be necessary for the heterodimerization of the E2Fs, but not responsible for the transactivation properties of the activators E2Fs. However, the association of E2F1 and E2F4 with either DP1 or DP2 influences the DNA bending properties of the heterodimeric complexes which may account for E2Fs binding site selectivity (Tao *et al.*, 1997).

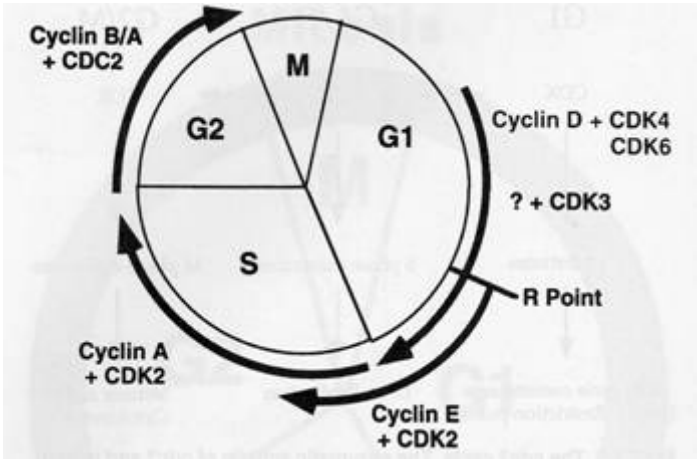
DP1 was the first isolated and since then is the most studied. but even very recently, new DP1 isoforms have been isolated (Ishida *et al.*, 2005). DP2 was cloned ten years ago and shown to be a novel dimerization partner of E2Fs (Zhang and Chellappan, 1995). Only a few reports are made on DP2 when compared to DP1. However, it has been interestingly shown recently that E2F1 and E2F4 can bind to the *aurora B* promoter both with DP2 but not DP1 (Kimura *et al.*, 2004). *Aurora B* gene expression is known to be cell-cycle regulated. The Aurora B chromosomal passenger protein function is to ensure proper chromosomes attachment to the mitotic spindle by localizing several crucial proteins to the centromere and kinetochore regions, by regulating chromosome bi-orientation through microtubules regulation and finally, Aurora B may also be involved in the spindle checkpoint surveillance with survivin (Ducat and Zheng, 2004).

Cyclin dependent Kinases (CDKs), Cyclins and CDK inhibitors (CDKIs)

Cyclin dependent kinases are key molecular players involved in the control of the cell cycle through phosphorylations of specific target proteins. CDKs usually are regulated by phopshoprylations / dephosphorylations. They form heterodimeric protein complexes through association to other proteins, the cyclins, and this association is required for their activation. Finally, Cyclin Dependent Kinase Inhibitors (CKIs) are.proteins that are also able to bind CDKs but inhibit their function.

CYCLINS/CDKs

Cyclins represent a family of related proteins who share the presence of a cyclin Box in their sequence and whose expression is generally restricted to specific cell cycle phases. Importantly, the expression of cyclins is rate limiting for CDKs activation and thus the control of cyclins expression at both transcriptional and post-transcirptional levels is a fundamental mechanism that allows CDK periodicity for the control of cell cycle. Each phase of the cell cycle is indeed characterized by a unique pattern of cyclin/CDK activity (figure 11).



<u>Figure 11.</u> Cyclins/CDK pattern of cell cycle phases
Taken from Clurman and Roberts (Cell cycle: an overview, in The Metabolic & Molecular Bases of Inherited Disease, 8th edition).

Each cell cycle phase is characterized by a specific pattern of molecular regulator complexes, the Cyclins / Cyclins Dependent Kinases (CDKs). In association with their respective CDK partners, Cyclins D and E control the progression towards S phase (« G1/S cyclins »), whereas Cyclins B and A are « G2/M Cyclins ». The R point mentioned refers to the Restriction point: going through it represents a real commitment for the cell towards achieving the whole cycle. Activation of both CyclinD/CDK-4/6 and CyclinE/CDK2 is believed to mark this crucial step.

Cyclin C has been identified as the partner of CDK3. They are involved in the control of the pRb dependent G0/G1 transition and thus control the re-entry in the cell cycle after resting.

G0/G1 cyclins/CDKs

Cyclin C/CDK3 seems to be the key player in stimulating cell cycle re-entry from quiescent cells through pRb phosphorylation (Ren and Rollins, 2004).

From an in vitro study of cells stimulated to re-enter the cell cycle from quiescent state by serum stimulation, cyclin L gene expression has been found to be transitory induced and as such proposed as a putative regulator of this transition (Iyer *et al.*, 1999).

G1/S cyclins/CDK

At the beginning of the cell cycle during early G1, D-type cyclins expression is induced by mitogens, and they associate with CDK-4/6. These complexes thus couple extracellular signals to the cellular machinery responsible of cell cycle progression through G1 phase. Three cyclins D (CCND1 to 3) exist and are expressed in a cell-type fashion. Their expression is maintained as long as cell-type specific mitogens are present and is necessary and crucial for G1 progression. The inhibition of CCND1 function leads to cell cycle block in G1 whereas its overexpression leads to the shortening of G1 duration for example.

Cyclin E activity may be required after cyclin D one during G1 but is necessary for the progression to G1/S transition. *CCNE* gene expression is controlled by E2Fs, and the cyclin E protein associates with CDK2. Importantly, CCNE protein level increase during G1 contributes to E2Fs activation by dissociation of E2Fs/pRB complexes. Its over-expression also leads to shortening of G1 and reduced dependency on mitogenic stimulation for the cell cycle to progress towards the G1 phase.

It is believed that both the activation of cyclins D and E associated kinases may biochemically constitute the restriction point in mammalian cells (please see hereafter for more details on the restriction point).

Cyclin A is required for the cell to enter S phase through its association with CDK2. Two cyclins A (1 and 2) do exist: cyclin A1 expression is restricted to meiotic, early embryonic and some tumour cells whereas cyclin A2 is on the contrary expressed by all proliferating somatic cells, and is an essential gene (Fung and Poon, 2005).

G2/M cyclins/CDKs

Cyclin B is the key mitotic cyclin and exclusively interacts with CDK1 (CDC2). They were together originally identified as the “M-phase promoting factor” (MPF). Cyclin A2 is also playing a role, but less clear, in association with CDK1 and CDK2 (Fung and Poon, 2005).

Three B-type cyclins exist: CCNB1 and B2 are co-expressed in dividing cells (CCNB3 is only expressed in germ cells and the adult testis). Activation of cyclinB/CDK1 by CDC25 phosphatases marks the G2/M transition. The predominant function of cyclinB/CDK1 is then to phosphorylate crucial effectors for mitosis, such as for example lamins subunits which will allow proper nuclear lamina breakdown (Smits and Medema, 2001). Complex regulations then take place for their degradation which is crucial for mitosis exit. These mainly involve the Anaphase Promoting Complex / Cyclosome, or APC/C (Fung and Poon, 2005).

Cyclins/ CDK INVOLVED IN TRANSCRIPTION REGULATION

Apart of all above mentioned cyclins directly controlling the cell cycle, others cyclins / CDK are involved in the regulation of transcription. These complexes include cyclin K/CDK9, cyclins T/CDK9, cyclin H/CDK7. For a review of their roles in the regulation of transcription, please refer to : Loyer *et al.* (2005).

CDK11 (formerly known as PITSLRE proteins) associate with cyclin L1. More than 20 protein isoforms of CDK11 are expressed from the use of different promoters and alternative splicing (Loyer *et al.*, 2005). Three different protein isoforms of cyclin L1 (L1 alpha, beta and gamma) are encoded by numerous transcript variants (Loyer *et al.*, 2005). Through its association to $CDK1^{p110}$, cyclin L1 is believed to participate to pre-mRNA splicing regulation (Loyer *et al.*, 2005). Of interest is that cyclin L1 gene expression is induced during G0/G1 transition, suggestive of a putative cell cycle related role as noted earlier.

CDK INHIBITORS

Two distinct classes of CDKIs are present in human.

The first class comprises the INK4 proteins that specifically inhibit G1 CDKs, namely CDK4 and CDK6 that associate with Cyclin D. Four related INK4 proteins have been described: p15, p16, p18 and P19, also respectively known as INK4B or CDKN2B, INK4A or CDKN2A, INK4C or CDKN2C and INK4D or CDKN2D. INK4 proteins are able to bind CDK4 and CDK6 monomers to prevent their association with CCND as well as to bind the CCND/CDK-4/6 complexes and inhibit their activity. The expression of p19 (INK4D) was found to oscillate with the cell cycle. Litte is known about their physiological role.

The second class is the CIP/KIP proteins that are universal CDKI able to inhibit all cyclins/CDKs (with the possible exception of cyclin B / CDK1). Three CIP/KIP proteins have been described: p21, p27, and p57, also respectively known as WAF1 or Cip1, Kip1. p21 and p27 seem to regulate the cell cycle at different levels but their primary role is to induce G1 arrest by inhibiting CCND/CDK-4/6 in case of DNA damage or mitogen deprivation, respectively.

An attempt towards an integrated, but simplistic, view for the understanding of global cell cycle control

Cyclins/CDKs control of the interactions of E2Fs with pocket proteins during G1

Some activating and repressors E2Fs are able to bind members of the pocket protein family, pRB, p107 and p130, whereas some others, E2F6 and E2F7 do not. The pRb protein is the sole pocket protein interacting with E2F1, E2F2, and E2F3. On the contrary, E2F4 is able to bind all three, when E2F5 binds only p130. These interactions are inhibiting E2Fs functions. They are abolished through phosphorylations of the pocket proteins by cyclinD/CDK-4/6, which have been previously induced after integration of mitogenic stimulation during early G1 through the RAS/MAPKinases pathway. E2Fs release allows them to transactivate the expression of several target genes that are needed for further cell cycle progression.

Among these, the induction of Cyclin E protein expression contributes (by its association to CDK2) through a positive feedback loop to even more activate E2Fs by phosphorylating pRb and thus dissociate the remaining E2Fs/pRb complexes (figure 12A). Importantly, this molecular mechanism allows a switch to a mitogens independence for cell cycle progression. It reinforces the expression of genes controlled by activating E2Fs that are needed for the G1/S transition and during S phase. Genes encoding proteins involved in DNA replication, cyclin A and its associated kinase CDK2 belong for example to this group.

Towards DNA replication: control of the G1/S phase transition and DNA replication

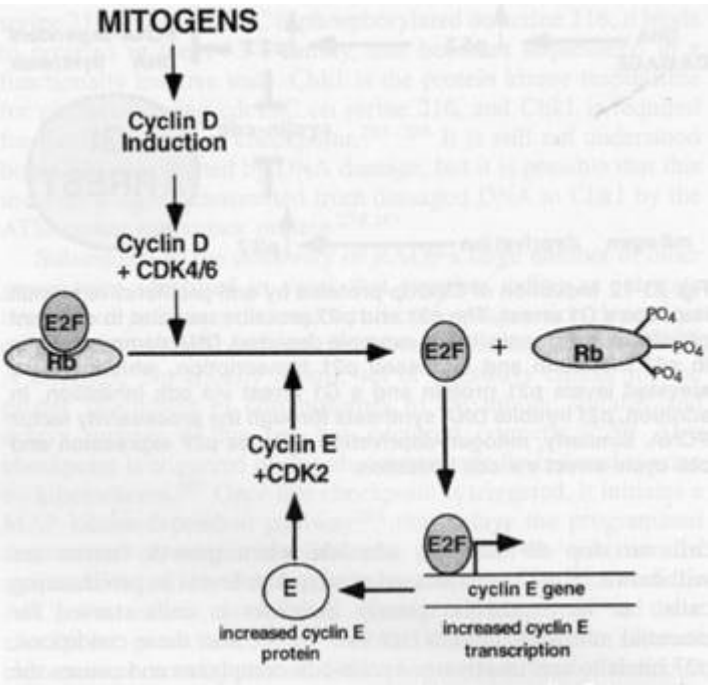
E2Fs have been historically characterized mainly as controlling the cell cycle progression towards and through the G1/S transition. This represents up to know the better characterized cell cycle phase controlled by E2Fs since they control well-known cell cycle regulators of this transition such as cyclin E or cyclin A both of which associate with CDK2 (Stevaux and Dyson, 2002; Woo and Poon, 2003). But as reported hereafter they may also act at other stages of the cell cycle.

At the G1/S transition, E2Fs are involved in the direct control of DNA repair and replication through control of the expression of several crucial genes such as proteins involved in the formation of pre-replication complexes at origins of replication (MCM proteins) and enzymes necessary for DNA synthesis (DNA polymerase alpha, replication factors and numerous others). At the same time, several genes encoding mismatch, excision and double strand break repair proteins are induced by E2Fs to ensure replication of DNA in an error-prone manner (Stevaux and Dyson, 2002).

Cyclin A/CDK2 is the master regulator involved in the control of DNA replication: it activates on one hand pre-existing replications complexes present at origins of replications and replication elongation on the other hand. Furthermore it prevents re-replication of DNA by preventing the formation of new pre-replication complexes through CDC6 control. (figure 12B; Woo and Poon, 2003).

Once DNA has been correctly replicated (which implies a control and repair of putative DNA damages, please see hereafter), the cell can exit S phase and prepare for its division through the G2 phase. In this context, it is now clear that E2Fs function is not only restricted to control the G1/S transition: through the list of their target genes identified by recent large scale approaches (Weinmann *et al.*, 2002; Ren *et al.*, 2002), they represent indeed the fundamental molecular integrators of cell cycle progression at several stages and related cell cycle checkpoints (Ren *et al.*, 2002). An example of that are the Smc2 and Smc4 E2Fs responsive genes: these proteins form the condensin complex readily after chromosome replication but is necessary for proper mitotic chromosome condensation.

A. B.



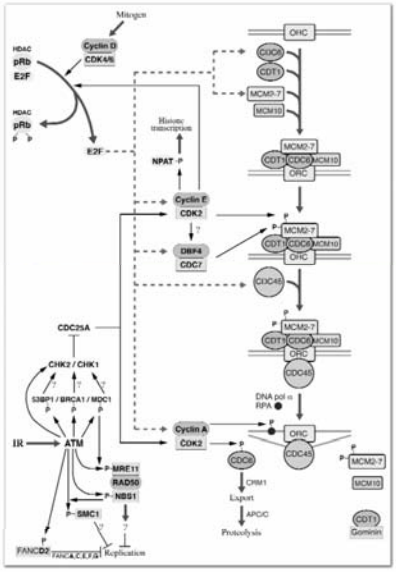


Figure 1. The Mammalian DNA replication pathway and the intra-S phase DNA damage checkpoint.

<u>Figure 12.</u> Some molecular events of G1 and S phases.
Taken from Clurman and Roberts (Cell cycle: an overview, in The Metabolic & Molecular Bases of Inherited Disease, 8th edition); Woo and Poo, Cell Cycle 2005.

A. Mitogenic (growth factors) stimulation in G1 activates *cyclinD1* transcription. Cyclin D1 binds to its CDK-4 or -6 partner and inactivates pRB through phosphorylations. This releases the activators E2Fs which can transactivate expression of their target genes. Among these, cyclin E expression is induced. Cyclin E binds to CDK2 and the complex contributes to pRB inactivation and sustains E2Fs activation in a positive feedback loop manner. Importantly, this represents a molecular switch for the cell that becomes independent of mitogens (growth factors) for cell cycle progression.

B. The mammalian DNA replication pathway (right of the picture) and the intra-S phase DNA damage checkpoint pathway (lower left part).
The picture is actually complex but partly illustrate how E2Fs control cell cycle progression in an integrated manner. The dashed gray arrows starting from E2Fs represent some of the proteins whose expression has been previously induced by E2Fs at the G1/S transition. E2Fs indeed control the expression of several crucial genes for DNA replication initiation (MCM proteins) and replication *per se* (DNA polymerase alpha) as well as for DNA repair (CDC25A, SMC1) to ensure DNA replication fidelity.

Towards mitosis entry: control of the G2/M transition

Several studies have shown that many of the E2Fs responsive genes are acting at different levels of mitosis, such as chromosomes condensation and segregation, or the spindle checkpoint (Stevaux and Dyson, 2002). As examples, key players of G2/M transition and mitosis such as cyclin B1, cyclin B2, Bub1 and CDK1 transcripts are induced by E2Fs. Analysis of large scale gene expression profiling of genes during the cell cycle of HeLa cells (Whitfield *et al.*, 2002; data available at http://genome-www.stanford.edu/Human-CellCycle/Hela/) reveal that the expression of the CCNB1, CCNB2, Bub1 and CDK1 genes takes place after the end of S phase, that is once the DNA replication has been completed, and is pursued during G2 and finally peaks in early mitosis. This highly suggests that E2Fs could also at some point and to a still unknown extent regulate the G2/M transition and mitosis (Ren *et al.*, 2002; Stevaux and Dyson 2002). This transition is actually mainly controlled by the mitotic cyclin B interacting with CDK1 whereas cyclin A seems also to play a role in mitosis, even if not fully understood yet (Fung and Poon, 2005).

Towards completion of the cell cycle: crucial points, checkpoints and mitosis exit

During the whole cell cycle, different so called "cell cycle checkpoints" and critical points (such as the restriction point) for the cell do exist. Checkpoints are actually molecular mechanisms ensuring the cell that a previous process, such as DNA replication or mitosis, was correctly completed. If not, they actively stop cell cycle progression (Kastan and Bartek, 2004).

The restriction point

The first bottleneck during normal cell cycle occurs during the G1 phase and is called the restriction point, or "R". This is not a checkpoint *per se*, at least as defined above. But it represents a real commitment since once the cell will go through, it will have to complete the cycle. After mitogenic stimulation integrated through the RAS / MAPK pathway, CCND1/CDK-4/6 are the key players who dissociate the E2Fs/pRB complexes. This leads to the activation of CCNE/CDK2. This restriction point mainly serves to ensure that enough growth factors are available for the cell to divide and thus prevents cell population overgrowth in too poor environmental conditions.

DNA integrity checkpoints

DNA damages are occurring continuously during cell life, provoked by several environmental factors such as UV light. Depending on the phase of the cell cycle in which these damages arise, they can trigger several different checkpoint pathways. Moreover, DNA replication, if not complete, is also able to activate the same checkpoint pathways (Melo and Toczyski, 2002; Kastan and Bartek, 2004).

DNA damages, uncompleted DNA replication or DNA replication blocks are sensed by different dedicated mediators proteins and transmitted to the two majors checkpoint kinase proteins, ATM (ataxia-telangiectasia mutated) or ATR (ATM-related), depending on the nature of the damage. In a schematic view, these activate then the downstream effector kinases Chk1 and Chk2 who are then able to transduce the signal through many different protein protein interactions with other effectors kinases by activating adapted downstream signalling pathways which lead to cell cycle progression arrest in order for the cell not to enter mitosis with damaged DNA. A simplified view of these signalling cascades with respect to DNA DSBs breaks is presented in figure 13. Ultimately it will lead either to the completion of DNA replication, repair of the DNA damage or to cell death.

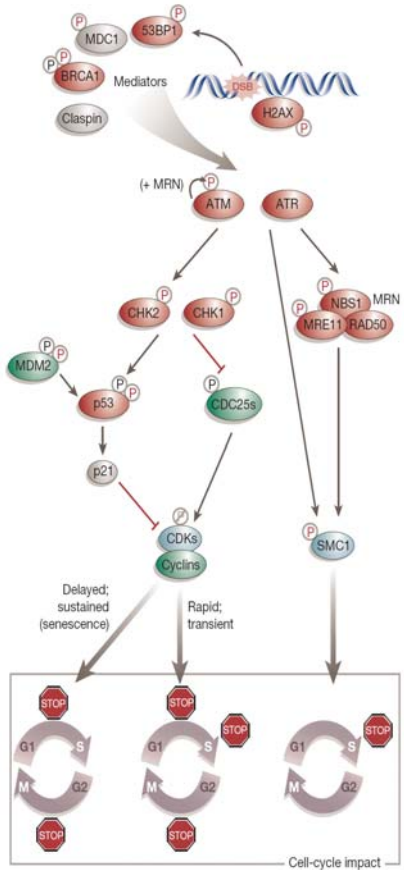


Figure 13. The double strand DNA breaks checkpoints
Taken from Kastan and Bartek, Nature 2004.

A simplified view of the cascade of molecular events taking place in case dsDNA breaks do occur. This scheme illustrate that DNA damages can have direct impact on the cell cycle progression at any point and actively stop it if needed, until the damage get repaired. This checkpoint role is accordingly to ensure the transmission of undamaged DNA to the daughter cells after cell cycle completion.

This is a way for the cell to protect against transmission of erroneous genetic information to its daughter cells either by correcting mutations when possible and appropriate or by triggering its programmed death, or so called apoptosis.

Mitotic checkpoint and control of mitosis exit

Another crucial cell cycle checkpoint consists in the mitotic checkpoint, also known as the spindle checkpoint. A huge number of proteins and interactions are involved since the beginning of mitosis. This checkpoint basically relies on a cell cycle arrest or progression delay until each chromosome is correctly attached through its kinetochores and microtubules to the mitotic spindle and is also well oriented with respect to both poles at the metaphase/anaphase transition and during early early anaphase. This is a way to ensure that each chromosome will properly segregate before cytokinesis and avoid any subsequent aneuploïdy in the daughter cells genomes (Kops *et al.*, 2005). Please see figure 14 for more details on how several mitotic checkpoint proteins, such as BubR1, through their binding to kinetochores are involved in regulating the correct segregation of chromatids and chromosomes.

Master regulators of **early mitosis and mitosis exit** are
(i) a multi-protein complex with E3 ubiquitin ligase activity, the Anaphase Promoting Complex / Cyclosome (APC/C). During mitosis, APC/C exists as two forms, depending on an additional binding partner: Cdc20 or Cdh1, allowing to form APC(Cdc20) or APC(Cdh1);
(ii) the dual specificity phosphatase CDC14 (Stegmeier and Amon, 2004).

The more precise description is at present done in yeast. Accordingly, the subtle interplay between these two forms of APC/C and CDC14 as well as the cascade of events that take place are described hereafter for the situation occurring in yeast (Stegmeier and Amon, 2004).

The APC(Cdc20) complex is responsible of anaphase initiation. It targets the securin protein for degradation by the proteasome. Separase is consequently released and thus becomes activated. The role of separase is then first to allow sisters chromatids of a chromosome to be separated through mitotic cohesin (composed of SMC1 and SMC3) degradation. With others proteins, separase second role is then to activate CDC14 through the FEAR protein network still at anaphase onset. FEAR stands for CDC Fourteen Early Anaphase Release since CDC14 is released from the nucleolus at this point, where it is sequestered during most of the cell cycle (Stegmeier *et al*., 2002; Sullivan and Uhlmann, 2003).

CDC14 activation seems primarily to be particularly important for the segregation of ribosomal and telomeric DNA (since yeast *cdc14* mutants have specific troubles of segregation of these), as well as for the translocation of chromosomal passenger proteins (Stegmeier and Amon, 2004; Torres-Rosell *et al.*, 2005). Very importantly, this activation of CDC14 was recently shown to be also both necessary and sufficient to silence spindle microtubule dynamics, which ensures proper chromosomes segregation in yeast. This silencing is achieved through CDC14 induced dephosphorylation of several targets such as INCENP from the Aurora B kinase complex, Ask1 from the DASH kinetochore complex, and putative regulations of other microtubule associated proteins (Higuchi and Uhlmann, 2005).

A.

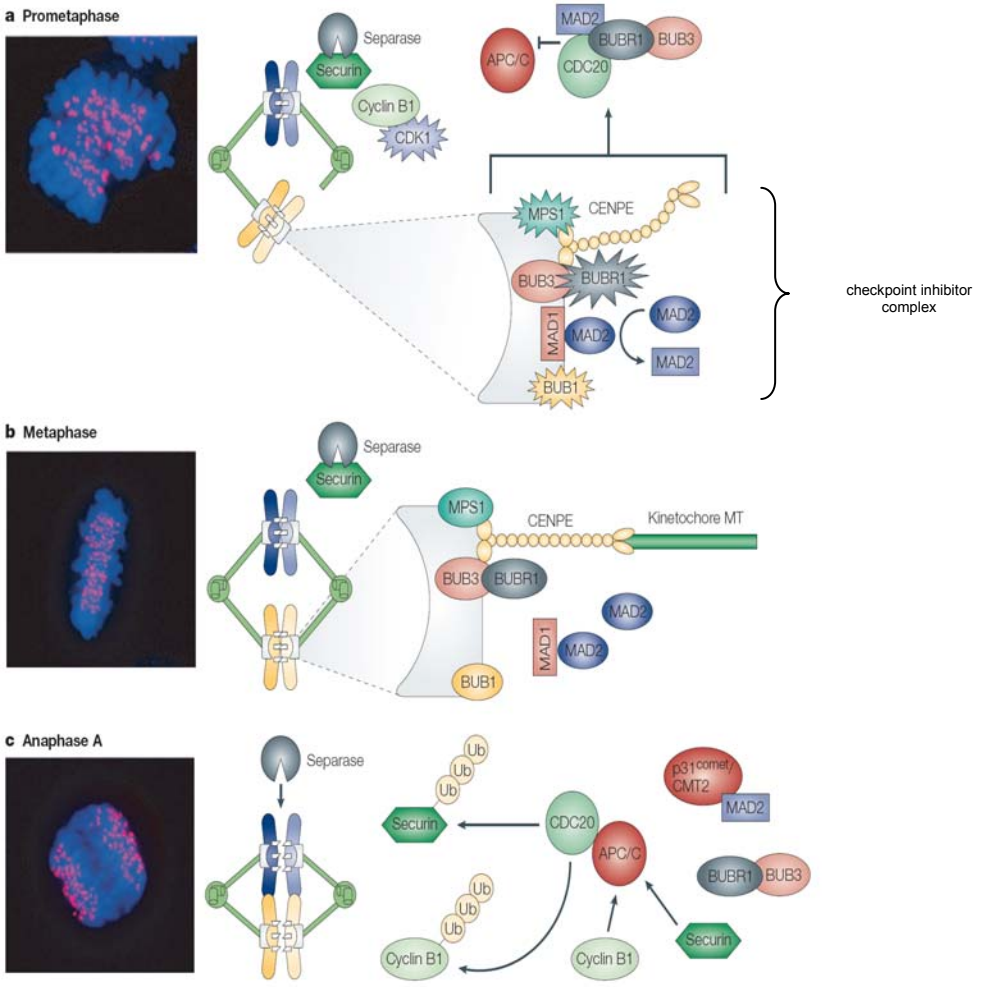


Figure 14. The mammalian mitotic checkpoint.
Taken from Kops *et al*., Nature Reviews Cancer 2005.

A. The kinetochore is assembled during prophase/early prometaphase and recruits the checkpoint complex composed of several proteins. BUBR1 kinase activity is essential and activated by its binding to the centromeric E protein (CENPE). Until all kinetochores are well attached to the spindle, the complex diffuses an inhibitory signal to wait for anaphase by sequestering CDC20. CDC20 is indeed an activator of the Anaphase Promoting Complex / Cyclosome (APC/C). At this point separase is inhibited by securin.

B.The attachment of the kinetochores to microtubules (MT) generates a tension which inhibits the chekpoint inhibitor complex. As figured, the centromeric protein E (CENPE) is one of the signal transducing protein involved in silencing BUBR1 kinase activity through microtubules capture.

C.After silencing of the checkpoint inhibitory complex at each kinetochore, the cell can enter anaphase. APC/C(CDC20) inactivates securin which leads to separase activation. Separase resolves sister chromatid cohesion through removal of the cohesin complex. This allows sister chromatids to get disjoined towards each pole by the spindle.

The photographs on the left represent prometaphase, metaphase and early anaphase figures of HeLa cells. Chromosomes are counterstained in blue with DAPI. The red dots correspond to kinetochores immunolabelled with anticentromere antiserum (ACA).

Furthermore CDC14 activation allows secondarly to link anaphase onset to mitosis exit by inactivating mitotic cyclins (indirectly) / CDKs (directly) and activating the Mitosis Exit Network.

First, CDC14 indeed then dephosporylates Cdh1. This activates the second form of APC/C, APC(Cdh1). Activated APC(Cdh1) then targets mitotic cyclins and principally cyclin B1 for degradation. This leads to the inactivation of CDK1 and thus contribute to mitosis exit (Bembenek and Yu, 2001).

Second, CDC14 activates through CDC15 also the MEN (Mitotic Exit Network). This is essential to mitosis exit and ultimately leads to sustain MEN and CDC14 activations. CDC14 can finally inhibit the pool of mitotic CDKs by rapidly directly dephoshorylating them, leading in turn to cytokinesis and rapid resetting of the cell cycle at the beginning of G1 (Stegmeier and Amon, 2004).

CDC14 involvement in the FEAR and MEN networks is summarized in figure 15.

To conclude, through these two networks, CDC14 is a master regulator involved in the temporal and spatial coordinations between segregation of chromosomes, disassembly of the mitotic spindle, and cytokinesis to successfully complete and exit mitosis (Trautmann and McCollum, 2002; Geymonat *et al.*, 2002; Stegmeier and Amon, 2004; Torres-Rosell *et al.*, 2005).

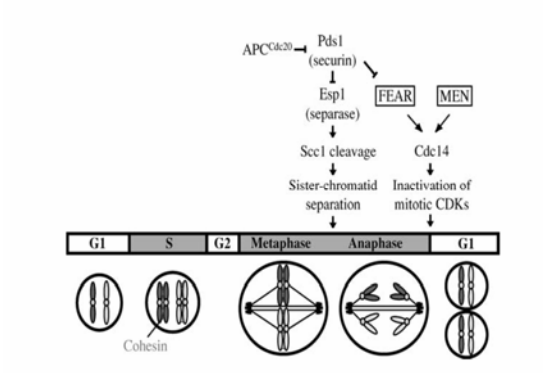
The yeast *CDC14* gene has two human homologs: *CDC14A* and *CDC14B*. Both of them seem to have functions that overlap yeast CDC14 ones. Hence, interference with the expression of CDC14A (but not CDC14B) either by down or up-regulation in two human tumour cell lines has been shown to induce diverse aberrant mitotic figures by disrupting both centrosomes separation and chromosomes segregation (Mailand *et al.*, 2002). This study indeed pointed CDC14A as a critical regulator of the centrosome duplication cycle, its mis-expression leading to the formation of multipolar spindles. CDC14B, the human paralog of CDC14A, has been less investigated but shown very recently to be able to bind, bundle and stabilize microtubules and to localize to centrosomes as well as to the central spindle in mitosis just as CDC14A (Cho *et al.*, 2005).

CONCLUSION

E2Fs, cyclins/CDKs and CDKIs play all together pivotal roles in the control of the whole cell cycle at numerous levels, through many subtle levels of regulations of both gene expression and protein activities. It seems a hard task to draw a complete integrated view of the global cell cycle regulation by these on the sole basis of current knowledge for several reasons. Among these are the functional redundancies and the antagonizing activities that can exist between several E2Fs at different points during the cell cycle, the only recent identification of some of their members and the still ongoing characterization and elucidation of some of their basic roles. A major obstacle is that cell cycle regulation by E2Fs, cyclins/CDKs, CDKIs and others imply a huge number of players as well as dynamic interplays between them. Moreover, these may change from cell type to cell type.

All of these questions anyhow represent major challenges at present. Their elucidation require better fundamental characterizations of the key regulators involved with respect of their normal physiological roles but also of their respective deregulations in cancer cells. These deregulations can indeed straightforward contribute to our understanding of cell cycle control.

A.



B.

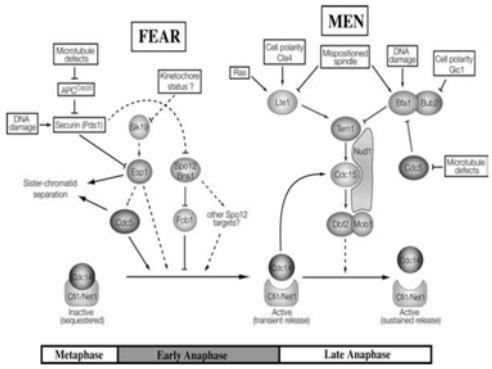


Figure 15. Regulation of mitosis in *S. cerevisiae*
Taken from Stegmeier and Amon, Annual Review of Genetics 2004.

A. Activation of CDC14 during anaphase is achieved through 2 protein networks: the FEAR (for CDC Fourteen Early Anaphase Release) and the MEN (for Mitotic Exit Network). Degradation of securin by APC/C induces the activation of the FEAR network and leads to CDC14 activation. CDC14 will mainly contribute to mitosis exit through inactivation of the mitotic CDKs.

B. Components of the FEAR and the MEN networks are presented together with the signalling events taking place between them during mitosis.

ii- Alterations of cell cycle regulators in cancer cells

Accordingly to the crucial roles of genes and proteins involved in normal cell cycle control, numbers of alterations of these have been described in human cancers. As already noted previously, *RB* was the first TSG to be isolated and loss of pRB causes cell proliferation deregulation in cancer cells. This part is thus dedicated to some studies that have established or participate to characterize some links between cell cycle regulators alterations and the cancer phenotype.

a) Alterations of E2Fs transcription factor family members

Members of the E2Fs transcription factors family have been found to be deregulated in a wide majority of cancers. The transcriptional program they regulate is prominent in human cancers (Rhodes *et al.*, 2005). Accordingly, this part is an overview of some important works underlining E2Fs involvement in cancers.

Oncogenic properties of E2F1 were shown in several independent studies (Johnson *et al.*, 1994) and are readily understood as *E2F1* encodes a transcriptional activator crucial to induce cell cycle progression. The picture is actually more complicated since tumours were shown to occur in mice models knocked outs for the *E2F1* gene and accordingly, this also gives a TSG status to *E2F1* (Weinberg, 1996).

Other E2Fs family members are slightly less characterized with respect to cancer since they were discovered more recently. Their putative involvement has been already assessed in a few examples and some evidences are reported hereafter.

It has been for example shown that transgenic mice with tissue specific promoter allowing high expression of E2F2 in the thymus develop thymic tumors, which demonstrates its oncogenic role *in vivo* (Scheijen *et al.*, 2004).

E2F3 locus was shown to be frequently amplified and consequently overexpressed in bladder cancers (Feber *et al.*, 2004; Oeggerli *et al.*, 2004). By immunohistochemistry, E2F3 protein overexpression was shown to be an independent predictor of poor clinical outcome in prostate cancers (Foster *et al.*, 2004). Of interest is also the finding that E2F3 inactivation leads to centrosome amplification, mitotic spindle defects and aneuploidy (Saavedra *et al.*, 2003). Importantly this study suggests that E2F3 could regulate the DNA replication and the centrosome duplication cycle.

E2F4 expression in invasive breast carcinomas is an independent predictor of poor prognosis (Rakha *et al.*, 2004). Its oncogenic role was also shown indirectly, since pituitary and thyroid tumours development in Rb heterozygous knockout mice is suppressed upon loss of E2F4 (Lee *et al.*, 2002).

E2F5 gene copy number has been shown to be recurrently increased in breast tumours, and oncogenic properties have been reported in cooperation with DP1 and activated RAS (Polanowska *et al.*, 2005). In ovarian tumours, E2F5 protein seem however to be down-regulated (Collins *et al.*, 2004)

E2F6 has been shown to be a negative regulator of the BRCA1 tumour suppressor gene (Oberley *et al.*, 2003). Nothing has been investigated concerning E2F6 direct role in cancers.

Finally, both E2F7 and E2F8 transcriptional repressors have been isolated recently and shown in the meantime to be able to block cellular proliferation upon ectopic expression (di Stefano *et al.*, 2003; de Bruin *et al.*, 2003; Maiti *et al.*, 2005; Logan *et al.*, 2004; Christensen *et al.*, 2005). This suggests a counteracting role for E2F-7 and E2F-8 to the cell cycle

promoting roles of the activating E2Fs. Further investigations are clearly needed to assess their putative involvement in cancers.

Proto-oncogenic activities have been shown for the two DP proteins (Jooss *et al.*, 1995). DP1 is a candidate driver gene of 13q34 amplifications in hepatocellular carcinomas (Yasui *et al.*, 2002) and its overexpression seems to be associated with progression of this cancer (Yasui *et al.*, 2003). In these tumors, DP2 and E2F3 have been shown to be upregulated (Liu *et al.*, 2003), as well as in chemically induced skin tumors in mouse with E2F1 to -5 (Balasubramanian *et al.*, 1999).

b) Alterations of Cyclins / CdK and CDKi

The *CDKN2A* (p16) gene is an established tumour suppressor gene located at 9p21, and allelic loss of *CDKN2A* has been retrieved in pre-malignant lesions in HNSCC (Califano *et al.*, 1996). Alterations at 9p21is indeed the most frequent event and may concern around 75% of these tumours (Gollin, 2001).

Cyclin D1 is an oncogene believed to be the driver oncogene of 11q13 gains and amplifications recurrently observed in SCC (Gollin, 2001).

A recent study assessed the prognosis implications of cell cycle (among others) biomarkers in Non Small Cell Lung Cancer by a wide and rigorous transversal literature analysis of numerous reported cases (Singhal *et al.*, 2005). The stronger evidences for a link with a poor prognosis were found for tumours with high expression of cyclin B1 and cyclin E. On the contrary, strong evidences for a link with favourable prognosis exist for tumours with high expression of p16, p21 and p27 CDK inhibitors (of the D-type associating CDK-4/6).

We have initially identified *Cyclin L1* as highly amplified and overexpressed in a head and neck cancer cell line and overexpressed in a series of tumours (Redon *et al.*, 2002). In a recent retrospective study, high levels amplifications of the *CCNL1* gene were found to concern a very few head and neck tumours but correlated with a poor prognosis (Sticht *et al.*, 2005).

CONCLUSION

The normal cell cycle is a highly controlled cellular process involving numerous sharpely tuned molecular interactions and regulations between the players involved. Accordingly, cell cycle deregulations represent one of the common features of tumour cells. Molecular regulators of the cell cycle underlie these alterations mainly through their impaired expression.

.

a-3 microRNAs : new critical contributors to the malignant phenotype of cancer cells ?

i- Introduction: a new class of non protein coding genes

MicroRNAs encoding genes represent a recently discovered and new class of genes as the first microRNA, lin-4, was identified in 1993 through studies of developmental defects in *C. elegans* (Lee *et al.*, 1993). But it is only very recently that several hundreds of microRNA genes were uncovered (Lagos Quintana *et al.*, 2001; Lau *et al.*, 2001; Lee and Ambros, 2001). MicroRNAs have been found until now in virus, plants, fungi and animals.

In mammalian cells, microRNA encoding genes are indeed believed to be transcribed by RNA polymerase II, into a hairpin precursor RNA, called "pre-miR", which is 70-100 nucleotides in length. This pre-miR is then further exported to the cytoplasm, where it is processed and cleaved to a small double stranded RNA of approximately 19-22 nucleotides length by the Dicer protein (a ribonuclease of type III) through the RISC complex. These dsRNA duplexes are very short-lived since one strand is rapidly degraded leaving the other strand as the mature microRNA, or "miR" available to be incorporated into the RISC complex to exert its function (Bartel, 2004; He and Hannon, 2004). These processes leading to microRNA biogenesis are schematized in figure 16A.

Several recent evaluations of the number of microRNA genes in the human genome seem importantly to point out some hundreds of them. At least, between 250 and 300 have been cloned and sequenced until now. More recent estimates even predict the existence of a thousand of them (Berezikov *et al.*, 2005).

The growing numbers of identified and isolated microRNAs in several species and of their respective putative targets has led to the establishment of a comprehensive microRNA registry (Griffiths Jones, 2004) which is now available as a database from the Sanger center, and called "miRBase" (http://microrna.sanger.ac.uk). Numerous other helpful online tools dedicated to microRNAs targets prediction are also available (as for example: http://www.microrna.org/ or http://pictar.bio.nyu.edu/).

ii- Molecular and physiological functions of microRNAs

a) Molecular Function

Their fundamental molecular function is to negatively regulate the expression of targets genes at the post-transcriptional level. This level of regulation is achieved through association by hybridization of a given microRNA to complementary sequences, or target sites, that are found in its target mRNAs.

Empirically determined rules seem to predict that these target sites are preferentially located in 3'UTR of the mRNAs, and furthermore that the binding is not necessary over the whole sequence of the microRNA but may be particularly important in its 5' sequence. Exemples of such associations are given in figure 16B for the founding member of the microRNA familiy, lin-4, binding to different sites in the 3'UTR of its target mRNA, lin-14. Binding of a microRNA to its target mRNA blocks protein expression either by inhibiting mRNA translation or by inducing its degradation (Bagga *et al.*, 2005; Pillai *et al.*, 2005). All these ways of achieving microRNA function are schematized in figure 16A.

A. B.

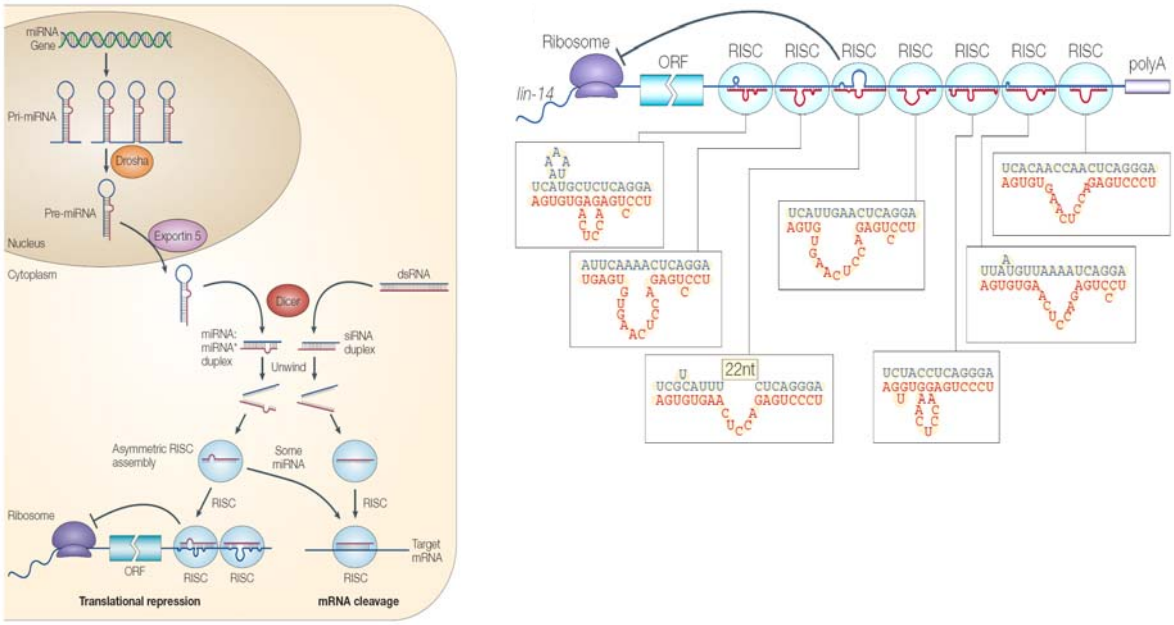


Figure 16. MicroRNA biogenesis and molecular interactions with their targets mRNAs
Taken from He and Hannon, Nature Reviews Genetics 2004.

A. MicroRNA encoding genes are transcribed to pre-microRNA (or pre-miR or pre-miRNA) that form hairpin structures and are exported to the cytoplasm. They are cleaved to microRNA by the Dicer protein and incorporated in RISC (RNA-induced silencing complex). MicroRNAs can finally bind to their target sites by complementary hybridizations to these in the 3'UTR region of their target mRNAs. This binding will ultimately lead to inhibition of protein expression either through translational repression (lower left part) or induction of mRNA degradation (lower right part).

B.Examples of associations of the founding member of the microRNA family, lin-4 (sequences in red) and its target mRNA (blue), lin-14. Seven sites in the 3'UTR of lin-14 mRNA are targeted by the lin-4 microRNA which binds to them in different manners leading to the repression of LIN-14 protein synthesis.

It has been postulated that microRNAs could potentially regulate up to one third of human mRNAs (Lewis *et al.*, 2005). The level of post-transcriptional regulation by microRNAs is even more complex since a given target mRNA could be regulated by several microRNAs at the same time (Krek *et al.*, 2005). This is believed to be a way of widespread fine tuning of genes expression (Bartel and Chen, 2004). In any case, it seems clear that each microRNA could regulate more than a single target at once, and even around 200 different targets (Krek *et al.*, 2005). A current view is that a given animal microRNA could fine tune the expression of hundreds of targets but only dramatically reduce a few (Bartel and Chen, 2004).

Typically, softwares that have been designed for the bioinformatic prediction of microRNA targets result in tens or hundreds of candidates for a given microRNA. As an example, the miR-28 microRNA encoding gene, located at 3q28 and that we began to characterize in this thesis, is predicted to have 638 targets genes (http://www.microRNA.org) for around a thousand of transcripts.

Thus, by regulating the expression level of several distinct proteins, microRNAs are believed to exert pleiotropic effects on cell physiology. It is from that clearly tempting to speculate that a given microRNA could regulate several targets involved in a same cellular process. This mechanism could provide a coordinated regulation of the expression of different proteins acting for examples in a given signalling pathway or cellular process.

b) Physiological functions

Examples of essential physiological microRNA functions include important developmental processes by effects on underlying fundamental cellular mechanisms such as differentiation, division, growth and death. For exhaustive reviews of their uptodate known roles, please see Bartel, 2004; He and Hannon, 2004; Harfe, 2005; Miska, 2005.

As an example, microRNAs of the let-7 family have been shown to be critical players of development timing in *C. elegans* (Abbot *et al.*, 2005). Furthermore, some miRNAs seem interestingly to be preferentially expressed in stem cells (Houbaviy *et al.*, 2003; Suh *et al.*, 2004) where they may have important regulatory roles. Moreover, a recent study showed an implication of microRNAs in the regulation of the cell cycle of germinal stem cells, or GSC (Hatfield *et al*., 2005). Analysis of Drosophila GSC, knocked-out for the Dicer gene, and thus unable to produce any mature and functional microRNA, revealed a specific defect in GSC cell cycle progression during G1/S transition. This indicates at least that one or more microRNAs are required for proper stem cell cycle progression in the GSC. Nevertheless it should be underlined that the genetic model used is equivalent to a knock-out for a huge number of different genes at once.

iii- microRNAs as TSG and oncogenes

Given their already shown crucial role in several processes, as well as the putative regulation of several targets by each microRNA, the question of their involvement in cancer is of compelling interest. Their deregulation could indeed be a shortcut to inhibit several genes/proteins at a glance: deregulation of a single microRNA encoding gene could have lots of consequences on cell physiology.

Several evidences for microRNAs being deregulated in human cancers have accordingly already been described, both for haematological malignancies and solid tumours.

The miR-15 and miR-16 genes, at 13q14, were found to be both frequently deleted and down regulated in CML (Calin *et al.*, 2002). Other microRNAs were found to be reduced during colorectal neoplasia (Michael *et al.*, 2003). Most striking was the finding that reduced levels of let-7 microRNAs in lung cancers seem to be well associated with a poor prognosis for the patient (Takamizawa *et al.*, 2004). It turned out that the let-7 microRNAs actually regulate the N-Ras and K-Ras oncogenic protein levels (Johnson *et al.*, 2005).

MicroRNAs expression profiling revealed different signatures in B-CLL samples (Calin *et al.*, 2004). This was recently prolonged by a study showing a microRNA signature associated with prognosis and disease progression in these tumours (Calin *et al.*, 2005). Moreover this latter study showed that mutations are, as for other genes, found in microRNAs transcripts and may be of importance with respect to the tumoural phenotype.

It was strinkingly shown a few months ago that the profiling of microRNAs expression enables to classify human tumours of various localizations (Lu *et al.*, 2005). Likewise, microRNA aberrant expression patterns in breast tumours have been reported, allowing first to distinguish the tumours from the normal tissue based on the expression pattern, and second to identify microRNAs whose expression is correlated with bio-pathologic features of the tumour, such as hormones receptors status, stage, proliferation index and vascular invasion (Iorio *et al.*, 2005).

A publication reporting the identification of a microRNAs encoding gene as an oncogene came out very recently also from a study of B cell lymphomas (He *et al.*, 2005). The involved gene is mapped to 13q31 and recurrent genomic amplifications of this locus were already described in this tumour type. Based on expression studies of the genes lying in the core amplicon, the putative driver gene was identified as *C13orf25* (Ota *et al.*, 2004). It turned out that this *C13orf25* gene is in fact a non protein coding gene, transcribed to a polycistronic RNA, and thus recalled the *miR-17-92* polycistron and proposed by the authors as "onco-miR-1" the first microRNAs encoding oncogene, which codes for 5 differents microRNAs (He *et al.*, 2005). An independent study published at the same time shows that c-Myc regulate the expression of this microRNA cluster and that 2 out of the 5 encoded microRNAs, namely miR-17-5p and miR-20, regulate the cell cycle related E2F1 protein expression (O'Donnell *et al.*, 2005). This miR-17-92 polycistron is furthermore over-expressed in lung cancers and enhances lung cancer cell proliferation (Hayashita *et al.*, 2005).

Another "onco-miR", miR-155 (encoded by the BIC non coding RNA) has also been proposed as an oncogene in B cell lymphomas (Eis *et al.*, 2005; Tam and Dahlberg, 2005).

In conclusion, for all of the above reasons, microRNAs encoding genes have now to be "sized up to cancer genes" (Caldas and Brenton, 2005). As such, their contribution as putative oncogenes or TSG is of high interest and has to be assessed in tumors (Chen, 2005; figure 17).

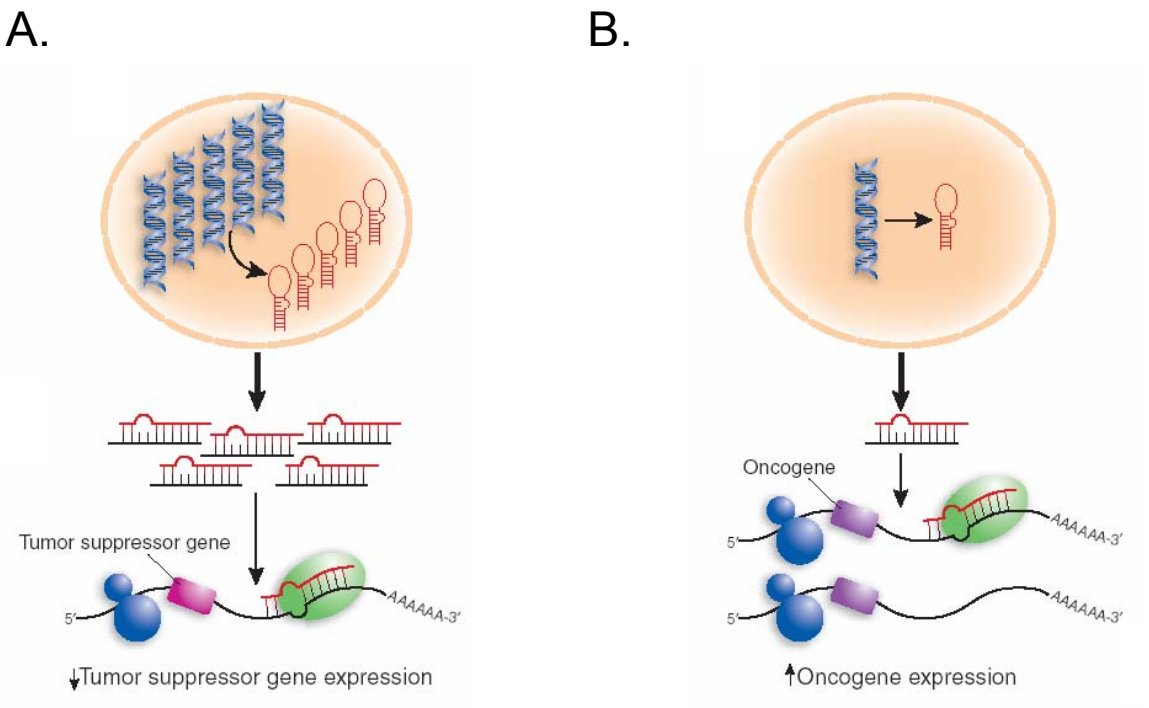


Figure 17. MicroRNA encoding genes as putative oncogenes and tumour suppressor genes. Taken from Caldas and Brenton, Nature Medicine 2005.

A. microRNA encoding genes have to be envisioned as putative oncogenes. Two microRNA encoding genes have already been shown to be amplified in lymphomas (numerous gene copies in blue representing gene amplification) and consequently over-expressed (numerous pre-microRNA red momecules representing over-expression consequent to gene amplification). Ultimately these « onco-miR » may downregulate the expression of tumour suppressor proteins and thus contribute to tumourigenesis.

B.Conversely, examples of microRNAs encoding genes as putative TSG have already been demonstrated. These may be deleted (only one gene copy in blue) or inactivated by mutations which will lead to the expression increase of oncogenic proteins. Accordingly tumour suppressor microRNAs may also contribute to carcinogenesis.

b- Modelling cancer by manipulating genes and their expression

Trying to define rules for making human tumour cells <u>in vitro</u>

Long standing efforts have been made to try to understand cellular transformation by trying to reproduce it in primary cells maintained under cell culture conditions. First important works were mainly done in rodent cells, which seem to require the sole activation of two cooperating oncogenes such as *c-myc* and *c-ras* (Land *et al*, 1983; Ruley, 1983).

Similarly, several attempts have been made in the past years to try to define the set of altered genetic elements that would lead to transform human non tumorigenic cells *in vitro*. Lots of efforts in this direction were unsuccessful, mainly and presumably due to one of the fundamental difference between human and murine somatic cells, that is their respective telomere biology. First impressive demonstration of human primary cell transformation was the work of W.C Hahn in R.A. Weinberg team, showing tumorigenicity induction in human primary epithelial cells as well as fibroblasts (Hahn *et al.*, 1999). They succeeded in experimentally transforming these cells by ectopic expression of a set of previously known or suspected human oncogenes upon retrovial introduction: the Simian Virus 40 Large T antigen encoding gene, or *SV40 LT*, in combination with *hTERT*, the gene encoding the catalytic subunit of human telomerase and finally an oncogenic allele of *H-Ras* (that is mutated, and consequently constitutively active). These elegant results were then further questioned on their interpretation by the group of P. Duesberg, since it seemed that the analyzed transformed cells rapidly became aneuploïd (Li *et al.*, 2000). One can in addition ask the rationale of ectopically over-expressing a mutated oncogene since this is surely not relevant for the *in vivo* observed situation (Hua *et al.*, 1997).

Further investigations in the team of R.A. Weinberg finally showed the direct derivation of cells from their previous work that were capable of inducing tumorigenicity without any widespread genomic instability (Zimonjic *et al.*, 2001). Once more, these results and their interpretation were contested and highly debated on the basis of the presence of chromosomal aneusomies and hybrid chromosomes in the cells when analyzed by the group of P. Duesberg to whom was rationally proposed in reply the explanation that improper cell handling could have been responsible of the described aberrations (Li *et al*., 2002 and reply of the authors, 2002).

Recent progresses have anyhow been made. A more comprehensive study investigated and reported species and cell type specific molecular requirements for transformation of the respective cell types (Rangarajan *et al.*, 2004). This *in vitro* study elegantly showed that the p53 pathway is necessarily inactivated in mouse and human cells to achieve transformation but these cells differ in the supplementary other events that are needed. As previously known, one single other event can lead to murine cell transformation whereas human ones need several additional events. In human cells, these subsequent events seem clearly different from cell type to cell type, suggestive of parallel pathways that may lead to transformation in vivo depending on the the cell type transformed. For a more extensive review of recent works in this field, see Boehm and Hahn (2005).

Clear limits of these studies aiming at the definition of a set of genes rely first on the use of cells in *in vitro* culture conditions which may very strongly differ from normal cells that have to be transformed when located in their normal tissue architecture to give rise to a malignant cell. Secondly, on the use of sustained aberrant oncogenes expression manipulation

at uncontrolled and very high level, which may not be relevant to more transitory and subtle gene dosage variations that are more likely to occur *in vivo*.

Taken together with the lessons we got from *in vivo* studies of tumour cells, these studies allow anyhow us to draw an integrated view on the cellular pathways and molecular functions that are important to be altered in a given cell to undergo transformation. Importantly, they give new insight to the understanding of cellular functions that can be altered at several levels in different cell types. From that, they finally allow to model try to deduce how this could be achieved in pre-neoplastic cells in vivo.

As a concluding remark, one should mention the extensive use of animal models since a few years that provides great and further insights to the understanding of how tumorigenesis may initiate and cancer progress *in vivo*. Among different models, the use of mice to model cancer is of practical use since gene knock-out, knock-in and conditional expression technologies have been developed and are easily used. Mouse models of lung cancers have been recently reviewed for example: Meuwissen and Berns (2005). It should be anyway underlined that, as already noted in the previous part, mouse cell and body normal physiology may significantly differ from human ones. This seems even to be more particularly true as regards cancer: commonly used laboratory mice strains do naturally and predominantly develop lymphomas, various sarcomas and leukemias tumours whereas the situation is completely different in Humans where carcinomas (*i.e.* epithelial tumours) are from far the more frequent type arising (Rangarajan and Weinberg, 2003; Vogelstein and Kinzler, 2004; Anisimov, 2005). As a consequence, lessons from the study of tumorigenesis in mice have also to be carefully considered.

c. Accumulation of mutations during tumour progression

c-1 A "mutator" phenotype acting as a tumourigenic catalyzer ?

Once a cancer cell has been initiated and present genetic instability, it will give rise to a progeny presenting genetic variation. These daughter cells will then very likely accumulate randomly further mutations. It has been proposed that a primarily acquired mutator phenotype may then act as a tumourigenic catalyzer (Loeb, 1991), which through an increase in the mutation rate in cancer cells will increase the likelihood of acquisition of new mutations that will confer a given cell with a growth advantage over the others. Whether tumour cells do in general present increased point mutation rates is still a matter of conflict (Loeb *et al.*, 2003).

As noted previously, defects in cellular DNA repair machineries can account for the accumulation of point mutations in cancer cells genomes. As an example, a 2 to 3 fold increase in the mutation rate is observed in MIN colorectal cancer cells (Rajagopolan *et al.*, 2003). MIN is however observed in approximately only 15% of colorectal cancers, whereas it is from far less frequent (<2%) in other tumour types (Lengauer *et al.*, 1998). Taken globally, this suggests that point mutations may accumulate in most of tumours cells but at more or less the same rates than in normal cells in most cases.

CIN has been proposed to explain the consistent aneuploïd state of cancer cells observed in the remaining majority of colons cancers (Rajagopolan *et al.*, 2003). Aneuploïdy is moreover observed in a wide majority of cancer types. However it is still unknown to what extent CIN may contribute to aneuploidy *in vivo*. Recent evidences in breast cancer suggest a telomere crisis at very early stages of cancer progression that may lead to transient chromosomal instability (Chin *et al.*, 2004).

Evidences that mutations are accumulating in cancer cells, both at sequence and chromosomal levels, have been clearly described for a long time. Apart for the tumours known to present MIN or for the cell lines for which CIN has been experimentally observed, these point mutations and chromosomal aberrations can at present not be attributed to a general increased mutation rate or so-called mutator phenotype in cancer cells. They may actually result also from other causes such as exposure to carcinogens for example.

The two following sections are accordingly dedicated to some mutated genes that have been found in tumour cells on one hand, and to chromosomal aberrations that have been described on the other.

c-2 At the NUCLEOTIDE level: mutated genes in cancer cells

Both point mutations inactivating TSG and gain of function ones for oncogenes have been widely reported. These mutations will confer the cancer cell in which they arise with a best fitted phenotype towards malignancy, by allowing the cell to escape apoptosis or to proliferate more rapidly for example. Consistent with that, numerous mutated TSG and oncogenes have been reported in tumour cells until now, leading to an estimate of 291 "cancer genes" (around 1% of human genes) that can be found mutated in tumour cells and are causally implicated in cancer development (Futreal *et al.*, 2004).

High-throughput sequencing approaches allow since recently for extensive and comprehensive analyses of gene families sequences in tumours cells. Recently, such large scale studies were for examples applied to the tyrosine kinome and phosphatome in colorectal cancers (Bardelli *et al.*, 2003; Wang *et al.*, 2004b; Parsons *et al.*, 2005). These analyses of genes coding for proteins involved in crucial cellular functions, such as tyrosine kinases and phosphatases suggest that a significant proportion of tumours may harbour a mutation in one of the numerous genes analyzed. This is not surprising since the most common domain in cancer genes is a kinase domain (Futreal *et al.*, 2004).

Similar approaches were recently also applied to the kinome genes in breast (Stephens *et al.*, 2005), lung (Davies *et al.*, 2005), and testicular germ cell tumours (Bignell *et al.*, 2006). These reports underline that mutations in the kinome of lung tumours is higher than in breast and testicular cancers but they seem to be distributed over several protein kinases in different tumours. No obvious candidate kinase encoding gene is designated as recurrently mutated. An excess of C:G>A:T transversions was observed, compatible with mutagenic effect of tobacco smoke in lung tumours.

All together, these reports however clearly underline an overall low mutation frequency in the tumours cohorts examined, except for some samples. It seems from these data that the mutator phenotype at the nucleotide level is thus not a general feature of tumour cells. In any case, these studies are of high interest for two main reasons. First by their extensiveness appraisal of a huge amount of genes sequences that encode proteins involved in phosphorylations and dephosphorylations which are fundamental cellular processes, previously known to be highly deregulated at several levels in cancer cells (Blume-Jensen and Hunter, 2001). Secondly, by their integrative approach in trying to elucidate functional redundancies of mutations in a given cellular pathway, which should be applied to other specific pathways in the future.

The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website have been built up recently to cure and make an available repertoire of these somatic mutations that can be found in human cancers (Bamford *et al.*, 2004). It is available from the Sanger Institute website at: http://www.sanger.ac.uk/genetics/CGP/cosmic/

Finally, several genes involved in chromosome segregation mechanisms have also been found mutated in various tumour cells (Draviam *et al.*, 2004). Among these, as noted in a previous part, several CIN causing genes have been identified but are only very rarely mutated in sporadic tumours. Actually, it seems that among the genetic aberrations that can be observed in cancer cells some chromosomal aberrations belongs to the most recurrent ones.

c-3 At the CHROMOSOMAL level : chromosomal aberrations in cancer cells

As noted previously, it is still hotly debated whether chromosomal aberrations could be initiating events and even driving forces of carcinogenesis. A majority of cancer cells do present chromosomal aberrations in advanced tumours. Some chromosomal aberrations have been found to be recurrent among tumours (Gebhart and Liehr, 2000) and thus may represent important features of cancer cells with major impact on cellular physiology and contribution to the tumoural phenotype (**see hereafter for example of the impact of some chromosomal aberrations on the transcriptome**). As such, efforts have been and are made to fully understand the mechanistic of their formation and their relevance to the malignant phenotype of the cells harbouring these aberrations.

i- Different types of chromosomal aberrations

Chromosomal aberrations observed in cancer cells can be basically divided into two categories: numerical and structural aberrations (Heim and Mitelman, 1995, Lengauer et al., 1998; Pihan and Doxsey, 2003).

Numerical aberrations refer to **gains and losses of whole chromosomes or large chromosomal regions** whereas **structural aberrations** refer to **translocations, inversions, insertions, deletions and amplifications of chromosomal material**. This is formal distinction since some structural aberrations can result in net gain or losses of chromosomal segement (as deletions or amplifications).

It has been proposed that both types of aberrations occur continuously in tumour cells and thus reflect a high degree of cancer cell genomes plasticity, induced by their intrinsic genomic instability. As such, they have been referred to as numerical and structural CIN (Pihan and Doxsey, 2003). It should be however noted that this appellation seems abusive and at least differing from the original definition of CIN (which refers to an accelerated rate of gains and losses of whole or large portions of chromosomes; Rajagopolan *et al.*, 2003), and used by others since then and which may be putatively involved in cancer initiation.

a) Numerical Aberrations

Gains and losses of whole chromosomes have been largely described in tumour cell karyotypes since they are easily cytogenetically detectable. It should be noted that the strict definition of aneuploïdy refer to such whole chromosomal gains and losses.

b) Structural Aberrations

Relocation of genetic material: Translocations, Insertions and Inversions

Translocations are DNA rearrangements that can be either **non-reciprocal** (unbalanced, involving a loss of a chromosomal segment) or **reciprocal** (balanced, neither gain nor loss of DNA). Importantly, non balanced translocations represent the most frequent aberration observed in human carcinomas (Gisselsson, 2003).

Another type of interchromosomal rearrangement corresponds to DNA **insertion** from a chromosomal segment into another. On the contrary, **inversions** are intrachromosomal aberrations.

These 3 types of aberrations have been shown to exert some of their **effects by oncogene activation**, either by creating chimeric genes encoding proteins with new "gain of function", or by activating oncogene through impaired expression regulation, **or by TSG inactivation,** either by gene breakage or loss of expression (Heim and Mitelman, 1995).

Copy Number changes of genetic material: Deletions, Isochromosomes and Amplifications

Interstitial deletions have clearly been documented as a way for a cancer cell to get rid of a functional allele of a TSG by simple physical elimination of it. Noteworthy, they can also have positional effects through the expression regulation of genes that are neighbouring the deleted segment (Pihan and Doxsey, 2003).

Isochromosomes

Gains and losses of large genomic regions such as entire chromosomal arms have been also characterized. Isochromosomes are among the more recurrently described structural aberrations present in tumour karyotypes (Jin *et al.*, 2000): they consist in a radially symmetric chromosome with a centromere linking two identical duplicated arms. Thus such a chromosome is constituted of two originally short or long arms of the chromosome, and the other arm get lost in the forming event.

They may arise through unbalanced translocation and constitute the majority of recurrent structural changes observed in skin and HN SCC for example (Jin *et al.*, 2000). Their formation may require centromere cleavage (Jin *et al.*, 2000) and it has been shown recently that the loss of a single telomere, apart of inducing global genomic instability, may directly provoke isochromosomes formation (Sabatier *et al.*, 2005). It is still unclear whether some originally thought isochromosomes from cytogenetic observations could in fact represent dicentric or pseudo-dicentric chromosomes that arose thanks to pericentromeric breaks. Whole arm translocations are more frequent in SCC than in adenocarcinoma and seem to involve centromeric cleavage (Hermsen *et al.*, 2005). Moreover, in most of SCC cases, the chromosome contains material from both partners suggesting that the formation of a pseudoisochromosome could lead to the loss of the other arm (Hermsen *et al.*, 2005). In this perspective, the most frequent gain in HNSCC is distal 3q gain and the most frequent loss concerns the 3p arm. This is believed to be achieved in many cases through isochromosome 3q formation (Gollin, 2001). An interesting correlated observation is that several genes located on the 3p arm have been characterized and proposed as TSG in SCC and others tumour types, whereas on the contrary, several proposed oncogenes have been shown to map on the 3q arm.

Gene amplifications

(Hogarty MD and Brodeur GM, Gene amplification in Human Cancers: Biological and Clinical Significance in The Metabolic & Molecular Bases of Inherited Disease, 8th edition)

Usually gene amplifications affect small sub-chromosomal genomic segments, which are referred as to **amplicons**, that can range in size from around 100kb to a few megabases.

From a structural point of view, gene amplifications have been described as intrachromosomal (Homogeneously Stained Region, or HSR) or extrachromosomal (double minute, or dmin chromosomes). It seems however that the predominant form of amplicons observed in tumour cells are dmin chromosomes (Benner *et al.*, 1991). This may possibly vary depending on tumour type and even tumours.

Cytogenetically observable examples of both dmin chromosomes and HSR are given in figure 18**.**

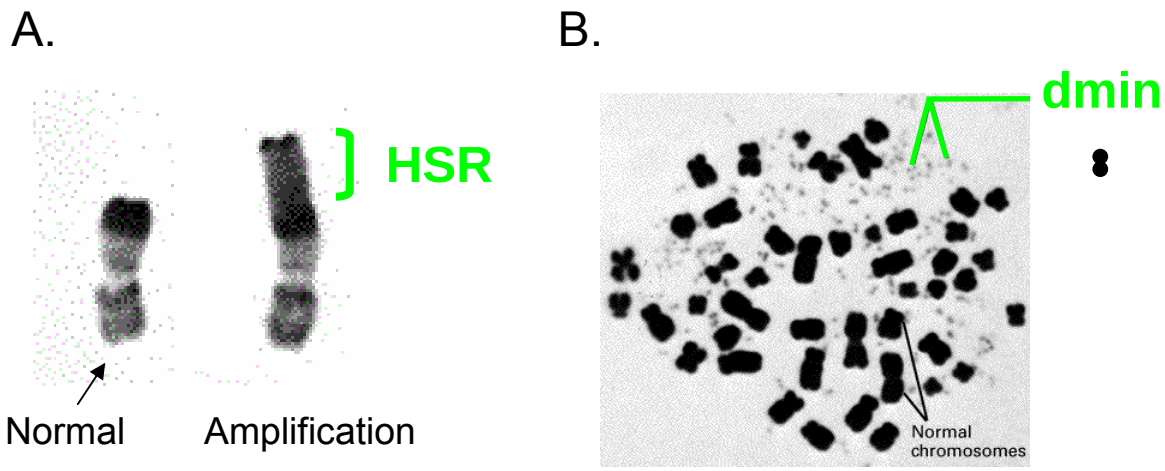


Figure 18. Cytogenetic examples of gene amplifications

A. The normal chromosome is represented on the left. On the right the homologous chromosome bears an amplification taking the form of a Homogeneously Stained Region (HSR) at its extermity.

B. Metaphase chromosome spread from a cell with an amplification taking the form of double minute (dmin) chromosomes. These dmin chromosomes are numerous and interspersed throughout the spread. Normal chromosomes are black due to light over-exposition to distinguish the dmin chromosomes. On the right is figured the typical appearance of dmin chromosomes at higher magnification (from which their name was derived)

ii- Models and mechanisms for the formation of chromosomal aberrations

a) Aberrant segregation of chromosomes during mitoses: towards aneuploïdy, the formation of numerical aberrations

As noted in the historical perspective, D. Hansemann already observed aberrant mitoses in tumour cells more than a century ago. Since then, recurrent observations that cancer cells do for example present mitotic spindle and centrosomes abnormalities that can generate numerical chromosomal aberrations have been made. As shown in figure 19, four main ways can lead to whole chromosome gains and losses all naturally linked to the handling of chromosomes during mitosis (Kops *et al.*, 2005): aberrant mitosis *per se*, chromosomes cohesion defects, improper centromere attachment and mitotic checkpoint deficiencies.

First, aneuploïdy may occur due to **aberrant mitotic divisions**; which can produce **polyploidizations** and **centrosomes amplifications**. The question of whether centrosomes aberrations could be causative of carcinogenesis, even if likely in some cases, is still debated (Nigg, 2002). Centrosomes amplification lead to the formation of a multipolar spindle which then provokes mis-segregation of chromosomes in the polyploid genome. **Defects in centrosomes duplication, segregation and maturation** without any prior aberrant mitosis can also lead to the formation of multipolar spindle and provoke subsequent aneuploïdy (Nigg, 2002; Kops *et al.*, 2005).

Second, **defects in chromosome cohesion** can also lead to aneuploïdy. Chromosome cohesion is ensured by the cohesin complex (SMC1 and -3 proteins), and resolved by a protease, called separase, at anaphase onset. Separase is previously inhibited by securin and deregulations of these proteins have been shown to lead to aneuploïdy.

Third, an **improper attachment of chromosomes to the spindle** can also lead to aneuploïdy. With this respect, merotelic centromere attachment (a kinetochore that is attached to the microtubules from both poles) has been shown causative of aneuploïdy (Cimini *et al.*, 2001).

Very recently, CDC14 was shown to play a crucial role in stabilizing the attachment of spindle microtubules to the chromosomes in yeast (Higuchi and Uhlmann, 2005). The APC (Adenomatous Polyposis Coli) protein seems to be an additional important player (Kaplan *et al.*, 2001; Fodde *et al.*, 2001).

Fourth, as exemplified by the CIN causing genes that can be found mutated in some tumour cells, **mitotic checkpoints defects** also engender numerical aberrations formation, through too early onset of anaphase in checkpoint deficient cells (Kops *et al.*, 2005).

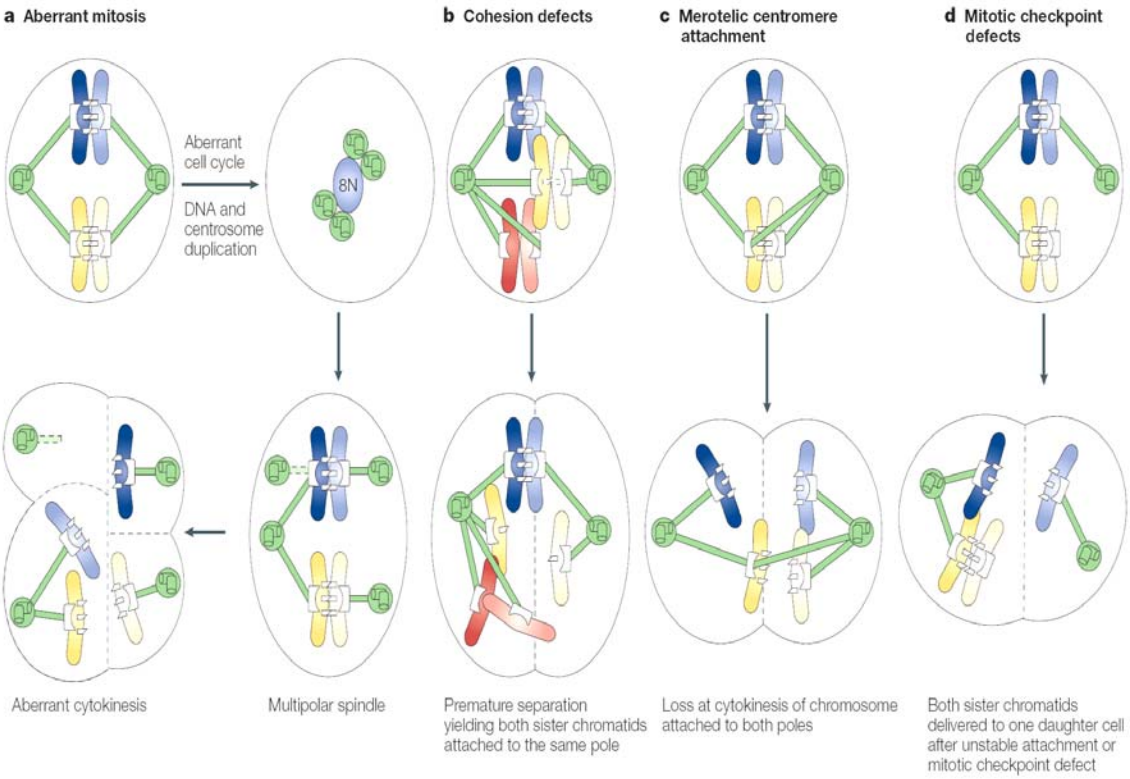


Figure 19. Formation of numerical aberrations (aneuploidy).
Taken from Kops *et al*., Nature Reviews Cancer 2005.

A. Multipolar spindles may lead to whole chromosome gains and losses. This can be achieved through previous aberrant mitosis leading to polyploidy or centrosome aberrations.

B. Chromatids cohesion defects can produce aneuploidy through premature loss of cohesion and subsequent mis-segregation of sisters chromatids.

C. Improper attachment of the kinetochore to both poles as figured (merotelic attachment) can provoke subsequent segregation defects and chromosomal loss at cytokinesis.

D. Mitotic checkpoint defects can cause too early anaphase onset and thus cause chromosomal gains and losses.

b) DSB breaks and telomere dysfunction: towards the formation of structural and numerical aberrations

As already noted previously, carcinomas do typically exhibit extensive genome scrambling. It is commonly admitted that the main causes are double stranded DNA (dsDNA) breaks that did not get repaired correctly as well as telomeres dysfunctions (Pihan and Doxsey, 2003).

DNA double stranded breaks and eroded telomeres ends are indeed unstable and can be fused to heterologous chromosome ends and thus may generate dicentric chromosomes or ring chromosomes which can initiate a "Breakage-Fusion-Bridge" (BFB) sequence during subsequent mitosis (Pihan and Doxsey, 2003). BFB (and associated BFB cycles) can generate extensive genome scrambling but among the many rearrangements it may lead to, they can notably generate non balanced translocations easily. Interestingly, these aberrations represent the most common aberration observed in human carcinomas (Gisselsson, 2003). The principle of BFB cycles is presented in figure 20. Importantly, B. McClintock, after having first observed ring chromosomes that may have arise after irradiations through telomere fusions, was the initial contributor in the late 1930's to BFB description through her study of irradiated strains of maize. She was the recipient of the 1983 Nobel Prize for her work on mobile genetic elements, or transposons.

The extensive rearrangements arising through BFB cycles in tumour cells already presuppose a loss of correct DNA repair and different cell cycle checkpoints to occur (as p53 inactivation). These are really more likely to occur in tumour cells of advanced tumour stages and this may partly explain the late occurrence of gene amplifications with respect to tumour progression in several tumour types (Ried *et al.*, 1999). Common fragile sites can have an influence for breakage near oncogene genomic localizations and lead to their subsequent amplifications through BFB cycles as exemplified for the amplicon of the *MET* oncogene in a gastric carcinoma cell line (Hellman *et al.*, 2002). Experimental works in specific mice models have underlined the importance of dsDNA breaks and their correct repair: Non Homolgous End Joining (NHEJ) and p53 deficient mice do develop lymphomas associated with unbalanced translocations that lead to subsequent gene amplifications through BFB cycles (Zhu *et al.*, 2002; Difilippantonio *et al.*, 2002). The NHEJ machinery is essential for the repair of dsDNA breaks naturally occuring during "V(D)J" recombinations allowing the assembly of antigen receptor variable region genes from different gene segments in developing lymphocytes. Finally, different studies point to a critical transient period of genomic instability during early breast cancer progression: the transition from hyperplasia (hyperproliferative epithelium) to the formation a carcinoma could be driven by a short telomere crisis associated with an increase in chromosomal aberrations and anaphase bridges (Chin *et al.*, 2004).

The mechanisms of BFB cycles can easily explain how gene amplification may occur in tumour cells by taking the form of intrachromosomal HSRs in a head to head conformation of the amplified sequence as shown in figure 21. But recent analyses of amplicons structures have pinpointed that these can be structured *in vivo* in a really more complex fashion (Volik *et al.*, 2003). Moreover, BFB cycles do not explain how other intrachromosomal amplicons in head to tail conformations and extrachromosomal ones (dmin chromosomes) do occur.

An experimental work has interestingly shown that HSR can lead to the formation of dmin chromosomes upon fragile site breakage induced by stresses such as hypoxia (Coquelle *et al.*, 2002). Another experimental study has shown that a circular DNA plasmid bearing a

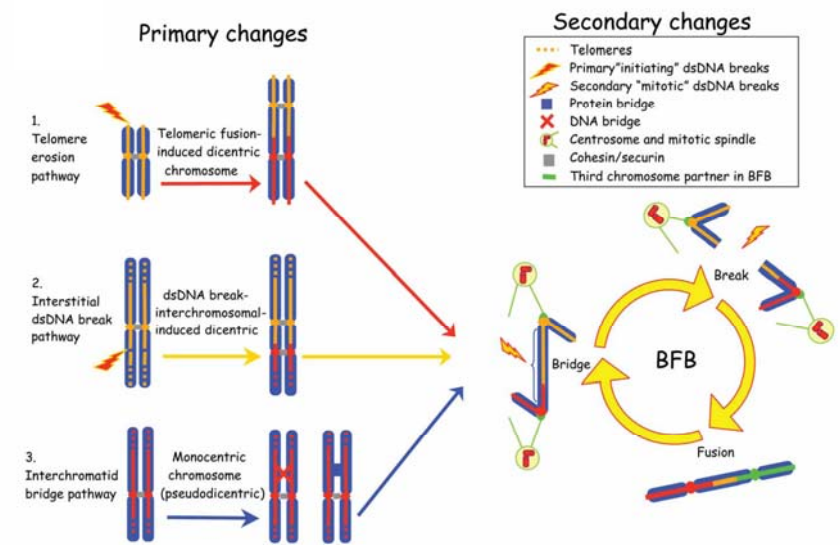


Figure 20. Formation of structural aberrations through Breakage-Fusion-Bridge (BFB) cycles.
Taken from Pihan and Doxsey, Cancer Cell 2003.

Primary chromosomal changes (telomere erosion, dsDNA breaks) may lead to fusion of unprotected chromosomal extremities. Acquisition of secondary chromosomal changes takes place during next mitotic division. At anaphase, the dicentric chromosome will form a bridge inbetween the two poles and will get broken. The newly broken chromosome (red and orange) may subsequently fuse with another (green) and generate a new dicentric chromosome which will form a new bridge at next anaphase. BFB cycles may hence take place and produce extensive genomic rearrangments. Importantly, they produce unbalanced translocations in a single round.

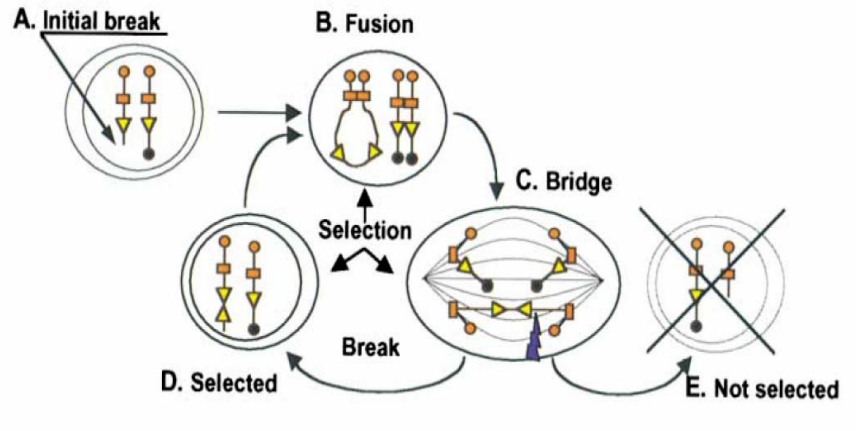


Figure 21. Formation of intrachromosomal gene amplification through BFB cycles.
Taken from Hellmann *et al.*, Cancer Cell 2002.

After an initial break (A), the 2 uncapped telomeric ends of sister chromatids may fuse and thus produce a dicentric chromosome (B). This will lead to the formation of a bridge at anaphase that can break (blue) (C) and give rise to two daughter cells (D and E) with unequal distribution of a region (yellow triangle). Under a selective pressure, the daughter cells with 3 copies could be selected (D) but not the cell with only one copy (E). The selective pressure would lead to selection of cells with higher gene copy number through BFB cycles.
Amplicons are denoted by yellow triangles, telomeres by circles and centromeres by rectangles.

replication origin and a matrix attachment region is able to integrate into the chromosomes of the transfected cell line and give rise to HSRs (Shimizu *et al.*, 2001). Others have accordingly analyzed some amplicons structures from chromosome 17 in various cell lines (Kuwahara *et al.*, 2004). They propose that dsDNA breaks at recombination hotspots located near matrix attachment regions may lead to the formation of circular episomes. These could finally over-replicate and reintegrate in chromosomes providing a mechanism to explain the homogeneous inter-amplicon junctions they have observed (Kuwahara *et al.*, 2004).

All together these studies suggest that there may be some dynamic sequences interconversions between different amplicons forms due to the high plasticity of cancer cell genomes.

iii- Methods to analyze chromosomal aberrations in tumour cells

a) Introduction

Unstained karyotype analyses have proven very useful in the early period of cytogenetics as exemplified by the discovery of the Philadelphia chromosome (Nowell and Hungerford, 1960). **Banding techniques** were then further great progresses towards analyses of chromosomal aberrations in tumour cell genomes at higher resolution but remained **limited to the obtention of metaphase spreads of high quality**. This remains really **difficult for solid tumour cells**. Eventhough, a pitfall is the resolution of chromosomal aberration detection that is limited to the chromosomal band level.

b) Molecular Cytogenetics approaches

To overcome these limitations, molecular cytogenetics methods have been developed thanks to the concomitant emergence of molecular biology techniques. Moreover, the sequencing of the human genome had a pivotal role to the introduction and further evolutions of molecular cytogenetics techniques that have clearly revolutioned the field at both fundamental research and molecular diagnostics levels (Speicher and Carter, 2005).

Fluorescence In Situ Hybridization (FISH)

In situ hybridization techniques were first developed using radioactivity (Gall and Pardue, 1969). Later on, fluorescence was introduced for labelling through the use of antibodies labelled with fluorochromes (Rudkin and Stollar, 1977). Direct labelling of the probes by fluorochromes was then an important step towards development of the FISH technique (Bauman *et al.*, 1980). FISH can be applied to metaphase spreads, interphase nuclei or chromatin fibers and has since then found many applications not only restricted to the cancer research field. It has also been very useful for gene mapping and for the analysis of oncogenes and TSG loci in cancer cells, and is a widely used technique in routine molecular diagnoses.

Several important evolutions of the FISH technique and the labelling methods used allow to perform multicolour experiments, and for example to determine all structural aberrations present in a cancer cell genome (Multiplex-FISH, Speicher *et al.*, 1996; Spectral Karyotyping, Schrock *et al.*, 1996). Coupled with the recent progress made in the construction of Tissue MicroArrays (TMA; Kononen *et al.*, 1998), FISH represent now a powerful tool to analyze in a single experiment the copy number of a preselected locus in a wide collection of archived tumour tissue sections at a glance, and to be able to directly correlate the results obtained on a large cohort of samples with associated anatomo and clinical parameters.

Comparative Genomic Hybridization (CGH)

CGH on chromosomes

Comparative Genomic Hybridization, or CGH, was more recently introduced as a method to overcome the difficulties of obtaining high quality metaphase spreads from different samples, including solid tumours, and in the meantime to be able to detect chromosomal or sub-chromosomal gains and losses for the whole genome in a single experiment, without any prior knowledge of these aberrations (Kallioniemi *et al.*, 1992, du Manoir *et al.*, 1993). CGH applied to chromosomes proved to be very useful and was extremely used for the analysis of chromosomal aberrations in cancer cells. In particular, it allowed to delineate recurrent aberrations associated with specific tumour types (for reviews see: Gebhart and Liehr 2000, Struski *et al.*, 2002), and particularly amplifications (for a review: Knuutila *et al.*, 1998).

However its limits mainly come from the spatial resolution, limiting the detection of aberrations that are around 5 Mb in size. Importantly, several progresses such as DOP-PCR (Telenius *et al.*, 1992) and other technical evolutions have enabled to apply the CGH technique to the analysis of a single cell genome (Klein *et al.*, 1999).

The principle of CGH is presented in figure 22, and can be applied either to metaphasic chromosomes spreads or microarrays (as explained below).

Array CGH

Advances of the Human Genome Project have led now to a nearby completion of the sequencing and assembly of the human genome (International Human Genome Sequencing Consortium, 2001). Concomitantly, the development of microarrays using large insert clones derived from human genomic DNA allowed in the late 90's the **translation of the original CGH technique to an array-based format** thus greatly improving CGH resolution power (Solinas-Toldo *et al.*, 1997; Pinkel *et al.*, 1998). This technique is now referred as to **matrix-CGH or array CGH**.

Different types of microarrays for array CGH

Depending on the coverage of the genome they offer, different types of arrays can be produced for different purposes: **"regional" arrays**, dedicated to one or more chromosomal regions, or **"pangenome" arrays**, covering the whole genome.

Tiling path arrays are made of clones that cover the whole genomic region considered: accordingly, **regional tiling path arrays** have been used and are dedicated to a specific chromosome or chromosomal arm/cytogenetic band for examples (Buckley *et al.*, 2005; van Duin *et al.*, 2005; our study) as well as **pangenomic tiling path** arrays consisting of more than 30000 BAC clones thus covering the whole genome without any gap (Ishkanian *et al.*, 2004; Baldwin *et al.*, 2005).

Importantly, array CGH by itself has already greatly evolved: starting from the use of arrays composed of large insert genomic clones such as BACs and PACs, the technique has been applied to gain further resolution to cDNAs microarrays (Pollack *et al.*, 1999) and even

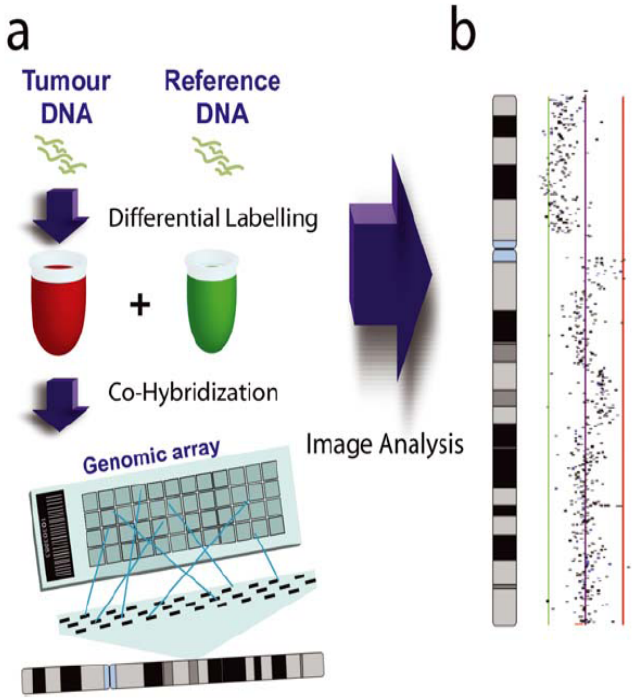


Figure 22. Principle of Comparative Genomic Hybridization (CGH) applied to microarrays. Taken from Davies *et al.*, Chromosome Research 2005.

A. Probes are made of tumour genomic DNA and reference normal genomic DNA after their extraction and differential labelling using different fluorochromes (red and green colors as symbols for the respective probes). Probes are finally mixed and co-hybridized onto a microarray composed of genomic clones of known chromosomal localizations (« genomic array »). The array is finally scanned and signals quantified using dedicated softwares.

B. For each clone, the calculated ratio of fluorescence can finally be represented as a function of the mapping of the clone. Ratios can be represented for a given chromosome as presented or for the whole genome when using pan-genome arrays.

The CGH on chromosomes technique relies on the same principle apart that the probes are co-hybridized onto normal chomosomes metaphases spreads, and ratios are calculated along each chromosome by a dedicated software.

more recently to several types of oligonucleotides microarrays with increasing resolutions. Finally, some of these oligonucleotides microarrays which were originally dedicated to the analysis of single nucleotide polymorphisms now allow to apply CGH for a huge number of oligonucleotides covering the genome and to analyze allelic imbalances of these loci at the same time (Zhou *et al.*, 2005). Reduction of the target size (from BAC clones to oligonucleotides) requires the averaging of several points from adjacent chromosomal localisation and the use of dedicated softwares to realize data smoothing.

Applications of array CGH

Cancer-Related applications

Array CGH has already been extensively used to study chromosomal aberrations in various cancer types. Basically, it has permitted to **identify new cancer-related genes** through mapping of deletions or amplifications, and has also proven to have potential for **tumour classification** and **patient prognostic prediction** in several cases (Pinkel and Albertson, 2005). These research efforts may translate in the near future in additional clinical tools to improve clinical management.

Examples of genes functionally linked to cancers that have been identified thanks to array CGH studies are *CYP24* (Albertson *et al.*, 2000) and *RAB25* (Cheng *et al.*, 2004). Numerous other genes have been proposed as putative oncogenes or TSG but still require additional functional investigations for validation of their status (Redon *et al.*, 2002; Tagawa *et al.*, 2005).

The use of genomic profiling also allows for the **classification of tumours into subtypes with distinct chromosomal aberrations signatures**. Examples concern lymphomas (Rubio-Moscardo *et al.*, 2005; Tagawa *et al.*, 2005), liposarcomas (Fritz *et al.*, 2002), gliomas (Roerig *et al.*, 2005) and breast (Loo *et al.*, 2004) tumours. This application is of high interest for tumours which are hard to classify because of their histological heterogeneity.

A few already published examples also describe the utility of array CGH to link tumour **genomic profiles with patient outcome**. Such retrospective studies have already been applied successfully to lymphomas (Rubio-Moscardo *et al.*, 2005), gastric (Weiss *et al.*, 2003; Weiss *et al.*, 2004), breast (Callagy *et al.*, 2005), hepatocellular carcinoma (Katoh *et al.*, 2005) and bladder (Blaveri *et al.*, 2005) tumours.

Others applications

Apart of its use to analyze cancer cell genomes, array CGH has also importantly shed new light on **several constitutional disorders.** Examples mainly concern patients with mental retardation or developmental abnormalities. In numerous cases, **interstitial chromosomal aberrations** such as duplications or deletions that could not be identified previously by karyotyping or classical CGH due to a lack of resolution have been delineated (for examples Shaw-Smith *et al.*, 2004; Redon *et al.*, 2005). A dedicated database, DECIPHER, has been built recently (http://www.sanger.ac.uk/PostGenomics/decipher/). These micro-alterations may affect crucial genes dosages during developmental processes (Pinkel and Albertson, 2005). Apparently, since the application of array CGH to these types

of constitutional diseases is very recent, one can easily bet that a small sub-fraction only has been characterized until now and that a lot more of findings lie ahead.

Finally, array CGH very importantly gave recently new insight into **human genetic diversity** by delineating the **very frequent presence of large copy number polymorphisms among phenotypically normal humans** (Sebat *et al.*, 2004; Iafrate *et al.*, 2004; Sharp *et al.*, 2005). This diversity was confirmed and moreover extended to an important level of structural variations by "sequencing a second human genome" (through end sequence profiling of fosmid clones) and comparison with the human reference sequence (Tuzun *et al.*, 2005).

c) Others methods for the detection of chromosomal imbalances and aberrations

Allelic imbalances: the "LOH technique" (Zhou *et al.*, 2005)

Analysis of LOH in tumour cell genomes has been widely used through a genotyping approach to identify regions harbouring TSG. Allelic losses may indeed not be detectable by CGH, and moreover loci can undergo reduplication after loss of one allele.

The LOH technique aims at detecting the loss of polymorphic heterozygous markers such as microsatellite or restriction fragment length polymorphism (RFLPs) markers. Its use is however laborious and limited to some pre-selected loci. Furthermore it can not detect amplifications. Finally one has to carefully interpret the results, particularly when analyzing highly unstable tumour genomes

Gene copy number determination by real time quantitative PCR

Real time quantitative PCR (qPCR) has been developed as a powerful tool for accurate and high throughput measurements of gene expression levels (Heid *et al.*, 1996) and led to adapted systems such as the LightCycler (Wittwer *et al.*, 1997). The interest of this method rely on the continuous measurement of PCR products accumulation during the reaction allowing a real measurement in the exponential phase of amplifications performed with primer pairs with comparable efficiencies. Thus, this technique highly contrasts with previously used PCR methods that rely on end-point measurements; its ease in high-throughput format makes of real time qPCR a technique of choice for molecular diagnostics approaches.

Among numerous applications, real time qPCR can be used to determine gene copy numbers. As examples, real time qPCR has been recently applied to copy number analyses for selected genes and oncogene candidates mapping at 3q in SCC from the oesophagus (Yen *et al.*, 2005) and from the oral cavity (Lin *et al.*, 2005).

Sequencing based approaches

Up to now, sequencing has been a heavy and costly method to analyze chromosomal aberrations in tumour genomes and thus only limited punctual uses were made of it. However, in the few published examples, it proved extremely powerful since it allowed for the characterization of amplicons, as examplified for the 20q13 amplcon which is recurrently observed in breast cancers (Collins *et al.*, 2001).

Methods such as End Sequence Profiling, or ESP, in combination with bioinformatics tools have been developed in an attempt to determine "tumor genome architectures" (Raphael *et al.*, 2003; Volik *et al.*, 2003). They basically rely on the construction of large insert clone libraries from tumour genomes and further sequencing of clones extremities. Finally, bioinformatics tools are used to align the sequences obtained with the human genome reference sequence. Data readout can pinpoint amplified genomic regions that are over-represented in the population of end sequences obtained. It also allows identification of translocated regions when the two end sequences of a given clone map to two different chromosomes. Finally, ESP can reveal interstitial deletions when the two end sequences map very close on the same chromosomal region.

The sole example published to date was applied to the breast cancer cell line MCF-7. This can be easily understood due to the huge amount of sequencing needed for such an approach. However, this study clearly illustrated the power of ESP as well as the high complexity that can be faced when trying to resolve an amplicon structure through the example of the 20q13.2 amplicon that was characterized in this cell line.

iv- Transcriptional consequences of large genomic gains and genic amplifications and their biological relevance for the malignant phenotype

a) Gains

Transcriptional consequences

The presence of gains (hereafter defined as low levels increases) of large genomic regions observed recurrently in cancer cells, such as those concerning a chromosomal arm or even a whole chromosome is intriguing in terms of the phenotypical consequences they can produce. In any case, it is believed that such gross changes, either losses or gains, may deregulate the expression of several genes which may participate to the tumoural phenotype.

Some clues can be obtained from a study of a series of commonly used yeast strains. Expression profiles across 300 different yeast gene deletion mutant strains were analyzed (Hughes *et al.*, 2000a). For 8% of these, the authors were surprised to find systematic chromosomal-wide gene dosage biases (Hughes *et al.*, 2000b). They postulated that theses biases might result from aneuploidy and further demonstrated the presence of segmental or complete chromosomal gains and losses using array CGH (Hughes *et al.*, 2000a). The genes that are located on the aneuploid segments lost or gained in these yeasts are nearby systematically under- or over-expressed, respectively. This suggests that no or very little dosage compensation may occur for most of these yeast genes.

A way to further address this question is to analyze the consequences of such large genomic aneuploïdies that can be observed in humans disorders.

Down syndrome is an archetypal constitutional disorder of autosomal trisomies. It is due to chromosome 21 trisomy and among numerous phenotypical traits, patients present severe mental retardation and learning disabilities. Several studies using different mouse models in particular tried to decipher the transcriptional consequences of chromosome 21 trisomy (Fitzpatrick, 2005). Taken together, these studies mainly suggest that at least 75-90% of the expressed trisomic genes seem to be submitted to primary gene dosage effect in most of the studies (Antonarakis *et al.*, 2004; Fitzpatrick, 2005). Lack of sensitivity of DNA array technology and differences in the thresholds used to define over-expression may account for some of the differences among different studies. However, they also clearly pinpoint that dosage compensation occur for a small fraction of them. This is well known for segmental aneuploidies in maize and *Drosophila* (Fitzpatrick, 2005).

Likewise, a recent study analyzed genome-wide gene expression profiles in three different human trisomic cell lines (compared to normal diploid cell lines of matched tissue origin; Zhou *et al.*, 2004). In each case, genes from the trisomic chromosomes (most of chromosome 9, whole chromosomes 18 or 21) are highly significantly over-represented among the over-expressed genes. The percentage of over-expressed genes/expressed genes from gained chromosome is not available for this work. By applying robust statistical analysis to the gene expression levels, the authors were able to accurately estimate the genomic breakpoint of the trisomy on chromosome 9.

To address the transcriptional consequences of large genomic gains in cancer, a similar approach was applied to hepatocellular carcinomas. It suggests that genowe-wide expression profiles when analysed by chromosomal arms may reflect underlying cytogenetic aberrations that can be described by CGH on chromosomes (Crawley and Furge, 2002).

Another approach is to focus on cases with few well defined cytogenetic aberrations. Accordingly, a study analyzed selected cases of Acute Myeloid Leukaemia (Virtaneva *et al.*, 2001). Most of AML cases have a normal karyotype (AML-CN), whereas a few do only present a trisomy of chromosome 8 (AML+8). By comparing gene expression profiles in these two AML types, the authors clearly showed on average a bias towards over-expression (1,32 fold the average for other chromosomes) of chromosome 8 genes in AML+8 cells when compared with AML-CN cases. A majority of the 41 chromosome 8 genes analyzed seem over-expressed in AML+8 cases when compared to AML-CN.

All together, these studies clearly point to a major impact of the large genomic gains on gene expression. Between 75 and 90% of expressed genes affected seems a reasonable estimate from studies with simple experimental design which are easy to interpret. Consequently, a notable fraction of the transcriptome deregulations in tumour cells may be attributable to aneuploidy since large genomic regions are frequently imbalanced in tumours.

Among these genes, some may be critical with respect to the tumour phenotype. Large genomic gains may also contribute to the tumour phenotype by leading to a general increased metabolism that may be needed by tumour cells and by deregulating the expression of some crucial genes.

Biological relevance

The biological consequences of slight variations in gene expression may vary from gene to gene. Some encoded proteins are indeed engaged in strict stoechiometric relationships (such as transcription factors and their co-factors) when others may have to act in a network of proteins (in a given signalling pathway for example).

Subtle gene dosage variations have indeed already been shown to influence drastically the tumoural phenotype for a single gene as exemplified for the *Oct-4* or *Plk-4* genes. The Oct-4 transcription factor has indeed been shown to be a dose-dependent oncogenic fate determinant of ES (Embryonic Stem) cells (Gidekel *et al.*, 2003). PLK4 region is frequently submitted to LOH in corresponding human tumours and as stated by the authors, “reduced PLK4 gene dosage increases the probability of mitotic errors and cancer development” (Ko *et al.*, 2005).

Post-transcriptional, traductional and post-traductional regulation mechanisms are key modifiers of the biological consequences of these large scale transcriptional changes. Expression of the proteins is obviously the ultimate level at which gene dosage effects should be evaluated. Development of proteomics techniques such as 2 dimensional gel electophoresis of proteins and their further identification by mass spectrometry allow now for such large scale studies. An example has already been published (Orntoft *et al.*, 2002). This pilot study shows that for at least 2-fold gain of DNA by CGH, a corresponding increase in mRNA transcripts is found. The relation seems to be weaker for deleted area for which either reduced or unaltered transcripts levels were observed. Interestingly, for few abundant proteins and of known location on the two-dimensional gels, a comparison of proteins and mRNA levels was made. In most of the cases a good correlation was obtained ($p < 0.005$) between transcript alterations and protein levels in this small size study. Caution has to be applied before generalizing this conclusion.

b) Gene amplifications

Relationship genomic gain - gene amplification ?

Genomic gains have been described as early events in the progression of several cancer types (breast, colorectal, cervical, head and neck carcinoma for example), whereas high-level amplifications are generally found in tumours of more advanced stages (Ried *et al.*, 1999). With respect to a putative relationship between genomic gains and high-level gene amplifications, it seems, at least for some examples, that recurrent high-level amplifications are observed in genomic regions that are already gained at low-level in tumours of less advanced stages.

The 3q arm gain is indeed an early event during SCC carcinogenesis (Ried *et al.*, 1999, Redon *et al.*, 2001). It has even been linked to a poor prognosis for HNSCC patients (Bockmuhl *et al.*, 2000). Amplifications at 3q24-qter are recurrently described later on during SCC progression (Hermsen *et al.*, 2001) and have also been linked to HNSCC patients survival based on the 3q26 locus copy number (Singh *et al.*, 2002).

Likewise, low-level gain of the 11q13 cytogenetic band is recurrently observed early during HNSCC carcinogenesis (27%, Redon *et al.*, 2001) whereas 11q13 high level amplifications are correlated with lymph node extension in advanced tumours (Hermsen *et al.*, 2005). CCND1 has been proposed as an oncogene at this locus but still other candidates seem to be recurrently co-amplified and may be important (Huang *et al.*, 2002; Freier *et al.*, 2005).

It is still unclear whether some of the recurrent high-level amplifications observed specifically in given tumour types do correlate with genomic gains that are gained at lower levels in "earlier" tumours. If this could be verified, it would suggest a mechanistic bias towards the formation of high-level amplifications in genomic regions that are gained. **BFB (breakage fusion bridge) mechanism is consistent with such statement, from which amplification appear in region previously rearranged as dicentric chromosomes.** This question could be experimentally addressed using TMA for example and analyzing specific loci copy numbers and comparing them for large cohorts of tumours of the various stages observed during progression.

Consequences of gene amplification

Introduction

(Hogarty MD and Brodeur GM, Gene amplification in Human Cancers: Biological and Clinical Significance, in The Metabolic & Molecular Bases of Inherited Disease, 8th edition)

On the contrary to large genomic gains, **gene amplification** is believed to reflect a means of rapid adaptative capability through high plasticity of cancer cell genomes in response to environmental selective pressures.

An amplicon usually contains **a few number of genes**, including in human tumours mainly **oncogenes**. A single oncogene may be the "target gene" of the amplification, but in some instances two or more genes may confer a selective advantage in consequence of their over-expression by gene dosage effect. The term "driver gene" should be preferred, since it is not the target of an active mechanism, but rather the advantage conferred that is naturally selected through clonal fixation of the mutation in the progeny.

The mechanisms by which oncogene amplifications confer these selective advantages to tumour cells depend on the gene(s) involved. High-level amplifications of *myc* genes which are recurrently observed in several tumour types are believed to be a way of continuously overproducing the myc oncoproteins which enables to overcome their short cellular half-life, in order to sustain proliferation. Other examples of amplified genes concern growth factors and their receptors, as well as signal transduction encoding proteins, which may all serve to mimic constitutive growth factor stimulation. Finally, amplifications of cell cycle regulatory genes may allow the cancer cell to loose cell cycle controls. Anyhow, it seems that in many cases, gene amplification is associated with an aggressive phenotype of the tumour, in terms of rapid progression and poor outcome.

Physiological consequences of gene amplifications

In vitro experimentally generated amplicons of drug resistance genes have shown a clear correlation between gene copy numbers and increases of the selective pressure: incremental increases of methotrexate lead to copy number increases of the *DHFR* gene which confers drug resistance to the cell. In these models, depending on the cell line the *DHFR* amplicons have initially been described as HSRs (Biedler and Spengler, 1976; Nunberg *et al.*, 1978; Dolnick *et al.*, 1979; Biedler *et al.*, 1980) or dmin chromosomes (Kaufman *et al.*, 1979; Brown *et al.*, 1981).

Removal of the selective pressure leads to a rapid loss of DHFR activity associated with a rapid loss of unstable dmin chromosomes: 50% of them are lost in 20 cell doublings (Brown *et al.*, 1981). For the HSRs amplicons, removal of the selective pressure can also lead to driver gene copy number decrease but at lower rates, which suggests that intrachromosomally amplified DNA may be excised and lost (Biedler *et al.,* 1980). On the contrary, continuous exposure to the selective pressure leads to the acquisition of stable resistance by the cells. This seems associated with almost no or a very few number of dmin chromosomes per cell, but rather intrachromosomal presence of the amplicon either taking the form of HSRs (Kaufman *et al.*, 1979) or "grain clusters" on one or a few chromosomes (Kaufman *et al.,* 1981).

All together, these *in vitro* studies suggest first that amplicons may be highly dynamic structures and secondly that some interconversions between extrachromosomal amplicons or dmin chromosomes and intrachromosomal amplicons or HSRs may be possible. Other experimental works have shown that such exchanges between the two amplicon structures can take place in cells *in vitro*. HSRs can actually give rise to dmin chromosomes upon fragile site activation induced by different stresses such as aphidicolin or hypoxia treatments for example (Coquelle *et al.*, 2002). Moreover, dmin chromosomes can be eliminated by the cell through formation of micronuclei (Shimizu *et al.*, 1996) but they can also be maintained in clonal cell population through the "hitchhike" mechanism: dmin chromosomes can adhere to and thus segregate with the normal mitotic chromosomes. Finally, dmin chromosomes can also integrate into chromosomes as exemplified by some DHFR stable resistance property acquisition (Kaufman *et al.,* 1981).

In vivo, tumour cells may also at some point have to face strong but transitory micro-environmental selective pressures. Vascularisation of tumours of notable size is a typical example of strong selective pressures during rapid tumour growth: some tumour areas that may still not be vascularized may be concerned by transient hypoxia or lack of nutrients. When such a crude lack of nutrients or oxygen would have to be faced, one can imagine that a majority of the cells will die. A few could survive only if they would acquire any capability

allowing them to overcome the selective pressure. One can imagine that the amplification of an oncogene and its subsequent over-expression could confer them with a protection against hypoxic conditions or the lack of nutrients in their microenvironment. This *per se* would not be sufficient to explain why and how the aberration will get fixed in the progeny, which is a prerequisite since we do observe recurrent amplifications in human tumour cells. An explanation for this could be found at three levels.

First, the amplification may concern a "polyvalent" driver oncogene.
By "polyvalent" oncogene, we mean a gene involved in more than one of the 6 traits characteristic of tumour cells as defined by Hahn and Weinberg (2000). In this theoretical model, the initial, transient and strong selective pressure would lead to the amplification of a driver oncogene. This would allow the cell and its progeny to survive and bypass the crisis. The clonal fixation of the aberration in the surviving progeny would then be provided and explained by any further positive action on the malignant phenotype exerted by the oncogene that initially drove the formation of the aberration.

An example giving support to this hypothesis can be found in the *KCNK9* gene (at 8q24.3) that has been found recently amplified and over-expressed in breast tumours (Mu *et al.*, 2003). The gene encodes a potassium channel that upon overexpression confers tumorigenicity in nude mice. Moreover its over-expression also enhances *in vitro* cell capacity to survive under culture in low-serum conditions as well as under hypoxic conditions in a *p53* WT background. It is possible that the amplifications and subsequent over-expressions of this gene observed in human breast tumours allowed the cells to bypass a transitory selective pressure represented by either hypoxia or lack of nutrients.

Secondly, the amplification may concern a driver oncogene and neighbour(s) gene(s) with additive functions.
In this model, the initial selective pressure leads to the selection of cells with high level amplification of the driver oncogene. The aberration may finally get fixed in the progeny if these cells are provided with an additional advantage from the neighbour(s) gene(s) (with regards to any trait of the tumoural phenotype), which could contribute to overcome other cells.
Consistent with the latter hypothesis, *N-myc* high-level amplifications at 2p24 occur in approximately one fourth of neuroblastomas, and half of these co-amplify the *DDX1* gene, genomic neighbour of *N-myc*. These tumours seem to be more aggressive, in terms of progression time and outcome (Hogarty MD and Brodeur GM, Gene amplification in Human Cancers: Biological and Clinical Significance in The Metabolic & Molecular Bases of Inherited Disease, 8th edition).

Similar situations could occur for the 17q23 amplicon in breast cancer, locus of the *ERBB2* oncogene (Barlund *et al.*, 2000; Kauraniemi *et al.*, 2001; Monni *et al.*, 2001).

Finally, one can imagine that tumour cells may not only have to face transitory micro-envirronmental selective pressures only once throughout evolution of the tumour. Subsequent crisis of the same nature than the initial one that drove the selection of cells with an oncogene amplification may lead to the direct re-selection of their progeny. Thus it may participate to reinforce the fixation of the aberration. It is tempting to speculate that in some cases initial low-level amplifications taking the form of a HSR could form through BFB. Under strong transient selective pressure they would rapidly give rise to higher copy number amplifications of the driver oncogene into dmin chromosomes. In this model, the dmin chromosomes would represent a rapid response due to cancer cells genome plasticity and allowing to select for cells that will bypass the crisis. After removal or diminution of the selective pressure, the

amplification could "get back" to an intrachromosomal form, or HSR, with a lower copy number but still allowing consistent over-expression. The coherence of this model is at least suggested by the *in vitro* situations observed for *DHFR* amplifications

Transcriptional consequences of genes amplifications

Gene amplifications may thus lead to the overexpression of several genes in their respective amplicons. These may potentially act as co-conspirators of the driver gene of the aberration. Accordingly, transcriptional consequences of high-level amplifications have already been addressed in breast and prostate cancers to evaluate the impact of high-level copy number increases in cancer cells (Hyman *et al.*, 2002; Pollack *et al.*, 2002; Wolf *et al.*, 2004). All these studies were performed using genome-wide cDNAs microarrays both for genome and transcriptome analyses, which allows for a direct comparison of gene copy number and expression levels using the same tools. However, limited use of these arrays for CGH have been done since they may only allow to detect accurately high copy number variations.

From a global point of view, correlations between genes copy number and their relative expression levels can be observed: a significant proportion of over-expressed genes in tumour cells are mapped in the genomic regions with increased copy numbers.

A first genome-wide study analyzed breast cancer cell lines and tumours (Pollack *et al.*, 2002). They found 91 highly amplified genes (ratios >4), of which around 40% are strongly over-expressed (ratios>4).

Another study was dedicated to 14 breast cancer cell lines (Hyman *et al.*, 2002). They have shown that 44% of the highly amplified genes (ratio >2.5) are strongly over-expressed (that is, among the upper 7% of expression ratios when compared to the reference RNA, constituted of a commercially available pool of numerous tumour cell lines RNAs). This value (44%) has to be compared with the few 6% of the genes with normal copy number levels that are over-expressed. Among the highly amplified genes, less than 10% seem under-expressed. Conversely, among all strongly over-expressed genes (ratio >10), 10.5% of genes are amplified, which underlines the importance of other means of regulation for the over-expression of genes but also points to a major impact of amplified genes.

An overall highly significant association between gene copy number and respective expression level was found in prostate cancer cell lines (Wolf *et al.*, 2004). In this study, among highly amplified genes (ratios >2), roughly 30% of them are over-expressed whereas less than 5% are under-expressed (same definition as in the previous study for over- and under-expression, but when compared here to a pool of 18 different cancer cell lines). Among genes with normal copy numbers, 6.8% are over-expressed.

A study analyzed the transcriptional consequences of large genomic regions that are amplified in liver metastasis from colon cancers (Platzer *et al.*, 2002). It addressed the question of these consequences for four sites they defined (7q, 8q, 13q and 20q) as highly amplified (ratio >2) by CGH on chromosomes. It should be first noted that these "amplifications" may vary from the gene amplifcations described above. These authors have anyway strikingly shown that solely 3.8 % of the genes that were mapped in the highly amplified regions were overexpressed leading them to conclude to "silent chromosomal amplifications in colon cancer" (Platzer *et al.*, 2002). It should be remarked that the aberrations were mapped on the basis of CGH on chromosomes which doesn't allow a strict comparison of the copy number and expression levels for each gene as for the previously

reported studies. Secondly, these amplifications may significantly vary from the gene amplification described above since they concern whole chromosomal arms in this study.

To conclude, it seems that high-level amplifications dynamics *in vivo* are still not fully understood. These amplifications produce major transcriptional consequences most of the time since they lead to strong over-expressions of several different genes. From these studies, a working estimate of 30-40% of the genes from amplified regions that are strongly over-expressed seems reasonable. Depending on cancer types and tumours, differences may be observed for the genes located in their respective amplicons and thus associated phenotypical consequences may be variable.

v- Rationale of positional cloning strategy for cancer-related genes identification

a) Strategy description

For all the reasons enumerated previously, it is of clear interest to make use of the chromosomal aberrations recurrently observed in tumour to identify cancer related-genes.

Such positional cloning strategies have already proven very useful when applied to a wide range of genetic disorders, comprising familial forms of cancer, to identify disease causing genes. This approach basically relies on the analysis of samples to define a common genomic interval linked with the disease and containing a limited number of candidate genes. Then, the discrimination between candidate genes located in the genomic interval of interest can be done by any means, i.e. by search for mutations associated with the disease or functional assays with respect to disease associated phenotypes

With respect to cancer, several TSG have been historically identified through mapping of LOH regions. TSG are ultimately given their status by description of loss of function mutations found in tumour cells and restoration of normal growth or induction of tumour cell death for examples when reintroducing a WT allele of the gene into cancer cells.

On the contrary, several candidate oncogenes have been identified through amplicons identification and mapping in tumour cells. Their oncogenic status can furthermore be given by the identification of activating mutations in tumour cells, and restoration of normal growth for example when inhibiting the oncogene in cancer cells. Functional assays of their transforming power can also be made.

b) Historical examples of oncogenes identified on the basis of gene amplification

Several genes were identified on the basis of amplicons identifications in tumour cells. In each case, their oncogenicity was only proven later on through expression and functional characterizations.

In that context, ***N-myc*** is the archetypal oncogene. It was indeed identified 22 years ago from study of amplifications in neuroblastoma cell lines and tumors (Schwab *et al.*, 1983; Kohl *et al*., 1983). The *N-myc* gene was further shown to be over-expressed in consequence of gene amplifications (Lee *et al.*, 1984). These were then correlated with advanced stages of

neuroblastomas and poor prognosis of patients (Brodeur *et al.*, 1984). Importantly, the *N-myc* gene was also shown to be over-expressed in neuroblastomas without amplification when compared to other tumour types suggestive of another type of regulation for its over-expression (Kohl *et al.*, 1984). Finally, N-myc was shown to cooperate with mutated Ras for transformation of cells in culture (Schwab *et al.*, 1985; Yancopoulos *et al.*, 1985). For a review on the myc family of transcription factors and some of their roles and targets, please refer to Patel *et al.* (2004).

More recently, the ***mdm-2*** gene was isolated from an amplicon present as double minute chromosomes in a spontaneously transformed 3T3 cell line (Cahilly-Snider *et al.*, 1987) and then found to be amplified in around a third of human sarcomas (Oliner *et al.*, 1992). The *hdm-2* gene encodes a protein which interacts with and inhibits the p53 protein (Momand *et al.*, 1992) and can upon over-expression immortalize primary fibroblasts, as well as transform them in cooperation with activated Ras (Finlay, 1993). The hdm-2 protein is actually an ubiquitin ligase directly inhibiting p53 but that also exerts master regulatory roles on the p53 pathway together with other partners (Harris and Levine, 2005).

vi- Chromosome 3 aberrations in MFH and SCC, significance and transcriptional consequences in SCC

a) Malignant Fibrous Histiocytoma

Chromosome 3 aberrations appear to be very rare cytogenetic events in MFH. A study (from our collaborators) using CGH on chromosomes to analyze MFH genomes have indeed reported only 4 cases out of 30 with high levels gains at 3q27 (Mairal *et al.*, 1999). Moreover others have repetitively shown a very minor prevalence of chromosome 3 aberrations in their MFH cohorts.

b) Squamous Cell Carcinoma

Chromosomal aberrations

In lung and HN Squamous Cell Carcinomas, chromosome 3 aberrations have been extensively described since losses at the 3p arm and gains of the 3q arm represent early events in the progression of these tumours.

Concerning the **losses on 3p**, numerous LOH studies have uncovered that several loci at 3p may be involved in SCC (Wistuba *et al.*, 2000; Hogg *et al.*, 2002).

Concerning lung SCC, LOH studies have suggested three different 3p sites (Wistuba *et al.*, 2000). One of the main region deleted in human non small cell lung cancers is indeed the LUCA region (for Lung Cancer) at 3p21.3 and is also one of the genome region with higher genes density. LOH studies dedicated to HNSCC suggest two frequent sites of alterations, 3p12 and 3p21.3 (Hogg *et al.*, 2002).

For both lung tumours and HNSCC, a high prevalence of whole 3p losses have been described by CGH studies (Gebhart and Liehr, 2000; Balsara and Testa, 2002; Struski *et al*, 2002) and even that the loss at 3p21 is the most frequent among several tumours (Gebhart and Liehr, 2000).

Whole genome array CGH applied to NSCLC (Non Small Cell Lung Carcinoma, of which 30 to 40% are SCC) tumours has shown that at least three different sites may be more frequently deleted at 3p (Massion *et al.*, 2002) which was confirmed for NSCLC tumour cell lines using tiling path arrays more recently (Garnis *et al.*, 2005). Tiling path arrays to study oral SCC have also confirmed an overall high frequency of losses at 3p (Baldwin *et al.*, 2005). Moreover, 3 different sites were also found to be more frequently lost at 3p in another study of oral tumours (Snidjers *et al.*, 2005).

On the contrary, large **gains of the 3q arm**, or even gain of the whole chromosome 3q arm, may be an early event in SCC progression and that is even correlated *per se* to HNSCC patient's poor prognosis (for the 3q21-q29 region; Bockmuhl *et al.*, 2000). Gain at 3q is the most common aberration that has been described for SCC (Gebhart and Liehr, 2000) and thus may represent a functional advantage through dosage increase of many genes that are important for SCC tumour progression specifically. Moreover, gains at 3q have been described as occurring early in the course of tumour progression, and as such proposed to play crucial roles (Redon *et al.*, 2001).

CGH studies have uncovered that 3q gains are present in a majority of SCC (Gebhart and Liehr, 2000; Balsara and Testa, 2002; Struski *et al*, 2002) and moreover that the 3q27-qter is very often gained in several tumour types (Gebhart and Liehr, 2000).

Whole genome array CGH applied to NSCLC has underlined the presence of at least two distinct 3q gain maxima (Massion *et al.*, 2002). Using tiling path arrays, the high frequency of whole 3q gains was confirmed in oral SCC and especially of the apparent 3q25-qter region in which 3 sub-zones seem to be gained at higher frequency (from their figure; Baldwin *et al.*, 2005). Similarly, 3 preferential sites of gain may be involved more frequently at 3q in another study of oral SCC (Snidjers *et al.*, 2005). Data analysis of this latter study reveals the position of these peaks of gains at 151 Mb, 172 Mb and 185-188 Mb on chromosome 3.

Concerning 3q gene amplifications, different studies have underlined that amplifications at 3q may target different sites at least in HNSCC (Redon *et al.*, 2001; Lin *et al.*, 2005). No high-resolution array CGH study has shown a consensus region of high level amplifications up to now in SCC at 3q26.

In accordance with all these results, several TSG mapping at 3p or oncogenes mapping at 3q have been identified or proposed until now. Chromosomal aberrations seem to pinpoint different chromosome 3 sites and their underlying genes that may be involved in SCC both at 3p and 3q. A way to reveal whether this is relevant for tumour phenotype is to analyze the transcriptional alterations for chromosome 3 genes in these SCC, which is now readily feasible through Oncomine or direct analyses of published data.

Transcriptional consequences of chromosome 3 aberrations

Analysis of genes expression in SCC through Oncomine reveals several peculiar features when compared with normal lung, especially for chromosome 3 genes.

First, the chromosomal arm harbouring genes that are consistently over-expressed most significantly over the whole genome is the 3q arm ($p<10^{-17}$ from the lung SCC data of Garber *et al.*, 2001; $p<10^{-10}$ from the lung SCC data of Bhattacharjee *et al.*, 2001; these p values are each time given hereafter in the same order). This is also observed in hypopharygeal HNSCC ($p<10^{-8}$ from the data of Cromer *et al.*, 2004).

Second for the lung SCC studies, this effect seem to be due to the genes located at 3q27 ($p<10^{-5}$; $p<10^{-4}$) and 3q29 ($p<10^{-5}$ and $p<10^{-3}$) which are the two cytogenetic bands most significantly harbouring over-expressed genes in these tumours. Combining the results of these two studies enable to draw a list of 20 genes mapping at 3q that are consistently and recurrently over-expressed in SCC *versus* normal lung. Twelve of them map in the 3q26-qter region (figure 23).

Third, the chromosomal arm harbouring genes that are consistently under-expressed most significantly over the whole genome is the 3p arm ($p<10^{-6}$; $p<10^{-6}$) Fourth, from the study of Garber *et al.* (2001) genes located at 3p14 and 3p21 seem to be major contributors ($p<0.01$). From the data of Bhattacharjee *et al.* (2001), 3p21 is the cytogenetic band harbouring genes most significantly under-expressed over the genome ($p<10^{-6}$). Combining the two sets of data for analysis underline consistent under-expression in lung SCC compared to normal for 28 genes located on the 3p arm. Half of them are located at 3p21 (43.6-54.4 Mb; figure 23).

A.

3p

Gene Symbol	Mapping (May 2004, Mb)
CNTN6	1,3
EDEM1	5,2
PPARG	12,4
IQSEC1	13,0
SHGBP5	15,3
RAFTLIN	16,4
RBMS3	29,7
OPA1L	32,2
MYD88	38,2
CXCR1	39,3
SIAH2	43,3
TNA	45,0
SACM1L	45,7
CCR1	46,2
PTHR1	46,9
LAMB2	48,1
USP4	49,3
RHOA	49,4
GNAI2	50,3
HYAL2	50,3
TUSC2	50,3
CACNA2D2	50,4
MAP4K4PK3	50,6
ALG1	52,2
PRKCD	53,2
ARLGP5	69,2
FRMD4B	69,5
MITF	70,1

3q

Gene Symbol	Mapping (May 2004, Mb)
TOMM70A	101,6
CSTA	123,5
MRPL3	132,7
ABCC5	135,2
RYK	135,4
PLOD2	147,3
PFN2	151,2
SIAH2	152,0
PDCD10	168,9
PRKCI	171,5
ACTL6A	180,3
NDUFB8	180,3
ATP11B	184,1
DVL3	185,4
ABCF3	185,4
ALG3	185,4
EPHB3	185,8
TP73L	191,0
NCBP2	196,2
BDH	196,7

B.

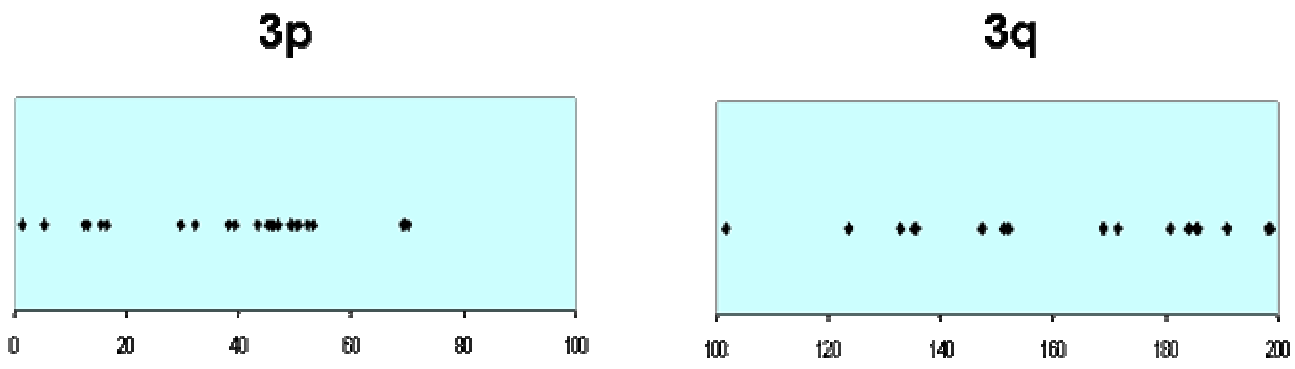


Figure 23. Chromosome 3 genes whose expression is recurrently altered in SCC vs normal lung.

A. Lists of the genes (28 at 3p; 20 at 3q) and their mapping values. Both lists were established by selecting chromosome 3 genes from Garber *et al.* (2001) and Bhattacharjee *et al.* (2001) analyzed through Oncomine. Genes were selected if they were significantly underexpressed in SCC vs normal lungs ($p<0,01$ in each study), and further mapped on chromosome 3.
B. Mapping of the under-expressed genes at 3p (left) and of the over-expressed genes at 3q (right). Half of 3p under-expressed genes are localized at 3p21, when 12 out of 20 over-expressed genes are mapped in the 3q26-qter region.

These results clearly underline the importance of genes located on chromsosome 3 with respect to lung squamous cell carcinogenesis. Moreover, they give a rationale support and allow a comprehensive view of the chromosome 3 aberrations that are recurrently observed in these tumours (losses of the short arm and gains/amplifications of the long arm). Finally they underline that many chromosome 3 genes are recurrently altered in lung SCC and shed new light on these genes that might be functionally important for squamous cell carcinogenesis.

A summary on the proposed TSG at 3p and ONCOGENES at 3q in SCC

This part is a rapid overview of some important genes, mapping on chromosome 3p and 3q arms that have been proposed as candidate TSG or oncogenes, respectively.

Tumour suppressor genes at 3p

Aberrations affecting the 3p arm seem to not only be restricted to SCC. Accordingly, several TSG have been proposed as the *VHL*, *OGG1*, *FHIT* and *RASSF1* genes. It should be underlined here that numerous LOH studies have uncovered that different discontinuous sites may be targeted at 3p. But the remaining allele most of the time do not present inactivating mutations as exemplified for VHL in lung cancers (Miyakis *et al.*, 2003). Accordingly, it has been proposed that the combined allelic deletions of multiple genes may be a way to initiate lung carcinogenesis (Hollander *et al.*, 2005). Several of the following TSG mapping at 3p may thus contribute to the tumoural phenotype through either drastic or subtle expression decrease of their expression due to 3p losses.

The *VHL* gene is located at 3p14 and encodes a protein part of an E3 ubiquitin ligase complex (with cullin2) whose function is to inhibit HIF-1 and thus downstream HIF1 targets such as VEGF (Kim and Kaelin, 2004). *VHL* however doesn't seem to be a general LOH target in HNSCC tumours (Waber *et al.*, 1996).

The *OGG1* gene at 3p12 encodes a DNA glycosylase involved in oxidized DNA repair, and specifically the 8-oxo-Guanine lesions. Little is known with respect to OGG1 putative involvement in lung and HNSCC, apart that some specific OGG1 genotypes may be at risk to develop orolaryngeal cancers in tobacco or alcohol consumers (Elahi *et al.*, 2002).

The *RARbeta* gene at 3p24 encodes a nuclear receptor of the RAR family. Retinoids are known to be crucial modulator determinants of epithelial cell fate, and thus RARbeta expression has been investigated in SCC: its protein expression seems to be downregulated during laryngeal tumour progression (Karamouzis *et al.*, 2004) and seems moreover downregulated in a majority of oral SCC (Ralhan *et al.*, 2005).

The *FHIT* (Fragile HIstidine Triad) *gene* is mapped at 3q14.2, and is a well-known tumour suppressor gene (Pekarsky *et al.*, 2002). Its genomic localization indeed encompass a common fragile site of the human genome (FRA3B) and *FHIT* loci is frequently homozygously deleted or affected by LOH in tumours (Pekarsky *et al.*, 2002). Its function remains unclear.

The ***RASSF1*** gene was isolated at 3p21.3 as a TSG whose genomic locus often undergoes LOH in lung cancers and its expression is recurrently silenced through epigenetic modifications of its promoter (Dammana *et al.*, 2000). Among several tumour types, it was further shown to be inactivated in squamous cell carcinomas also through promoter

hypermethylation in a majority of oesophagus SCC (Kuroki *et al.*, 2003) and subsets of Head and Neck SCC (Hogg *et al.*, 2002; Dong *et al.*, 2003). Of interest is that *RASSF1* seems to exert its normal TSG role by encoding a mitotic regulator inhibiting the APC(Cdc20) complex (Song *et al.*, 2004) and through microtubule stabilization by localizations at the mitotic spindle for two of its isoforms that are able to suppress Ras induced genomic instability (Vos *et al.*, 2004).

The *XPC* gene at 3p25 is involved in NER and DNA repair capacity are impaired in consequence of loss of a single *XPC* allele. It was furthermore shown that upon homozygous deletions of XPC in mice, 100% of them do develop adenomas, which evolve to NSCLC upon loss of another specific gene, *Gadd45a,* located at 1p in human which is also frequently deleted in human lung tumours (Hollander *et al.*, 2005). It thus seems that XPC could represent, with respect to RASSF1, a specific lung TSG (Hollander *et al.*, 2005).

In the same line, it has also interestingly been shown that *VHL* haploinsufficiency enhances lung tumourigenesis susceptibility (*i.e.* induced by carcinogens) in *FHIT* deficient mice (Zanesi *et al.*, 2005).

Finally, several genes mapped at 3p have functions that upon alteration may directly provoke appearance of chromosomal aberrations: examples may concern *TOPOIIB* at 3p24, and *RAD18* at 3p12 (neighbour of VHL).

The *TOPIIB* gene encodes the topoisomerase II isoform that is expressed in somatic tissues. Its inactivation or reduced expression may lead to translocations (Hermsen *et al.*, 2005).

The RAD18 protein is part of the RAD6/RAD18 DNA damage tolerance complex involved in post-replication repair, whose mechanism has just been uncovered experimentally in yeast and shown to involve sister chromatid recombination (Zhang and Lawrence, 2005). It has moreover been shown that this complex may also be important for the prevention of specific DNA transversions: mutations in OGG1 and RAD18 have indeed synergistic effects on lesions and subsequent transversions accumulations (de Padula *et al.*, 2004). Finally, RAD18 has also been shown to be a suppressor of gross chromosomal rearrangements in yeast (Smith *et al.*, 2004).

An overall intermediate conclusion concerning the 3p arm and its genes is that it seems quite clear that early loss of the 3p arm in a majority of SCC may cause expression changes affecting several genes which could in turn act in a synergistic manner towards mutation accumulation or tumour progression as exemplified from yeast and mice models.

Oncogenes and oncogene candidates at 3q

Given that 3q gains seem to be more specific to SCC than other tumour types as noted previously, only genes that have been proposed as oncogenes or candidate oncogenes in SCC are presented hereafter.

We have proposed the *CCNL1* gene at 3q25.3 as an oncogene candidate in HNSCC, initially as highly amplified and consequently over-expressed in a head and neck cancer cell line and also over-expressed in HNSCC tumours (Redon *et al.*, 2002). Since then, we have characterized *cyclin L1* gene alterations in a large series of HNSCC showing recurrent copy number increases and overexpression but without association to clinical parameters (Muller *et al.*, submitted; see the results part of this thesis). Others have shown a clear association

between cyclin L1 gene high-level amplifications, although very rare, and a poor prognosis (Sticht *et al.*, 2005).

The *TERC* gene encodes the RNA component of the telomerase enzyme (which serves as a template for the telomerase reverse transcriptase). Given that telomerase is reactivated in a wide majority of tumour cells, it has been rationally proposed that *TERC* could contribute to tumourigenesis. TERC was shown to be over-expressed in NSCLC (comprising SCC; Yokoi *et al.*, 2003) and has been proposed as target of 3q amplifications.

SNO / *SKIL* encodes a TGFbeta regulator that has also been proposed as an oncogene target of amplifications at 3q in oesophageal SCC (Imoto *et al.*, 2001).

The *PIK3CA* gene encoding the catalytic subunit of the PhosphoInositide-3 catalytic subunit of a serine/threonine kinase and has been clearly demonstrated as an oncogene. The PIK3CA protein is furthermore implicated at the crossroads of numerous signalling pathways in the cell. It is recurrently gained and overexpressed in SCC (Redon *et al.*, 2001). Moreover it was found recently to be mutated in other cancer types (Weir *et al.*, 2004; Samuels *et al.*, 2004; Wu *et al.,* 2005) and importantly, the most recurrent mutations were shown to be oncogenic (Kang *et al.*, 2005; Samuels *et al.*, 2005). Interestingly, it seems that some breast tumours may have both PIK3CA copy number increase and mutations (Wu *et al.,* 2005).

The *TP63* gene (now also known as TP73L) is a member of the *p53* gene family which encodes different transcripts with transactivating (TAp63) or repressive (DNp63) activities. TP63 is clearly not a TSG just as p53, since no mutations have been found in human tumours. The p63 protein is an epithelial stem cell marker and is required for epithelium homeostasis. Accordingly, *TP63* has been recurrently shown to be amplified in SCC as well as over-expressed with the DNp63 being responsible of its oncogenic role (Benard *et al.*, 2003).

The *EIF4G1* gene encodes a translation initiation factor with RNA binding capacity and acting as the scaffold of the preinitiation complex which binds mRNAs in their 5'UTR and scans their sequence until they found an initiator codon. With respect to SCC, *EIF4G1* copy numbers have been shown not to vary in oesophageal SCC (Yen *et al.*, 2005), whereas it is amplified in hypopharyngeal SCC (Cromer *et al.*, 2004). Over-expression of EIF4G has been accordingly shown to be able to transform cells *in vitro* (Fukushi *et al.*, 1997) and moreover apoptosis induction rapidly provokes EIF4G proteic cleavages (Clemens *et al*, 1998). Thus, over-expression of EIF4G1 in cancer cells may provide them with different advantages (at least increased global translation capabilities and enhanced apoptosis resistance);

The *DVL3* and *RFC4* genes are among the 67 commonly deregulated genes in the "cancer signature" established from meta-analyses of expression profiles from several tumour types (Rhodes *et al.*, 2004). Surprisingly, little is known until now concerning their respective functions and peculiar contributions to SCC cancer, even if their already deciphered functions are of compelling interest with respect to cancer. *DVL3* indeed encodes a *Drosophila* dishevelled protein homolog (Bui *et al.*, 1997) and is thus believed to be an activator of the Wnt / beta-Catenine pathway (Uematsu *et al.*, 2003). ***RFC4*** is mapped at 3q29 and encodes the Replication Factor C 4 belonging to the Replication Factor C complex which is required for the loading of PCNA (Proliferating Cell Nuclear Antigen) onto primer-templates junctions for DNA replication. It moreover seems to play important roles as sensor

of different DNA damage checkpoints (Kim and Brill, 2001) and apparently being involved in chromatids cohesion in *Drosophila* and yeast (Dobie *et al.*, 2001; Krause *et al.*, 2001; Mayer *et al.*, 2001).

Finally, the *RP42 homolog* gene, which was named *SCCRO* (for Squamous Cell Carcinoma Related Oncogene; Estilo *et al.*, 2003) has also been proposed as an oncogene at 3q26.3.

In a first study, Singh *et al.* determined by FISH (probe made of 3 YACs) a strong correlation between 3q26.3 copy number and patient poor outcome depending on 3q26.3 copy number (Singh *et al.*, 2002a). This led them to further propose the SCCRO and PIK3CA genes as novel "oncogenes" in oral tongue carcinoma (Estilo *et al.*, 2003) by first referring to another work investigating PIK3CA oncogenic role (Singh *et al.*, 2002b) and second to an abstract whose title refers to SCCRO (their reference 14, Estilo *et al.*, 2003). This abstract is not present in the Proceedings of the American Association of Cancer Research where it is claimed to have been published (personal observation). Moreover this group has shown a maximum for 3q26.3 amplifications among HNSCC cell lines for "BAC clone 202B22" (Singh *et al.*, 2002; Estilo *et al.*, 2003) they claim to contain the SCCRO gene (Estilo *et al.*, 2003) without at any time giving any library origin of this clone. Analysis of the CloneRegistry at NCBI (http://www.ncbi.nlm.nih.gov/genome/clone/query.cgi) only reveals one human BAC clone from the RPCI11 library, RP11-202B22, which seem to be mapped at 2p24.

In any case, the more robust data they have presented for SCCRO as a putative oncogene are a putative association of its over-expression with clinical parameters in oral tongue carcinomas (all p values being comprised between 0.1 and 0.01; Estilo *et al.*, 2003). More doubtful is a putative correlation of SCCRO expression with VEGF-A in lung SCC (Talbot *et al.*, 2004). One could remark that in this latter paper is stated "we have previously established an association between SCCRO and both aggressive tumor and clinical outcome" and "we have previously shown the association between SCCRO and survival" referring to their reference 1, which has been deleted in proof.

In conclusion, lots of putative functional data concerning SCCRO have been described and even results largely stated and discussed in several of their papers as unpublished or personal observations (Estilo *et al.*, 2003; Talbot *et al.*, 2004), but no consistent functional data has ever been presented.

Really more interestingly, very recent work concerning *RP42* has been done on some homologs (*Dcn1p* in yeast and *DCN1* in *C. elegans*) showing that it is crucial for correct mitotic spindle orientation during the first mitotic embryonic division. This role is achieved through post-meiotic neddylation of cullins which leads to the targeting of other proteins for degradation (Kurz *et al.*, 2005). Accordingly, the human gene get renamed recently and approved by the HUGO committee and will be thus referred as to *DCUN1D1*, for "defective in cullin neddylation 1, domain containing 1".

Conclusion

Many of the enumerated chromosome 3 genes may contribute to SCC whereas others may still have to be identified. With this respect, several zones of gains have been described at 3q in SCC. Recently, very small micro-alterations were uncovered (deletions, gains and or amplifications) thanks to high resolution array CGH as in oral cancers genomes for example (Baldwin *et al.*, 2005).

In accordance this thesis work was dedicated to the analysis of chromosome 3 aberrations using different home made arrays in two tumour types, MFH and SCC. We mostly concentrated our efforts in triyng to define common regions of high-level amplifications to identify the putative driver oncogenes of these aberrations.

Part 2

Analysis of CHROMOSOMAL ABERRATIONS

in Malignant Fibrous Histiocytomas

and Squamous Cell Carcinomas :

towards the identification of amplified and overxpressed

ONCOGENES

-I- INTRODUCTION

On the use of tumoural cell lines as models

Tumoural cell lines were used in several studies presented in this thesis. With this respect, several important observations should be made in introduction.

These cell lines are of real practical utility since they are immortalized, growing exponentially in culture, homogeneous and stable at least at short term. Thus they represent an easy manner to rapidly get huge amounts of biological extracts (DNA, RNA, proteins…) to further analyze their tumoural traits.

It should be however stressed that several important issues have to be addressed when using cancer cell lines.

The first drawback in their use lies in the possible clonal selection of cells when these are initially established in culture from the original sample. It is known for years that during this process, a wide majority of cancer cells die and only a very few maybe survive to finally expand and give rise to the cell line. Accordingly, an important question is whether the alterations that one can observe in cancer cell lines is relevant for the *in vivo* malignant phenotype or for a best fitted phenotype to survive in culture *in vitro* that has been selected.

A second drawback in their use lies in the potential variability that may arise in the cell line after long term culturing periods. It should be first stressed here that we never culture our cell lines for very long term periods. Secondly, all the cell lines are provided to us through the IGBMC cell culture core service where specialists ensure freezing and storage of cell lines from their arrival in the institute.

We have good reasons to think that at least for the malignant fibrous histiocytoma COMA cell line, hereafter used in the MFH results parts, the two previous arguments should not apply. First, as the cell line was established, its chromosomal aberrations were characterized by CGH on chromosomes and said to be overall similar between the primary tumour and the COMA cell line after long term culture of 1 and 6 months (Mairal *et al.*, 2000). Secondly, our focus in this study concerned a high-level amplification of the 3q28 locus. This amplicon was already present in the malignant cells of the primary tumour as shown by CGH on chromosomes for the primary tumour (Mairal *et al.*, 2000). Accordingly, we strongly believe that the study of this amplification in the derived cell line is relevant for the *in vivo* situation.

We also analyzed several head and neck cancer cell lines. The respective chromosomal aberrations present in the primary tumours have unfortunately not been determined. Accordingly, we can not advance that the high-level amplification that we have mapped at 3q25.3 in the Cal-27 HNSCC cell line was already present in the initial tumour. However, amplifications of this locus have been also recently found in some HNSCC primary tumours (Sticht *et al.*, 2005).

-II- RESULTS

a. Malignant Fibrous Histiocytoma

a-1 3q28 amplifications overlap in COMA cell line and two primary sarcomas

(manuscript submitted to Cancer Genetics and Cytogenetics)

i- INTRODUCTION

Our entry point in this study came from the presentation at a meeting by Alain Aurias (Institut Curie, Paris) of a cell line of potential interest for us since it seemed to contain a narrow amplification of the long arm of chromosome 3. This cell line, named COMA, was established in his laboratory (Mairal *et al.*, 2000).

Their initial cytogenetic characterization of the COMA cell line by various means (karyotypying, SKY, FISH and CGH on chromosomes) indicated several high level amplifications. Two distinct amplicons are present as two exclusive populations of double minute (dmin) chromosomes: one amplicon contains the *c-myc* locus (at 8q24), and the other harbours chromosome 3 material (as shown by chromosome painting experiment) and likely from the 3q26-qter region (from the CGH profile).

These prompted us to first map the 3q28 amplicon of the COMA cell line using our chromosome 3 dedicated array. We then further extended the study to some selected sarcomas that were initially diagnosed as MFH and for which CGH on chromosomes were available from A. Aurias and indicative of chromosome 3 aberrations, especially 3q gains.

ii- RESULTS **(see the manuscript hereafter)**

Title: Overlapping amplified 3q28 region in COMA cell line and malignant fibrous histiocytoma primary tumors

Running title: Overlapping amplified 3q28 region in COMA cell line and MFH primary tumors

Corresponding author: du Manoir Stanislas, PhD
INSERM U 596
IGBMC
1 rue Laurent Fries
67404 Illkirch-CU de Strasbourg
Fax 33 (0)3 88 65 32 46

Tel 33 3 88 65 35 25 office
Tel 33 (0)3 88 65 34 17 lab

dumanoir@igbmc.u-strasbg.fr

Keywords: malignant fibrous histiocytoma, array-CGH, chromosome 3 amplification.

Abbreviations:

Array-CGH: Microarray-based Comparative Genomic Hybridization
BAC: Bacterial Artificial Chromosome
Chromosome CGH: Comparative Genomic Hybridization on chromosomes
DOP-PCR: Degenerated OligoPrimed Polymerase Chain Reaction
DM: Double Minute (chromosomes)
EMA: Epithelial Membrane Antigen
ESTs: Expressed Sequence Tags
FISH :Fluorescent In Situ Hybridization
HMGIC: High-mobility group protein C gene
kb: kilobase
LASP-1: LIM and SH3 protein 1 gene
LPP: Lipoma-Preferred Partner gene
LPS: LiPoSarcoma
MFH: Malignant Fibrous Histiocytoma
Mb: Megabase
PAC: P1 Artificial Chromosome
RT-PCR: Reverse transcriptase Polymerase Chain Reaction
SSC: Sodium Chloride-Sodium Citrate buffer

Overlapping 3q28 amplifications in COMA cell line and undifferentiated primary sarcoma

Thomas Hussenet[1], Nedjoua Mallem[1], Richard Redon[1,3], Bernard Jost[1], Alain Aurias[2] and Stanislas du Manoir[1]

1: Molecular Pathology Department, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/Collège de France, 1 rue Laurent Fries, BP10142 67404 Illkirch Cedex, C.U. de Strasbourg, France ; 2: INSERM U509, Molecular Pathology of Cancers, Institut Curie, Paris, France ; 3: Present address: The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

ABSTRACT

Historically, amplicons mapping and characterization of Double Minute (DM) chromosomes content have been the way to pinpoint important oncogenes. Interestingly, the COMA cell line established from a sarcoma contains DMs some of them composed of material of the long arm of chromosome 3. To identify putative oncogenes on 3q that may be included in these DMs, we have analyzed the COMA cell line by array-CGH. We have detected the amplification of 1-Mb segment at 3q28, which contains the genes *LPP*, *FLJ42393* and *hsa-mir-28*. FISH experiments confirmed the presence of numerous copies of 3q28 segment included in DM chromosomes. Further screening of 8 undifferentiated primary sarcomas with 3q gains previously detected by chromosome CGH disclosed, in two cases, amplifications at 3q28 overlapping the 1-Mb segment amplified in COMA. To isolate target genes up-regulated by gene dosage effect, we measured the transcription levels of every Refseq gene located in the common region of amplification, selected ESTs and the micro-RNA *hsa-mir-28* in the COMA cell line compared to one MFH cell line without alteration at 3q28. Expression levels of all transcripts were almost similar in both cell lines, except for two ESTs (AI338598 and BX118304) showing a 20-fold increase. Since these two transcripts are poorly characterized, their contribution to MFH carcinogenesis is difficult to evaluate.

INTRODUCTION

Amplicon mapping is a traditional positional cloning strategy to identify candidate oncogenes. Following this line, characterization of 12q amplifications in human soft tissue tumors pinpointed a genomic region encompassing *mdm2* gene [1]. This gene was first isolated in a spontaneously transformed cell line as included in DM chromosomes [2] and later on its oncogenic properties and involvement in sarcoma were established. Improvement of spatial resolution of DNA copy number variation methods by the introduction of array CGH [3, 4] has streamlined this strategy.

The COMA cell line has a complex karyotype including DM chromosomes [5]. Previous FISH analysis revealed that the *c-myc* gene is present in some DMs when others contain material from chromosome 3. Precise delineation of the segment amplified on chromosome 3 may uncover new candidate oncogenes for soft tissue tumors or even others tumors types as exemplified by *mdm2* discovery in sarcomas followed by demonstration of its implication in other cancer types [6]. The COMA cell line was established from a storiform pleomorphic tumor from the thigh, initially diagnosed as a malignant fibrous histiocytoma (MFH). MFH were formerly one of the most commonly diagnosed soft tissue sarcoma. However, this particular tumor type is progressively less considered as a diagnostic entity *per se* [7, 8]. MFH are now rather regarded in the new WHO classification as undifferentiated forms of various other soft tissue sarcomas [8].

Transcriptome analysis of sarcoma tumors [9-11] have indeed repetitively uncovered that classifications derived from gene expression profiles failed to separate some MFH samples from other soft tissue sarcomas, in particular leiomyosarcomas (LMS ref 9,10) or liposarcomas (LPS;11). In addition, cytogenetic analysis of MFH has revealed very complex karyotypes frequently including homogeneously stained regions or DM chromosomes [12]. A survey of 30 tumors using chromosome CGH showed frequent losses on chromosome 13

(17/30) and a set of 11 chromosomal regions commonly gained [13]. This pattern of aberrations is broadly overlapping with the pattern found in leiomyosarcomas (LMS) and thus suggests that a notable proportion of MFH could actually be undifferentiated LMS [14]. In contrast, another sub-group of MFH - with simple karyotypes and recurrent high-level 12q14-q15 amplifications - corresponds to undifferentiated liposarcomas (LPS, 15]. This region contains the *mdm2* gene, initially isolated as amplified in DM chromosomes [2]. Although some MFH are now clearly considered as LMS or undifferentiated LPS [16], other tumor subtypes may still exist among initially diagnosed MFH. The discovery of recurrent genetic abnormalities, amplifications in particular, may participate to their identifications and thus to the definition of new subgroups, just as recently achieved for LPS [17].

In this study, we found a narrow amplification at 3q28 in the COMA cell line by array-CGH. Analysis of 8 additional undifferentiated primary sarcomas (initially diagnosed as MFHs and with gains on chromosome 3q) revealed two amplifications at 3q28 and allow to define a common region of about one megabase. Expression analysis of the genes located in this region demonstrated that two transcripts are strongly over-expressed in the COMA cell line compared to a MFH cell line without 3q amplification. The possible role of these two transcripts in the pathogenesis of undifferentiated sarcomas remains difficult to envision as very few information on their function are available *in silico*.

MATERIAL AND METHODS

Web resources

Genome browser, May 2004 version (NCBI build 35), UCSC: http://genome.ucsc.edu/
RepeatMasker software, Institute for Systems Biology: http://www.repeatmasker.org/
Primer3 software, MIT: http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi

Samples

The COMA cell line was established [5] from a storiform pleomorphic MFH tumor localized in the thigh. The tumor was positive for vimentin and slightly positive for EMA, desmin and negative for S100 antigen. The MFH19 cell line was established by A. Aurias group from a stage III and poorly differentiated storiform pleomorphic MFH tumor localized in the psoas. Eight tumors samples with chromosome CGH profiles showing gain of 3q were selected among the collection of A. Aurias (no RNA was available for the primary tumors). Main clinical and histological features are summarized in Table I.

Array-CGH

Pan-genome chip
IntegraChip™ were obtained from IntegraGen (http://www.integragen.com) and contain 3172 clones.

Chromosome 3 chip
Clones: BAC and PAC clones are coming from the RPCI-11, RPCI-4 and RPCI-5 collection. This version of the chip contains 223 chromosome 3 clones (listed with additional information in table 2) and is an upgrade of the previous one (85 clones; ref 18). Additional clones have been selected (http:/www.chori.org/bacpac/) to obtain a better final coverage of the short arm and the 3q26-qter region (98 clones for 37 Mb). Clones from sub-telomeric regions [19] were included to estimate the "normal genomic ratios" (see Data analysis).

Array preparation

Two level PCR amplifications were used to prepare DNA for spotting as described previously [20] with modifications. BAC DNAs were extracted using Millipore Montage™ BAC96 Miniprep Kit following supplier instructions after culture in 96 well plates with 1.5mL 2XLB medium. Each BAC was then amplified by two independent DOP-PCR [20] using two different primers: 6MW [21] and DOP2 [20]. About 75 ng of DNA of the primary DOP-PCR products were then amplified through a secondary PCR [20]. Finally, PCR products corresponding to a single BAC were pooled, ethanol-precipitated, dried and resuspended in 75% formamid and 25% water for spotting using a Microgrid II Arrayer (Biorobotics) on GAPS II slides (Corning). Spotted slides were processed as described [18].

Detailed conditions of the two PCR steps (buffers, volumes and reaction cycles) are available as supplementary information.

Array design

Two identical arrays each containing triplicate spots for each BAC were spotted by slide, hybridized simultaneously and subsequently averaged.

Sample DNA labeling

For each DNA, 300 ng of sonicated DNA were labeled by Random-Priming (Bioprime Labeling kit, Invitrogen) according to the supplier instructions but for 18h. Nucleotides final concentrations were 0.04 mM of dCTP-Cy5 for tumor DNA or dCTP-Cy3 for control DNA (Amersham Biosciences), 0.05 mM of dATP, dGTP, dTTP and 0.01 mM dCTP. Labeled DNAs were purified using MicroSpin-G50 columns (Amersham Biosciences). COT Human DNA (67.5 µg, Roche) and Salmon Testes DNA (30 µg, Sigma Chemical Co.) were added to the mixed tumor and reference DNAs. They were subsequently concentrated using a microcon YM-30 (Millipore) and resuspended to obtain 200 µl of hybridization solution (final concentration: 50% formamide, 0.8 x SCC, 16% W/V dextran sulfate, 2% SDS, pH=7). Finally, the probes were pre-annealed for 3 hours at 37°C.

Hybridization & post-hybridization

Every step was carried out in the HS 4800 hybridization station (Tecan). Slides pre-treatment included: Stripping: 2.5 mM Na2HPO4, 0.1% SDS at 76 °C for 1 min.; Slide blocking: 3% BSA (fraction V, Sigma Chemical Co.), 5 x SSC, 0.2% SDS at 42 °C for 20 min.; Pre-hybridization: 400 µg of Salmon Testes DNA and 67,5 µg of COT Human DNA in 200 µl of hybridization solution incubated 1h at 37°C.

Hybridization was performed during 18h at 37°C under low agitation. Four post-hybridization washes were carried out: 50% formamide, 2 x SSC, 0.1% SDS at 45 °C for 37 sec, 4 times; 2 x SSC, 0.1% SDS at 45 °C for 37 sec, 4 times; 4 x SSC, 0.1 Tween 20 at 45 °C for 37 sec, 4 times; slides were finally nitrogen dried and scanned using a ScanArray 4000 (Packard Bioscience). Images were quantified using Imagene 5.1 software as described previously [18].

Data analysis

For each array, we excluded from analysis the clones that displayed signal intensities for control DNA hybridization lower than twice the local background values for more than one of three replicates. After subtraction of local background, we calculated the ratio of Cy5 and Cy3 intensities. Values were considered as valid only if the coefficient of variation (SD/mean) of the replicates was lower than 20%. For each clone, the geometric mean of the ratios for the upper and lower arrays was then calculated. To determine the "normal genomic ratios", we calculated the median of the ratios obtained for the sub-telomeric clones. Then, individual ratios were normalized to 1 using this median. We used the values of 2, 1.25 and 0.75 as thresholds for amplifications, gains and losses, respectively.

FISH

For FISH analysis, the clones 103, 104, 105 (respectively RP11-426J22, RP11-67E18, RP11-297K7, see table 2) and RP1-80K22 (containing the *c-myc* gene) were labeled by random priming with digoxigenine-11-dUTP or biotin-16-dUTP. The probes were then hybridized at 37°C to COMA metaphase chromosomes or interphase nuclei for 24 h and then detected as previously described [22] using Cy3 and FITC based detection system. Chromosomes were counterstained with 4,6-diamidino-2-phenylindole.

PCR

Semi-quantitative PCR on genomic DNAs

Experiments were carried out as we previously described [18]. Primers were designed firstly on exons of the two Refseq genes and secondly to cover as much as possible the amplified region found by array-CGH with minimal gaps. Briefly, the Primer3 software was used to choose primer pairs targeting 16 loci in the region. We first set up PCR conditions on normal genomic DNA to correct for different amplifications efficiencies between primer pairs: varying amounts for each PCR primer pair were tested to obtain the minimum of variation between different genes after quantification of their respective bands intensities. Final primers concentration ranged from 225 nM to 1 μM (primer sequences and final concentrations used in the reactions are available in table 3). PCR products were mixed with PicoGreen® (Molecular Probes), run on 2% agarose gels, and gel images were acquired using the Typhoon™ 9200 imager (Amersham Biosciences). Finally, band intensities were quantified using the TotalLab® software (nonlinear dynamics) and normalized to GAPDH band

intensity obtained for the COMA and the normal genomic DNAs prior to ratios calculation for each locus. Reactions were performed 3 times independently and ratios subsequently averaged.

RT-PCR

RNA extractions, reverse-transcription and PCR were performed as previously described [18]. Putative contaminating genomic DNA was removed by digestion of total RNA for 1h at 25°C by digestion with 10U RNase-free DNaseI (Roche Diagnostics) and subsequent RNA purification (RNeasy mini kit, Qiagen) prior to reverse-transcription. Semi-quantitative PCR was performed, using primer pairs listed in table 4, for 30 cycles (10" at 95°C, 10" at 60°C, 10" at 72°C) for all genes or ESTs. For *GAPDH* the reference gene, 20 cycles were applied to avoid saturation. Quantifications were then done as described above for PCR on genomic DNA.

Northern Blots for microRNA expression

Analysis was performed following the guidelines described in the protocol from Nelson C. Lau [23], the detailed updated protocol is available at http://web.wi.mit.edu/bartel/pub/protocols/miRNA_Nrthrns_Protocol.pdf). Briefly, equal loading of total RNA from each cell line were verified by ethidium bromide staining of PAGE gels prior to their electrophoretic transfer to nitrocellulose membranes (Hybond N+, Amersham). Membranes were then hybridized with a 5' labelled oligonucleotide, of 21-nucleotides sequence composition antisense to hsa-miR-28 (sequence of the oligonucleotide used: 5'-CTCAATAGACTGTGAGCTCCTT-3'; miR-28 sequence: 5'-AAGGAGCUCACAGUCUAUUGAG-3', ref 24). After the washes and revelation using an imaging plate (type BAS-III, Fuji), the blots were scanned using a Typhoon™ 9200 imager (Amersham Biosciences).

RESULTS

Characterization of chromosomal imbalances in the COMA cell line by array-CGH

Chromosome CGH analysis of the COMA cell line showed the partial gain of chromosome 3q. Interestingly, the intensity ratio profile suggests the presence of DNA amplification in the terminal 3q region (see Fig. 1A and 5). To confirm this hypothesis, we have further investigated DNA copy number changes on chromosome 3 in COMA by CGH using one dedicated BAC/PAC microarray. In accordance with chromosome-CGH, the chromosome 3 array-CGH profile presents one region of gain on the long arm (156-163 Mb) with ratios higher than 1.5 (Figure 1B). In addition, a narrow high-level amplification centered at 190 Mb is revealed by ratio values higher to 2 for the three adjacent clones 103, 104 and 105 (Fig. 1B and C). The extent of this amplification is 1.5 Mb (central positions of the first excluded clones: 188.9 and 190.4 Mb). The *Lipoma-Preferred Partner* gene (*LPP*) is located in this interval.

Using a commercially available pan-genome chip, we also detected the presence of the 8q24 amplification with ratio values higher than 4 for two adjacent clones flanking the *c-myc* locus consistent with earlier chromosome CGH results (Fig. 1D and E).

FISH experiments on COMA metaphase spreads and interphase nuclei with probes from amplified regions

To explore the structural organization of these amplifications, we carried out FISH experiments on COMA nuclei and metaphase chromosomes with probes from amplified region detected by array-CGH on chromosomes 3 and 8 (BAC clones 103, 104 and 105 ; clone RP1-80K22 containing *c-myc*). For each of the three chromosome 3 clones the array-CGH results were confirmed: more than 20 hybridization signals were found dispersed inside nuclei in more than 40% of the cells (Fig. 1F and 1G). We noted that the DM chromosomes painted by the chromosome 3 clones are different from those highlighted containing *c-myc* and thus concluded that two exclusive populations of DM chromosomes coexist in the COMA cell line likely resulting of two independent cytogenetic events. The 3q amplified region encompasses the *LPP* gene, which is frequently implicated in t(3;12) translocation in lipomas resulting in *LPP-HMGIC* fusion [25]. Thus, we co-hybridized the chromosome 3 clones with one clone representing the *HMGIC* locus on chromosome 12 on COMA metaphase spreads: no evidence of this translocation was observed (Fig. 1H).

Characterization by array-CGH of 8 selected undifferentiated sarcomas with 3q gains

To determine whether this amplification at 3q28 (at 190Mb) is present in other MFH, we screened 8 selected primary soft-tissue sarcomas - initially diagnosed as MFH tumors -, with gains on the long arm of chromosome 3 detected on chromosome CGH profiles (Fig. 2A). Partial 3q gains were found by array-CGH in all of them except one (MFH93). The frequency of gain per 2 Mb genomic interval is maximum for the class 188-190 Mb (Fig. 2B). Two preferential regions of gain are observed, at 152 Mb and from 180 Mb to the telomere. In the terminal segment, amplifications (defined as ratio higher than 2) were measured in three samples (Fig. 2C). The amplified 3q28 region in MFH20 is complex with at least four ratio maxima between 181 and 195 Mb, whereas it is simpler in BER18 and restricted to two adjacent clones with ratio higher than 2 (around 190 Mb, Fig 2C). The extent of this amplification is 1.7 Mb (central positions of the first excluded clones: 188.9 and 190.6). The MFH114 tumor displays a narrow amplification (between 193.6 and 194.4) encompassing one clone (position 193.7), more distal than for the two other tumors.

In summary, gains around 190 Mb occur in about 60% of the screened cases and amplifications at this locus were found three times (two tumors and the COMA cell line). These amplifications overlap, defining a consensus region of 1.5 Mb in size (Fig. 2D).

Fine delineation of the 3q28 amplification in COMA by semi-quantitative genomic PCR

We carried out semi-quantitative genomic PCR for 16 loci located between the two first excluded BAC clones in the array-CGH experiment (Fig. 1C, for details see Materials and Methods). Sharp variations found at 189.3 and 190.25 Mb allow to define the extend of the amplicon size to 0.88-1.09 Mb. The borders are localized within 150 kb and 62 kb DNA segments. Sequence analysis using UCSC genome browser and Repeatmasker software reveals peculiar features only in the second segment, with a high-interspersed repeat content (62%) and particularly of LINE1 (35%). A small sub-region from 190.1 to 190.25 Mb inside the amplicon shows higher copy number (see Figure 3 A).

Transcriptional levels of genes included in the amplicon by RT-PCR

We investigated the expression levels of all Refseq genes mapped in the amplicon (namely *LPP* and *FLJ42393*) as well as one microRNA encoding gene (*hsa-mir-28*) and 5 selected ESTs (including BG778254 and AW277126) located in the sub-region with apparent higher DNA copy number. We were able to obtain specific products for the Refseq genes and

two ESTs (AI338598 and BX118304). The three others ESTs were found not to be expressed or gave unspecific PCR amplification products.

Comparison of COMA and the MFH19 cell line - which has no 3q28 copy number variation detected by array-CGH (data not shown) - indicates that LPP expression level is not changed while FLJ42393 is slightly over-represented (Fig 3B). Northern blot analyses of hsa-mir-28 reveal it is expressed at similar levels in both cell lines (Figure 3C). Interestingly, for AI338598 and BX118304 ESTs, a twenty-fold increase was observed in COMA (Fig 3B).

In conclusion, this strong over-expression of AI338598 and BX118304 ESTs in COMA cell line, is in the same order of magnitude than the number of DM chromosomes as seen by FISH (some tens per cell), and is thus very likely to be due to gene dosage effect.

DISCUSSION

We have mapped at high resolution a narrow amplification of the 3q28 locus contained in double minute chromosomes in the COMA cell line. Furthermore, we have confirmed the presence of two different populations of DM chromosomes as previously shown by Mairal et al [5]. Some DM chromosomes contain the gene *c-myc* while the others are composed of material from the 3q28 locus. This excludes the possibility that amplification of 3q material is a side effect of *c-myc* amplification - with the 3q sequence hitchhiking by contingence to the *c-myc* segment, as shown for other co-amplified segments [26, 27].

Although amplification at 3q28 appears uncommon in MFH, 4 tumors were found to contain high-level amplifications at 3q26-qter in a series of 25 tumors analyzed by chromosome CGH [13]. Thus, the question arises whether a rare subgroup of MFH or sarcomas may contain high copy number of the particular 3q28 segment amplified in COMA. By using array-CGH for the analysis of 8 tumors with 3q gain previously described by chromosome CGH, we uncovered two tumors with amplifications at 3q28 overlapping with the amplicon delineated in COMA cells. Considering the small amplified segments (less than 1.7 Mb) identified in both COMA and BER18 cells, this overlap was unlikely to occur by chance (Fig. 2D).

For the COMA cell line, mapping of this 3q28 amplicon was refined by genomic PCR performed on 16 loci and demonstrated that its actual size is 0.88-1.09 Mb. No large-scale copy number polymorphism is reported within this region in the Database of Genomic Variants (http://projects.tcag.ca/variation/). Volik *et al.* have recently made a detailed description of an amplicon containing 20q and 3p segments in the MCF7 cell line [27]. They reported that three of the four characterized breakpoints lie in a region with high repetitive density in particular L1 elements suggesting that recombination between repetitive elements may play a role in amplicon formation. Interestingly, the telomeric breakpoint is located in a L1 repeat enriched region. In addition, we found an internal segment of about 150 kb (delineated by the BG778254 and AW277126 ESTs) that seems to be present in higher copy number in COMA cell line. Previous studies on the structural organization of DNA amplifications in cancer cells have shown their unstable nature and that secondary rearrangements can for instance lead to higher level amplifications of internal sub-regions eventually favored by structural DNA features such as repetitive sequences [27], matrix attachment [28] and /or replication initiation sites [29].

Bioinformatic analysis of the 3q28 amplicon sequence indicates a relatively poor gene content. Two Refseq genes are mapped in this interval, *LPP* and *FLJ42393*, as well as one microRNA-encoding gene, *hsa-mir-28*.

LPP gene was first isolated in a translocation t(3;12), which is very common in benign lipomas [25]. This translocation was not present in COMA cell line according to our experiments. Moreover, no increase in *LPP* expression was found in COMA compared to MFH19, another MFH cell line but without amplification at 3q28. Thus, neither normal nor rearranged *LPP* transcripts are over-expressed in COMA.

The second candidate, *FLJ42393*, is a Refseq gene predicted solely on mRNA evidence. Transcriptional analysis showed only a slight over-expression but not at the level expected when considering the number of DM chromosomes per cell.

The third candidate, *hsa-mir-28*, encodes a microRNA and is located in an intron of *LPP*. This microRNA is expressed at similar levels in COMA and MFH19 cell lines. However, human miRNA genes are frequently located in genomic regions involved in cancers [30] and there is growing evidence of a strong functional link between miRNAs and cancer [31-34]. In consequence, it remains of interest to investigate whether *hsa-mir-28* may participate to the tumor phenotype in MFH or sarcomas.

As the histogenesis of these tumors is still unknown, we were unable to investigate expression levels of the three candidate genes in corresponding normal cells. Given that their expression was observed in both MFH cell lines and assuming that expression could be activated by an other mechanism than gene amplification in the control cell line used in our study, we cannot exclude that one of them could be critical for this tumor phenotype.

Finally, we evaluated the expression levels of 5 ESTs located in the higher copy segment in the genomic PCR experiments. We found a 20-fold increased expression in COMA for two of them, AI338598 and BX118304. Both ESTs are supported by a unique Unigene cluster containing other EST sequences (Hs.333208 and Hs.54542, respectively). Analyses of sequence databases and gene/CDS predictions tools did not show any evidence of protein-coding sequence. Although they are not included in the current mammalian noncoding RNA databases [35,36], we cannot exclude that they represent non-coding RNA. Moreover, both ESTs to be expressed in two different MFH cell lines, making them unlikely to represent cloning artifacts or contaminating sequences, which are frequently observed in ESTs databases. Further experimental investigations are still needed to decipher the structure and the expression regulation of these two putative genes. Finally, we cannot exclude the presence of another non-annotated gene in this region.

ACKNOWLEDGMENTS

We thank Jean-Louis Mandel for fruitful discussions. We are grateful to A. Baguet and C. Tomasetto for technical advice and thank in particular the DNA chip, cell culture and image facilities from the IGBMC. This study was supported by grants from the 'Comité Régional du Haut-Rhin de la Ligue contre le Cancer' and by funds from INSERM, CNRS and the 'Hôpital Universitaire de Strasbourg' and Identity card program (CIT) from 'Ligue Nationale contre le Cancer'. Equipments necessary for array-CGH was funded by 'Association pour la Recherche sur le Cancer'. Thomas Hussenet is recipient of a bursary from 'Association pour la Recherche sur le Cancer'.

REFERENCES

[1] Nilbert M, Rydholm A, Mitelman F, Meltzer PS, Mandahl N. Characterization of the 12q13-15 amplicon in soft tissue tumors. Cancer Genet Cytogenet 1995;83:32-6.

[2] Fakharzadeh SS, Trusko SP, George DL. Tumorigenic potential associated with enhanced expression of a gene that is amplified in a mouse tumor cell line. EMBO J 1991;10:1565-9.

[3] Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer 1997;20:399-407.

[4] Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 1998;20:207-11.

[5] Mairal A, Chibon F, Rousselet A, Couturier J, Terrier P, Aurias A. Establishment of a human malignant fibrous histiocytoma cell line, COMA. Characterization By conventional cytogenetics, comparative genomic hybridization, and multiplex fluorescence In situ hybridization. Cancer Genet Cytogenet 2000;121:117-23.

[6] Courjal F, Cuny M, Rodriguez C, Louason G, Speiser P, Katsaros D, Tanner MM, Zeillinger R, Theillet C. DNA amplifications at 20q13 and MDM2 define distinct subsets of evolved breast and ovarian tumours. Br J Cancer 1996 74:1984-9

[7] Weiss SW, Goldblum JR eds. Enzinger and Weiss's Soft Tissue Tumors, 4th ed. St Louis: Mosby, 2001. 1632p.

[8] Fletcher CD, van denBerg E, Molenaar WM. Tumors of soft tissue and bone, WHO:Pleomorphic malignant fibrous histiocytoma/undifferentiated highgrade pleomorphic sarcoma. In: Fletcher CD, Unni KK, Mertens F, editors : World Health Organization Classification of tumors, Pathology of Tumors of Soft Tissue and Bone. Lyon: IARC Press, 2002; pp. 120-122

[9] Skubitz KM, Skubitz AP. Characterization of sarcomas by means of gene expression. J Lab Clin Med 2004;144:78-91.

[10] Lee YF, John M, Edwards S, Clark J, Flohr P, Maillard K, Edema M, Baker L, Mangham DC, Grimer R, Wooster R, Thomas JM, Fisher C, Judson I, Cooper CS. Molecular classification of synovial sarcomas, leiomyosarcomas and malignant fibrous histiocytomas by gene expression profiling. Br J Cancer 2003;88:510-5.

[11] Segal NH, Pavlidis P, Antonescu CR, Maki RG, Noble WS, DeSantis D, Woodruff JM, Lewis JJ, Brennan MF, Houghton AN, Cordon-Cardo C. Classification and subtype prediction of adult soft tissue sarcoma by functional genomics. Am J Pathol 2003;163:691-700.

[12] Mertens F, Fletcher CD, Dal Cin P, De Wever I, Mandahl N, Mitelman F, Rosai J,

Rydholm A, Sciot R, Tallini G, Van den Berghe H, Vanni R, Willen H. Cytogenetic analysis of 46 pleomorphic soft tissue sarcomas and correlation with morphologic and clinical features: a report of the CHAMP Study Group. Chromosomes and MorPhology. Genes Chromosomes Cancer 1998;22:16-25.

[13] Mairal A, Terrier P, Chibon F, Sastre X, Lecesne A, Aurias A. Loss of chromosome 13 is the most frequent genomic imbalance in malignant fibrous histiocytomas. A comparative genomic hybridization analysis of a series of 30 cases. Cancer Genet Cytogenet 1999;111:134-8.

[14] Derre J, Lagace R, Nicolas A, Mairal A, Chibon F, Coindre JM, Terrier P, Sastre X, Aurias A. Leiomyosarcomas and most malignant fibrous histiocytomas share very similar comparative genomic hybridization imbalances: an analysis of a series of 27 leiomyosarcomas. Lab Invest 2001;81:211-5.

[15] Chibon F, Mariani O, Derre J, Malinge S, Coindre JM, Guillou L, Lagace R, Aurias A. A subgroup of malignant fibrous histiocytomas is associated with genetic changes similar to those of well-differentiated liposarcomas. Cancer Genet Cytogenet 2002;139:24-9.

[16] Chibon F, Mariani O, Mairal A, Derre J, Coindre JM, Terrier P, Lagace R, Sastre X, Aurias A. The use of clustering software for the classification of comparative genomic hybridization data. an analysis of 109 malignant fibrous histiocytomas. Cancer Genet Cytogenet 2003;141:75-8.

[17] Fritz B, Schubert F, Wrobel G, Schwaenen C, Wessendorf S, Nessling M, Korz C, Rieker RJ, Montgomery K, Kucherlapati R, Mechtersheimer G, Eils R, Joos S, Lichter P. Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. Cancer Res 2002 62:2993-8.

[18] Redon R, Hussenet T, Bour G, Caulee K, Jost B, Muller D, Abecassis J, du Manoir S. Amplicon mapping and transcriptional analysis pinpoint cyclin L as a candidate oncogene in head and neck cancer. Cancer Res 2002;62:6211-7.

[19] Knight SJ, Lese CM, Precht KS, Kuc J, Ning Y, Lucas S, Regan R, Brenan M, Nicod A, Lawrie NM, Cardy DL, Nguyen H, Hudson TJ, Riethman HC, Ledbetter DH, Flint J. An optimized set of human telomere clones for studying telomere integrity and architecture. Am J Hum Genet 2000;67:320-32.

[20] Fiegler H, Carr P, Douglas EJ, Burford DC, Hunt S, Scott CE, Smith J, Vetrie D, Gorman P, Tomlinson IP, Carter NP. DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. Genes Chromosomes Cancer 2003;36:361-74.

[21] .Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA, Tunnacliffe A. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics 1992;13:718-25.

[22] Redon R, Muller D, Caulee K, Wanherdrick K, Abecassis J, du Manoir S. A simple specific pattern of chromosomal aberrations at early stages of head and neck squamous cell

carcinomas: PIK3CA but not p63 gene as a likely target of 3q26-qter gains. Cancer Res 2001;61:4122-9.

[23] Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 2001;294:858-62.

[24] Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science 2001;294(5543):853-8.

[25] Petit MM, Mols R, Schoenmakers EF, Mandahl N, Van de Ven WJ. LPP, the preferred fusion partner gene of HMGIC in lipomas, is a novel member of the LIM protein gene family. Genomics 1996;36:118-29.

[26] MacLeod RA, Spitzer D, Bar-Am I, Sylvester JE, Kaufmann M, Wernich A, Drexler HG. Karyotypic dissection of Hodgkin's disease cell lines reveals ectopic subtelomeres and ribosomal DNA at sites of multiple jumping translocations and genomic amplification. Leukemia 2000;14:1803-14.

[27] Volik S, Zhao S, Chin K, Brebner JH, Herndon DR, Tao Q, Kowbel D, Huang G, Lapuk A, Kuo WL, Magrane G, De Jong P, Gray JW, Collins C. End-sequence profiling: sequence-based analysis of aberrant genomes. Proc Natl Acad Sci USA 2003;100:7696-701.

[28] Kuwahara Y, Tanabe C, Ikeuchi T, Aoyagi K, Nishigaki M, Sakamoto H, Hoshinaga K, Yoshida T, Sasaki H, Terada M. Alternative mechanisms of gene amplification in human cancers. Genes Chromosomes Cancer 2004;41:125-32.

[29] Shimizu N, Hashizume T, Shingaki K, Kawamoto JK. Amplification of plasmids containing a mammalian replication initiation region is mediated by controllable conflict between replication and transcription. Cancer Res 2003;63:5281-90.

[30] Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, Croce CM. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci USA 2004;101:2999-3004.

[31] He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM. A microRNA polycistron as a potential human oncogene. Nature 2005;435:828-33.

[32] Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. MicroRNA expression profiles classify human cancers. Nature 2005;435:834-8.

[33] Eis PS, Tam W, Sun L, Chadburn A, Li Z, Gomez MF, Lund E, Dahlberg JE. Accumulation of miR-155 and BIC RNA in human B cell lymphomas. Proc Natl Acad Sci USA 2005;102:3627-32.

[34] Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ. RAS is regulated by the let-7 microRNA family. Cell 2005;120:635-47.

[35] Pang KC, Stephen S, Engstrom PG, Tajul-Arifin K, Chen W, Wahlestedt C, Lenhard B, Hayashizaki Y, Mattick JS. RNAdb--a comprehensive mammalian noncoding RNA database. Nucleic Acids Res 2005;33 Database Issue:D125-30.

[36] Rfam: annotating non-coding RNAs in complete genomes. Sam Griffiths-Jones, Simon Moxon, Mhairi Marshall, Ajay Khanna, Sean R. Eddy and Alex Bateman.Nucleic Acids Res 2005 33:D121-D124.

TABLE 1

NAME	GENDER	AGE	PRIMARY	GRADE	LOCALISAT.	CHIMIO.
MFH56	F	28	Relapse	2	cheek	no
MFH53	F	59	Prim.	3	breast	no
LMS41	F	52	Prim.	3	os ischii	no
MFH93	F	82	Prim.	3	thigh	no
BER18	F	60	Prim.	3	buttock	no
MFH20	F	53	Prim.	3	thigh	no
MFH114	M	78	Prim.	3	thigh	no
MFH51	M	NA	Prim.	3	thorax	no
MFH59	M	60	Prim.	2	collarbone region	no

Prim. : Primary tumor; Localisat : localization,; Chimio. : previous chemotherapy treatment.

TABLE 2

BAC_clone	Local_name	Band	FISH	Fish_Valid	Bac_Acc_num	Contig_Acc	Bac_End_Acc	STS_Id	MAY2004-FINAL
RP4-632A13	3	3p24	F	UCSF		NT_005825		D3S3658	40878228,5
RP4-702I22	4							AFMb047xc1	72447435,5
RP4-616B15	6							AFM031xa1	105101560
RP4-710G11	7							AFMc028ye5	110604620
RP4-671C12	8							AFM321xf5	114555098
RP5-954O7	9							AFM116xh2	124525986
RP5-1043H21	10							AFMb025we5	128755753
RP4-730E1	12							AFM268vc9	141116102
RP5-1164J18	13							AFMA106YF9	144198404
RP5-924I19	14							AFM292ye5	146810876
RP4-727D1	15							AFM220YH12	151666643
RP5-854L10	16							AFM157xb4	157629042
RP4-701G18	18							AFM224ya3	171671691
RP5-1020C14	20							AFM206zf4	178878418
RP5-870D10	21							AFMa239zb5	185767969
RP5-875F16	22							AFM308zh9	187514568
RP4-756C5	23							AFM207yd8	191144958
RP4-714P24	24							AFM308yf1	193160144
RP11-245C23	32				AC076966	NT_022396	AQ487241	RH70978;GDB:596369;SHGC-132833;SHGC-12912;SHGC-77571;RH41892;RH65632;SHGC-53062;AFM	180397435
RP11-414N11	33							AFMb354xb1	177871287
RP11-310L1	35							WI-1347;D3S2346	178128419
RP11-120G16	36						AQ350347;AQ350348		178224653
RP11-238P22	37							WI-3743	178243419
RP11-566L8	39							WI-2020	190844456
RP11-352A18	40				AC024103		AQ530586;AQ530589	AFM164zf8	190694987
RP11-3K16	43				AC007849	NT_034563	B48820;B48819	D11S3226;D3S2982	170831395
RP11-198D7	45	3q26	F	UCSF		NT_025667	AQ416866;AQ416864	AFM164TC7;D3S1548	174653348
RP5-1038I8	46							AFM224zb12	174965539
RP11-333L12	50	3q26	F	UCSF				D3S3037	178924373
RP11-80E16	51						AQ282230	RH12776	185345683
RP11-517A22	52							WI-3008	173153975
RP11-204C23	55	3p26	F	FHCRC	AC024216;AC018814	NT_005927;NT_035205	AQ413785;AQ413782	AFM217XD2	1942962,5
RP11-28P14	56	3p25-3p26	F	FHCRC	AC067736	NT_005927	B87467;AQ003475	AFM234TF4	6916497,88
RP11-121D3	57	3p25	F	FHCRC	AC016963	NT_005927	AQ350922	AFM205VA9	2296199,5
RP11-245E5	58	3p24	F	FHCRC	AC018542	NT_022517	AQ487267;AQ487269	AFMB331YE9	22839543,4
RP11-11L6	59	3p23-3p24	F	FHCRC	AC021751;AC093412	NT_005646	B74911;B74912	AFM183YC3	28735130,5
RP11-103N21	60	3p23-3p24	F	FHCRC	AC018351;AC099048	NT_005646	AQ314093;AQ314091	AFM320WC9	29942873,2
RP11-286G5	61	3p22	F	FHCRC	AC020624;AC092051	NT_005580		AFM220XG3	36459077
RP11-111P21	62	3p21.3	F	FHCRC	AC063916;AC092500	NT_034534	AQ341909;AQ341906	AFMA231XE9	45406679,3
RP11-122D19	63	3p21.2	F	FHCRC	AC018354	NT_033038	AQ343573;AQ383795	AFM287YD9	54097919,5
RP11-124O2	64	3p21.2	F	FHCRC	AC018352		AQ343374;AQ343375	AFM198YF2;D3S1271;D3S4006	54435072,6
RP11-169G24	65	3p21.2	F	FHCRC	AC018920;AC099775	NT_005787	AQ413589	AFMA082WH1	56262269,5
RP11-94D19	66	3p14	F	FHCRC	AC016965;AC096917	NT_005607	AQ282117;AQ315786	AFM211ZG11	60842018,8
RP11-225J10	67							AFM294ZF9	86253613
RP11-203L15	68	3q25-3q26.1	F	FHCRC	AC016920	NT_034561	AQ419335;AQ419337	AFMA218YG1	162123535
RP11-198G24	69	3q26.1	F	FHCRC	AC018457	NT_022421		AFM116XH4	165377394
RP11-229I14	70						AQ487654;AQ487655	RH12631	180857358
RP11-441O11	71							WI-16592	181001671
RP11-398D6	73							RH10511	191506368
RP11-303B22	74							AFM260yb1	191574760
RP11-436M6	75				AC112775;AC009099	NT_033013;NT_033280		D3S1265;SHGC-143979;SHGC-77677;A008C12;SGC33690;SHGC-77679;SHGC-77692;RH48112;RH480	197013138
RP11-171N2	76				AC069513	NT_033013		D3S1265;SHGC-143979;SHGC-77677;A008C12;SHGC-77680;SGC33690;SHGC-77679;SHGC-77692;RH	196955625
RP11-496H1	77				AC024560	NT_022508		RH98664;SGC32359;WI-18114;STS-H75485;SHGC-77708;SHGC-56692;SGC34513;SHGC-77703;RH661	198903881
RP11-79L9	79	3q22	F	FHCRC	AC020890	NT_025664	AQ284970;AQ284966		139812325
RP11-4H14	80				AC068429;AC132032	NT_005832	B49041	WI-22240;SHGC-144240;RH103624;STS-H98013;SHGC-64739;D3S2376;RH92630;DXS7953;D3S1301	142147644
RP11-464E15	81				AC021059	NT_005616		D3S3098;D3S4076;STS-X79537;WI-19179;G59812;A002A11;G19843;D3S3493;SHGC-12604	150251221
RP11-385G14	82	3q25.1c	F	NCI	AC011103;AC108726			G54347;G54371;G54365;D3S3882;RH25242;STS-D13626;G54359;G54346;G54369;G54358;D3S1315;G5	152295786
RP11-529G21	83				AC015991;AC117394	NT_005849		SHGC-110392;SHGC-77468;SHGC-77466;RH79703	153599991
RP11-227J5	84				AC024221	NT_034561		HSC0TF112;RH46490;STS-N58827;STS-N58827;SHGC-52508;STS-W93650;SHGC-77500;STS-N90779;	161622185
RP11-16N11	85				AC069224	NT_034561	B82557;B83640	RH47985;SHGC-81842;SHGC-78467;G59303	161944552
RP11-526M23	86				AC048332			G64459;STS-M16541;D3S3886WI-18944;GDB:632866;GDB:580581;GDB:363553;GDB:363550;SHGC-10	166994038
RP11-89J17	87	3q26.3	F	FHCRC	AC016936	NT_022458	AQ284773;AQ284771		175858216
RP11-809F4	88				AC104640;AC068566	NT_022458	AQ598300;AQ710738	D3S2427;SHGC-18115;DXS7944	177185063
RP11-496B10	89				AC068298			RH66248;WI-11544;G49561;RH94147;STS-N22566;D3S3662;G49560	181874989
RP11-259I19	90	3q26.33c	F	NCI	AC024104	NT_005952	AQ484803;AQ484806	D3S3957;D3S2314;DXS7944;116XC5	183530635
RP11-139K4	91				AC021055		AQ383030;AQ383020		183734437
RP11-657G5	92				AC134503		AQ512213		181419076
RP11-14I2	93				AC068769	NT_022676	B85286	SHGC-77598;SHGC-77591;SHGC-33134;SHGC-146579;SHGC-77593;STS-R06585;SHGC-149652;RH93	184898518
RP11-12L14	94	3q27-3q28	F	FHCRC	AC013631	NT_022676	B75882;B75883	T69473	184955793
RP11-329B9	95				AC107294;AC016975	NT_005910		SHGC-84430;G66857;RH120888;RH93134;RH104275;RH12083;A006Q20;SHGC-77621	185993748
RP11-683J13	96				AC069420	NT_022549		SHGC-77628;RH81085;GDB:1317732;SHGC-36455;RH46339;RH93790;RH80449;RH92734	186755151
RP11-410F19	97				AC009247	NT_033009	AQ549693;AQ535500	RH44937;STS-T76997;WI-14755;SHGC-77608;SHGC-173174	187092173
RP11-379C23	98	3q27.2a	F	NCI	AC007917	NT_005962		G30123;G09397;D3S1602;SHGC-3840;WI-15878;SHGC-77630;SHGC-77615	187461731
RP11-119E13	99				AC012149	NT_005962		GDB:313513;G29832;D3S3935;STS-M11437;D3S3899;SHGC-35523;SHGC-77594;UTR-9640;SGC31473	187992697
RP11-208N14	100				AC007920;AC046154	NT_005962;NT_034634	AQ419439	RH123734;D3S2492;D3S3600;D3S3315;RH12357;WI-16893;RH12435;RH16233;SHGC-34843;STS-D175	188475578
RP11-211G3	101				AC072022	NT_005962		SHGC-110034;SHGC-112481;RH47862;SHGC-34056;RH69801;SHGC-64334;STS-W46183;RH70520;SH	188920816
RP11-30O15	102				AC108681;AC021414	NT_005962	B87920;B87921	DXS7944;D3S3651;SHGC-142224;SHGC-145538	189021608
RP11-426J22	103				AC069226	NT_005962	AQ554092;AQ554088	SHGC-7226;RH122128;G16600;SHGC-144996	189640224
RP11-67E18	104	3q28	F	FHCRC	AC055711	NT_005962	AQ201693	AFM270ZG9	189997148
RP11-297K7	105				AC009319	NT_005962		SHGC-84029;STS-T96174;RH93795;D3S2973;D3S3104;SHGC-77806;RH99187;SHGC-56873	190157896
RP11-115I8	106	3q28a	F	NCI	AC104669;AC018516	NT_005962	AQ345490;AQ345493		190393203
RP11-525C11	107				AC068645	NT_034569		D3S3321;Cda0ce01;D3S4224;SHGC-77656;RH92964;SHGC-32957;A001T04;RH45920;SHGC-77660;STS	192541332
RP11-608P9	108	3q28d	F	NCI	AC112913;AC025512		AQ359925;AQ367090	SHGC-149697;RH123908;RH120321;SHGC-81817	192993499
RP11-91M9	109	3q28	F	FHCRC	AC026671	NT_005571	AQ281615;AQ281611		193709683
RP11-532L22	110				AC016947	NT_034583		D3S2490;RH123300;DXS7953;D3S2491;D3S3642	9832610,5
RP11-143P4	111				AC092966;AC021846	NT_005571	AQ387254;AQ387251	SHGC-31081;RH39292;WI-19455;RH92253;G42341;D3S3642	194372946
RP11-528A4	112				AC069421	NT_005571		RH123557;STS-H99909;RH78614;SHGC-77667;SHGC-37414;SGC38140;RH103296;SHGC-33708;SHGC	194975072
RP11-252K11	114				AC026308	NT_029928	AQ490767;AQ490768;AZ517314	STS-U09609;RH79114;RH104251;A007I23;SHGC-36945;SHGC-77670;SGC33919;RH92567;SHGC-1493	197439873
RP11-23M2	115				AC022621	NT_034601	AQ014334;AQ014336	Cda0ba07;RH12742;RH99143;STS-T76997;SHGC-52923;SHGC-57092;RH103233;RH80465;STS-W7475	199072636
RP11-649A16	116				AC027060;AC092959	NT_005775	AQ437925	D3S1744;D3S2439;SHGC-31137;A009Z14;G33011;D3S3991;RH91704;SGC30596;SHGC-32454;RH1226	148629376
RP11-316N24	117				AC041012;AC134026	NT_005678	AQ541326;AQ507853	SHGC-35942;RH102541;A004B27;SHGC-30578;Cda09b06;SHGC-79130;SHGC-77472;Cda1ig01;SHGC-	155530994
RP11-197N19	118				AC020630	NT_005612	AQ416801	GDB:252088;WI-7070;SHGC-77488;RH98708;D3S1605;SHGC-77484	158655127
RP11-249J17	119				AC067725	NT_015150	AQ484902	AFMB337XE5	117079420
RP4-736H12	120	3p21	F	UCSF	AC006060	NT_005825		D3S2319	42392679,5
RP11-2L8	121				AC007953		B47827;B47826		179155798
RP11-7M24	122	3p26	F	FHCRC	AC012136	NT_005927	B72232;B72231		6267910,88
RP11-17C24	123	3p24.3	F	FHCRC	AC022049	NT_022517	B82630;B82629		21557007
RP11-21C11	124				AC013808;AC130412		B85804;B85805;AQ008848	SHGC-77838;SHGC-82717;SHGC-79939;SHGC-78824;SHGC-82702;D3S2974	179582315
RP11-21E14	125				AC015495;AC104467		B83110;B85853	D3S3389;G63125;D3S2974;SHGC-82702;SHGC-78824	179663085
RP11-32F23	126	3p26	F	FHCRC	AC069264	NT_005927	AQ047334	AFM217XD6	4107499,33
RP11-33A14	127				AC024220	NT_022509	AQ046296	D3S2974;SHGC-78824;SHGC-82702;G63125;SHGC-79939	179647298
RP11-41F5	128	3p24.3	F	FHCRC	AC020626	NT_005564	AQ029210;AZ518154	AFMA346XA9	23795120,9
RP11-51A1	129	3q26	F	UCSF	AC007653	NT_022396	AQ052255	D3S3565	181020757
RP11-53D15	130	3q28	F	FHCRC	AC016966	NT_005883	AQ053018	AFM207YD8	191200746
RP11-54L9	131	3q28	F	FHCRC	AC073963	NT_005883	AQ115238;AQ081612	AFM260YB1	191535526
RP11-57D6	132	3p25	F	FHCRC	AC017001	NT_005927	AQ194759;AQ116103		13229794,9
RP11-59J16	133	3q21	F	FHCRC	AC016968	NT_005588	AQ200043	AFMB020ZB9	128675761
RP11-70G21	134				AC110992;AC024387	NT_022509;NT_033261	AQ239108;AZ519697		179310562
RP11-71G7	135				AC092938;AC019160	NT_022458	AQ267761;AQ239210		177767289
RP11-72E23	136	3q24	F	FHCRC	AC016967	NT_005591	AQ267021;AQ285099	AFM210VE7	146616470
RP11-78O10	137	3p21	F	FHCRC	AC099330	NT_006014	AQ285668	AA129089	49925523,5
RP11-79A14	138	3q25	F	FHCRC	AC063937	NT_005612	AQ282386;AQ282389;AQ580454;AQ628779		159066912
RP11-79C12	139	3p12-3p13	F	FHCRC	AC092036	NT_022473	AQ282455;AQ282458;AQ581277;AQ629575;AZ519148		68077526,5
RP11-79E6	140	3p24	F	FHCRC	AC107621	NT_022517	AQ282584;AQ282587;AQ581307;AQ581310;AQ629581;AZ519155		20506145,1
RP11-79G24	141	3q26.1	F	FHCRC	AC024905	NT_022463	AQ283083;AQ581401;AQ629620;AZ519173		163094791
RP11-79K12	142	3q11.2	F	FHCRC	AC104303;AC076969		AQ283176;AQ283173;AQ581447;AZ519184;AQ581445	D3S1450;RH102173;RH93562;SHGC-14756;DXS7944	58810244,3
RP11-79H17	143	3q13.2-3q13.3	F	FHCRC	AC026354	NT_005795	AQ284249;AQ284252		113112448
RP11-79K17	144	3p14	F	FHCRC	AC121249	NT_022481	AQ281989;AQ317608;AZ519098		58823279,5
RP11-79L21	145	3q22	F	FHCRC	AC018916	NT_022393	AQ285013;AQ285027		134052035
RP11-80D24	147	3p24.3-3p25	F	FHCRC	AC090959	NT_005927	AQ317757;AQ317736		17263333,1
RP11-80H18	148	3p21-3p14	F	FHCRC	AC098479	NT_005787	AQ284089;AQ284092;AZ521462		58238207,6
RP11-80H8	149	3q23-3q24	F	FHCRC	AC018450	NT_005772	AQ284062;AQ284065		144298660
RP11-80L14	150	3q26.1	F	FHCRC	AC018917	NT_022421	AQ317796;AQ284165		165297625
RP11-81P15	151	3p12	F	FHCRC	AC012221	NT_022420	AQ282058;AQ282060		87134141,5
RP11-81D17	152	3p12	F	FHCRC	AC018918	NT_022459	AQ281061;AQ281063;AZ521538;AZ521540		83541904
RP11-81N13	153	3p14.1	F	FHCRC	AC016935	NT_022473	AQ281897;AQ282006;AZ516063		68091187,8
RP11-88B8	154	3p21.32	F	FHCRC	AC016928	NT_022567	AQ286022;AQ286020	D3S3729	47952748,8
RP11-88H10	155	3q24	F	FHCRC	AC055758	NT_005775	AQ286778;AQ286792;AZ516229		147098112
RP11-88H12	156	3p14.1	F	FHCRC	AC121493	NT_005530	AQ286794;AQ286797;AZ516230;AZ516242		65271702,4
RP11-88P6	157	3q27	F	FHCRC	AC018919	NT_005962	AQ286886;AQ286888;AZ516906;AZ516908		188776818
RP11-89E16	159	3q23	F	FHCRC	AC117396	NT_005655	AQ282753;AQ282756;AZ517032		141348557
RP11-89F17	160	3p21.2	F	FHCRC	AC092037	NT_005986	AQ284681;AQ284683;AZ517755		51474844,6
RP11-89F18	161	3p24.3	F	FHCRC	AC092038	NT_005564	AQ285843		23490277
RP11-89H10	162	3p12.3	F	FHCRC	AC021866	NT_022446	AQ285889;AQ285891	WI-11200;D3S2484;SHGC-112056	76183659,5
RP11-89O2	163	3p14	F	FHCRC	AC118548	NT_005999	AQ283526;AQ283529;AZ517689	AFM198TE7	64154475,6
RP11-90B15	164	3p22	F	FHCRC	AC099331	NT_005498	AQ283534;AQ283537;AZ517980		40190143,3
RP11-90H15	165	3p13-3p14	F	FHCRC	AC097632	NT_005769	AQ284299;AQ284302;AZ518485		71238740,4
RP11-90I9	166	3p24	F	FHCRC	AC099538	NT_022517	AQ281667;AQ281670		19118531,6
RP11-90M23	167	3p22	F	FHCRC	AC092417	NT_005580	AQ284410;AQ284407		38421293,6
RP11-90M7	168	3q26.1	F	FHCRC	AC024906		AQ283740;AQ283737	DXS7944;STS-T76997;DXS7568;DXS8042;SHGC-8793	164441652
RP11-90N21	169	3q26.1	F	FHCRC	AC026306	NT_005612	AQ285208;AQ285210;AZ518641		158898315
RP11-91A15	170	3p11-3q11.1	F	FHCRC	AC026305	NT_022497	AQ286316;AQ286318		89691227,8
RP11-91B3	171	3q13.1	F	FHCRC	AC016138	NT_030622	AQ283215;AQ283216		106876675
RP11-91E8	172	3p22	F	FHCRC	AC026349	NT_034534	AQ284420;AQ284417		46070042
RP11-91F9	173	3q13.3	F	FHCRC	AC023793	NT_005994	AQ283336;AQ283338;AZ519287		119337027
RP11-91K4	174	3p25-3p26	F	FHCRC	AC048350	NT_005927	AQ282613;AQ282611		9204336,13
RP11-91K8	175	3q22	F	FHCRC	AC022296	NT_022393	AQ282622;AQ282625		134688316
RP11-91K9	176	3q26.3	F	FHCRC	AC012031		AQ281524;AQ281521		179050368
RP11-91L9	177	3q25	F	FHCRC	AC012027	NT_034560	AQ284142;AQ284143		159747534
RP11-91M15	178	3q11.2	F	FHCRC	AC067740	NT_022401	AQ282283;AQ286875		96518696,9
RP11-91O5	179	3q22	F	FHCRC	AC015539	NT_025663	AQ282505;AQ282507		136615963
RP11-95H16	180	3q21	F	FHCRC	AC069233	NT_005543	AQ315994	AFM119YH2	125629891
RP11-108A8	181	3p14	F	FHCRC	AC072020	NT_005999	AQ318249;AQ318245	AFM308XC9	63325632,8
RP11-114M1	182	3q26.3	F	FHCRC	AC026355	NT_034564	AQ345488;AQ345485	AFMA064YC1	178834287
RP11-114P15	183	3p14	F	FHCRC	AC092039	NT_005999	AQ383673;AQ342459	AFMA163YF1	62556527,6
RP11-118F4	184	3q26.3-3q27	F	FHCRC	AC048335	NT_022396	AQ349290	AFMA162XB9	180992055
RP11-119L2	185	3p21-3p22	F	FHCRC	AC048329	NT_005825	AQ341776;AQ341775	AFMC006YH1	42513834
RP11-120C2	186	3p21.2	F	FHCRC	AC121250	NT_005787	AQ349823;AQ349825	AFMB059WF1	57353586,8
RP11-125E8	187	3q26.3-3q27	F	FHCRC	AC078797	NT_029253		AFMB031ZG5	185542477
RP11-128A5	188	3p25-3p26	F	FHCRC	AC068312	NT_005927		AFMC023XB5	8771667,5
RP11-129B22	189	3p14	F	FHCRC	AC092040;AC021874	NT_005999		AFM198TE7	64098770,5
RP11-129G16	190	3q11.2	F	FHCRC	AC018353	NT_022552		AFM126ZC5	54435133,5
RP11-129K1	191	3p26	F	FHCRC	AC069522	NT_005927		AFMA064WH9;SHGC-56908;D3S3525	2296199,5
RP11-141C22	192	3q26.1	F	FHCRC	AC067720	NT_034563	AQ385885;AQ385887	AFM179XH10	170448755
RP11-145F16	193	3q24	F	FHCRC	AC018545	NT_022555	AQ375287;AQ375289	AFM220YH12	151593651
RP11-146E16	194	3p13-3p14	F	FHCRC	AC092044	NT_022473	AQ372336;AQ372333	AFMA059XB5	67776296,6
RP11-151A21	195	3q26	F	UCSF	AC026312	NT_025667	AQ376269;AQ376270	D3S2433;AFMA082ZD9	173226876
RP11-153K19	196	3q13.2-3q13.3	F	FHCRC	AC021655	NT_005795	AQ386412;AQ386414	AFMB353ZE1	114274137
RP11-154H23	197	3p14.1	F	FHCRC	AC097634	NT_005769	AQ385849	AFMA176YG9	71719733
RP11-160A13	198	3q23-3q24	F	FHCRC	AC020648	NT_005772	AQ374615;AQ374613	AFMA106YF9	144233225
RP11-162J10	199	3q23	F	FHCRC	AC020788	NT_005567	AQ377322;AQ377319	AFM267XD9	138983249
RP11-163H6	200	3q26	F	UCSF	AC020649;AC069259	NT_025667	AQ380830;AQ380833	WI-18033;AFMA082ZD9	173264521
RP11-165B13	201	3q13.3	F	FHCRC	AC020650	NT_005994	AQ415251;AQ382248	AFMB322ZF1	119153369
RP11-165M11	202	3q23-3q24	F	FHCRC	AC016934	NT_019358	AQ414594	AFM240XB2	144879365
RP11-166C10	203	3q13.1-3q13.2	F	FHCRC	AC016932	NT_022434	AQ379897	AFM225XB12	110013972
RP11-166D18	204	3q23-3q24	F	FHCRC	AC016933	NT_005655	AQ380950;AQ419849	AFMA121WD5	141019499
RP11-169N13	205	3q13.3	F	FHCRC	AC069444	NT_033022	AQ416519	AFMB280XB1	121009026
RP11-170K19	206				AC093413;AC023562	NT_005607		RH16423;SHGC-33999;AFMA285ZB5;G59345;G59179 ;D3S3378;G49403;SHGC-77059;D3S3361;D3S3	59788656
RP11-172G5	207	3q26.3	F	FHCRC	AC018456	NT_025667	AQ422718	AFM164TC7	174628684
RP11-178K17	208	3q28	F	FHCRC	AC018898	NT_005962		AFM164ZF8;RH122820;D3S3530	190646636
RP11-189H19	209	3p21.3	F	FHCRC	AC092046;AC069071	NT_034534		AFMB291ZC9	44652681
RP11-189K9	210	3p21.2	F	FHCRC	AC068753	NT_005787	AQ421648	AFM274YD5	54653984,5
RP11-190C21	211				AC072028	NT_034556		RH46321;SHGC-12872;DXS7953;RH120168;RH46027	143899282
RP11-196F13	212	3q26.2c-3q26.31a	F	NCI	AC016938	NT_025667		AFM333YA5	173691933
RP11-205A6	213	3q21	F	FHCRC	AC016924	NT_005588	AQ419300	AFM290ZF1	127456380
RP11-208G16	214	3p24.3	F	FHCRC	AC121253	NT_022517	AQ418708	AFM301ZE9	22443724,5
RP11-209H21	215	3q24-3q25	F	FHCRC	AC012401	NT_034560		AFM200ZE11	160814767
RP11-209O16	216	3p22	F	FHCRC	AC016156	NT_005580	AQ420322	AFM038XC1	38326045
RP11-220G20	217	3q13.3	F	FHCRC	AC021867	NT_022445		AFM057XA5	118704507
RP11-220J13	218	3q22	F	FHCRC	AC016919	NT_025663		AFM347YG1	136882686
RP11-229A12	219	3p21.2	F	FHCRC	AC092418	NT_005787		AFM191VA5	57452683
RP11-233L3	220	3q13.3	F	FHCRC	AC016918	NT_005994;NT_034546		AFM225YD6	120844517
RP11-234D2	221	3q26	F	UCSF	AC026322	NT_034564		AFMA064YC1;D3S1754	173089897
RP11-241P3	222	3p22	F	FHCRC	AC092048	NT_005825	AQ481742;AQ482364	AFMA162WD9	42362195,1
RP11-253H10	223	3q26.31d-3q26.32a	F	NCI	AC026317	NT_034564		D3S1754;D3S3715;D3S3511;STS-T76997	178843705
RP11-255G21	224				AC092920;AC024233	NT_022458		D3S3041;DXS7944	178038291
RP11-255O19	225	3p24.3-3p25	F	FHCRC	AC090945	NT_005927		AFM197XG11	15860501,5
RP11-264D7	226	3q26.1	F	FHCRC	AC069530	NT_022397	AQ487231;AQ487229	AFMB353YA9	168660578
RP11-266J6	227				AC022382;AC011610	NT_005927	AQ479047;AQ479045		9759300,25
RP11-278A4	229	3q26	F	UCSF	AC117445	NT_022458	AQ506157;AQ506156		178156363
RP11-279L16	230	3p24.3-3p25	F	FHCRC	AC069521	NT_005927	AQ508067;AQ508068	AFMB046ZB5	15300255,5
RP11-335L9	233				AC026325	NT_034564		STS-T76997;SHGC-105734;D3S2381;G59491;D3S3096	178676516
RP11-372E1	234	3q23-3q24	F	FHCRC	AC021074	NT_005772		WI-18153;SHGC-36471;RH94214;RH66190;SHGC-52430;RH70276;SGC34062;WI-18261;G59306;D3S35	144108679
RP11-382A21	235				AC018828;AC023236	NT_005927		SHGC-76704;RH92698;SGC34603;WI-14152;GDB:361144;RH71303;GDB:532945;GDB:270439;STS-L15	10155572
RP11-438D8	237				AC010184	NT_005832		WI-17024;RH65567;SHGC-58314;RH48927;Cda01f02;D3S3106;A006U06;STS-D60661	142646058
RP11-477P16	238				AC008120	NT_022458	AQ637431;AQ637427	RH48016;RH104007;DXS7944;RH123586;RH122102;WI-11234;SHGC-105283;SHGC-77556	175428644
RP11-484D18	239				AC023235	NT_022458	AQ631785;AQ631787		178233392
RP11-488M12	240				AC061705	NT_034639		A007B05;G29218;IB397;UTR-04950;SHGC-7485;RH92621;SHGC-77602;STS-R80399;1436;RH77956;RH	185426878
RP11-493K19	241				AC104450;AC022975	NT_006014		D3S3178;WI-18724;STS-AA001762	50079886
RP11-499P15	242				AC078799	NT_022458		SHGC-149434;RH122844;SHGC-77561;SHGC-77560;A009S15;G32804;G63142;RH122782	178360267
RP11-544D12	243				AC068638	NT_022396		RH123299;RH99303	182096853
RP11-569L16	245				AC092929;AC021975	NT_022509		D3S3389;G63125;SHGC-84001;DXS7302;SHGC-77567	179796798
RP11-571C23	246				AC092970;AC023144	NT_022458			177507950
RP11-590F24	247				AC079847		AQ826433	RH123959;RH16209;SHGC-77655;SHGC-77661	193627990
RP11-644C3	249				AC024473;AC126469	NT_022458		D3S3041;DXS7944;G16623;G60455;D3S3676;SHGC-14925;SHGC-144758	177913464
RP11-789N15	250				AC080006	NT_005883		SHGC-82139;D3S1314;WI-19957;A001W08;G30141;SHGC-153733	191442263
RP11-795G18	251				AC103549	NT_022509	AQ522167;AQ467501		179211751
RP11-816J6	252				AC078802	NT_034563	AQ556554;AQ525901	DXS7944;RH120520;D3S2982	170929793
RP11-875H7	253				AC121764;AC121248	NT_005787	AQ819924	RH80790	55437134
RP11-271E2	255	3p24.3-3p25	F	FHCRC	AC024151;AC069462	NT_005927;NT_034608		AFM283VB5	13971844,2
RP11-545F2	256				AC026311	NT_034564		D3S1754;D3S3715;STS-T76997	178921425
RP11-45I24	257	3q26	F	UCSF	AC015620	NT_022396	AQ201243;AQ201245	WI-16592;AFMA162XB9	180947956
RP11-79M2	258	3q13.3-3q21	F	FHCRC	AC068754	NT_005543	AQ283212;AQ283219;AQ581485;AZ519189		123445424

TABLE 3

Gene Name or Genbank accession number	Forward primer sequence (5' -> 3')	Reverse primer sequence (5' -> 3')	Final concentrations used (nM) for PCR on genomic DNA
BC032141	AAGTTTCCCAGACCCTCGTG	TGGTGGCGGTTATAGTGCTG	250
AA255943	CCCTCTGCTCGCACGTTT	TGGACTAAATCCCCGGTCTC	375
AA969336	TCAGGACGGGTCAGAAGAAAA	CTTGGGGTTCCAGAGGACAG	1000
BX096801	GGGAGGTCTCAGCCAAGCTAT	ACAGTGGGGGCCAGTCTCT	225
LPP (5')	TATTCAGGGCTCGCTGTCAA	GTCAGAGCTGGCAAAGACCA	500
AL080232	TGGGCAGGGATGGTATTTTC	AAAAGCAGGGCACACACAAA	500
BF759914	GCAAAGCCCCTCCCTTACA	TGGGGCAAGGCTCTTTAGAA	375
LPP (middle)	CCCCCTCTAACAGCAACCAA	CGTGTTGAAGGAGGTGGACA	375
BQ425887	TGACTTGGAGCTGTGGCTGT	TTAGGGTGGGCAGAGGAATG	250
AK023352	CCTCTTTGGGAGGCATTACG	AGCTTCCTTGGGAATCAGCA	750
BF336179	TGGAACGAGGCTATGGGAAC	TTGGCTGCTGAATGAGAAACA	625
LPP (3')	TCTGCCAACTCACAGGTGCT	TGGGCCGATAACTCTTTTCC	250
BG778254	TGTCACGCACGCACACTTAC	GAGGCTGTCAGGAGACCAGAA	350
AW277126	TTTGTCAGCCTCACCTACTGGA	ATTCCAGTTCCCCCTTTAGCA	875
BQ432978	GCACTGTTTACAAACCACTGGAA	TTATTTGCTCTGACTGGGAAACC	1000
BG188653	GAACTGTCAGGCGGGAAAAC	GTACTGGCCCCGCTTCCT	375

TABLE 4

Gene Name or Genbank accession number	Forward primer sequence (5' -> 3')	Reverse primer sequence (5' -> 3')
LPP (first 5' exon)	TATTCAGGGCTCGCTGTCAA	GTCAGAGCTGGCAAAGACCA
LPP (middle CDS)	CCCCCTCTAACAGCAACCAA	CGTGTTGAAGGAGGTGGACA
LPP (last 3' exon)	TCTGCCAACTCACAGGTGCT	TGGGCCGATAACTCTTTTCC
FLJ42393	TTGAATTTGCCACCGTGAAG	CAAGTGTGTCGCAGATGTGG
AI338598	TCTTTTTCAAGGTCCTGCTCAA	AACTCCAAACCCCTCCTCCT
BX118304	CTCCTGAGCCCCAGGTTACA	GTTAGAGTAGGTCCCTTCACCTGATT

FIGURE LEGENDS

Figure 1. Mapping of 3q and 8q high level amplifications of COMA cell line by chromosome-CGH, array-CGH and FISH

DNA amplifications on chromosome 3 (A) and on chromosome 8 (D) detected by chromosome-CGH [12] are confirmed and more precisely delineated by array-CGH (B and E respectively). The interval amplified on 3q encompasses three adjacent clones (103, 104 and 105) using our high-density CGH-array dedicated to chromosome 3 (B) and its extent is about 1.5 Mb (see enlargement in C; central positions of the first excluded clones =188.9-190.4 Mb). Analysis with a pan-genome CGH microarray reveals that the 8q amplicon covers two adjacent clones (E). Interestingly, *c-myc* mapping position is situated 60 kb after the first clone.

FISH experiments on COMA nucleus (F) and metaphase spread (G) with one clone reporting the amplification on chromosome 3 by array-CGH (clone 105, red signals) and one containing the c-myc gene (RP1-80K22, green signals). Multiple sites of hybridization are visible in the nucleus and highlight numerous DM chromosomes on metaphase spreads. DNA segments from chromosome 3 and chromosome 8 are carried by separate DM chromosomes. Similar results were obtained when using clones 103 and 104 in red. (H) Hybridization of clone 104 (red signals) and chromosome 12 clone (containing *HMGIC* gene) on COMA metaphase spreads shows no co-localization of the signals, indicating the absence of translocation

Figure 2. Mapping of 3q gains and amplifications in selected MFH samples.

A. Chromosome 3 profiles of 8 selected tumors (chromosome-CGH experiments). All tumors present 3q aberrations, either 3qter low-level gains or high-level amplifications.

B. Frequency of gains for the 8 tumors along chromosome 3 (divided by classes of 2 Mb) found by array-CGH. Gains are defined as ratio higher than 1.25. Gains in more than 40% of the tumors are found for a class at 152 Mb and several classes between 180 Mb-telomere. The class with the higher percentage is 188-190 Mb intervals.

C. Individual chromosome 3 array-CGH profiles for the three tumors presenting 3q high-level amplification (ratio > 2). All high level amplifications are located at 3q27.3-q28. Arrows denote the position of the amplified region in the COMA cell line (see text for details).

D. Consensus region of high-level amplifications between the COMA cell line and BER18 and MFH20 tumors.

Figure 3. Fine mapping of the COMA 3q28 amplification by genomic semi-quantitative PCR (A) and expression level of transcripts from the 3q27.3-q28 region (B and C).

A. 16 loci located in the interval defined by the array-CGH experiment (Fig. 1C) were amplified by PCR using genomic DNAs from COMA and normal cells. COMA vs. normal ratio values calculated after band quantification in densitometry are represented in Figure 3A as a function of primer pair positions in the 3q27.3-q28 region.

This analysis reveals that the amplicon size is less than 1.1 Mb and that breakpoint regions are included in 150 and 62 kb intervals, respectively. In addition, a small part of the amplified segment (from 190.1 to 190.25 Mb) presents higher DNA copy ratio (p values [using unpaired t-test with other markers] <0.01 and <0.02, for markers BG778254-1 and AW277126-1 respectively).

Each value is the average of 3 independent experiments. Error bars represent the SD.

Triangle link by bold line represents the position of the BAC clones used in array-CGH experiments. Disks refers to positions of the ESTs used in expression analysis in Figure 3B (1: FLJ42393, 2:LPP 5', 3:LPP middle, 4:LPP 3', 5:AI338598, 6:BX118304).

B. The expression levels of 2 Refseq genes (3 different primer pairs for LPP, one for FLJ42393) and 2 ESTs (AI338598 and BX118304) were compared between COMA and MFH19 cell lines (the latter harboring normal 3q copy number). For each cell line, band intensities were quantified and normalized to GAPDH. Ratios COMA/MFH19 were calculated and averaged after 3 independent RT-PCR experiments. Error bars represent the SD. Two ESTs (AI338598 and BX118304) are likely to be strongly over-expressed (more than 20 fold) in consequence of the 3q amplification present in the COMA cell line (no difference was found for LPP).

C. Expression analysis by Northern blot. The expression level of the microRNA encoding gene *hsa-miR-28* was measured in the two MFH cell lines. The microRNA level is not increased in COMA when compared to MFH19 and seem expressed at similar levels from a repeated experiment.

Supplementary information

PCR conditions for BAC DNA amplification for CGH-array production

From: Overlapping amplified 3q28 region in COMA cell line and MFH primary tumors
Thomas Hussenet, Nedjoua Mallem, Richard Redon, Bernard Jost, Alain Aurias and Stanislas du Manoir

DOP-PCR amplification (1st step)

DOP-PCR composition

20 µl 5X Mix
(composition: 50% 10X PCR Buffer, Sigma; 5 mM for each dNTP; 25mM $MgCl_2$, Sigma)

+ 2 µl primer ([stock] = 100 µM)
NB: either 6MW primer (CCGACTCGAGNNNNNNATGTGG; Telenius *et al.*,1992)
OR* DOP2 primer (CCGACTCGAGNNNNNNTAGGAG; Fiegler *et al.*, 2003)

+ 5U GoTaq® DNA polymerase (Promega)

+ 75 ng BAC DNA

+ H_2O qsp 100 µl

*** each BAC clone DNA was independently amplified using 6MW AND DOP2 primer.**

DOP-PCR cycling conditions:

1) 5min - 95°C
2) 1min - 94°C
3) 1min 30" - 30°C
4) Temperature Gradient de jusqu'à 72°C (slope = 0,2 °C / sec)
5) 3min - 72°C
6) Goto 2) 4 times
7) 1min - 94°C
8) 1min - 62 °C
9) 3min - 72 °C (+1" every cycle)
10) Goto 7) 34 fois
11) 10min - 72°C

PCR amplification (2nd step)

PCR composition

12 µl Mix 5X
(composition : 50% 10X PCR Buffer, Sigma; 5 mM for each dNTP; 25mM $MgCl_2$, Sigma)

+ 1 µl primer (GGAAACAGCCCGACTCGAG, Fiegler *et al.*, 2003 ; [stock] = 100 µM)

+ 3 U GoTaq® DNA polymerase (Promega)

+ 3 µl DNA (DOP-PCR product)

+ 44.4 µl H2O

during the PCR amplification (2nd step), each DOP-PCR product was independently amplified.

PCR cycling conditions:

1) 10min - 95°C
2) 1min - 94°C
3) 1min 30" - 60°C
4) 7min - 72°C
5) Goto 2) 34 times
6) 10min - 72°C

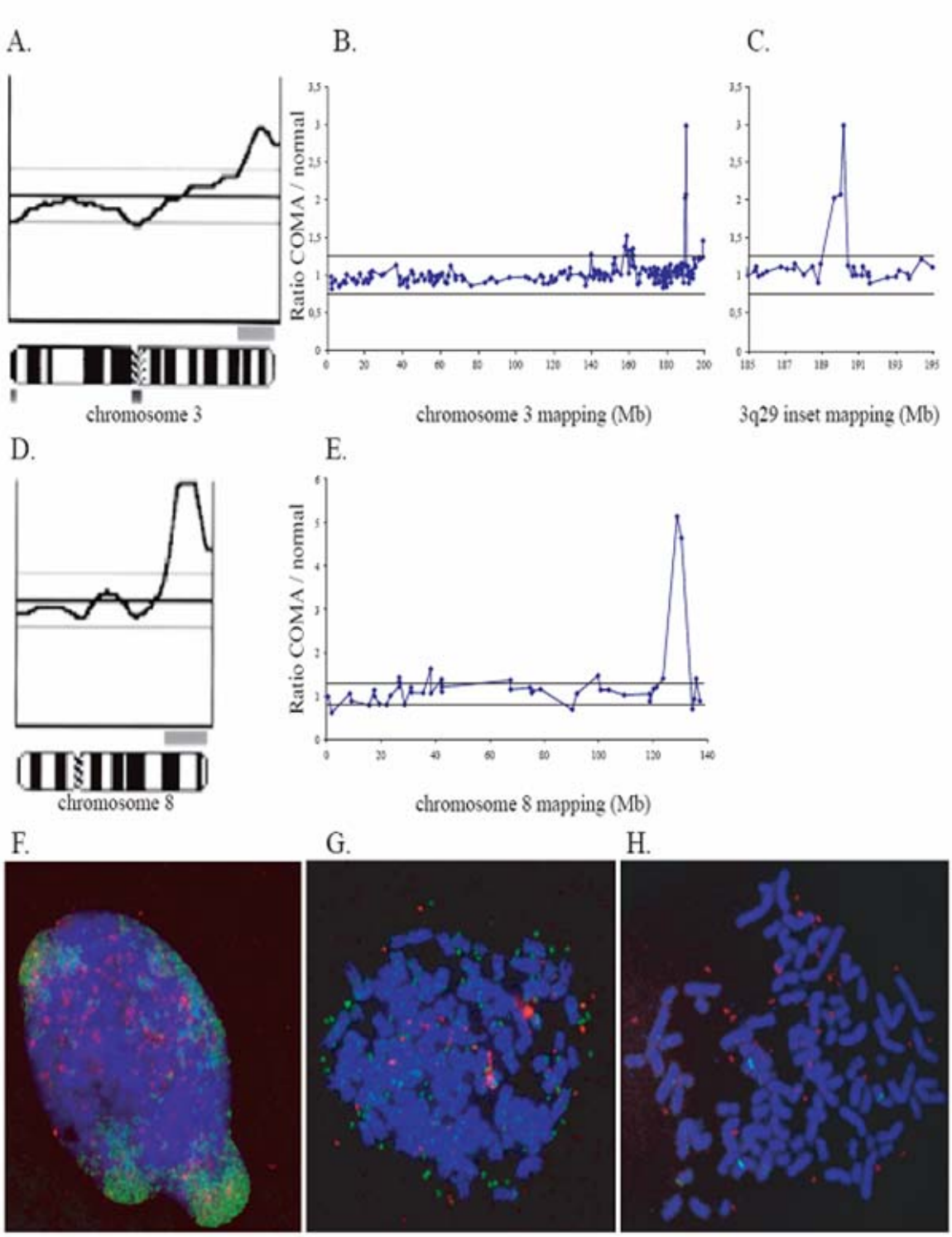


Figure 1 from Hussenet *et al.*
(manuscript submitted)

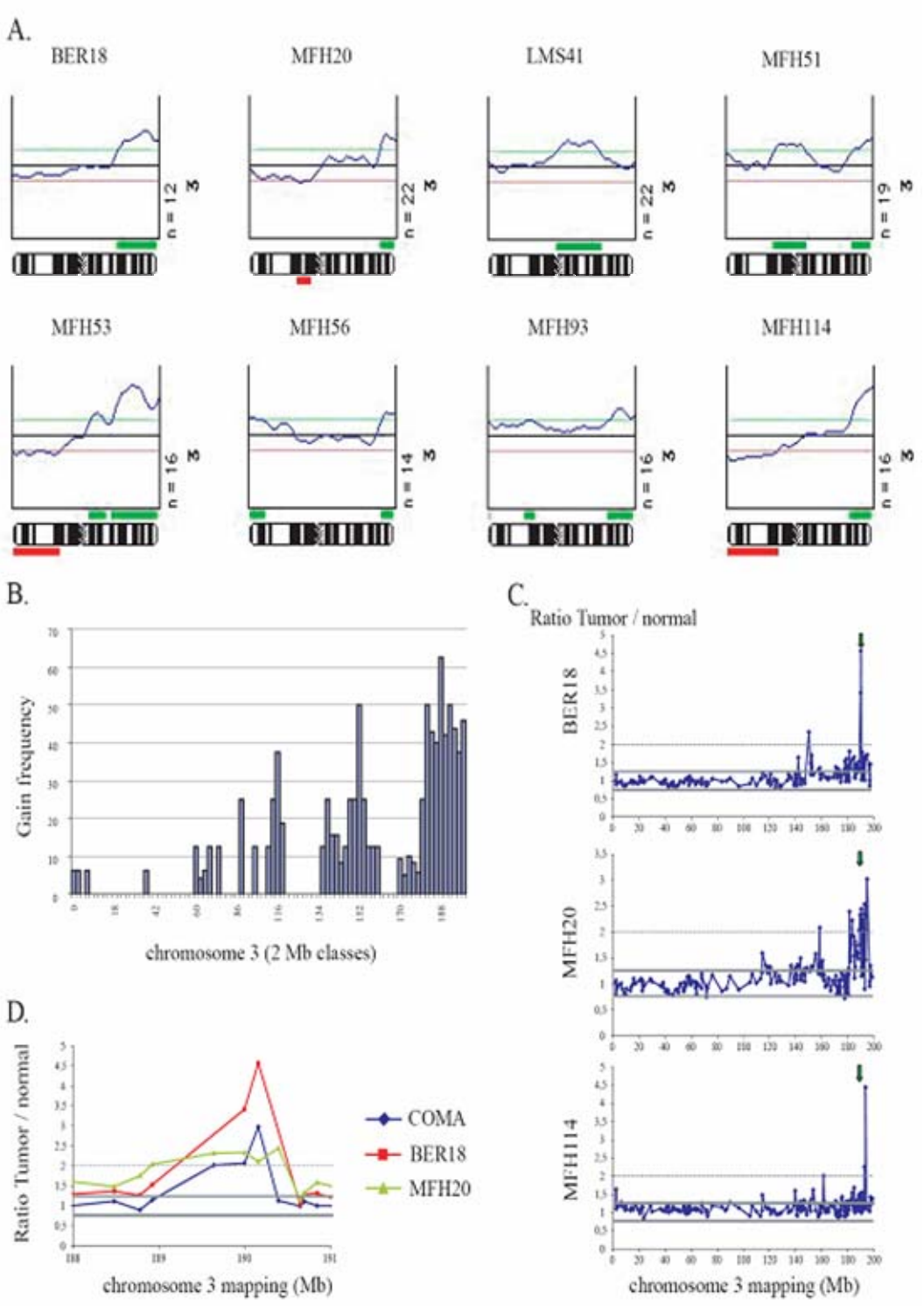


Figure 2 from Hussenet *et al.*
(manuscript submitted)

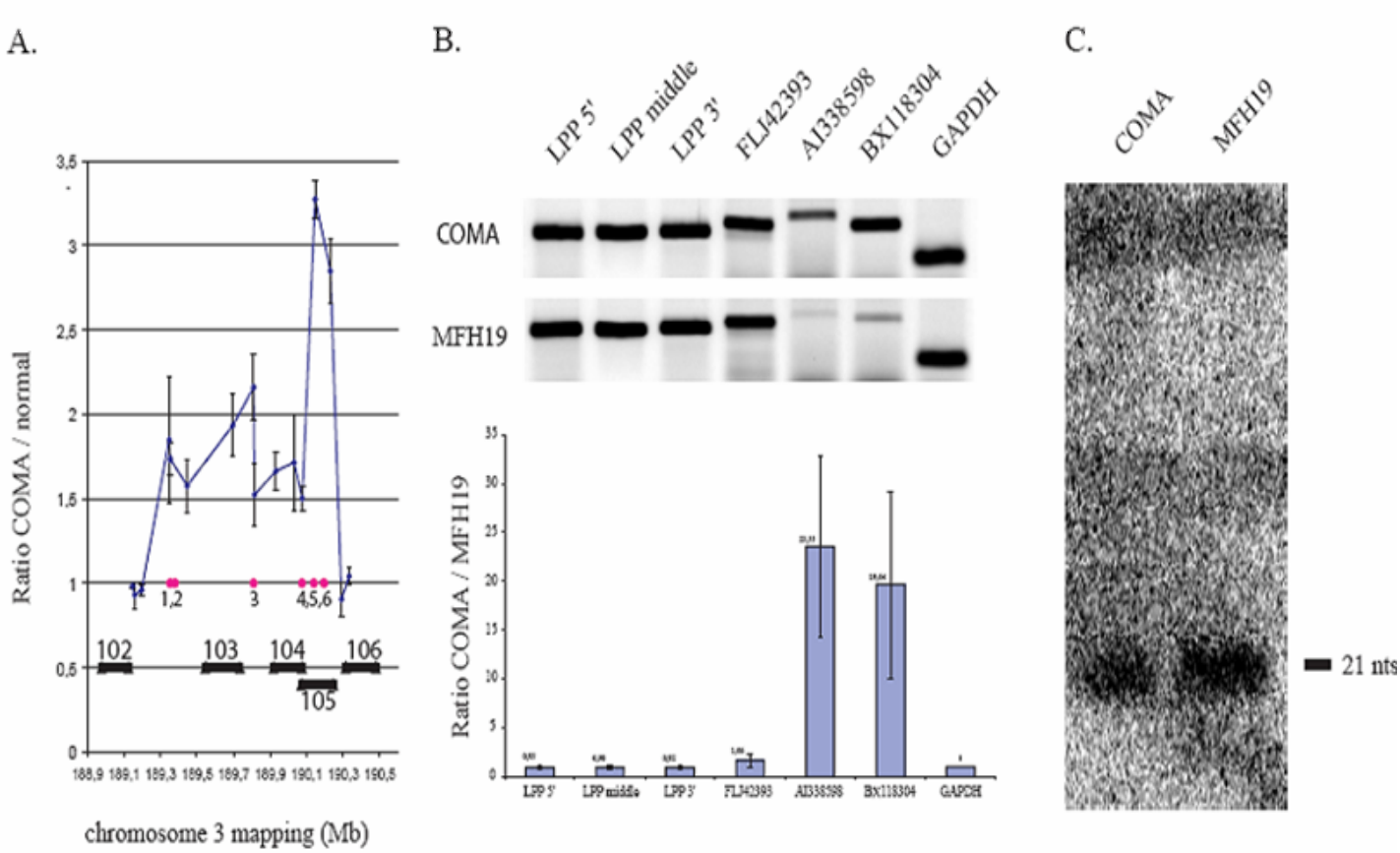


Figure 3 from Hussenet *et al.*
(manuscript in preparation)

iii- CONCLUSIONS / PERSPECTIVES

Even if MFH has been a common diagnosis for sarcoma, the existence of MFH as an entity *per se* has been recently challenged. Their histological origin is indeed unknown and these tumour cells are highly pleiomorphic.

As no normal tissue corresponding to MFH tumours is available, we have thus been limited for the gene expression analyses to the use of another MFH tumour cell line (MFH19) with no aberration of the 3q28 locus (amplified in the COMA cell line). We found the two known genes of this region expressed at similar levels in both cell lines, as well as two ESTs strongly over-expressed in COMA cell line. Unfortunately, since they are very poorly characterized, we were unable to predict any CDS nor functional domain from these EST sequences, nor to include them in an EST contig by bioinformatic exploration of human transcribed sequences.

Finally we analyzed the expression level of *hsa-miR-28*, a microRNA encoding gene, and found it to be similarly expressed in the two MFH cell lines Following our strategy of oncogene identification, correlating gene copy number increase and subsequent over-expression, we were unable to propose any protein-coding gene as a likely driver of the 3q28 amplicon in the COMA cell line.

Nevertheless, we cannot exclude that among the three known genes of the 3q28 region we have found to be expressed at similar levels in both MFH cell lines, one or more could be important for the malignant phenotype of these two MFH cell lines. In this case, we would assume that expression of the driver gene of the 3q28 amplicon in the COMA cell line is activated by an alternative mean in the MFH19 cell line, which has no copy number increase of this locus.

Finally, further studies would be necessary to assess whether amplifications at 3q28 could define a sarcoma/MFH subgroup since this cannot be concluded directly from our study.

a-2 miR-28 microRNA as an "onco-miR" candidate in MFH

(manuscript in preparation)

i- INTRODUCTION

MicroRNAs encoding genes represent a recently identified and new category of non-protein coding genes. They encode tiny RNAs of 19-22 nucleotides in length, or microRNAs, that are involved in the post-transcriptional regulation of the expression of numerous other genes. This large scale regulation of genes introduces a new level of complexity that coordinates expression in complex processes such as development (see the general introduction part of this thesis). Current estimate of their number in Human is about three hundreds. Littlle is known about their roles and recent works highlighted putative implication for some of them in carcinogenesis. For example, profiling of microRNAs expression enables to classify human tumours of various localizations (Lu *et al.*, 2005). Furthermore, a specific microRNA signature associated with prognosis and disease progression was recently depicted in B-CLL samples (Calin *et al.*, 2005).

Since in a previous study, we found a microRNA encoding gene, *hsa-miR-28*, located at 3q28 in a consensus region of amplifications between two MFH tumours and a MFH cell line, it was appealing to further characterize the function of this gene. Interestingly, the expression level of *hsa-miR-28* was notably higher in both MFH (soft-tissue) tumour cell lines when compared to its relative expression in HNSCC (epithelial) tumour cell lines (see figure1 hereafter).

Taken together, all these arguments prompted us to speculate that *hsa-miR-28* could represent a gene contributing to the tumour phenotype in these MFH tumour cell lines. Accordingly, we decided to evaluate its involvement in cellular proliferation of the two MFH cell lines and attempt to identify some of its targets to begin to elucidate its basic function.

ii- MATERIAL and METHODS

miR-28 and pre-miR-28 expression measurements

Northern Blot was carried out to determine miR-28 expression level in MFH and HNSCC cell lines (see the previous submitted manuscript for Northern Blot protocol details).

To overcome the need of high amounts of total RNA to assess microRNAs encoding genes expression, we set up the new method allowing pre-miR expression measurement using PCR (Schmittgen *et al.*, 2004). Recently, primer pairs suitable for real time qPCR (with amplification efficiency values around 2) became available for most of known microRNAs genes including miR-28 (Jiang *et al.*, 2005). Reverse transcriptions were performed using these primers and associated protocol. PCR conditions were identical to published conditions except that the final concentration of MgCl2 was 1 mM (Redon *et al.*, 2002). Annealing temperatures were adapted for each primer pair according to their melting characcateristics (available with the primer sequences in the supplemental data from Jiang *et al.,* 2005).

Anti-miRs transfections

Transfections of "antimiRs" were done with SiPortNeoFx (Ambion), following manufacturer's specifications, both for cell proliferation assay and for proteins extractions.

Cell proliferation assay

For each transfection, 5.10^4 trypsinized cells in suspension were mixed either with "anti-miR-28" (Ambion) or "anti-miR negative control #1" (random sequence, Ambion) at a 10 nM final concentration of anti-miRs. Cells were then allowed to attach in 6-well plates for 18h and rinsed 3 times with PBS. For each cell line, cells from one well were subsequently trypsinized 24h after and counted (J1). Cells in one other well were allowed to grow for another 48h by adding 4 ml of fresh medium per well, and medium was replaced each day, up to 72h after the end of transfections (J3).

Preparation of protein extracts

Transfections were made as described above, except that 5.10^5 trypsinized cells were transfected in 6 cm round plates, and then allowed to recover for 24h in fresh medium before preparation of protein extracts according to standard procedures.

mir-28 targets selection

Candidate proteins were selected from the analysis of miR-28 targets prediction list available online (http://www.microrna.org/; John *et al.*, 2004). E2F6 and DP2 were first selected for their role in the cell cycle, secondly their known mutual interaction and thirdly the putative antagonizing role of E2F6 with respect to c-myc.

Antibodies and western blotting

Western Blots were carried out on these proteins extracts using standard protocols. Protein extracts were quantified using Bradford method and same amounts of total proteins were used for electrophoresis. Proteins were transferred to nitrocellulose membranes which were blocked for 1h in 1X PBS / 10% non fat dry milk prior to incubation with primary antibodies.

The table 1 recapitulates the primary antibodies sources and dilutions used. All primary antibodies were incubated in 1X PBS / 10% dry milk for 1 to 3h at room temperature and then overnight at 4°C under shaking conditions. Membranes were washed 3 times for 5 min in PBS 1X / 0,1% Tween20.

Secondary antibodies coupled to HorseRadish Peroxidase (HRP), Goat anti Mouse-HRP and Goat anti Rabbit-HRP (Jackson Immunoresearch Laboratories, Inc.), were then applied at 1:20000 dilutions in 1X PBS / 0,05%Tween20 for 1h at room temperature. After 3 washes in 1X PBS / 0,05% Tween20, detection was done using chemiluminescence (ECL, Pierce).

Several independent experiments, from transfections of anti-miRs to western blotting, were performed and representative images of results are presented.

Primary Antibody @	Organism	Manufacturer	Reference	Dilution
E2F1	Rabbit	SantaCruzBiotech	SC-193	1/500
E2F4	Rabbit	SantaCruzBiotech	SC-866	1/500
E2F6	Mouse	ActiveMotif	39509	1/250
DP1	Rabbit	SantaCruzBiotech	SC-610	1/250
DP2	Rabbit	SantaCruzBiotech	SC-830	1/500
CDC14A	Rabbit	Zymed Laboratories	34-8100	3/500
CDC14A	Mouse	Abcam	AB10536	1/2000
Ki-67	Rabbit	Novocastra Laboratories	NCL-Ki67p	1/2000
gammaTubulin	Rabbit	Sigma	T3320	1/1000

Table 1. Primary antibodies sources and dilutions used for western blotting.

HeLa cell synchronization

HeLa cells were synchronized at G2/M by nocodazole treatment during 20h, then stimulated to re-enter cell cycle by nocodazole removal and finally harvested every 2 hours during 18h for RNA extraction using Trizol.

iii- RESULTS

hsa-miR-28 expression in MFH and HNSCC cell lines

Northern Blot analysis shows *hsa-miR-28* expression at somewhat similar levels in 2 MFH cancer cell lines, (figure 1). In comparison, *hsa-miR-28* is expressed at very low levels (FaDu cell line) or undetectable (Cal-27 and SCC116) in the three HNSCC cell lines tested.

miR-28 inhibition slows down MFH cancer cell lines proliferation

To investigate whether *hsa-miR-28* expression could be important for the tumoural phenotype of MFH cancer cell lines, we then decided to analyze their proliferation rate after miR-28 inhibition by antisense RNA oligonucleotide. For both MFH cell lines, proliferation was actually slowed down in anti-miR-28 transfected cells in comparison to control cells, which were transfected with an "anti-miR negative control" composed of a random sequence (figure 2A). In sharp contrast, proliferation of HNSCC cell lines was not diminished after miR-28 inhibition (figure 2B). This result is concordant with the lower levels of *hsa-miR-28* expression in those cell lines shown by Northern blot previously. Moreover, in anti-miR-28 transfected MFH cancer cells, doublets of dying cells were very often observed for COMA and MFH19 cell lines (data not shown). These intriguing pictures were evocative of a putative incomplete cytokinesis leading to cell death.

DP2, E2F6 and CDC14 protein levels are regulated by mir-28 in MFH cancer cell lines

In an effort to decipher which endogenous mRNAs could be regulated by *hsa-miR-28*, we extensively analyzed the recent literature concerning microRNA targets prediction and reviewed different web resources (see Material and Methods and introduction part of this thesis). From these and our data onn proliferation, we first selected E2F6 and DP2 as interesting potential regulated mRNAs for their function in the cell cycle.

Mammalian microRNAs regulate protein expression levels either by inhibiting translation of their target mRNAs or by targeting them for degradation. Accordingly, we evaluated, in cell lines, the effect of miR-28 inhibition on E2F6 and DP2 at the protein level (figure3A).

Western blots analyses on MFH (COMA, MFH19) clearly demonstrate that E2F6 and DP2 protein levels are upregulated after miR-28 knock-down in MFH (figure 3B). On the contrary miR-28 knockdown has no effect on the levels of these proteins in HNSCC cell lines, consistent with the low level of expression of miR-28 in these cell lines. Thus, E2F6 and DP2 mRNAs are two targets regulated by hsa-miR-28 in human living cells. We found CDC14A (also among predicted targets) protein level to be regulated by miR-28, validating CDC14A as an additional target (figure 3B). The mRNA encoding the proliferation marker Ki-67 is also among miR-28 predicted targets, but, contrary to E2F-6, DP-2 and CDC14A, its protein level doesn't seem to be regulated by miR-28 (data not shown).

To assess whether miR-28 could play a more general role in the regulation of the E2Fs transcription factors family, we then extended this analysis to other E2Fs family members, E2F1, E2F4 and DP1. Their protein levels are not regulated by miR-28 in MFH nor HNSCC cell lines (figure 3C and 4, respectively).

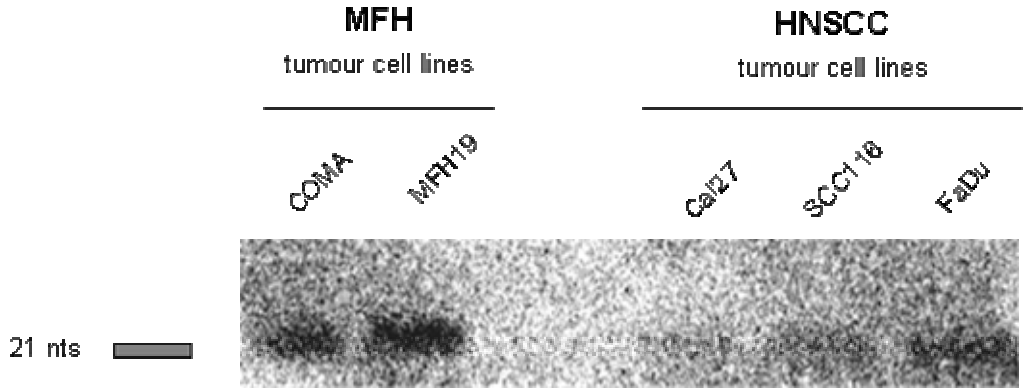


Figure 1. Expression of miR-28 in various cancer cell lines.

Northern Blot analysis of miR-28 expression level in 2 MFH and 3 HNSCC cell lines. MiR-28 is expressed at nearby similar levels in both MFH cell lines but higher than in HNSCC cell lines. Prior to transfer to nitrocellulose membrane, equal amounts of total RNA loading per lane were verified by Ethidum Bromide staining of the gel.

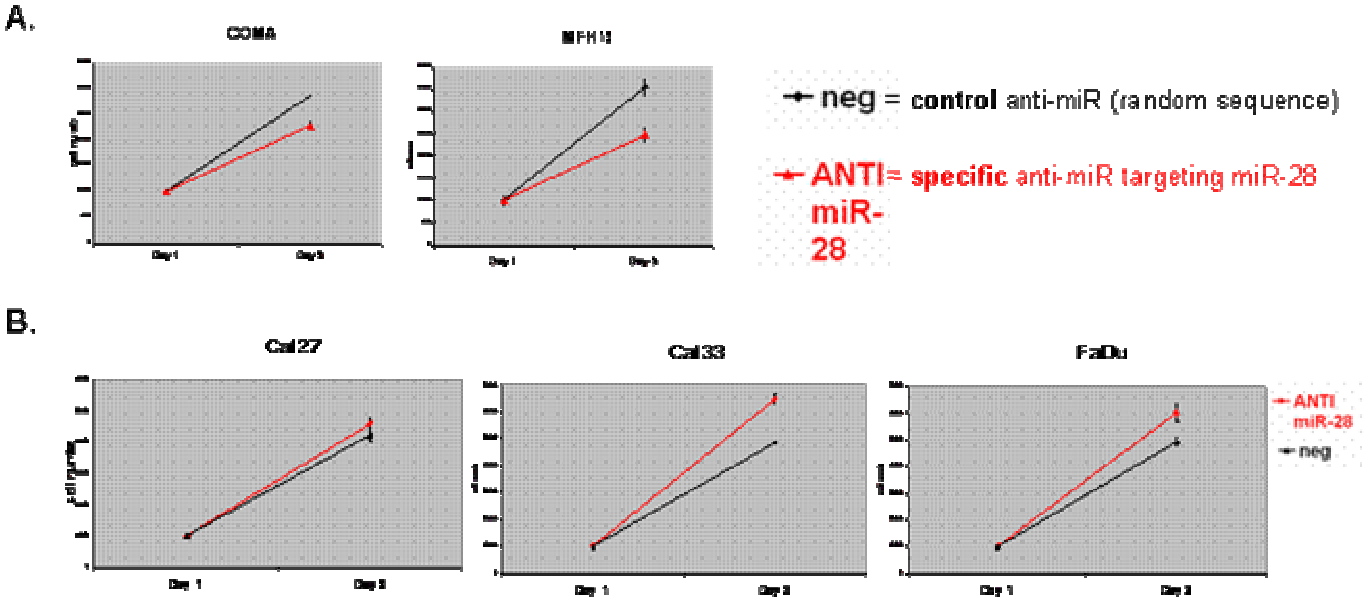


Figure 2. Effect of the inhibition of miR-28 on the proliferation of MFH and HNSCC cancer cell lines.

A. Two MFH cancer cell lines, COMA and MFH19 were transfected by a chemically modified RNA oligonucleotide either of specific sequence targeting miR-28 (« Anti-miR-28 ») or of random sequence as a negative control («neg »). Proliferation of the two MFH cell lines is slowed down when inhibiting miR-28

B. On the contrary, the proliferation of three HNSCC cell lines is not diminished when inhibiting miR-28

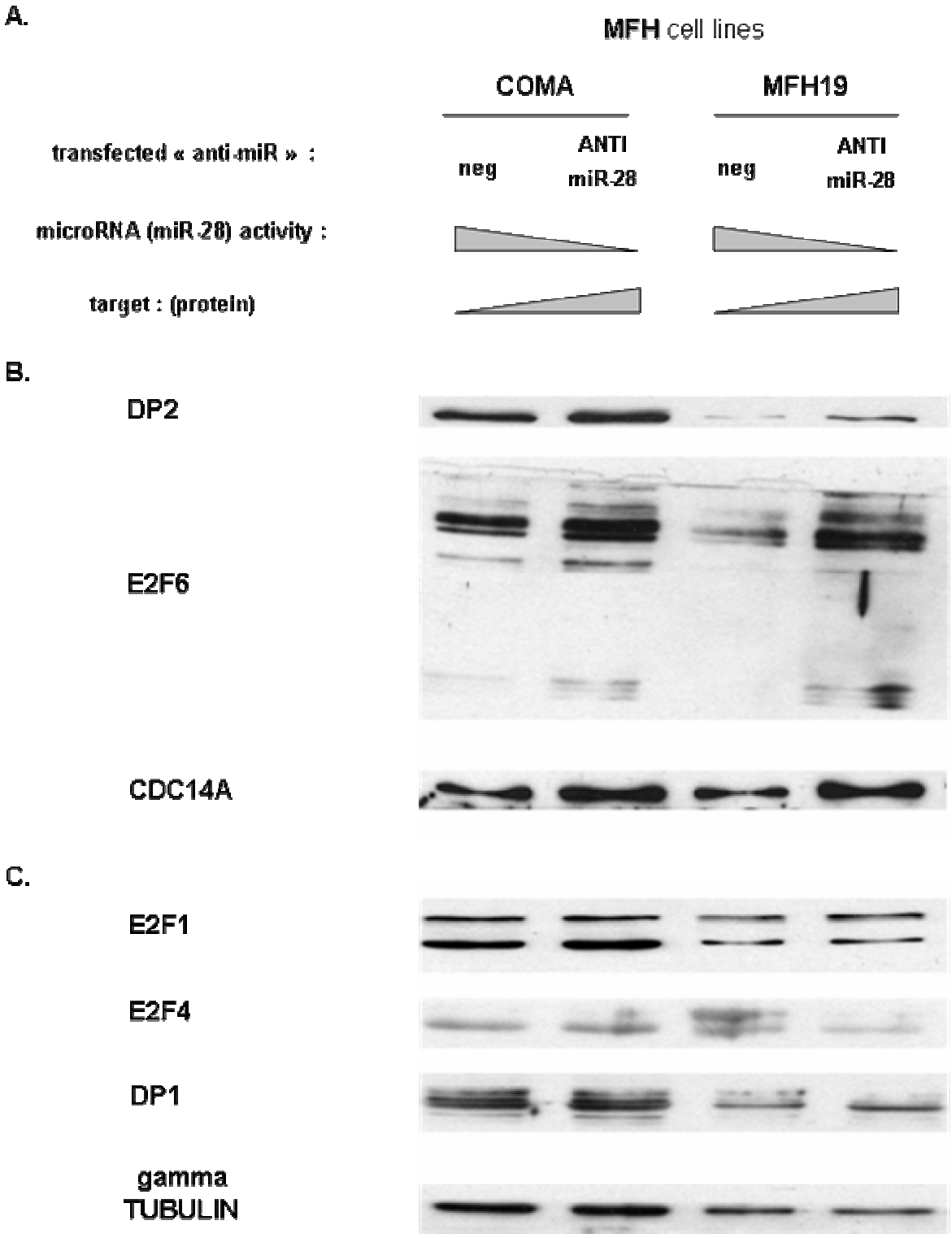


Figure 3. MiR-28 inhibition to identify its targets in MFH cancer cell lines.

A. To identify targets of a given microRNA, the simplest way is to downregulate its expression which leads to up-regulation of the protein level encoded by the targeted mRNA. This was achieved through transfection of a specific « anti-miR-28 » oligonucleotide in comparison with a negative control (« neg »).

B. Among predicted targets of miR-28, DP2, E2F6 and CDC14A protein levels are upregulatd as a consequence of miR-28 inhibition. Representative western blot scanned images of several independent experiments (from transfection of anti-miRs to western blotting) are presented.

C. On the contrary, E2F1, E2F4, DP1 levels are not regulated by miR-28.

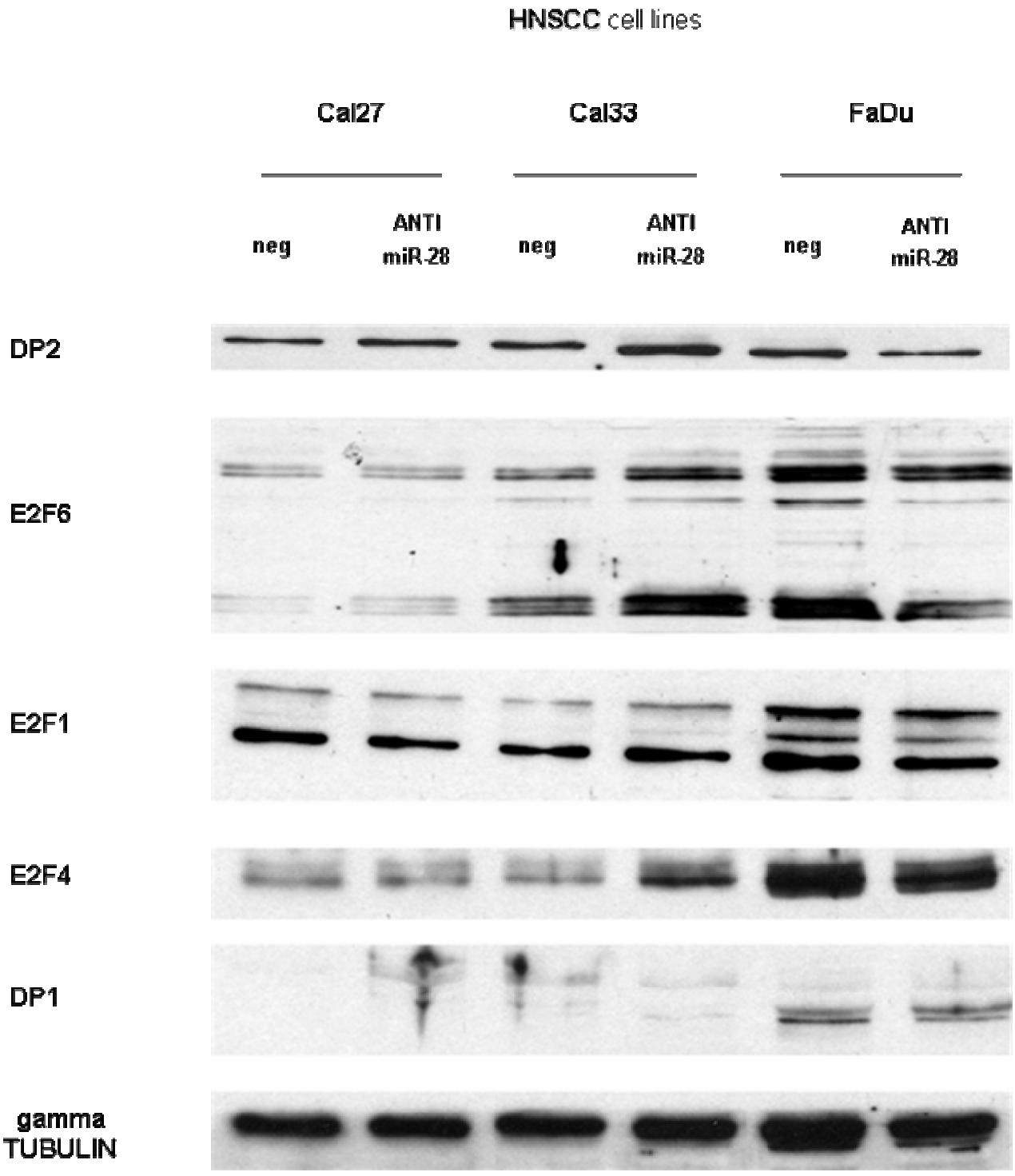


Figure 4. miR-28 inhibition in HNSCC cell lines

No consistent protein level upregulation is consistently observed for E2F1, E2F4, E2F6, DP1 nor DP2 when inhibiting miR-28 in the 3 HNSCC cell lines tested.

hsa-mir-28 expression along HeLa cell cycle

Seeing that miR-28 indeed regulates the protein level of several proteins involved in the cell cycle, we addressed the question of *hsa-miR-28* gene expression itself along the cell cycle using semi-quantitative RT-PCR in a time course experiment. Our preliminary result indicates clearly that pre-miR-28 level vary along HeLa cell cycle, peaking at 8h and 18h after mitotic release. These times may actually correspond respectively to around G1/S and G2/M transitions. These results will be confirmed by an independent experiment and by using several phase markers to assess the previous assumptions.

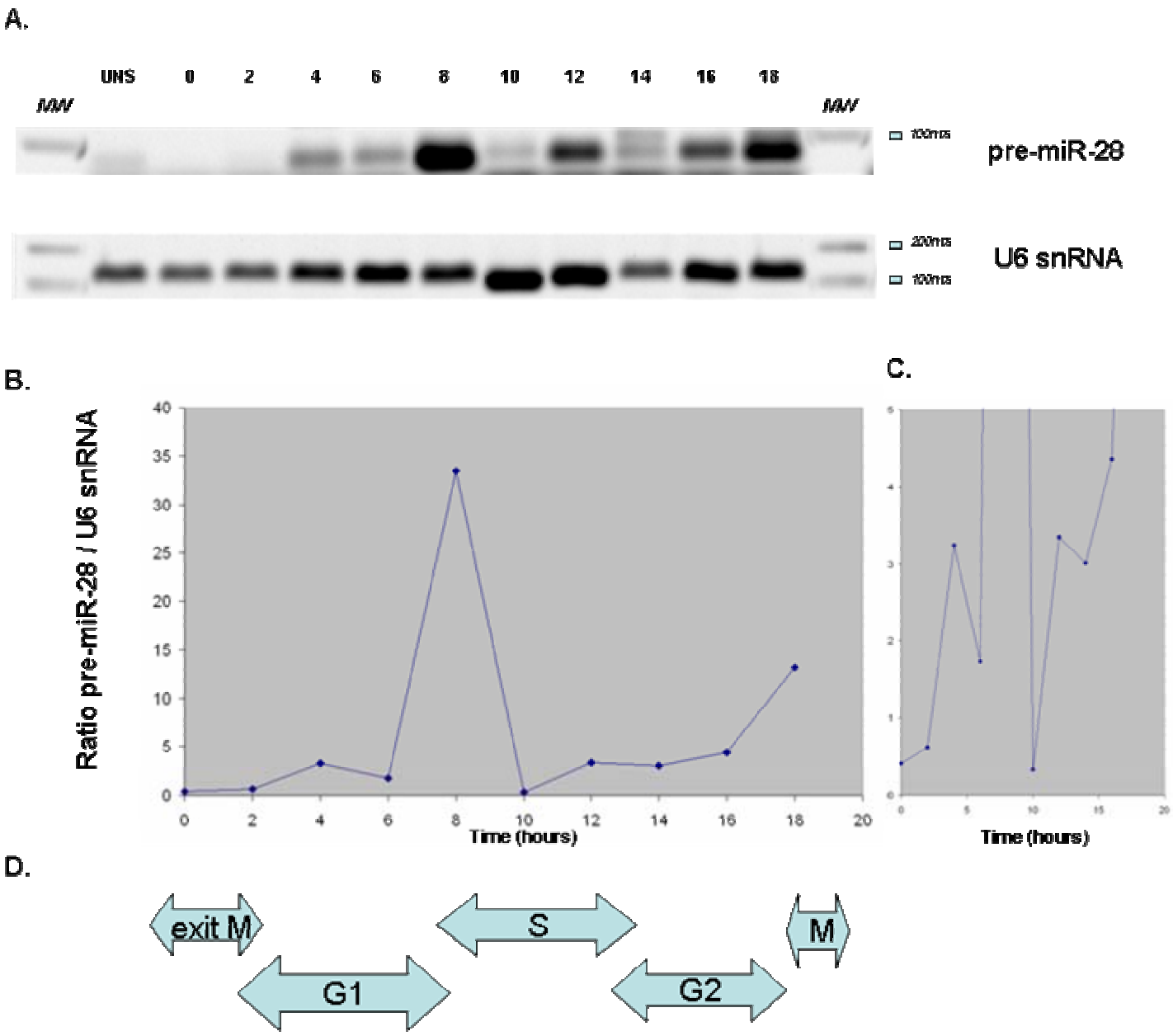


Figure 5. miR-28 expression along cell cycle of HeLa cells.

A. RT-PCR products obtained for pre-miR-28 and U6 snRNA (internal control) along Hela cell cycle. Hours after nocodazole removal are indicated from 0 to 18 above the corresponding lanes. The lane noted UNS corresponds to unsychronized (*i.e.* not nocodazole treated) HeLa cells that were used to determine pre-miR-28 basal expression level. Molecular Weights marker lanes are indicated by *MW*, and corresponding sizes are indicated on the rigtht of the figure.

B. Semi-quantitative measurement of pre-miR 28 level (relative to U6 snRNA used as internal control)

Levels of pre-miR-28 vary along cell cycle from very low levels at 0, 2 and 10 h to a very high level at 8h. It seems that the level of expression of the *hsa-miR-28* gene is cell cycle regulated.

C. Inset of the graph presented in B to a lower y-axis scale (pre-mir-28/U6snRNA ratios from 0 to 5). It clearly appears that pre-miR-28 level is minimum 10h after nocodazole removal.

D. Inferred corresponding cell cycle phases.

Comparison with published cell cycle phases durations of HeLa cells synchronised by double thymidine-nocodazole block (Whitfield *et al* , 2002) suggests that the pre-mir-28 expression maxima 8 hours after nocodazole removal may corresponsd to G1/S transition. The pre-miR-28 expression minima observed 10h after nocodazole removal may correspond to the middle of S phase.

iv- CONCLUSIONS

Recently, first demonstrations of microRNAs with oncogenic functions have been uncovered. Together with the ability of classifying different tumour types on the basis of the expression of several microRNAs suggest that this new class of genes may participate to the tumour phenotype. Accordingly, we decided to assess the effect of miR-28 inhibition by a specific antisense RNA on two MFH (COMA and MFH19) and three HNSCC cell lines. The miR-28 microRNA is highly expressed in the MFH cell lines compared to three HNSCC cell lines used as controls in which its expression is barely detectable as shown by Northern Blot.

A decrease of cellular proliferation was observed for MFH cell lines but not for the others which is concordant with their respective levels of *hsa-miR-28* expression and suggests that we are revealing a process mediated by miR-28. Thus, *hsa-miR-28* seems to promote cell proliferation in MFH cell lines. In addition, we have observed a peculiar phenotype of numerous doublets of dead cells in the miR-28 inhibited MFH cell lines compared to cells transfected with a scramble sequence anti-miR negative control (data not shown; two independent observers). This phenotype suggests that some of the transfected cells could have undergo aberrant mitoses that they were not able to fully complete and triggering subsequent death for the two daughter cells which remained attached after cytokinesis failure. This phenotype could be in part responsible for the slow down of the global MFH cell population proliferation rate in our experiments. In any case, we are planning to perform live cell videomicroscopy experiments to document this phenotype.

We have looked at cleaved caspase-3 and caspase-7 levels by western blots (hallmarks of apoptosis induction) but no evidence of apoptosis induction was observed (data not shown). This is in accordance with the fact that cell death occurring through mitotic catastrophe seems to be a caspase independent process (Okada and Mak, 2004).

Finally, this phenotype could be in part responsible for the slow down of the global MFH cell population proliferation rate as we have initially quantified for these cells. In any case, we are planning to perform live cell videomicroscopy experiments to document this phenotype.

Among 638 bionformatically predicted targets genes (and around 1000 transcripts encoded by these genes; http://www.microRNA.org), we first chose two cell cycles related proteins, namely E2F6 and DP2. *E2F6 en*codes a transcriptional repressor of the E2F protein family, inducing S phase exit delay when over-expressed in NIH-3T3 cells (Cartwright *et al.*, 1998). More recently, it was found to be part of several multimeric protein complexes with repressive functions, including one capable of antagonizing c-myc function in quiescent cells (Ogawa *et al.*, 2002). DP2 protein is an E2F6 partner *in vivo* (Cartwright *et al.*, 1998) and more generally an E2F partner. We showed that E2F6 and DP2 expression levels are regulated by miR-28.

We revealed also that CDC14A (among the predicted targets) is an actual target of miR-28. CDC14A is a key mitotic player from anaphase onset to mitosis exit (Stegmeier and Amon, 2004) and aberrant kinesis, similar to “dead cell doublets” phenotype that we observed when inhibiting miR-28, has been described when expression of CDC14A is manipulated (Mailand *et al.*, 2002). In fact, upregulation in U2OS (osteosarcoma) cells or downregulation by siRNA in Hela (uterine cervix carcinoma) cells produced in each case diverse types of aberrant mitoses apparently due to the formation of multipolar spindles (Mailand *et al.*, 2002; please see also their supplementary live cell imaging movies available from Nature’s website).

On the contrary, we found that neither Ki-67 (another predicted target), E2F1, E2F4 nor DP1 (non predicted targets) expression levels are regulated by miR-28. These results point in direction of *hsa-miR-28* implication in cell proliferation particularly, in S phase or G2-M and prompt us to monitor *hsa-miR-28* expression during cell cycle in synchronized cell culture. Our preliminary experiment analyzing *hsa-miR-28* expression along HeLa cell cycle (for which cell cycle phases duration were known) indicates that *hsa-miR-28* gene expression is likely cell cycle regulated.

This study brings several new informations: first, *hsa-miR-28* may promote cell proliferation and be involved in cytokinesis control as revealed by the "dead cell doublets" phenotype. Videomicrocopy monitoring experiment should validate this statement. Secondly, to the best of our knowledge, it is the first time that the expression level of a microRNA is shown to evolve during the cell cycle (furthermore with outstanding variations). Some of miR-28 mediated effects may rely on these important quantitative variations. Thirdly, we already have identified 3 targets of miR-28 which in the context of this emerging field contributes to our understanding of the physiological role of microRNAs with respect to the targets they regulate. From a global point of view, this is still in infancy since microRNAs are expected to act on several targets (and perhaps even some hundreds). Complete understanding of *hsa-miR-28* function will need further investigations including particular methodology as outlined below.

v- PERSPECTIVES

miR-28 or its targets as, respectively oncogene / TSGs in tumors

With concern to a potential role of *hsa-miR-28* in carcinogenesis of certain tumour types, the first question is related to its expression level in a series of primary tumours, to determine whether it could be a marker of a sarcoma subtype or any other tumour type. The recently introduced RT-PCR protocol will help to document this point. Up to now, miR-28 expression has been measured in normal organs (in mice) and some human tumours: among several mouse organs, its higher expression levels are found in ovary, uterus, bladder and skeletal muscle (Baskerville and Bartel, 2005). In human tumours, higher expression levels are found in uterin, ovarian and prostate types and was not evaluated for MFH tumours (Lu *et al.*, 2005).

The analysis of the phenotypical consequences after deregulation of miR-28 expression in cellular models is another avenue to elucidate its function. Commercially available vectors allowing over-expression of a given pre-miR in mammalian cells upon transient transfections (Ambion) and new strategies to silence microRNAs are useful tools to address these questions. In fact, silencing of microRNA wit the "antagomiRs" method (chemically modified oligonucleotides) has been recently proven to be efficient *in vivo* (Krutzfeldt *et al.*, 2005). This method paves the way to analyze the effect of silencing miR-28 *in vivo*, for example in mice xenografts from tumour cells that do express this microRNA.

Fundamental characterization of miR-28 role

We have experimentally validated that miR-28 modulate the protein levels of three targets with cell cycle related functions. Assuming that miR-28 may coordinate the expression of several proteins involved in this process, it is interesting to notice that several others genes cell cycle related are present in the very large list of predicted targets (compilation of three sources). These represent interesting candidates to be tested.

Our preliminary result (that need to be repeated) showing the variation of miR-28 expression level along the cell cycle (figure 5) prompt the question of the physiological regulation of the *hsa-miR-28* gene expression itself and the transcription factor(s) involved. Analysis of the 5' genomic upstream sequence of *hsa-miR-28* through the UCSC genome browser reveals that the proximal transcription factor binding site conserved among Human up to Mouse and Rat species is a binding site for the CDP (CCAAT displacement protein, also known as Cut or CUTL1) transcription repressor. Very interestingly, CDP activity seems to be cell cycle regulated in early S phase when CDP becomes able to bind DNA to exert its repressive activity (Moon *et al.*, 2001). This could easily explain the down-regulation of pre-miR-28 level in our experiment between 8 hours and 10 hours after nocodazole removal.

b. Squamous Cell Carcinomas

b-1 Lung SCC (manuscript in preparation)

i- INTRODUCTION

Descriptions of the chromosome 3 aberrations in lung SCC are summarized in the Introduction section of this thesis (see the part "Chromosome 3 aberrations in MFH and SCC"). Briefly, recurrent deletions involving several sites on the chromosome 3p arm have been reported. Gains of the long arm (or parts of the long arm) are one of the most common aberrations (present in the majority) of the SCC in studies using CGH on chromosomes (Balsara and Testa, 2002; Gebhart and Liehr, 2000).

Conventional metaphase CGH studies frequently report that amplifications cluster in the 3q26-q29 regions. Using a BAC array of 348 encoding genes loci (including 20 mapped on the long arm of chromosome 3) that may contribute to lung cancer formation, Massion and collaborators described a large region of high-level of gains (about 30 Mb in size), with two maxima at 3q26.33 (locus of PIK3CA) and 3q27.3 (locusi of STT) separated by a region that seems less amplified (Massion *et al.*, 2002).

In this study, we screened a series of 25 lung SCC for chromosome 3 aberrations using high-resolution chromosome 3 dedicated arrays to delineate precisely the common regions of alterations in lung SCC. Advanced stages (T3 and T4 tumours) were selected to increase the likelihood to pick up high-level gene amplifications.

This strategy was fruitful since we uncovered five tumours with narrow amplifications, defining a common region of 2Mb. This relatively gene poor segment contains 9 known genes (Refseq) and some Genbank mRNAs. We monitored the expression level of 10 of these genes in tumours with high-level amplification of the locus. Two genes, *SOX2* and *SOX2OT* are consistently overexpressed and represent putative driver oncogenes of the high-level amplifications of this locus observed in lung SCC.

ii- MATERIAL AND METHODS

Samples

Tumour genomic DNA from 25 lung Squamous Cell Carcinomas (T3 and T4 tumours) were extracted using classical phenol/chloroform protocol from frozen blocks. Tumour RNA from the corresponding tumors were extracted from frozen block cuts (10 slices of 10 µm thickness per tumor) using Trizol (Invitrogen) reagent following manufacturer's instructions.

Web resource and mapping informations

All mapping values are from UCSC genome browser (http://genome.ucsc.edu) July 2003 version corresponding to NCBI build 34 of the human reference sequence produced by the International Human Genome Sequencing Consortium.

Array-Comparative Genomic Hybridization

Arrays were prepared and used as previously described ("COMA cell line" manuscript submitted). Two different high-resolution chromosome 3 arrays were used in this study: the previously described "chromosome 3" array and an updated version, the "contig" array. The first one is composed of 220 clones and was used to analyse the whole series of 25 SCC. The second consists in a modified version of the previous one, with an addition of 103 clones (clones names are given in table 1). This latter array has an overall higher chromosome 3 coverage and a BAC clone tiling coverage of the 3q26.3 locus over more than 5 megabases (figure 3). It was used to reanalyze tumour DNAs with 3q26.3 amplifications found in the first screen (see results). Some tumours were also reanalyzed using IntegraChips™ pangenomic arrays composed of 3172 clones covering the whole genome**,** available from IntegraGen company (http://www.integragen.com; Evry, France).

Real time qPCR

Total RNA were submitted to DNase I (Roche) digestion for 1h at 25°C to avoid any risk of contaminating genomic DNA. Reverse Transcription reactions were performed on 500 ng RNA strictly following manufacturer's instructions (Superscript II RT Kit, Invitrogen). PCR Primers were designed using Primer 3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and their sequences are available in Table 2. The *RPLP0* gene was used as internal control for normalization purposes. We were not able to validate primer pairs for the MCF2L2 gene, AK127955 and BC036236 mRNAs. At least two primer pairs were tested in each case. Quantitative PCR were done using the Lightcycler machine (Roche) as previously described (Redon *et al.*, 2002).

Experiments (RNA extractions and reverse transcription and real time quantitative PCR reactions) were performed three times independently and results subsequently averaged.

iii- RESULTS

Array Comparative Genomic Hybridization screen of chromosome 3 aberrations

We first analyzed 25 lung SCC to delineate the chromosome 3 consensus regions of deletions and gains/amplifications in these tumours. We detect losses on the 3p arm, gains of large 3q regions and high-level amplifications of 3q26-qter genomic regions, as shown in example for two tumours (figure 1).

We calculated the frequency of losses and gains for each 2 megabases interval along chromosome 3 (figure 2A). Deletions mostly occur on the 3p arm and seem to affect several sites. The 2Mb interval (from 8 to 10 Mb) is predominantly lost in 60% of tumours, and this region contains the *OGG1* and *RAD18* genes. Large gains often target the 3q arm, with a global gain of the 3q26-qter (176-196 Mb) region in approximately 60% of tumours. Two intervals, from 180 to 182 Mb and from 188 to 190 Mb are gained in more than 80% of tumours. The first interval contains the *PIK3CA* gene whereas the second is the *RFC4, SST* and *BCL6* genes locius.

Analysis of individual clone amplification frequency in our series of experiments (ratio higher than 2 in an individual experiment) reveal frequent high-level amplifications only in the 3q26-qter region with a maximum for clone RPCI11-259I19 (figure 2B). This clone is located at 3q26.3 between *SOX2* and *DCUN1D1* genes (clone position: 183.45-183.64 Mb). We found that locus amplified in 20% of our samples (5/25 tumours).

High resolution 3q26.3 amplicons mapping and pangenomic analysis of aberrations

In an effort to depict the consensus minimal region of amplifications at 3q26.3, we decided to precisely map amplicons borders for all lung SCC with 3q26.3 amplification found in the first screen. Therefore, we built an updated version of our chromosome 3 array that contains additional clones to obtain a tiling resolution of the 3q26.3 region over more than 5 megabases (figure 3). This array is hereafter referred as to the contig array.

Analysis of the 5 tumours using the contig array confirmed the presence of 3q26.3 amplicons in each case. Moreover their extents were precisely mapped for each tumour as exemplified for 2 cases (figure 4). Thus, the common minimum 3q26.3 amplified segment in these tumours is a segment of slightly less than 2 Mb in size (figure 5A). A strict consensus region (around 1.9 Mb in size) could have been deduced on the sole basis of amplicon mapping for tumour #15. We took into account that direct neighbouring regions are gained in most of the other tumours and decided to slightly extend this interval to define the consensus region as 2.7 Mb in size, from 181.9 to 184.6 Mb on chromosome 3. UCSC genome browser genomic map is shown in figure 5B and reveals that this region contains only 9 known genes (Refseq). All these genes stand as candidate oncogenes potentially activated by gains and amplifications since their copy number is frequently increased in the series of lung SCC investigated.

We confirmed the presence of 3q26.3 amplification for the two lung SCC with the highest levels amplicons using a completely independently build array commercially available. Moreover, 3q26 amplifications in these tumours represent the amplification of highest level over the genome in both tumours (data not shown).

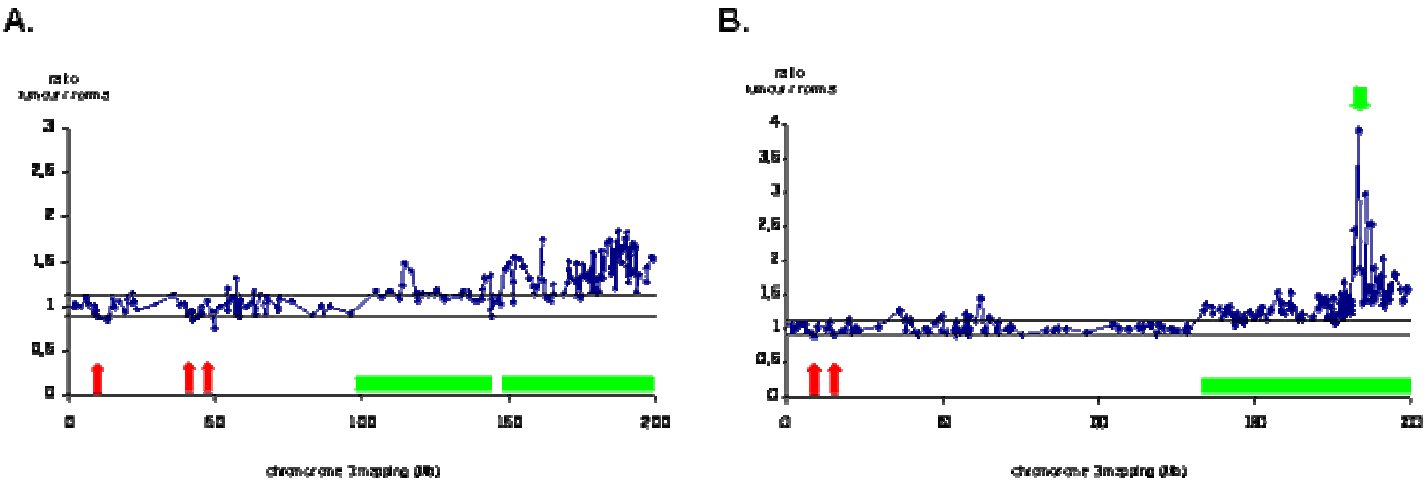


Figure 1. Examples of chromosome 3 array CGH profiles obtained for 2 lung SCC.

A. Chromosome 3 array CGH profile for lung tumour #41 Horizontal black lines represent the interval of normal copy number variations. Red arrows denote deviations under the lower limit of the confidence interval, revealing 3 interstitial deletions on the 3p arm (0-95 Mb). Deviations are observed above the upper limit of the confidence interval for most of the 3q arm (95-200 Mb). This is interpreted as a global low-level gain of most of the 3q arm as two large regions (green bars). Individual data points are arbitrarly joined by a line.

B. Chromosome 3 array CGH profile for lung tumour #35. Two interstitial deletions are detected at 3p (red arrows). Low-level gain of the 3q22-qter region is detected (green bar) Several high level amplifications are detected at 3q26-qter (160-200 Mb). The highest amplification level is noted by a green arrow (BAC clone RP11-259I29).

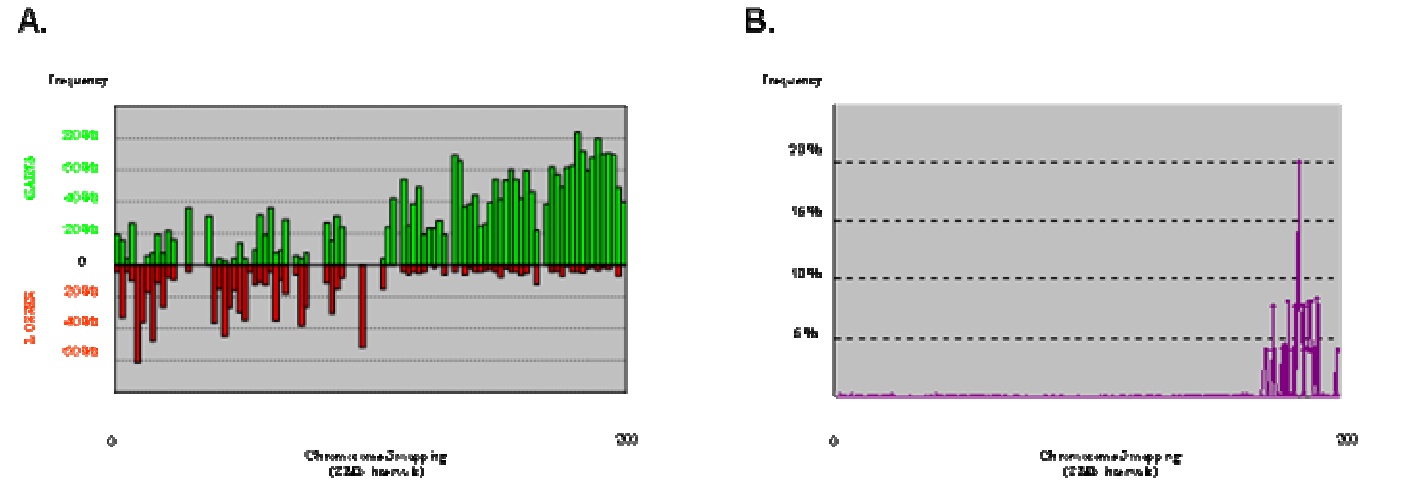


Figure 2. Frequency of chromosome 3 aberrations in 25 lung SCC.

A. Chromosome 3 losses and gains frequency. For each 2 Mb interval along chromosome 3, frequencies were calculated by dividing the total number of occurrences of losses or gains observed in the series of lung SCC by the total number of possible occurences. They were expressed as percentages and represented along the chromosomal order of the corresponding classes. Frequent losses are more often observed for the 3p arm than for the 3q arm. Losses at 3p seem to affect discontinuous zones. The most frequently affected region is the 6-10 Mb interval, lost in around 60% of samples. Gains are more frequently detected for the 3q arm than for the 3p arm The 3q26-qter is gained in at least 50% of the samples with two discrete intervals (180-182 and 188-190Mb) gained in around 80% of lung SCC.

B. Chromosome 3 high level amplifications frequency For a given clone, a ratio superior to 2 (arbitrary threshold) in a single experiment was considered to represent a high-level amplification. Frequencies were calculated for the series of 25 lung SCC and plotted according to the localization of each clone. Individual data points are arbitratly joined by a line. High level amplifications cluster in the 3q26-qter region. A peak maxima is observed for BAC clone RP11-259I19. High level amplifications of this locus were found in 20% of the samples

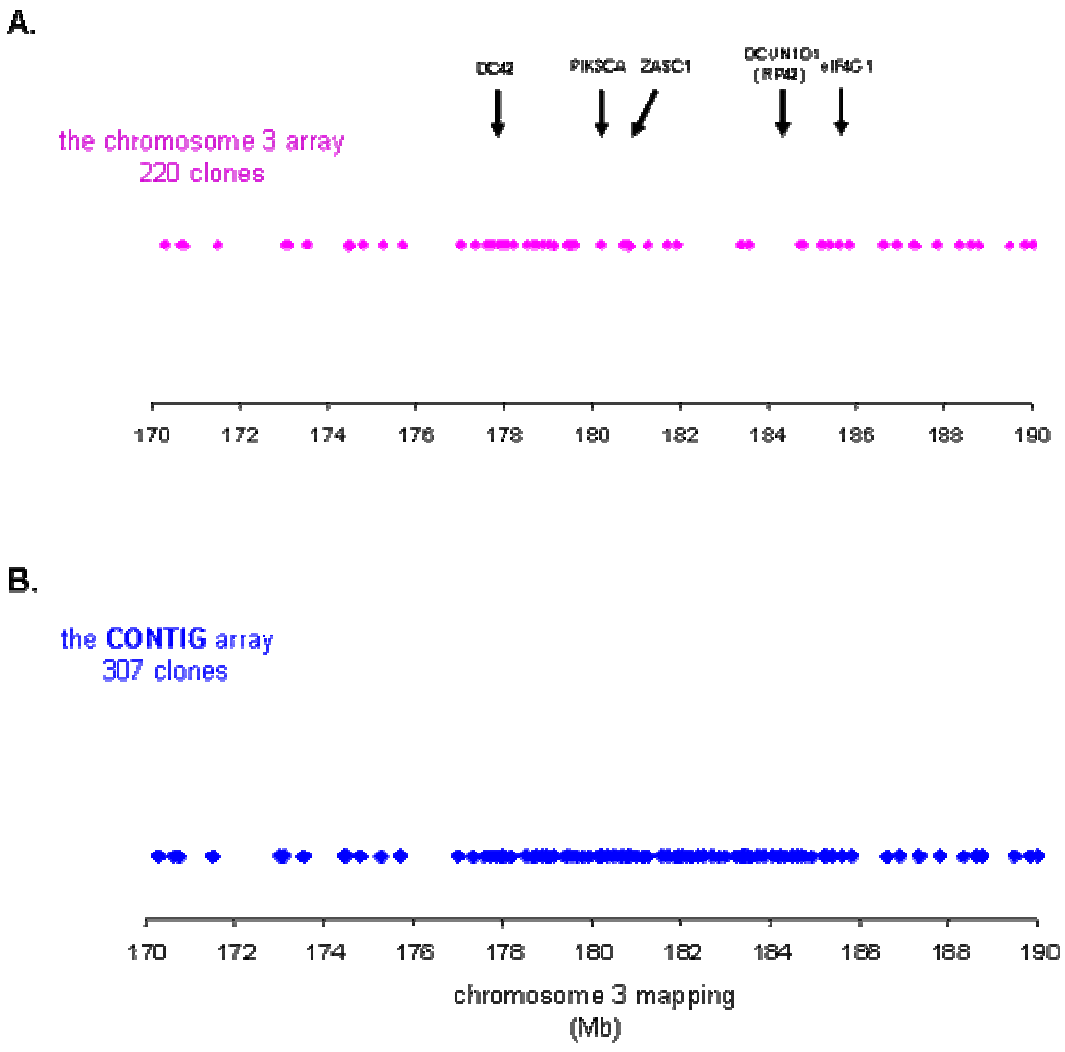


Figure 3. Coverage of the 3q26.3 region by the two different chromosome 3 arrays used for array CGH.

A. Coverage of the 3q26.3 locus by the **« chromosome 3 » array**. This array is composed of 220 BAC clones and was used for the initial screening of chromosome 3 aberrations in the series of 25 lung SCC

Arrows denote the position of some known genes whose names are indicated.

B. Coverage of the 3q26.3 locus by the **« contig » array**. This array is composed of 307 BAC clones. The 3q26.3 region is fully covered by a set of overlapping BAC clones representing more than 5 Mb of sequence.

Mapping scale is given at the lower part of both panels.

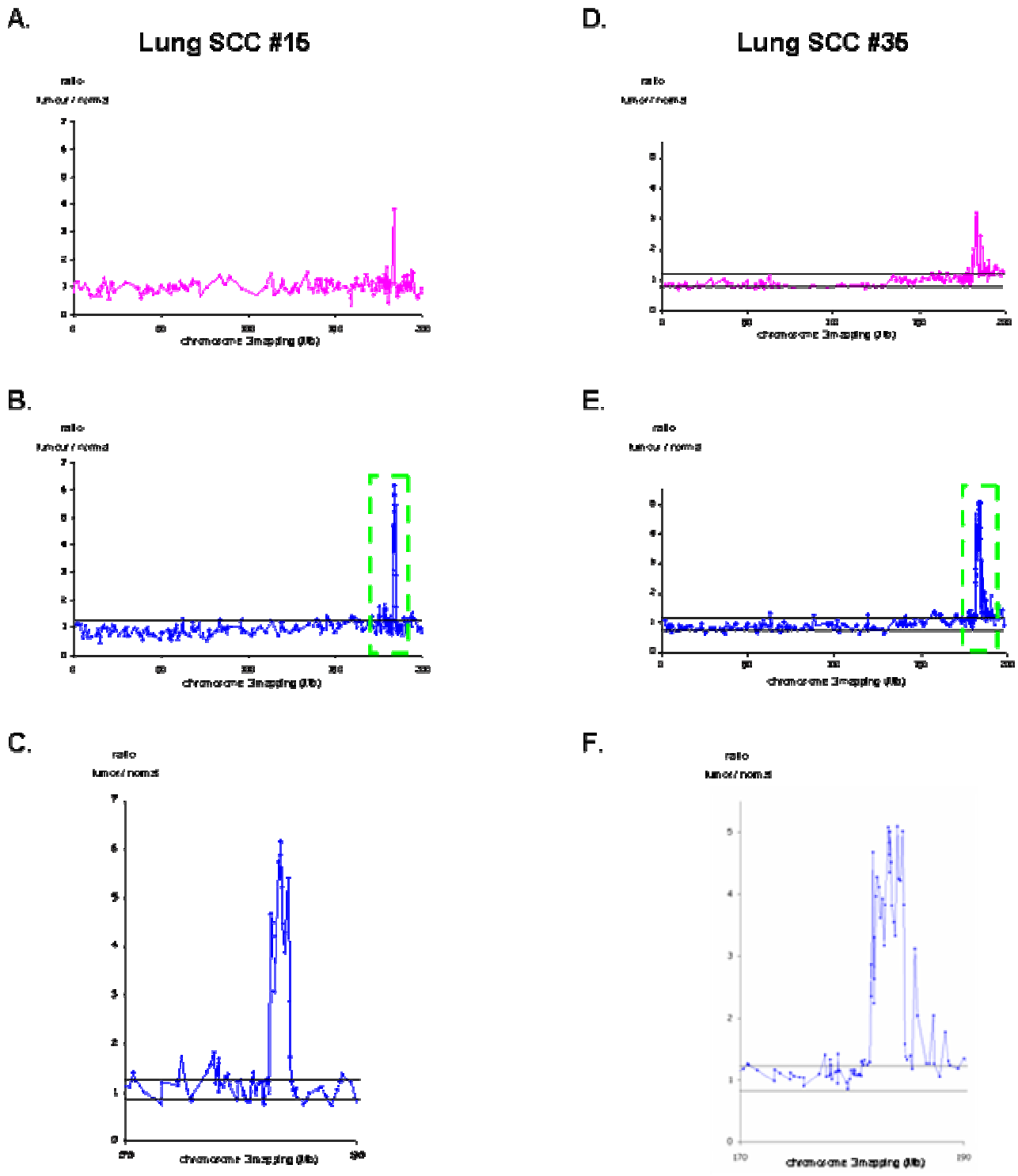


Figure 4. Reanalysis of lung SCC with 3q26.3 high-level amplifications using the contig array: examples of 2 tumours.

Results for lung SCC #15 are included in panels A, B and C, for lung SCC#35 in panels D, E and F.

The initial profiles obtained with the chromosome 3 array for tumours #15 and #35 are presented in panels A and D, respectively.

The corresponding chromosome 3 profiles obtained after re-analysis with the contig array are presented in panels B and E, respectively.

For both tumours, the profiles for the region framed by a green dashed line in panels B and E is presented in panels C and F, respectively. These tumours present high-level amplifications that are relatively small in size (1.9 and 3.1 Mb) with well defined amplicons boundaries.

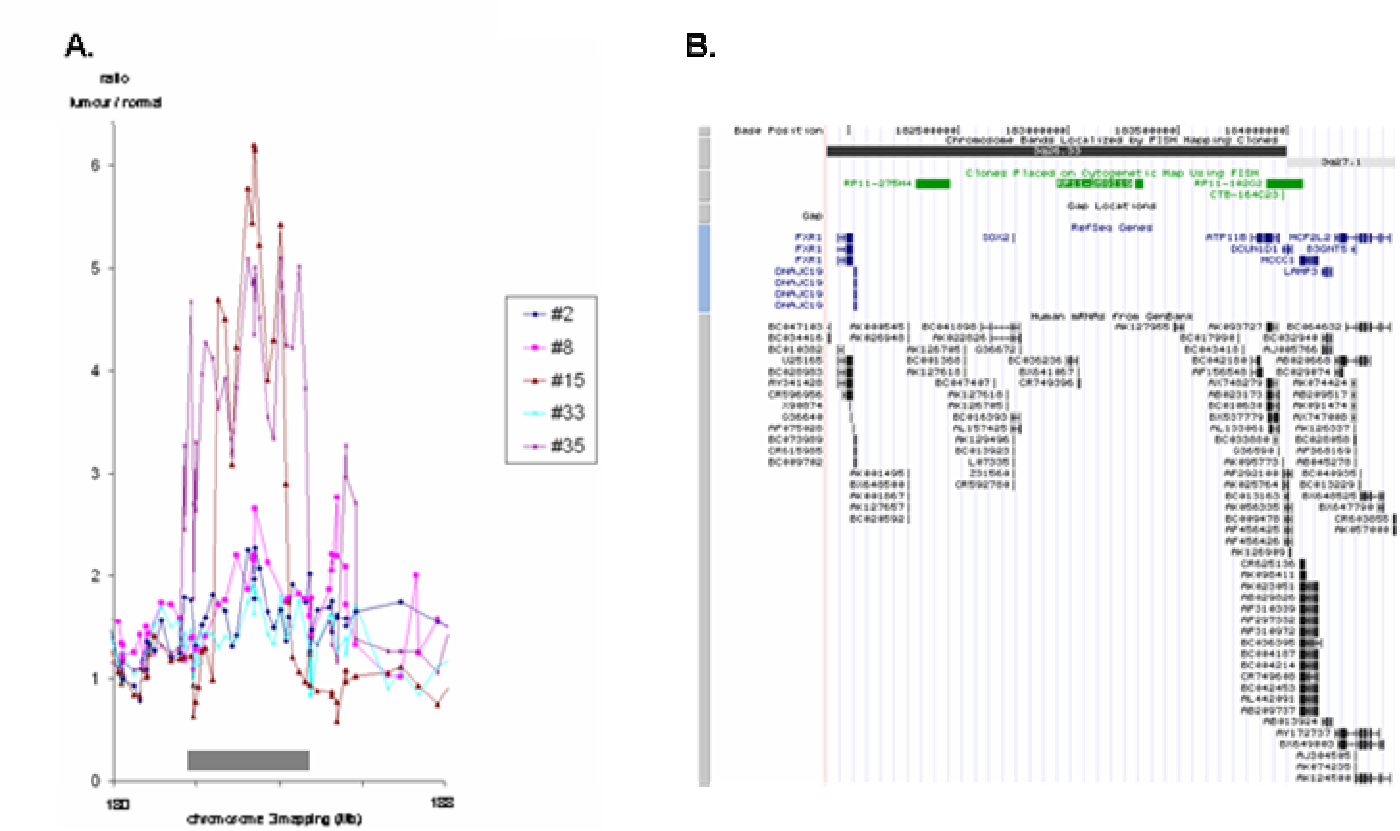


Figure 5. The consensus region of 3q26.3 high-level amplifications in 5 lung SCC

A. Individual array CGH profiles for the five tumours with 3q26.3 high-level amplifications. High ratios deviations are observed for two tumours with straight amplicons boundaries (#15 and #35). Their respective amplicons sizes are around 1.9 Mb and 3.1 Mb. A strict consensus region could have been defined on the sole basis of amplicon mapping in tumour #15, spanning 1.9 Mb. Since regions around this amplicon are gained in the other tumours, we extended the consensus region to a 2.7 Mb region spanning the 181.9-184.6 Mb interval denoted by the grey rectangle.

B. UCSC genome browser view of the consensus region (181.9-184.6 Mb). This genomic region is fully sequenced and assembled. Nine genes (Refseq) are mapped in the interval together with Genbank mRNAs.

Transcriptional consequences of 3q26.3 amplifications in Lung SCC

To discriminate between the candidate oncogenes mapped in the consensus region of amplifications, we decided to investigate by quantitative RT-PCR their respective expression levels in the 5 tumours with 3q26.3 genomic amplifications in comparison to 2 tumours with low level gains of the region (tumours #10 and #14) and to 2 tumours with normal 3q26.3 copy number (tumours #19 and #21). All of these had been submitted to array CGH analysis during the initial screen. The previously mentioned 3q26.3 copy number status (normal, low-level gain or amplification) were inferred from these results (data not shown). We were able to validate primer pairs for 8/9 Refseq genes and one Genbank mRNA. Results are presented in figure 5. In the five tumours with 3q26.3 amplifications, all the genes are recurrently over-expressed (4/5 tumours with a ratio higher than 2) except *LAMP3*. Recurrent strong over-expression (4/5 tumours with a ratio higher than 5) is found only for *SOX2* and *SOX2OT*. In the two tumours with gain (panels F and G), the two genes highly over-expressed are *SOX2* and *SOX2OT*. In conclusion, 8 out of the 9 tested genes seem to be recurrently over-expressed, but *SOX2* and *SOX2OT* genes seem strikingly the most consistently overexpressed.

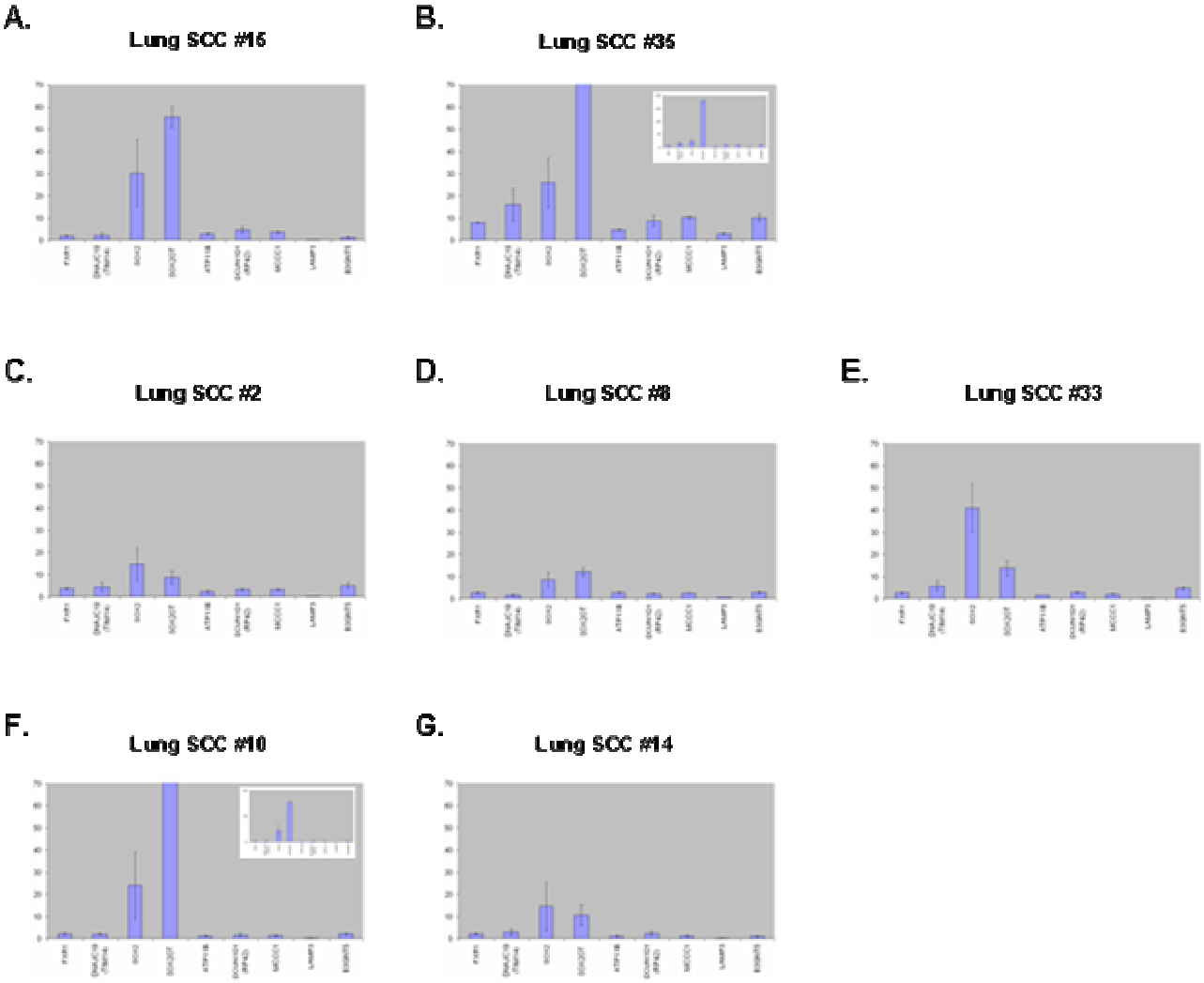


Figure 6. Transcriptional consequences of 3q26.3 high-level amplifications

Expression levels of genes from the consensus region were measured for the 5 tumours with high levels amplifications at 3q26 3 found in the initial array CGH screening (panels A to E). Two tumours with low-level copy number gains of the 3q26.3 locus found in the initial screen were also analyzed (tumours #10 and #14 presented in panels F and G, respectively).

Identical y-scale (from 0 to 70) for all graphs represents the expression ratio for a given gene when compared to its relative average expression level in two lung SCC without copy number change of the 3q26.3 locus as we have determined (tumours #19 and #21). Genes are ordered according to their genomic localization in the consensus region of amplifications. Insets in panels B and F correspond to extended scale of the y-axis to values of 200 and 100, respectively.

Two genes, *SOX2* and *SOX2OT*, are consistently over-expressed in tumours with copy number changes at 3q26.3 when compared tp tumours with normal copy number of the locus. Other genes are recurrently over-expressed but at lower levels when compared to *SOX2* and *SOX2OT*.

iv- CONCLUSIONS

Using array CGH, we have characterized chromosome 3 aberrations in a series of 25 lung SCC using a high-resolution dedicated array.

First, this enabled us to delineate several discontinuous sites of recurrent losses at 3p. These may involve different tumour suppressor genes and represent combined allelic losses that altogether could contribute to lung tumourigenesis (Hollander *et al.*, 2005). The genomic locus of *OGG1* and *RAD18* genes is the most frequently lost (60% of cases) lost in the series we investigated. These genes would represent good TSG candidate in lung SCC based on their functions.

Secondly, we have shown a high prevalence (around 50-60 %) of low-level gains of the whole 3q26-qter region in lung SCC. This is in accordance with the recurrent imbalances at 3q in SCC of various localizations and particularly from the lung (Gebhart and Liehr, 2000; Balsara and Testa, 2002). Moreover, our study pinpoint two loci from the 3q26-qter region as gained in a wide majority of advanced lung SCC. The first concerns the *PIK3CA* gene locus and the second points to the genomic locus of *RFC4*, *SST* and *BCL6* genes. *PIK3CA* is a known oncogene in various tumour types. The *PIK3CA* gene is gained and consequently over-expressed in head and neck SCC as we have previously shown (Redon *et al.*, 2001). RFC4 is also recurrently deregulated in human tumours and stands among the 67 genes consistently over-expressed in various tumour types when compared to normal counterparts (Rhodes *et al.*, 2004; *see also figure 12 of the introduction part of this thesis*)

Lastly, we delineated a common region of high-level amplifications in 20% of the investigated cases. This represents a high frequency with regards to other well known amplicons.

Re-investigations of these cases using a newly produced chromosome 3 array providing tiling coverage of the 3q26.3 region locus considered confirmed in each case the presence of the amplicon. Furthermore, the accurate mapping of amplicons borders allowed us to delineate a 2 Mb common region of high-level amplifications. Nine genes and one non-coding RNA (SOX2OT) are localized in this interval. Most of these genes seem to be over-expressed in consequence of copy number increases. The striking increases in *SOX2* and *SOX2OT* genes expression suggest that they may be the driver genes and participate to the tumour phenotype by their deregulated expression.

v- PERSPECTIVES and DISCUSSION

We found most of the 9 genes tested over-expressed except *LAMP3*. The *LAMP3* gene has just been very recently demonstrated as playing a crucial role towards invasion in cervical cancer (Kanao *et al.*, 2005). Among all our candidate genes, the *Lamp3* gene thus remains the best characterized with respect to its implication in cancer. Data mining of expression through Oncomine, the "cancer profiling database", for example is a complementary way to determine to which extend our 9 candidate genes are upregulated in other studies of lung SCC in comparison to normal tissue. From this data, *Lamp3* gene is not over-expressed when compared to normal lung tissue. The same comparison for the 8 others genes we have analysed, demonstrate an over-expression of *SOX2*, *DCUN1D1* (formerly *RP42* or *SCCRO*), *FXR1* ($p<=10^{-5}$) and *ATP11B* ($p=10^{-3}$) genes in at least one of the two lung datasets present in Oncomine and that have profiled both lung SCC and normal lung samples (Bhattacharjee *et al.*, 2001; Garber *et al.*, 2001). The other genes were either not in these transcriptome survey or not significantly over-expressed.

Functions of these genes may also help to understand their putative contribution to carcinogenesis. *DCUN1D1* (formerly *RP42* or *SCCRO*) was recurrently claimed to be an oncogene but without bringing well documented functional evidences. Interestingly, a beginning of elucidation of the function of its yeast and nematode homolog as playing a role in the mitotic spindle regulation (Kurz *et al.*, 2005) is compatible with a putative oncogenic role for *DCUN1D1*. The *ATP11B* gene is poorly characterized and *FXR1* involvement in tumorigenesis is not supported by any data.

In this study, we demonstrate that up-regulation of two genes, namely *SOX2* and *SOX2OT* were striking and recurrent among tumours with either 3q26.3 amplifications or gains. These large expression variations are very likely to have important biological consequences.

SOX2 is contained in an intron of the SOX2 Overlapping Transcript encoding gene, *SOX2OT*. This gene is transcribed in the same sense than SOX2 and thus does not represent an antisense RNA. Its function remains unknown at present. but it would remain of interest to determine it. On the contrary, the *SOX2* gene represents an oustanding oncogene candidate based on its function. The Sox-2 protein is indeed a cofactor of the Oct-4 transcription factor. Oct-4 oncogenicity has been shown recently in testicular germ cell tumours (Gidekel *et al.*, 2003) and both are crucial markers of embryonic stem cells, involved in the control of their proliferation and differentiation by regulating the downstream expression of a huge number of target genes (Boyer *et al.*, 2005). Finally, *in vitro* experiments have shown the involvement of Sox2 in apoptosis: hypoxia and glucose deprivations were shown to induce early Sox2 proteolysis as well as double strand DNA breaks that classically mediate apoptosis induction (Lei *et al.*, 2005). Furthermore, Sox2 binds to genomic DNA at a scaffold/matrix attachment region, where DNA breaks are known to initiate after oxygen and glucose deprivations treatment (Lei et al. 2005). Consequently, the authors propose that early proteolysis of complexes (including SOX2) at DNA sites represent a universal mechanism that renders these DNA sites vulnerable to endonucleolysis. It is from that tempting to speculate that *SOX2* overexpression in tumour cells could protect them from apoptosis induced by oxygen and nutrient starvations.

Taken together, our data and those from the literature suggest that the genes located at 3q26.3 that are recurrently amplified and consequently over-expressed in lung SCC may act as co-conspirators in these tumours through modulations of several tumoural traits. The next step towards the elucidation of their respective contributions to the tumour phenotype when amplified and over-expressed relies on a functional analysis. This will be achieved experimentally through dedicated *in vitro* assays. These aim at analyzing the effects of

candidate oncogene over-expression. We have subcloned the cDNAs encoded by these genes in a common mammalian expression vector and already performed set-up experiments to monitor the effects of their over-expression on three tumoural traits, proliferation, invasion and resistance to oxygen-glucose deprivation.

b-2 Head and Neck Squamous Cell Carcinomas

i- CCNL1 as a candidate oncogene in Head and Neck SCC

a) INTRODUCTION

Our entry point in this study was a HNSCC cell line, CAL27 that we screened using one of the first chromosome 3 dedicated BAC array we developed. We were able to delineate a narrow amplification comprising the *cyclinL* / *ania 6a* gene (now known as the *cyclinL1* or *CCNL1* gene) and showed that it was the most consistent oncogene candidate (Redon *et al.*, 2002; see this manuscript hereafter).

To further assess *CCNL1* putative involvement in HNSCC carcinogenesis, we decided to decipher CCNL1 gene alterations in a large series of tumours. Accordingly, the real time qPCR method was chosen to explore CCNL1 gene copy number status as well as *CCNL1* gene expression level determination in two sets of 38 and 96 tumours, respectively (manuscript submitted to British Journal of Cancer: Muller *et al.*, 2005; see the manuscript 2 hereafter).

b) RESULTS

Manuscript 1
Amplicon mapping and transcriptional analysis pinpoint *cyclin L* as a candidate oncogene in Head and Neck cancer.
Redon *et al.*, Cancer Research, 2002. Co-first author.

Manuscript 2
Analysis of CCNL1 gene copy number and expression level alterations in a large series of HNSCC by qPCR.
Muller *et al.*, submitted to British Journal of Cancer, 2005. Fourth author.

Manuscript 1

Amplicon mapping and transcriptional analysis pinpoint *cyclin L* as a candidate oncogene in Head and Neck cancer

Redon *et al.*, Cancer Research, 2002.

Co-first author

Richard Redon, **Thomas Hussenet**, Gaétan Bour, Krishna Caulee, Bernard Jost, Danièle Muller, Joseph Abecassis and Stanislas du Manoir

Amplicon mapping and transcriptional analysis pinpoint *Cyclin L* as a candidate oncogene in head and neck cancer.

Cancer research 62, 6211-6217 (2002)

Pages 6211-6217:

Manuscript 2

Analysis of *CCNL1*

gene copy number and expression level alterations

in a large series of HNSCC by qPCR.

Muller *et al.*, submitted to British Journal of Cancer (2005),
under favourable revisions.

4th author

Cyclin L1 (CCNL1) gene alterations in human head and neck squamous cell carcinoma.

D. Muller[1], R. Millon[1], S. Théobald[2], T. Hussenet[3], B. Wasylyk[3], S. du Manoir[3], J. Abecassis[1].

[1]Laboratoire de biologie tumorale, [2] Service de statistique et d'épidémiologie, Centre de lutte contre le cancer Paul Strauss, 3 rue de la Porte de l'Hôpital, 67085 STRASBOURG cedex, France. [3] Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/Collège de France, 1 rue Laurent Fries, BP10142 67404 Illkirch Cedex, C.U. de Strasbourg, France.

Correspondence to: D. Muller[1] ; E-mail : btumorale@strasbourg.fnclcc.fr

Running title: *Cyclin L1* alterations in HNSSC.

ABSTRACT

We evaluated the expression and amplification of *cyclin L1* (*CCNL1)* gene, a potential oncogene localized in the commonly amplified 3q25-28 region, in human head and neck squamous cell carcinomas (HNSCCs). Overexpression was observed in 55/96 cases (57%) and amplification in 9/35 tumours (26%) with no relationships to the clinico-pathological parameters. The Cyclin L1 antibody we developed labels nuclear speckles in tumour cells compatible with a role for CCNL1 in RNA splicing.

Keywords: Head and neck, squamous cell carcinoma, cyclin L1, CCNL1, chromosome 3q, HNSCC

INTRODUCTION

Definition of the genetic changes associated with head and neck squamous cell carcinoma (HNSCC), in particular by comparative genomic hybridization (CGH), pinpointed gain of the 3q chromosome as one of the most frequent abnormalities (Bockmuhl et al, 2000). We demonstrated that the 3q25-28 region is a consistent site of gain in early HNSCC (Redon *et al*, 2001), and suggested that the gain is associated with shortened disease survival (Redon *et al*, 2001). This association was confirmed in other larger studies underlying the importance of this region in the development of HNSCC (Ashman *et al*, 2003, Bockmuhl et al, 2000, and publications therein). Several putative oncogenes from the 3q25-28 region have been proposed to be potentiel targets of amplification, including *PIK3CA* (Estilo *et al*, 2003), *SCCRO* (Estilo *et al*, 2003), *p63* (Thurfjell *et al*, 2005), and *cyclin L1 (CCNL1),* which we

showed by microarray-based CGH to be amplified in a HNSSC cell-line (Redon *et al*, 2002). The main □ isoform of Cyclin L1 contains an RS domain (Ser-Arg rich proteins) that may confer splicing activity to cyclin L1 (Dickinson *et al*, 2002). *CCNL1* is an immediate early gene induced by several growth factors (Berke et al., 2001) that may regulate G0-G1 cell-cycle progression (Redon *et al*, 2002 and publications herein).

In the present study, we measured *CCNL1* gene expression in 96 human HNSCCs by real time quantitative PCR. *CCNL1* gene copy number was evaluated to examine the relationship between RNA expression and DNA copy number alterations. The relationship between *CCNL1* gene alterations and pathological features as well as clinical outcome is reported. Finally, the localization of the cyclin L1 protein was determined by immunohistochemistry on normal and neoplastic head and neck tissues.

MATERIAL AND METHODS

- **Patients and samples**

96 HNSCC samples were selected from consenting patients who underwent surgery as first treatment, without previous radiation or chemotherapy. Matched normal mucosa samples were obtained for 82 cases by resection at least 5 cm from the tumour. They were immediately frozen and stored in liquid nitrogen. Clinico-pathological features are summarized in Table 1 using the UICC TNM system. The mean clinical follow-up was 62 months (range 0.5 – 199) .

Cyclin L1 gene expression and copy number

Total RNA isolation and cDNA preparation were performed as described previously (Redon *et al*, 2001). Genomic DNA was isolated by phenol/chloroform extraction. *CCNL1* gene amplification and expression were assessed by quantitative PCR performed with the LightCycler system, using LC Fast start DNA master SYBR green I mixture, and version 3.0 software (Roche Diagnostics, Meylan, France).

The specific oligonucleotide primers, designed by Primer3 software, were: (a) for gene expression analysis: *CCNL1,* 5'-ACTCCAAGCCCCCTGATCCT-3' and 5'-TGGCAACGGAATCTGAAGTG-3' which amplify the □ and □ isoforms; the ubiquitous gene *RPLP0,* 5'- GAAGGCTGTGGTGCTGATGG-3' and 5'-CCGGATATGAGGCAGCAGTT-3', (b) for gene amplification analysis; *CCNL1,* 5'-TAGGCGGAGTCGATCTGGAA-3' and 5'- CCATGGTGCTTGCTTTTATGG-3'; and two genes located on respectively chromosome 15q15-21 and 11p15, CAPN3 (calpain3/p94), 5'- GCTGGTAGGAGACCCCCAAG-3' and 5'-CCACAGATGCGCTAATGACG-3', HBB (beta-globin), 5'-GAAGAGCCAAGGACAGGT-3' and 5'- TGGTGTCTGTTTGAGGAAGC-3'.

In order to obtain for each gene results in ng for expression and copy number, respectively, calibration curves were constructed using pools of cDNAs from 10 normal head and neck tissue samples and DNA from peripheral blood samples from 10 healthy individuals. *CCNL1* gene expression was evaluated for 96 tumours and 82 corresponding normal tissues using the mean value from 3 independent experiments. The relative *CCNL1* gene expression level was calculated by successive normalisation to the RPLP0 (Ribosomal phospho-protein P0) internal control and then to the mean expression of all the normal tissue samples. A relative value ≥ 1.7 was considered to represent overexpression. *Cyclin L1* copy number changes were

analyzed for 35 tumours and 14 normal samples by independent duplicate PCR reactions with two DNA inputs, 6 and 2 ng. Copy number values for a tumour sample were calculated in relative units adjusted to the mean values of CAPN3 and HBB internal controls. The results were expressed as fold differences in target gene copy number in tumours relative to normal samples. A relative value ≥1.5 was considered to represent amplification.

- **Cyclin L1 protein expression**

Cyclin L1 protein expression was assessed by immunohistochemistry, using a mouse monoclonal antibody raised against the peptide SKHHGGRSGHCRHRR. This sequence is found only at the C-terminal of the isoform □. For retrieval, 3 µm de-waxed sections were pressure-cooked for 3 min in 0.1 M citrate buffer pH 6.0. They were incubated overnight at 4°C with a 1/1000 dilution of the primary antibody, and revealed by the avidin-biotin peroxydase complex method (DAKO labelled streptavidin-biotin LSAB). Negative controls lacked the primary antibody.

- **Statistics**

Associations with clinico-pathological features and *CCNL1* expression or amplification were analysed at the 5% significance level using the Anova or Kruskal Wallis tests where appropriate. Survival was analysed by the Kaplan-Meier method and the log-rank test with MedCalc software (MedCalc, Belgium).

RESULTS

Real time quantitative PCR analysis of 96 HNSCCs and 82 of their counterpart normal tissues showed that *Cyclin L1* gene expression levels were significantly

different between tumours and normal tissue samples [T-test; p<0.0001] (Figure 1A). Overexpression was found in 55 tumours (57%). Expression levels in tumours were not significantly associated with their clinico-pathological features (tumour site, size or differentiation, lymph node involvement, Table 1).

CCNL1 gene amplification, evaluated in 35 HNSCCs, was observed in 26% of the tumours and does not exceed 3 fold (relative to the mean of 14 normal tissue samples) (Figure 1B). Overexpression was observed in almost all tumours with amplification (7/9 cases with amplification), while lack of overexpression in tumours without amplification was observed in 65% (17/26) of the cases, suggesting correlation between *CCNL1* overexpression and amplification (Table 2, Fisher's test p=0.049). Amplification was not statistically related to the clinico-pathological features of the tumours (not shown).

Disease progression was observed for 60 patients. No statistically significant relationship was observed between either *CCNL1* overexpression or amplification and crude survival or relapse-free survival of the patients, even for tumours with high expression levels (highest quartile, data not shown). Gene amplification evaluation was mainly performed on tumours of pharyngeal origin (86%). For the tumours with concomitant overexpression and amplification, no association with either histo-clinical features or survival outcome was observed.

We developed a monoclonal antibody against the long form of the protein (isoform)and used it for Western blotting (Figure 2A). The nuclear staining (Figure 2B) observed by immunocytofluorescence is similar to that observed in COS cells transfected with flagged CCNL1 and revealed by an anti-FLAG antibody (Berke *et al*, 2001). In tumours, Cyclin L1 staining, is mostly restricted to the nuclei of the epithelial

cells with discrete additional cytoplasmic staining in some cells (Figure 3). No staining was observed in stromal cells. Heterogeneous staining was observed throughout the tumour sections and between different tumours. Normal epithelium displays weak nuclear staining of the parabasal cells (arrow).

DISCUSSION

In this large series of HNSCCs, we observed a high prevalence of *CCNL1* gene overexpression, occurring in more than half of the tumours. *CCNL1* gene copy number was increased in 26% of the cases, in accordance with the 34 % of gains reported by FISH analysis with a *CCNL1* gene probe (Sticht *et al*, 2005). Furthermore, we demonstrated gene overexpression in most HNSCCs with amplification of the *CCNL1* gene, supporting the hypothesis that its expression level partly follows a gene dosage effect. However, *CCNL1* overexpression was also detected in 34% (9/26) of tumours without *CCNL1* gene amplification, showing that other transcriptionnal regulation factors are involved. Similar discrepancy between overexpression and amplification has been reported for other oncogenes. For example overexpression of cyclin D1 in oral squamous cell carcinomas is twice as frequent as *CCND1* amplification (Miyamoto *et al*, 2003).

A recent study, comparing DNA copy number to gene expression levels over large chromosomal regions in HNSCCs, concludes that chromosomal alterations affect the expression of many genes (Masayesva *et al*, 2004). The discrimination between genes that are important for tumour evolution from bystander ones within regions of copy number changes remains complex. Only some of these overexpressed genes

may be important for tumorigenesis, and identifying them replies on several clues such as gene amplification/expression correlations, functional characterization or association with clinical features. Multiple genes within the commonly gained 3q region appear to be relevant for head and neck cancer progression. The gain of chromosomal region 3q25-28 is associated with aggressive clinical behavior (Bockmuhl *et al*, 2000, Ashman *et al*, 2003 and Singh *et al*, 2002 cited therein). Overexpression of several genes at 3q26 have prognostic value, including *SCCRO* in oral tongue (Estilo *et al*, 2003) and *ZASC1* in oesophageal (Imoto *et al,* 2003) tumours. Recently, high levels of *CCNL1* gene amplification, assessed by FISH (> 8 signals/cell) have been found in 3% of a large HNSCC series and were shown to be associated with shorter survival (Sticht *et al*, 2005). In our series, high-level amplification was not found, and neither *CCNL1* gene amplification nor overexpression was associated with unfavourable tumour phenotype. Difference concerning amplification between both studies may be due to sampling differences since our population is mainly composed of tumours derived from pharynx (86 %) while Sticht *et al*. investigated mostly oral cavity tumours.

Inappropriate expression of multiple genes at the 3q25-28 chromosomal locus, including *CCNL1*, could contribute to disease progression, providing growth advantage by additive effects in different cellular pathways. *CCNL1* (isoform □ could be a regulator of the G0 to G1 cell-cycle transition. Analysis of the relationships with cell cycle regulators specially relevant in head and neck cancers, such as Cyclin D1 or growth factors, would be of interest. *CCNL1* may have also a role in transcription or RNA splicing, since it colocalizes within nuclear speckles with splicing factors SC-35 and 9G8 in cells transfected by *cyclin L1*isoform □ (Yang *et al*, 2004). Interestingly, cyclin L1 was also observed in the nuclei of malignant cells of head and

neck tumours, by histological immunodetection using our antibody. CCNL1 overexpression may modify the splicing pattern of specific genes, in particular genes involved in apoptosis for which splicing variants have been shown to have agonist or antagonist effects (Schwerk *et al*, 2005). Altogether, these findings suggest that cyclin L1 might have a role in the RNA processing complex and could participate to tumour progression of HNSCC. It thus remains of interest to investigate further the physiological functions of cyclin L1 and its link to head and neck carcinogenesis.

REFERENCES

Ashman JN, Patmore HS, Condon LT, Cawkwell L, Stafford ND, Greenman J. (2003). Prognostic value of genomic alterations in head and neck squamous cell carcinoma detected by comparative genomic hybridisation. *Br J Cancer* **89** : 864-869.

Berke JD, Sgambato V, Zhu PP, Lavoie B, Vincent M, Krause M, Hyman SE. (2001). Dopamine and glutamate induce distinct striatal splice forms of Ania-6, an RNA polymerase II-associated cyclin. *Neuron* **32** : 277-87.

Bockmuhl U, Schluns K, Kuchler I, Petersen S, Petersen I. (2000). Genetic imbalances with impact on survival in head and neck cancer patients. *Am J Pathol.* **157** : 369-375.

Dickinson LA, Edgar AJ, Ehley J, Gottesfeld JM. (2002). Cyclin L is an RS domain protein involved in pre-mRNA splicing. *J Biol Chem*. **277:** 25465-25473

Estilo CL, O-Charoenrat P, Ngai I, Patel SG, Reddy PG, Dao S, Shaha AR, Kraus DH, Boyle JO, Wong RJ, Pfister DG, Huryn JM, Zlotolow IM, Shah JP, Singh B. (2003). The role of novel oncogenes squamous cell carcinoma-related oncogene and phosphatidylinositol 3-kinase p110alpha in squamous cell carcinoma of the oral tongue. *Clin.Cancer Res*. **9** : 2300-2306.

Imoto I, Yuki Y, Sonoda I, Ito T, Shimada Y, Imamura M, Inazawa J. (2003).

Identification of ZASC1 encoding a Kruppel-like zinc finger protein as a novel target for 3q26 amplification in esophageal squamous cell carcinomas**.** *Cancer Res*. **63** : 5691-5696.

Masayesva BG, Ha P, Garrett-Mayer E, Pilkington T, Mao R, Pevsner J, Speed T, Benoit N, Moon CS, Sidransky D, Westra WH, Califano J. (2004). Gene expression alterations over large chromosomal regions in cancers include multiple genes unrelated to malignant progression. *Proc. Natl. Acad. Sci. USA* **101** : 8715-8720.

Miyamoto R, Uzawa N, Nagaoka S, Hirata Y, Amagasa T (2003). Prognostic significance of cyclin D1 amplification and overexpression in oral squamous cell carcinomas. *Oral Oncol.* **39** : 610-618.

Redon R, Muller D, Caulee K, Wanherdrick K, Abecassis J, du Manoir S (2001). A simple specific pattern of chromosomal aberrations at early stages of head and neck squamous cell carcinomas: PIK3CA but not p63 gene as a likely target of 3q26-qter gains. *Cancer Res*. **61** : 4122-4129.

Redon R, Hussenet T, Bour G, Caulee K, Jost B, Muller D, Abecassis J, du Manoir S. (2002). Amplicon mapping and transcriptional analysis pinpoint cyclin L as a candidate oncogene in head and neck cancer. *Cancer Res*. **62** : 6211-6217.

Schwerk C, Schulze-Osthoff K (2005). Regulation of apoptosis by alternative pre-mRNA splicing. Mol Cell. **19** : 1-13.

Sticht C, Hofele C, Flechtenmacher C, Bosch FX, Freier K, Lichter P, Joos S. (2005). Amplification of Cyclin L1 is associated with lymph node metastases in head and neck squamous cell carcinoma (HNSCC*). Br J Cancer* **92**: 770-774.

Thurfjell N, Coates PJ, Boldrup L, Lindgren B, Backlund B, Uusitalo T, Mahani D, Dabelsteen E, Dahlqvist A, Sjostrom B, Roos G, Vojtesek B, Nenutil R, Nylander K. (2005). Function and importance of p63 in normal oral mucosa and squamous cell carcinoma of the head and neck*. Adv Otorhinolaryngol*. **62** :49-57.

Yang L, Li N, Wang C, Yu Y, Yuan L, Zhang M, Cao X. (2004). Cyclin L2, a novel RNA polymerase II-associated cyclin, is involved in pre-mRNA splicing and induces apoptosis of human hepatocellular carcinoma cells. *J Biol Chem*. **279** : 11639-648.

ACKNOWLEDGEMENTS

We are indebted to Christine Macabre for technical support. We are grateful to Mustapha Oulad from the IGBMC for the preparation of the cyclin L1 antibody and technical advice. This study was supported by grants from the “Comité Régional du Haut-Rhin de la Ligue contre le Cancer” (Equipe Labellisée), and by funds from INSERM and the CNRS. Thomas Hussenet is recipient of a bursary from “Association pour la Recherche sur le Cancer”.

LEGENDS TO FIGURES

Figure 1 : *CyclinL1* gene alterations assessed by quantitative real-time PCR.

A. Relative *cyclin L1* gene expression level in 96 head and neck tumors (T) and 82 normal tissues (N) ; T-test, p<0.0001, mean value for N : 1 by construction (95%Ci=0.9-1.09), for T : 2.01 (95%Ci=1.8-2.2),

B. Relative *cyclin L1* gene copy level in 14 normal (N) and 35 tumor tissues (T) ; T-test, p=0.02, mean value for N : 1 by construction (95%Ci=0.95-1.05), for T : 1.26 (95%Ci=1.12-1.39**).**

Figure 2 : Validation of anti Cyclin L1 alpha mouse monoclonal antibody.

A. Western Blot validation. The monoclonal antibody was validated with a Cal-27 HNSCC cell line protein extract. It detects a protein (indicated by the arrow) of size between 54 and 80 kDa, compatible with predicted size of CCNL1 alpha (526 aa, around 60 kDa). Molecular marker sizes are indicated on the left.

B. Immunocytofluorescence validation. The CCNL1 alpha antibody (red color) labels nuclear speckles of the two transfected COS cells presented. No cytoplasmic staining was observed. Blue color corresponds to DAPI staining of nuclei. This pattern is identical to a previously published localisation using an anti-FLAG labelling on transfected COS cells (Berke *et al.*, 2001).

Figure 3 : Immunohistochemical detection of cyclin L1 protein in normal and neoplastic head and neck tissues from HNSCC patients**.**

A. Slight positive staining of parabasal cells (arrow) of normal epithelium from the upper aero-digestive tract (Bar represents 200 µm).

B. Strong nuclear staining of tumour cells in a moderately differentiated head and neck tumour, tract (Bar represents 50 µm).

C. Enlargement of image B, positive staining of speckles in the nuclei tract (bar represents 25 µm).

Figure 1 from Muller *et al.* (2005).

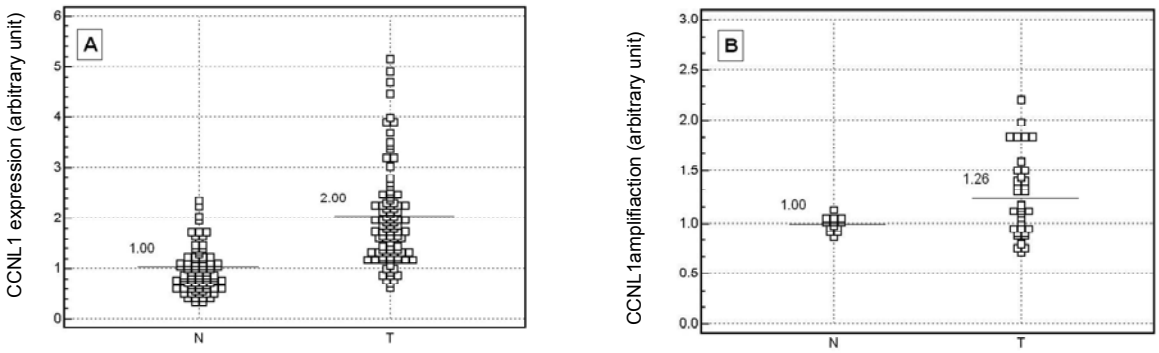


Figure 2 from Muller *et al.* (2005).

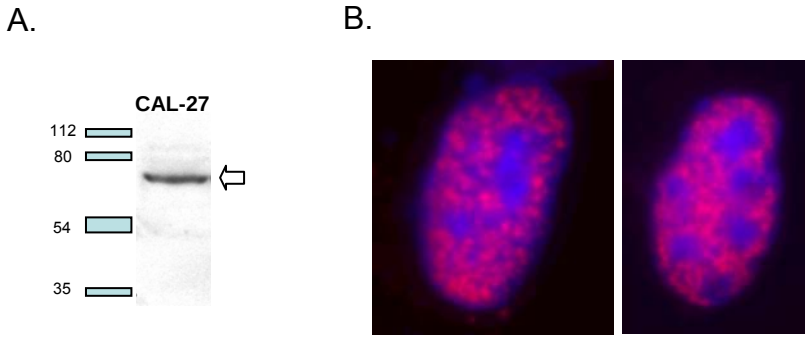
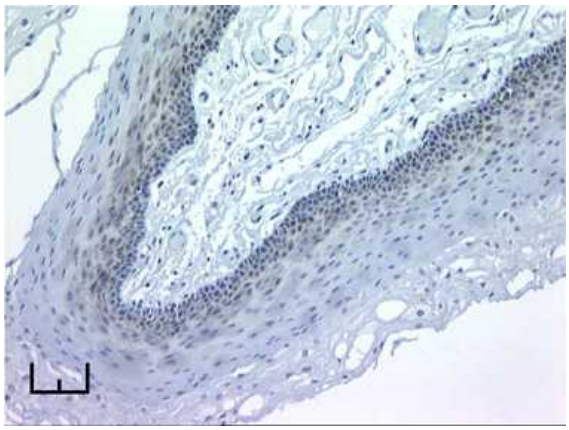
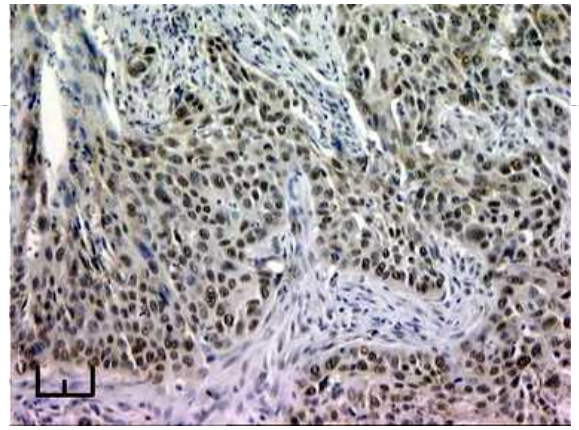


Figure 3 from Muller *et al.* (2005).

A.



B.



C.

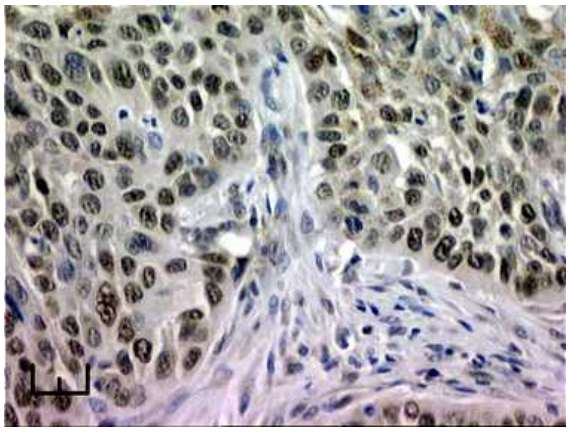


Table 1 : **Relation between *Cyclin L1* gene expression and clinico-pathological features of tumours**

	Number of cases	Overexpression Number of cases (%)	P value*
Localization			
All	96	55 (57%)	
Hypopharynx	64	38 (59%)	
Oropharynx	20	13 (65%)	0.245
Oral cavity	12	4 (33%)	
Tumour size			
T1	12	7 (58%)	
T2	49	33 (67%)	0.065
T3-4	35	15 (43%)	
Lymph node involvement			
N0	44	22 (50%)	0.59
N+	52	33 (63%)	
Tumour Differentiation			
Well	30	13 (43%)	0.34
Moderately	34	22 (65%)	
Poorly to un-differentiated	32	20 (62%)	

* ANOVA test on relative expression values

Table2 : **Relationship between *cyclin L1* gene amplification and expression in head and neck SCC. Fisher's test, p=0.0497**

	Not amplified	**Amplified**	*total*
No overexpression	**17**	**2**	*19*
Overexpression	**9**	**7**	*16*
total	*26*	*9*	*35*

c) CONCLUSIONS / PERSPECTIVES

The first perspective concerns the basic function of CCNL1. One open question is still to determine if CCNL1 has any role in the cell cycle, in particular during G0/G1 transition as we have proposed (Redon *et al.*, 2002). We have already established NIH-3T3 stable clones for the two main isoforms of cyclin L1, the gene being under the control of an inducible promoter. We intend to monitor the effects of its over-expression on proliferation parameters in cell culture experiments.

The second perspectives is related to the expression of CCNL1 in HNSCC tumours and its relationship to clinical parameters. A recent publication from Sticht *et al.* (2005) using Tissue MicroArrays and FISH with probes for four 3q potential oncogenes (*CCNL1*, *PIK3CA*, *SNO* et *p63*) demonstrated (i) a relative similar frequency of gains for these four sites (ii) an association between CCNL1 gains and lymph nodes invasion (iii) an association with reduced survival for tumours with high-level gains (more than 8 copies). These data confirmed that amplification of *CCNL1* gene may be correlated with clinical parameters. The anti-CCNL1 antibody that we have produce may help to reply to this question in large series of HNSCC and may become a useful new biomarker for clinical purposes.

ii- Chromosome 3 aberrations in Head and Neck SCC

a) INTRODUCTION

Survey of chromosome 3 aberrations for HNSCC is included in the Introduction part of this thesis and in our previously published paper (Redon *et al.*, 2002). The best reviews are those of Bockmuhl *et al.* (2000) and of Gebhart and Liehr (2000). Overall picture is very similar to lung SCC with several regions of losses on the short arm and recurrent gains on the long arm. Nevertheless, precise mapping may pinpoint cytogenetic bands for which amplification and losses frequency may differ. Consequently, it remains of interest to depict theses copy number changes in HNSCC in order to compare these two tumours types of similar histological origin and different localisations.

We undertook a screen of chromosome 3 aberrations in a series of 25 HN SCC using a high-resolution chromosome 3 dedicated array (referred as to the contig array in the lung SCC chapter). We hypothesized that tumours of advanced stages will contain more amplifications and will be more suitable to delineate a narrow genetic interval to identify candidate oncogenes.

b) MATERIAL AND METHODS

Samples

Tumour genomic DNA from 25 head and neck Squamous Cell Carcinomas (T3 and T4 tumours) were extracted using standard protocols.

Mapping values

All mapping values given hereafter were from UCSC genome browser July 2003 version corresponding to NCBI build 34 of the human reference sequence produced by the International Human Genome Sequencing Consortium.

Array-Comparative Genomic Hybridization

The "contig" array was prepared and used as described previously (see the lung SCC study) for the analysis of the 25 HNSCC cases.

c) RESULTS

Array Comparative Genomic Hybridization screen of chromosome 3 aberrations in HNSCC

We analyzed 25 HNSCC to delineate the chromosome 3 consensus regions of deletions and gains/amplifications in these tumours. We were able to frequently detect losses on the 3p arm and gains of large 3q regions. Three tumours with high-level amplifications at 3q26-qter were also observed (figure 1). In the three cases more than one peak was present. Finally the three amplicons with the higher ratio maxima were included in the 182-185 segment (arrow).

We calculated the frequency of deletions and gains for each 2 megabases interval along chromosome 3 (figure 2A).

Losses of large blocks of chromosome 3p were observed. Four different genomic regions are prominently lost, in more than 60% of the investigated cases: these are the intervals from 2 to 4 Mb, from 18 to 20 Mb, from 44 to 48 Mb and from 84 to 88 Mb along chromosome 3.

An overall gain of the 3q26-qter is observed in approximately 40% of tumours. Two intervals, from 156-158 and 182-186 Mb have frequency of gain of about 80 %. The first one is the *GMPS-CCNL1* genomic region. The second is the *FXR1-MAGEF1* interval which contains numerous genes including *SOX2* and *EIF4G1*.

High level amplifications were found to be restricted to the 3q26-qter region (figure 2B).

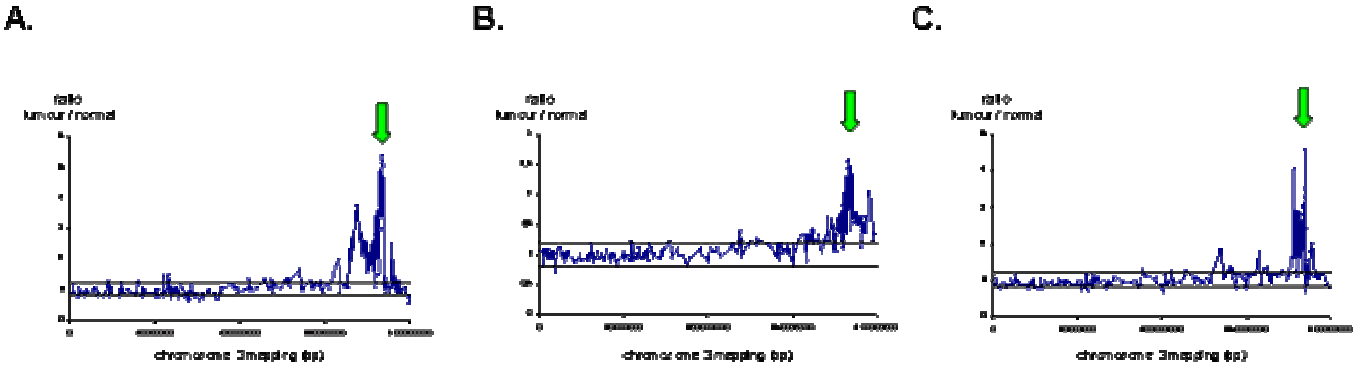


Figure 1. Examples of chromosome 3 array CGH profiles for 3 HNSCC with high-level amplifications at 3q26-qter.

Chromosome 3 array CGH profiles are included for HNSCC tumours #217, #331 and #677 (in panel A, B and C, respectively Horizontal lines represent the interval of normal copy number variations

High-level amplifications are clustered in the 3q26-qter region. These seem to include large genomic regions and three different sites among tumours. The amplification maxima is in each case located in the 182-185 Mb genomic segment, denoted by the green arrow.

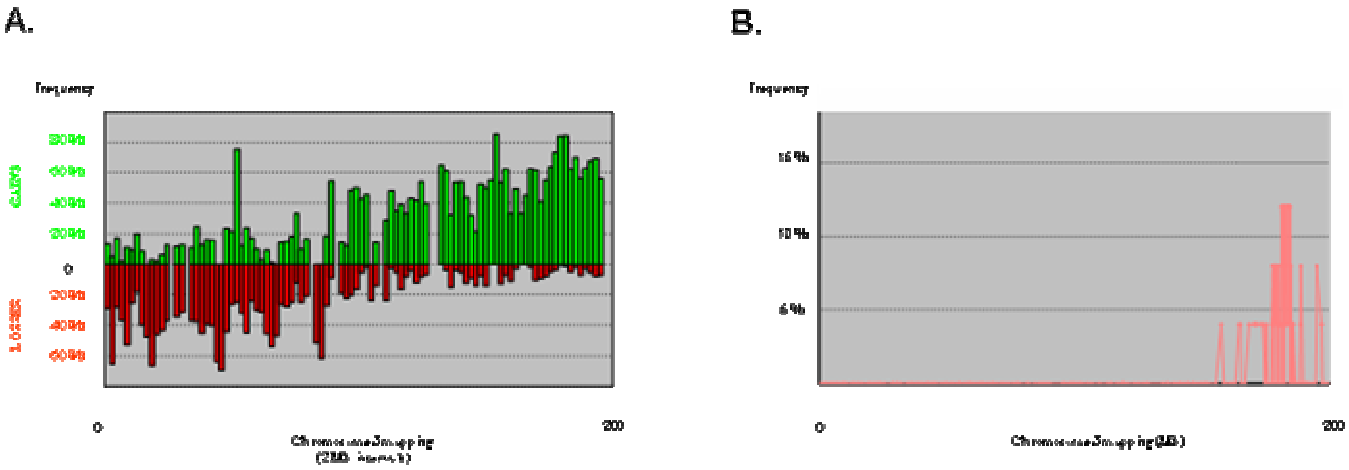


Figure 2. Frequency of chromosome 3 aberrations in 25 HNSCC.

A. Losses and Gains frequency for each 2Mb interval along chromosome 3. For each interval, frequencies were calculated by dividing the total number of occurrences of losses or gains observed in the series of 25 HNSCC by the total number of possible occurrences. They were expressed as percentages and represented along the chromosomal order of the corresponding classes.

B. Chromosome 3 high level amplifications frequency. A ratio superior than 2 (arbitrary threshold) in a single experiment was considered to represent a high level amplification for the individual clone. Frequencies were subsequently calculated for the series of HNSCC and plotted as a function of the localization of each clone. Individual data points are arbitrarily joined by a line.

High level amplifications are restricted to the 3q26-qter region .

d) CONCLUSIONS / PERSPECTIVES

Complex pattern of discontinuous deletions is observed on the 3p arm. Gain of the GMPS-CCNL1 region at 3q25.3 seems to be more frequent in this series of HNSCC when compared to lung SCC.

Amplifications seem to affect larger genomic region than in lung SCC. Individual narrow amplifications occur at least at three sites in the 3q26-qter region. Nevertheless, one of this three sites, is common to the three HNSCC tumours with amplifications and located around 183 Mb.

Interestingly, the common region of amplifications defined for lung SCC by our previous study corresponds to the 181.9 to 184.6 Mb inetrval. It is likely that amplified genes from this region participate to the SCC tumour phenotype of the two localizations. Detailed analysis of the datas is currently on going.

Part 3

General Discussion

The immediate perspectives of each individual study have been presented in the corresponding chapter and thus will not be exposed in this general discussion part.

Genomic instability is a hallmark of cancer cells. Among them, MIN (microsatellite instability) and CIN (chromosomal instability) are the best documented (Rajagopalan *et al*., 2004). Even if chromosomal instability role as causal or circumstantial in carcinogenesis is still controversial, most of the solid tumours depict chromosomal aberrations. Furthermore, recurrent chromosomal patterns are strongly associated with tumour type and stage (Gebhart and Liehr, 2000). They likely reflect Darwinian selection of genetic changes conferring proliferative advantage to cells carrying these alterations, among mainly randomly generated aberrations. Mapping of recurrent aberrations paves the way to the identification of DNA copy number biomarkers in relation to clinical features and to the discovery of cancer genes following a positional cloning strategy.

Comparative genomic hybridization (CGH) was the first efficient approach to scan the entire genome for segmental DNA copy number changes (Kallioniemi *et al.*, 1992; du Manoir *et al.*, 1993). Initial method on chromosomal spreads had a limited resolution to cytogenetic scale (a segmental deletion or gain about 5 to 10 Mb can be detected). To overcome this limitation, array-CGH also called matrix-CGH (made of DNA segments of known chromosomal location that are ordered on a slide) was introduced (Solinas-Toldo *et al.*, 1997; Pinkel *et al.*, 1998). Its resolution is basically only dependent of the size of the DNA segments present on the slide and represents a 100 to 1000 increase in resolution compared to CGH on chromosomes (Pinkel *et al.*, 2005). In the past five years, incremental technological progresses have lead to chromosomal dedicated arrays, genome wide arrays with a clone at every megabase and genome wide contig arrays. DNA fragments ranging from oligonucleotides to 100 Kb (BAC clones) have been employed and some of these arrays are now commercially available (Pinkel *et al.*, 2005). Accordingly, the mapping results can be rapidly translated in candidate cancer gene selection blurring the limits between this new cytogenetic method and molecular biology (Speicher and Carter, 2005).

In this thesis, we followed such a positional clonal strategy dedicated to malignant fibrous histiocytomas, lung and head & neck cancer squamous cell carcinomas. This strategy has uncovered some interesting candidate genes as the *cyclin L1* in head and neck SCC or a microRNA encoding gene in malignant fibrous histiocytomas.

Long CCNL1 protein isoform, which contains a cyclin box domain, seems clearly implicated in RNA splicing. Function of the classical short protein isoform is unknown (and does not comprise a cyclin box). A recent publication depicts an even more complicated situation, since a high number of RNA isoforms may exist (Loyer *et al.*, 2005). Possible role of CCNL1 in cell cycle regulation remains elusive. Finally, Sticht's study (2005) showing correlation between high level gene copy number of CCNL1 genomic region and survival in oral squamous carcinoma asks the question if this is a result of some CCNL1 (or isoform) protein overexpression or if a close neighbouring gene in this chromosomal segment may explain this result. GMPS gene coding for an enzyme participating to nucleotide synthesis, a process commonly unregulated as seen in tumor transcriptome data, is a good candidate even if our previous study reveals that it seems poorly submitted to gene dosage (Redon *et al.*, 2002).

The microRNA encoding gene, *hsa-mir28*, that was pinpointed by our characterization of an amplicon in the COMA cell line is another output of our positional cloning strategy. As outlined in the introduction part of this thesis, microRNA encoding genes belong to a new class of post-transcriptional regulators that may coordinate large number of genes in complex processes such as embryogenesis development or cancer pathogenesis. New data in this unexplored field are clearly scientifically exciting. Our result that mir28 may regulate some cell cycle related proteins and seems to be expressed differently during cell cycle phases is clearly innovative.

The part of this thesis which concerns amplification mapping at 3q26.3 in squamous carcinoma is probably significant towards the identification of a core process in squamous cell carcinogenesis. Because gains or amplifications of this long arm are one of the most common aberrations of these tumours and have been associated with pejorative outcome, the identification of major player oncogenes from this region may uncover gene pivotal for the development/progression of such tumours. In this sense, our work delineation of a 2 Mb amplified segment in some lung and head and neck SCC is a step forward toward this identification. Finally, we propose that the *SOX2* gene at 3q26.3 (among others in a 2Mb interval) may be important to lung and head & neck cancer squamous cell carcinogenesis.

Such positional cloning strategy relies on the assumption that biological effects of copy number are mediated by corresponding variations at the transcriptional level. Our knowledge about the consequences of low level gains has rapidly increased in the past years, consequently to the introduction of technologies allowing to analyze the whole transcriptome in DNA chip experiments. It appears that the transcriptome of cancer cells is largely modulated by the large genomic gains and losses found in tumours. Estimation of the percentage of genes submitted to gene dosage in these regions lies from 75 to 90%, in simple experimental system (a review of the original studies underlying this statement is contained in the introduction part). Accordingly, the gain of the long arm of the chromosome 3 in lung or HN SCC should upregulate slightly the expression of numerous genes. The biological consequences of such small variations may vary from gene to gene. Some genes, as it has been shown for *PIK3CA*, have notable phenotypic consequences even when submitted to slight variation. This gene, which maps to an interval that we found gained in 80 % of the lung SCC, is likely to be pivotal in squamous cell carcinoma as previously proposed by several authors. Another major player in squamous cell carcinoma initiation/progression is the p63 gene at 3q29, but its transcriptional level seems to be mainly regulated by other means than gene dosage (Redon *et al.*, 2001). Slight variations of the expression level of many other genes may be neutral if taken individually but could have cumulated effect if involved in common pathways.

Deducing the global biological output of the slight deregulation of numerous genes is clearly challenging. Analysis of transcriptome data in relation with gene copy dosage evaluation may be a way to tackle this problem. Accordingly, we are involved in a large-scale combined transcriptome/genomic study of 96 HNSCC in a project of Bohdan Wasylyk and Joseph Abecassis (Strasbourg) that will give new insights in this regulation: direct correlation of chromosome 3 gene expression and copy number is a reasonable initial strategy. We believe that activation of 'modules of expression" (Rhodes *et al.*, 2005; these modules are transcriptionally co-regulated sets of genes that are co-acting in specific biological functions) in groups of tumours selected for their content in a specific chromosomal aberration is even a more promising avenue to understand their biological output.

Experimentally, the effect of low level gene copy changes is difficult to address in cellular models and appropriate technological evolution to add or remove one chosen chromosomal segment in an otherwise balanced genome are clearly needed to progress towards the solving of this question. Such model may furthermore help to understand how peculiar DNA sequences (as fragile sites, presence of local homology or presence of sequences repeat) may participate to the mechanistic formation of these chromosomal aberrations (Shimizu *et al.*, 2001).

In addition, in our SCC study, more accentuated DNA copy changes, namely amplifications, were also found. From the pattern of the array CGH profile, they likely represent gene amplifications under the form of HSR or double minutes. A typical array CGH profile corresponding to dmin chromosomes is included in COMA cell line study associated with direct FISH evidence of dmin chromosomes. Effects of such high gene copy number changes are expected to be more accentuated at the transcriptional level.

In our work, most of the 9 genes in the 2Mb amplicon were found to be overexpressed. Amplification is a mean of rapid adaptative capability in response to increased selection pressure. An example of striking selection barrier is the oxygen and nutriment deprivations when the tumour reaches a size at which simple diffusion is not sufficient anymore to feed tumour cells. This results in massive hypoxia-induced apoptosis. SOX2 has been recently demonstrated as directly involved in this process. This proposed SOX2 role together with our data strongly suggests that its overexpression in tumour cells could protect them from hypoxia-induced apoptosis. Experimental *in vitro* assays of overexpression of this gene in cellular model under hypoxic conditions are obviously the next step to carry out.

Nevertheless, even if this hypoxia related hypothesis is our preferred one, SOX2 is also a main cofactor of transcription regulators in cooperation with other transcription factors and supposed to be stem cell markers. SOX2 is found fixed to several 5' upstream gene sequence but specific responsive genes are poorly described. Again, analysis of tumour transcriptome data, using Gene Ontology and 'modules of expression" may reveal if SOX2 belongs in different tumours to hypoxia related modules, stem cell related ones or other relevant functions for tumour biology.

Our current vision is that *Sox2* is the gene driver of the 3q26.3 amplifications that we described but that neighbour(s) gene(s) may have additive functions. Examples of such synergistically effect exist in the scientific literature (see Introduction). Taken together, our data and those from the literature suggest that the genes located at 3q26.3 that are recurrently amplified and consequently over-expressed in lung SCC may act as co-conspirators in these tumours through modulations of several tumour traits. The next step towards the elucidation of their respective contributions to the tumour phenotype when amplified and over-expressed relies on a functional analysis. This will be achieved experimentally through dedicated *in vitro* assays. This aims at analyzing the effects of candidate oncogene over-expression. We have sub-cloned the cDNAs encoded by these genes in a common mammalian expression vector and already performed set-up experiments to monitor the effects of their over-expression on tumour traits which are hallmarks of cancer.

Mapping of recurrent aberrations also paves the way to the identification of DNA copy number biomarkers in relation to clinical features. Identification of such biomarkers is now facilitated by array CGH and has begun to produce prognosis-related classification scheme for few cancer types (several publications, for example see Weiss *et al.*, 2004). Historically, for different regions of the 3q, gains have been coupled with a pejorative prognosis in SCC. New technological development as tissue microarrays should streamline the marker validation

process. Translation of a biomarker from research setting to clinical practice has been difficult. Robustness and low inter-operator variability of the technique is main issue for this process. DNA based biomarkers, assuming their sufficient informativity and by the relative stability of this molecule has an intrinsic advantage for the design of such test.

Part 4

Bibliography

Books

Cancer Cytogenetics, 2nd edition. Heim S, Mitelman F. Wiley-Liss Inc ©1995.

Cancer, Principles and Practice of Oncology, 6th Edition. DeVita VT, Hellman S, Rosenberg SA. LWW company ©2001.

The Metabolic & Molecular Bases of Inherited Disease, 8th edition. Scriver CR, Beaudet AL, Sly WS, Valle D. The McGraw-Hill Companies Inc ©2001.

Cited monographies

Abbott AL, Alvarez-Saavedra E, Miska EA, Lau NC, Bartel DP, Horvitz HR, Ambros V. (**2005**) The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans. *Dev Cell.,* 9:403-14.

Albertson DG, Ylstra B, Segraves R, Collins C, Dairkee SH, Kowbel D, Kuo WL, Gray JW, Pinkel D. (**2000**) Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nat Genet.,* 25:144-6.

Anisimov VN, Ukraintseva SV, Yashin AI. (**2005**) Cancer in rodents: does it tell us about cancer in humans? *Nat Rev Cancer,* 5:807-19.

Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. (**2004**) Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nat Rev Genet.,* 5:725-38.

Aslanian A, Iaquinta PJ, Verona R, Lees JA. (**2004**) Repression of the Arf tumor suppressor by E2F3 is required for normal cell cycle kinetics. *Genes Dev.,* 18:1413-22.

Attwooll C, Oddi S, Cartwright P, Prosperini E, Agger K, Steensgaard P, Wagener C, Sardet C, Moroni MC, Helin K. (**2005**) A novel repressive E2F6 complex containing the polycomb group protein, EPC1, that interacts with EZH2 in a proliferation-specific manner. *J Biol Chem.,* 280:1199-208.

Bagga S, Bracht J, Hunter S, Massirer K, Holtz J, Eachus R, Pasquinelli AE. (**2005**) Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell,* 122:553-63.

Baker SJ, Fearon ER, Nigro JM, Hamilton SR, Preisinger AC, Jessup JM, vanTuinen P, Ledbetter DH, Barker DF, Nakamura Y, White R, Vogelstein B. (**1989**) Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science,* 244:217-21.

Baker SJ, Markowitz S, Fearon ER, Willson JK, Vogelstein B. (**1990**) Suppression of human colorectal carcinoma cell growth by wild-type p53. *Science,* 249:912-5.

Balasubramanian S, Ahmad N, Mukhtar H. (**1999**) Upregulation of E2F transcription factors in chemically induced mouse skin tumors. *Int J Oncol.,* 15:387-90.

Baldwin C, Garnis C, Zhang L, Rosin MP, Lam WL. (**2005**) Multiple microalterations detected at high frequency in oral cancer. *Cancer Res.,* 65:7561-7.

Balmain A. (**2001**) Cancer genetics: from Boveri and Mendel to microarrays. *Nat Rev Cancer,*1:77-82.

Balsara BR, Testa JR. (**2002**) Chromosomal imbalances in human lung cancer. *Oncogene,* 21:6877-83.

Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, Flanagan A, Teague J, Futreal PA, Stratton MR, Wooster R. (**2004**) The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer.*, 91:355-8.

Bardelli A, Parsons DW, Silliman N, Ptak J, Szabo S, Saha S, Markowitz S, Willson JK, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE. (**2003**) Mutational analysis of the tyrosine kinome in colorectal cancers. *Science,* 300:949.

Barlund M, Monni O, Kononen J, Cornelison R, Torhorst J, Sauter G, Kallioniemi OLLI-P, Kallioniemi A. (**2000**) Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. *Cancer Res.,* 60:5340-4.

Bartel DP. (**2004**) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116:281-97.

Bartel DP, Chen CZ. (**2004**) Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet.,* 5:396-400;

Baskerville S, Bartel DP. (**2005**) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA,* 11:241-7.

Bauer C, Brass N, Diesinger I, Kayser K, Grasser FA, Meese E. (**2002**) Overexpression of the eukaryotic translation initiation factor 4G (eIF4G-1) in squamous cell lung carcinoma. *Int J Cancer.,* 98:181-5.

Bauman JG, Wiegant J, Borst P, van Duijn P. (**1980**) A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. *Exp Cell Res.,* 128:485-90.

Bembenek J, Yu H. (**2001**) Regulation of the anaphase-promoting complex by the dual specificity phosphatase human Cdc14a. *J Biol Chem.,* 276:48237-42

Benard J, Douc-Rasy S, Ahomadegbe JC. (**2003**) TP53 family members and human cancers. *Hum Mutat.,* 21:182-91.

Benner SE, Wahl GM, Von Hoff DD. (**1991**) Double minute chromosomes and homogeneously staining regions in tumors taken directly from patients versus in human tumor cell lines. *Anticancer Drugs,* 2:11-25.

Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E. (**2005**) Phylogenetic shadowing and computational identification of human microRNA genes. *Cell,* 120:21-4.

Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ, Meyerson M. (**2001**) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A.,* 2001 98:13790-5.

Biedler JL, Spengler BA. (**1976**) Metaphase chromosome anomaly: association with drug resistance and cell-specific products. *Science,* 191:185-7.

Biedler JL, Melera PW, Spengler BA. (**1980**) Specifically altered metaphase chromosomes in antifolate-resistant Chinese hamster cells that overproduce dihydrofolate reductase. *Cancer Genet Cytogenet.,* Volume 2, Issue 1, Pages 47-60 (June 1980)

Bignell G, Smith R, Hunter C, Stephens P, Davies H, Greenman C, Teague J, Butler A, Edkins S, Stevens C, O'meara S, Parker A, Avis T, Barthorpe S, Brackenbury L, Buck G, Clements J, Cole J, Dicks E, Edwards K, Forbes S, Gorton M, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jones D, Kosmidou V, Laman R, Lugg R, Menzies A, Perry J, Petty R, Raine K, Shepherd R, Small A, Solomon H, Stephens Y, Tofts C, Varian J, Webb A, West S, Widaa S, Yates A, Gillis AJ, Stoop HJ, van Gurp RJ, Oosterhuis JW, Looijenga LH, Futreal PA, Wooster R, Stratton MR. (**2006**) Sequence analysis of the protein kinase gene family in human testicular germ-cell tumors of adolescents and adults. *Genes Chromosomes Cancer,* 45:42-6.

Bishop JM. (**1990**) Nobel Lecture. Retroviruses and oncogenes II. *Biosci Rep.,* 10:473-91.

Black EP, Hallstrom T, Dressman HK, West M, Nevins JR. (**2005**) Distinctions in the specificity of E2F function revealed by gene expression signatures. *Proc Natl Acad Sci U S A.,* 102:15948-53.

Blaveri E, Brewer JL, Roydasgupta R, Fridlyand J, DeVries S, Koppie T, Pejavar S, Mehta K, Carroll P, Simko JP, Waldman FM. (**2005**) Bladder cancer stage and outcome by array-based comparative genomic hybridization. *Clin Cancer Res.,* 11:7012-22.

Blume-Jensen P, Hunter T. (**2001**) Oncogenic kinase signalling. *Nature,* 411:355-65.

Bockmuhl U, Schluns K, Kuchler I, Petersen S, Petersen I. (**2000**) Genetic imbalances with impact on survival in head and neck cancer patients. *Am J Pathol.,* 157:369-75.

Boehm JS, Hahn WC. (**2005**)Understanding transformation: progress and gaps. *Curr Opin Genet Dev.,* 15(1):13-7.

Boveri T. (**1914**) Zur Frage der Entstehung Maligner Tumoren. *Gustav Fischer, Jena*.
[English translation *The Origin of Malignant Tumors* by Boveri, M. (Williams and Wilkins Baltimore, 1929)]

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. (**2005**) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell,* 122:947-56.

Braga E, Senchenko V, Bazov I, Loginov W, Liu J, Ermilova V, Kazubskaya T, Garkavtseva R, Mazurenko N, Kisseljov F, Lerman MI, Klein G, Kisselev L, Zabarovsky ER. (**2002**) Critical tumor-suppressor gene regions on chromosome 3P in major human epithelial malignancies: allelotyping and quantitative real-time PCR. *Int J Cancer.,*100:534-41.

Brat DJ, Hahn SA, Griffin CA, Yeo CJ, Kern SE, Hruban RH. (**1997**) The structural basis of molecular genetic deletions. An integration of classical cytogenetic and molecular analyses in pancreatic adenocarcinoma. *Am J Pathol.,* 150:383-91.

Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM. (**1985**) Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science,* 224:1121-4.

Brown PC, Beverley SM, Schimke RT. (**1981**) Relationship of amplified dihydrofolate reductase genes to double minute chromosomes in unstably resistant mouse fibroblast cell lines. *Mol Cell Biol.,* 1:1077-83.

Bryan TM, Englezou A, Dalla-Pozza L, Dunham MA, Reddel RR. (**1997**) Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat Med.,* 3:1271-4.

Buckley PG, Jarbo C, Menzel U, Mathiesen T, Scott C, Gregory SG, Langford CF, Dumanski JP. (**2005**) Comprehensive DNA copy number profiling of meningioma using a chromosome 1 tiling path microarray identifies novel candidate tumor suppressor loci. *Cancer Res.,* 65:2653-61.

Bui TD, Beier DR, Jonssen M, Smith K, Dorrington SM, Kaklamanis L, Kearney L, Regan R, Sussman DJ, Harris AL.(**1997**) cDNA cloning of a human dishevelled DVL-3 gene, mapping to 3q27, and expression in human breast and colon carcinomas. *Biochem Biophys Res Commun.,* 239:510-6.

Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JK, Markowitz SD, Kinzler KW, Vogelstein B. (**1998**) Mutations of mitotic checkpoint genes in human cancers. *Nature,* 392:300-3.

Cahill DP, Kinzler KW, Vogelstein B, Lengauer C. (**1999**) Genetic instability and darwinian selection in tumours. *Trends Cell Biol.,* 9(12):M57-60.

Cahilly-Snyder L, Yang-Feng T, Francke U, George DL. (**1987**) Molecular analysis and chromosomal mapping of amplified genes isolated from a transformed mouse 3T3 cell line. *Somat Cell Mol Genet.,* 13:235-44.

Caldas C, Brenton JD. (**2005**) Sizing up miRNAs as cancer genes. *Nat Med.,* 11:712-4.

Califano J, van der Riet P, Westra W, Nawroz H, Clayman G, Piantadosi S, Corio R, Lee D, Greenberg B, Koch W, Sidransky D. (**1996**) Genetic progression model for head and neck cancer: implications for field cancerization. *Cancer Res.,* 56:2488-92.

Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, Bullrich F, Croce CM. (**2002**) Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A.,* 99:15524-9.

Calin GA, Liu CG, Sevignani C, Ferracin M, Felli N, Dumitru CD, Shimizu M, Cimmino A, Zupo S, Dono M, Dell'Aquila ML, Alder H, Rassenti L, Kipps TJ, Bullrich F, Negrini M, Croce CM. (**2004**) MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc Natl Acad Sci U S A.,* 101:11755-60.

Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, Iorio MV, Visone R, Sever NI, Fabbri M, Iuliano R, Palumbo T, Pichiorri F, Roldo C, Garzon R, Sevignani C, Rassenti L, Alder H, Volinia S, Liu CG, Kipps TJ, Negrini M, Croce CM. (**2005**) A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med.,* 353:1793-801.

Callagy G, Pharoah P, Chin SF, Sangan T, Daigo Y, Jackson L, Caldas C. (**2005**) Identification and validation of prognostic markers in breast cancer with the complementary use of array-CGH and tissue microarrays. *J Pathol.*, 205:388-96.

Caretti G, Salsi V, Vecchi C, Imbriano C, Mantovani R. (**2003**) Dynamic recruitment of NF-Y and histone acetyltransferases on cell-cycle promoters. *J Biol Chem.,* 278:30435-40

Cartwright P, Muller H, Wagener C, Holm K, Helin K. (**1998**) E2F-6: a novel member of the E2F family is an inhibitor of E2F-dependent transcription. *Oncogene,* 17:611-23.

Cavenee WK, Dryja TP, Phillips RA, Benedict WF, Godbout R, Gallie BL, Murphree AL, Strong LC, White RL. (**1983**) Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature,* 305:779-84.

Chen CZ. (**2005**) MicroRNAs as oncogenes and tumor suppressors. *N Engl J Med.,* 353:1768-71.

Cheng KW, Lahad JP, Kuo WL, Lapuk A, Yamada K, Auersperg N, Liu J, Smith-McCune K, Lu KH, Fishman D, Gray JW, Mills GB. (**2004**) The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers. *Nat Med.,* 10:1251-6.

Chin K, de Solorzano CO, Knowles D, Jones A, Chou W, Rodriguez EG, Kuo WL, Ljung BM, Chew K, Myambo K, Miranda M, Krig S, Garbe J, Stampfer M, Yaswen P, Gray JW, Lockett SJ. (**2004**) In situ analyses of genome instability in breast cancer. *Nat Genet.,* 36:984-8.

Cho HP, Liu Y, Gomez M, Dunlap J, Tyers M, Wang Y. (**2005**) The dual-specificity phosphatase CDC14B bundles and stabilizes microtubules. *Mol Cell Biol.,* 25:4541-51.

Christensen J, Cloos P, Toftegaard U, Klinkenberg D, Bracken AP, Trinh E, Heeran M, Di Stefano L, Helin K. (**2005**) Characterization of E2F8, a novel E2F-like cell-cycle regulated repressor of E2F-activated transcription. *Nucleic Acids Res.,* 33:5458-70.

Cimini D, Howell B, Maddox P, Khodjakov A, Degrassi F, Salmon ED. (**2001**) Merotelic kinetochore orientation is a major mechanism of aneuploidy in mitotic mammalian tissue cells. *J Cell Biol.,* 153:517-27.

Cleaver JE. (**1968**) Defective repair replication of DNA in xeroderma pigmentosum. *Nature,* 218:652-6.

Cleaver JE. (**2005**) Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat Rev Cancer,* 5:564-73.

Clemens MJ, Bushell M, Morley SJ. (**1998**) Degradation of eukaryotic polypeptide chain initiation factor (eIF) 4G in response to induction of apoptosis in human lymphoma cell lines. *Oncogene,* 17:2921-31.

Cole LJ, Nowell PC. (**1965**) Radiation carcinogenesis: the sequence of events. *Science,* 150:1782-6.

Collins C, Volik S, Kowbel D, Ginzinger D, Ylstra B, Cloutier T, Hawkins T, Predki P, Martin C, Wernick M, Kuo WL, Alberts A, Gray JW. (**2001**) Comprehensive genome sequence analysis of a breast cancer amplicon.
Genome Res., 11:1034-42.

Collins Y, Tan DF, Pejovic T, Mor G, Qian F, Rutherford T, Varma R, McQuaid D, Driscoll D, Jiang M, Deeb G, Lele S, Nowak N, Odunsi K. (**2004**) Identification of differentially expressed genes in clinically distinct groups of serous ovarian carcinomas using cDNA microarray. *Int J Mol Med.,* 14:43-53.

Coquelle A, Rozier L, Dutrillaux B, Debatisse M. (**2002**) Induction of multiple double-strand breaks within an hsr by meganucleaseI-SceI expression or fragile site activation leads to formation of double minutes and other chromosomal rearrangements. *Oncogene*, 21:7671-9.

Crawley JJ, Furge KA. (**2002**) Identification of frequent cytogenetic aberrations in hepatocellular carcinoma using gene-expression microarray data. *Genome Biol.,* 3:RESEARCH0075.

Cromer A, Carles A, Millon R, Ganguli G, Chalmel F, Lemaire F, Young J, Dembele D, Thibault C, Muller D, Poch O, Abecassis J, Wasylyk B. (**2004**) Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis. *Oncogene,* 23:2484-98.

Dammann R**,** Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP. (**2000**) Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. *Nat Genet.,* 25:315-9.

Davies H, Hunter C, Smith R, Stephens P, Greenman C, Bignell G, Teague J, Butler A, Edkins S, Stevens C, Parker A, O'Meara S, Avis T, Barthorpe S, Brackenbury L, Buck G, Clements J, Cole J, Dicks E, Edwards K, Forbes S, Gorton M, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jones D, Kosmidou V, Laman R, Lugg R, Menzies A, Perry J, Petty R, Raine K, Shepherd R, Small A, Solomon H, Stephens Y, Tofts C, Varian J, Webb A, West S, Widaa S, Yates A, Brasseur F, Cooper CS, Flanagan AM, Green A, Knowles M, Leung SY, Looijenga LH, Malkowicz B, Pierotti MA, Teh BT, Yuen ST, Lakhani SR, Easton DF, Weber BL, Goldstraw P, Nicholson AG, Wooster R, Stratton MR, Futreal PA. (**2005**) Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer Res.,* 65:7591-5

de Bruin A, Maiti B, Jakoi L, Timmers C, Buerki R, Leone G. (**2003**) Identification and characterization of E2F7, a novel mammalian E2F family member capable of blocking cellular proliferation. *J Biol Chem.,* 278:42041-9.

DeGregori J. (**2002**) The genetics of the E2F family of transcription factors: shared functions and unique roles. *Biochim Biophys Acta.,* 1602:131-50.

de Padula M, Slezak G, Auffret van Der Kemp P, Boiteux S. (**2004**) The post-replication repair RAD18 and RAD6 genes are involved in the prevention of spontaneous mutations caused by 7,8-dihydro-8-oxoguanine in Saccharomyces cerevisiae. *Nucleic Acids Res.,* 32:5003-10.

De Weerd-Kastelein EA, Keijzer W, Bootsma D. (**1972**) Genetic heterogeneity of xeroderma pigmentosum demonstrated by somatic cell hybridization. *Nat New Biol.,* 238:80-3.

Dimova DK, Dyson NJ. (**2005**) The E2F transcriptional network: old acquaintances with new faces. *Oncogene,* 24:2810-26.

Di Stefano L, Jensen MR, Helin K. (**2003**) E2F7, a novel E2F featuring DP-independent repression of a subset of E2F-regulated genes. *EMBO J.*, 22:6289-98.

Dobie KW, Kennedy CD, Velasco VM, McGrath TL, Weko J, Patterson RW, Karpen GH.(**2001**) Identification of chromosome inheritance modifiers in Drosophila melanogaster. *Genetics ,* 157:1623-37

Dolnick BJ, Berenson RJ, Bertino JR, Kaufman RJ, Nunberg JH, Schimke RT. (**1979**) Correlation of dihydrofolate reductase elevation with gene amplification in a homogeneously staining chromosomal region in L5178Y cells. *J Cell Biol.,* 83:394-402.

Dong SM, Sun DI, Benoit NE, Kuzmin I, Lerman MI, Sidransky D. (**2003**) Epigenetic inactivation of RASSF1A in head and neck cancer. *Clin Cancer Res.,* 9:3635-40.

Draviam VM, Xie S, Sorger PK. (**2004**) Chromosome segregation and genomic stability. *Curr Opin Genet Dev.,* 14:120-5.

Ducat D, Zheng Y. (**2004**) Aurora kinases in spindle assembly and chromosome segregation. *Exp Cell Res.,* 301(1):60-7.

Duesberg P, Rasnick D. (**2000**) Aneuploidy, the somatic mutation that makes cancer a species of its own. *Cell Motil Cytoskeleton,* 47:81-107.

Duesberg P, Fabarius A, Hehlmann R. (**2004**) Aneuploidy, the primary cause of the multilateral genomic instability of neoplastic and preneoplastic cells. *IUBMB Life,* 56:65-81.

du Manoir S, Speicher MR, Joos S, Schrock E, Popp S, Dohner H, Kovacs G, Robert-Nicoud M, Lichter P, Cremer T. (**1993**) Detection of complete and partial chromosome gains and losses by comparative genomic in situ hybridization. *Hum Genet.,* 90:590-610.

Elahi A, Zheng Z, Park J, Eyring K, McCaffrey T, Lazarus P. (**2002**) The human OGG1 DNA repair enzyme and its association with orolaryngeal cancer risk. *Carcinogenesis,* ;23:1229-34.

Eis PS, Tam W, Sun L, Chadburn A, Li Z, Gomez MF, Lund E, Dahlberg JE. (**2005**) Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proc Natl Acad Sci U S A.,* 102:3627-32.

Estilo CL, O-Charoenrat P, Ngai I, Patel SG, Reddy PG, Dao S, Shaha AR, Kraus DH, Boyle JO, Wong RJ, Pfister DG, Huryn JM, Zlotolow IM, Shah JP, Singh B. (**2003**) The role of novel oncogenes squamous cell carcinoma-related oncogene and phosphatidylinositol 3-kinase p110alpha in squamous cell carcinoma of the oral tongue. *Clin Cancer Res.,* 9:2300-6.

Feber A, Clark J, Goodwin G, Dodson AR, Smith PH, Fletcher A, Edwards S, Flohr P, Falconer A, Roe T, Kovacs G, Dennis N, Fisher C, Wooster R, Huddart R, Foster CS, Cooper CS. (**2004**) Amplification and overexpression of E2F3 in human bladder cancer. *Oncogene,* 23:1627-30.

Feldser DM, Hackett JA, Greider CW. (**2003**) Telomere dysfunction and the initiation of genome instability. *Nat Rev Cancer,* 3:623-7.

Finlay CA. (**1993**) The mdm-2 oncogene can overcome wild-type p53 suppression of transformed cell growth.
Mol Cell Biol., 13:301-6.

Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R. (**1993**) The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell,* 75:1027-38.
Erratum in: *Cell,* 77:167.

FitzPatrick DR. (**2005**). Transcriptional consequences of autosomal trisomy: primary gene dosage with complex downstream effects. *Trends Genet.,* 21:249-53.

Fodde R, Kuipers J, Rosenberg C, Smits R, Kielman M, Gaspar C, van Es JH, Breukel C, Wiegant J, Giles RH, Clevers H. (**2001a**) Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nat Cell Biol.,* 3:433-8.

Fodde R, Smits R, Clevers H. (**2001b**) APC, signal transduction and genetic instability in colorectal cancer. *Nat Rev Cancer,* 1:55-67.

Fodde R. (**2002**) The APC gene in colorectal cancer. *Eur J Cancer,* 38:867-71.

Fodde R. (**2003**) The multiple functions of tumour suppressors: it's all in APC. *Nat Cell Biol.,* 5:190-2.

Forsyth NR, Wright WE, Shay JW. (**2002**) Telomerase and differentiation in multicellular organisms: turn it off, turn it on, and turn it off again. *Differentiation,* 69:188-97;

Foster CS, Falconer A, Dodson AR, Norman AR, Dennis N, Fletcher A, Southgate C, Dowe A, Dearnaley D, Jhavar S, Eeles R, Feber A, Cooper CS. (**2004**) Transcription factor E2F3 overexpressed in prostate cancer independently predicts clinical outcome. *Oncogene,* 23:5871-9.

Francke U, Kung F. (**1976**) Sporadic bilateral retinoblastoma and 13q- chromosomal deletion. *Med Pediatr Oncol.,* 2:379-85.

Freier K, Sticht C, Hofele C, Flechtenmacher C, Stange D, Puccio L, Toedt G, Radlwimmer B, Lichter P, Joos S. (**2005**) Recurrent coamplification of cytoskeleton-associated genes EMS1 and SHANK2 with CCND1 in oral squamous cell carcinoma. *Genes Chromosomes Cancer,* 2005 advance online publication.

Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, Dryja TP. (**1986**) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature,* 323:643-6.

Fritz B, Schubert F, Wrobel G, Schwaenen C, Wessendorf S, Nessling M, Korz C, Rieker RJ, Montgomery K, Kucherlapati R, Mechtersheimer G, Eils R, Joos S, Lichter P. (**2002**) Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. *Cancer Res.,* 62:2993-8.

Fukuchi-Shimogori T, Ishii I, Kashiwagi K, Mashiba H, Ekimoto H, Igarashi K. (**1997**) Malignant transformation by overproduction of translation initiation factor eIF4G. *Cancer Res.,* 57:5041-4.

Fung TK, Poon RY. (**2005**) A roller coaster ride with the mitotic cyclins. *Semin Cell Dev Biol*., 16:335-42

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. (**2004**) A census of human cancer genes. *Nat Rev Cancer,* 4:177-83.

Gall JG, Pardue ML. (**1969**) Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci U S A.,* 63:378-83.

Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de Rijn M, Rosen GD, Perou CM, Whyte RI, Altman RB, Brown PO, Botstein D, Petersen I. (**2001**) Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A.,* 98:13784-9.
Erratum in: *Proc Natl Acad Sci U S A,* 99:1098.

Garnis C, Lockwood WW, Vucic E, Ge Y, Girard L, Minna JD, Gazdar AF, Lam S, Macaulay C, Lam WL. (**2005**) High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH.
Int J Cancer, advance online publication.

Gebhart E, Liehr T. (**2000**) Patterns of genomic imbalances in human solid tumors (Review). *Int J Oncol.,* 16:383-99.

Gene Ontology Consortium. (**2001**) Creating the gene ontology resource: design and implementation. Genome Res., 11:1425-33.

Geymonat M, Jensen S, Johnston LH. (**2002**) Mitotic exit: the Cdc14 double cross. *Curr Biol.,* 12:R482-84.

Giangrande PH, Zhu W, Schlisio S, Sun X, Mori S, Gaubatz S, Nevins JR. (**2004**) A role for E2F6 in distinguishing G1/S- and G2/M-specific transcription. *Genes Dev.,* 18:2941-51.

Giaretti W, Venesio T, Prevosto C, Lombardo F, Ceccarelli J, Molinu S, Risio M. (**2004**) Chromosomal instability and APC gene mutations in human sporadic colorectal adenomas*. J Pathol.,* 204:193-9.

Gidekel S, Pizov G, Bergman Y, Pikarsky E. (**2003**) Oct-3/4 is a dose-dependent oncogenic fate determinant. *Cancer Cell,* 4:361-70.

Girling R, Partridge JF, Bandara LR, Burden N, Totty NF, Hsuan JJ, La Thangue NB. (**1993**) A new component of the transcription factor DRTF1/E2F. *Nature,* 362:83-7. Erratum in: *Nature,* 365:468.

Gisselsson D. (**2003**) Chromosome instability in cancer: how, when, and why? *Adv Cancer Res.,* 87:1-29.

Gollin SM. (**2001**) Chromosomal alterations in squamous cell carcinomas of the head and neck: window to the biology of disease. *Head Neck.,* 23:238-53.

Green RA, Kaplan KB. (**2003**) Chromosome instability in colorectal tumor cells is associated with defects in microtubule plus-end attachments caused by a dominant mutation in APC. *J Cell Biol.,* 163:949-61.

Griffiths-Jones S. (**2004**) The microRNA Registry. *Nucleic Acids Res.,* 32 (Database issue):D109-11.

Groffen J, Stephenson JR, Heisterkamp N, Bartram C, de Klein A, Grosveld G. (**1984**) The human c-abl oncogene in the Philadelphia translocation. *J Cell Physiol Suppl.,* 3:179-91.

Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, Weinberg RA. (**1999**) Creation of human tumour cells with defined genetic elements. *Nature,* 400:464-8.

Hahn WC, Weinberg RA. (**2002a**) Modelling the molecular circuitry of cancer. *Nat Rev Cancer,* 2:331-41.

Hahn WC, Weinberg RA. (**2002b**) Rules for making human tumor cells. *N Engl J Med.,* 347:1593-603.
Erratum in: *N Engl J Med.,* 348:674.

Hanahan D, Weinberg RA. (**2000**) The hallmarks of cancer. *Cell,* 100:57-70.

Hanash S. (**2004**) Integrated global profiling of cancer. *Nat Rev Cancer,* 4:638-44.

Harfe BD. (**2005**) MicroRNAs in vertebrate development. *Curr Opin Genet Dev.,* 15:410-5.

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R; Gene Ontology Consortium. (**2004**) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.,* 32(Database issue):D258-61

Harris SL, Levine AJ. (**2005**) The p53 pathway: positive and negative feedback loops. *Oncogene,* 24:2899-908.

Hatfield SD, Shcherbata HR, Fischer KA, Nakahara K, Carthew RW, Ruohola-Baker H. (**2005**) Stem cell division is regulated by the microRNA pathway. *Nature,* 435:974-8;

Hayashita Y, Osada H, Tatematsu Y, Yamada H, Yanagisawa K, Tomida S, Yatabe Y, Kawahara K, Sekido Y, Takahashi T. (**2005**) A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res.,* 65:9628-32.

He L, Hannon GJ. (**2004**) MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet.,* 5:522-31.

He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM. (**2005**) A microRNA polycistron as a potential human oncogene. *Nature,* 435:828-33.

Heid CA, Stevens J, Livak KJ, Williams PM. (**1996**) Real time quantitative PCR. *Genome Res.,* 6:986-94.

Helin K, Wu CL, Fattaey AR, Lees JA, Dynlacht BD, Ngwu C, Harlow E. (**1993**) Heterodimerization of the transcription factors E2F-1 and DP-1 leads to cooperative trans-activation. *Genes Dev.,* 7:1850-61.

Hellman A, Zlotorynski E, Scherer SW, Cheung J, Vincent JB, Smith DI, Trakhtenbrot L, Kerem B. (**2002**) A role for common fragile site induction in amplification of human oncogenes. Cancer Cell, 1:89-97.

Hermsen M, Guervos MA, Meijer G, Baak J, van Diest P, Marcos CA, Sampedro A. (**2001**) New chromosomal regions with high-level amplifications in squamous cell carcinomas of the larynx and pharynx, identified by comparative genomic hybridization. *J Pathol.,* 194:177-82.

Hermsen M, Alonso Guervos M, Meijer G, van Diest P, Suarez Nieto C, Marcos CA, Sampedro A. (**2005**) Chromosomal changes in relation to clinical outcome in larynx and pharynx squamous cell carcinoma*. Cell Oncol.,* 27:191-8.

Higuchi T, Uhlmann F. (**2005**) Stabilization of microtubule dynamics at anaphase onset promotes chromosome segregation. *Nature,* 433:171-6.

Hogg RP**,** Honorio S, Martinez A, Agathanggelou A, Dallol A, Fullwood P, Weichselbaum R, Kuo MJ, Maher ER, Latif F. (**2002**) Frequent 3p allele loss and epigenetic inactivation of the RASSF1A tumour suppressor gene from region 3p21.3 in head and neck squamous cell carcinoma. *Eur J Cancer,* 38:1585-92.

Hollander MC, Philburn RT, Patterson AD, Velasco-Miguel S, Friedberg EC, Linnoila RI, Fornace AJ Jr. (**2005**) Deletion of XPC leads to lung tumors in mice and is associated with early events in human lung carcinogenesis. *Proc Natl Acad Sci U S A.,* 102:13200-5.

Houbaviy HB, Murray MF, Sharp PA. (**2003**) Embryonic stem cell-specific MicroRNAs. *Dev Cell.,* 5:351-8.

Hua VY, Wang WK, Duesberg PH. (**1997**) Dominant transformation by mutated human ras genes in vitro requires more than 100 times higher expression than is observed in cancers. *Proc Natl Acad Sci U S A.,* 94(18):9614-9.

Huang X, Gollin SM, Raja S, Godfrey TE. (**2002**) High-resolution mapping of the 11q13 amplicon and identification of a gene, TAOS1, that is amplified and overexpressed in oral cancer cells. Proc *Natl Acad Sci USA*, 99:11369-74.

Huebner RJ, Todaro GJ. (**1969**) Oncogenes of RNA tumor viruses as determinants of cancer. *Proc Natl Acad Sci U S A,* 64:1087-94.

Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburtty K, Simon J, Bard M, Friend SH. (**2000a**) Functional discovery via a compendium of expression profiles. *Cell,* 102:109-26.

Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ, Friend SH, Marton MJ. (**2000b**) Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat Genet.,* 25:333-7.

Humbert PO, Verona R, Trimarchi JM, Rogers C, Dandapani S, Lees JA. (**2000**) E2f3 is critical for normal cellular proliferation. *Genes Dev.,* 14:690-703.

Hunter KD, Parkinson EK, Harrison PR. (**2005**) Profiling early head and neck cancer. *Nat Rev Cancer.,* 5:127-35.

Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, Ringner M, Sauter G, Monni O, Elkahloun A, Kallioniemi OP, Kallioniemi A. (**2002**) Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res.,* 62:6240-5.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. (**2004**) Detection of large-scale variation in the human genome. *Nat Genet.,* 36:949-51.

Imoto I, Pimkhaokham A, Fukuda Y, Yang ZQ, Shimada Y, Nomura N, Hirai H, Imamura M, Inazawa J.(**2001**) SNO is a probable target for gene amplification at 3q26 in squamous-cell carcinomas of the esophagus. *Biochem Biophys Res Commun.,* 286:559-65.

International Human Genome Sequencing Consortium.
Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ. (**2001**) Initial sequencing and analysis of the human genome. *Nature,* 409:860-921.
Erratum in: *Nature*, 412:565. *Nature*, 411:720. Szustakowki, J [corrected to Szustakowski, J]

Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. (**1993**) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature,* 363:558-61.

Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M, Menard S, Palazzo JP, Rosenberg A, Musiani P, Volinia S, Nenci I, Calin GA, Querzoli P, Negrini M, Croce CM. (**2005**) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.,* 65:7065-70.

Ishida H, Masuhiro Y, Fukushima A, Argueta JG, Yamaguchi N, Shiota S, Hanazawa S. (**2005**) Identification and characterization of novel isoforms of human DP-1: DP-1{alpha} regulates the transcriptional activity of E2F1 as well as cell cycle progression in a dominant-negative manner. *J Biol Chem.,* 280:24642-8.

Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, Ling V, MacAulay C, Lam WL. (**2004**) A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet.,* 36:299-303.

Ivey-Hoyle M, Conroy R, Huber HE, Goodhart PJ, Oliff A, Heimbrook DC. (**1993**) Cloning and characterization of E2F-2, a novel protein with the biochemical properties of transcription factor E2F. *Mol Cell Biol.,* 13:7802-12.

.

Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JC, Trent JM, Staudt LM, Hudson J Jr, Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO. (**1999**) The transcriptional program in the response of human fibroblasts to serum. *Science,* 283:83-7

Jiang J, Lee EJ, Gusev Y, Schmittgen TD. (**2005**) Real-time expression profiling of microRNA precursors in human cancer cell lines. *Nucleic Acids Res.,* 33:5394-403.

Jin Y, Jin C, Salemark L, Martins C, Wennerberg J, Mertens F. (**2000**) Centromere cleavage is a mechanism underlying isochromosome formation in skin and head and neck carcinomas. *Chromosoma,* 109:476-81;

John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. (**2004**) Human MicroRNA targets. *PLoS Biol.,* 2:e363

Johnson DG, Cress WD, Jakoi L, Nevins JR. (**1994**) Oncogenic capacity of the E2F1 gene. *Proc Natl Acad Sci U S A.,* 91:12823-7.

Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ. (**2005**) RAS is regulated by the let-7 microRNA family. *Cell,* 120(5):635-47.

Jooss K, Lam EW, Bybee A, Girling R, Muller R, La Thangue NB. (**1995**) Proto-oncogenic properties of the DP family of proteins. *Oncogene,* 10:1529-36.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. (**1992**) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science,* 258:818-21.

Kanao H, Enomoto T, Kimura T, Fujita M, Nakashima R, Ueda Y, Ueno Y, Miyatake T, Yoshizaki T, Buzard GS, Tanigami A, Yoshino K, Murata Y. (**2005**) Overexpression of LAMP3/TSC403/DC-LAMP promotes metastasis in uterine cervical cancer. *Cancer Res.,* 65:8640-5.

Kang S, Bader AG, Vogt PK. (**2005**) Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. *Proc Natl Acad Sci U S A.,* 102:802-7.

Kaplan KB, Burds AA, Swedlow JR, Bekir SS, Sorger PK, Nathke IS. (**2001**) A role for the Adenomatous Polyposis Coli protein in chromosome segregation. *Nat Cell Biol.,* 3:429-32.

Karamouzis MV, Sotiropoulou-Bonikou G, Vandoros G, Varakis I, Papavassiliou AG. (**2004**) Differential expression of retinoic acid receptor beta (RARbeta) and the AP-1 transcription factor in normal, premalignant and malignant human laryngeal tissues. *Eur J Cancer.,* 40:761-73.

Kastan MB, Bartek J. (**2004**) Cell-cycle checkpoints and cancer. *Nature,* 432:316-23.

Katoh H, Shibata T, Kokubu A, Ojima H, Loukopoulos P, Kanai Y, Kosuge T, Fukayama M, Kondo T, Sakamoto M, Hosoda F, Ohki M, Imoto I, Inazawa J, Hirohashi S. (**2005**) Genetic profile of hepatocellular carcinoma revealed by array-based comparative genomic hybridization: Identification of genetic indicators to predict patient outcome. J Hepatol. 43:863-74.

Kaufman RJ, Brown PC, Schimke RT. (**1979**) Amplified dihydrofolate reductase genes in unstably methotrexate-resistant cells are associated with double minute chromosomes. *Proc Natl Acad Sci U S A.,* 76:5669-73.

Kaufman RJ, Brown PC, Schimke RT. (**1981**) Loss and stabilization of amplified dihydrofolate reductase genes in mouse sarcoma S-180 cell lines. *Mol Cell Biol.,* 1:1084-93.

Kauraniemi P, Barlund M, Monni O, Kallioniemi A. (**2001**) New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res.,* 61:8235-40.

Kim HS, Brill SJ. (**2001**) Rfc4 interacts with Rpa1 and is required for both DNA replication and DNA damage checkpoints in Saccharomyces cerevisiae. *Mol Cell Biol.,* 21:3725-37.

Kim WY, Kaelin WG. (**2004**) Role of VHL gene mutation in human cancer. *J Clin Oncol.,* 22:4991-5004.

Kimura M, Uchida C, Takano Y, Kitagawa M, Okano Y. (**2004**) Cell cycle-dependent regulation of the human aurora B promoter. *Biochem Biophys Res Commun.,* 316:930-6.

Kinzler KW, Vogelstein B. (**1997**) Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature,* 386:761, 763.

Klein CA, Schmidt-Kittler O, Schardt JA, Pantel K, Speicher MR, Riethmuller G. (**1999**) Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl Acad Sci U S A,.* 96(8):4494-9.

Knudson AG Jr. (**1971**) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A.,* 68:820-3.

Knudson AG Jr, Meadows AT, Nichols WW, Hill R. (**1976**) Chromosomal deletion and retinoblastoma. N *Engl J Med.*, 295:1120-3.

Knudson AG. (**2001**) Two genetic hits (more or less) to cancer. *Nat Rev Cancer*, 1:157-62.

Ko MA, Rosario CO, Hudson JW, Kulkarni S, Pollett A, Dennis JW, Swallow CJ. (**2005**) Plk4 haploinsufficiency causes mitotic infidelity and carcinogenesis. *Nat Genet.,* 37:883-8.

Kohl NE, Kanda N, Schreck RR, Bruns G, Latt SA, Gilbert F, Alt FW. (**1983**) Transposition and amplification of oncogene-related sequences in human neuroblastomas. *Cell,* 35:359-67.

Kolodner RD, Putnam CD, Myung K. (**2002**) Maintenance of genome stability in Saccharomyces cerevisiae. *Science,* 297:552-7.

Komarova NL, Wodarz D. (**2004**) The optimal rate of chromosome loss for the inactivation of tumor suppressor genes in cancer. *Proc Natl Acad Sci U S A.,* 101:7017-21.

Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP. (**1998**) Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med.,* 4:844-7.

Kops GJ, Weaver BA, Cleveland DW. (**2005**) On the road to cancer: aneuploidy and the mitotic checkpoint. *Nat Rev Cancer,* 5:773-85.

Kosugi S, Ohashi Y. (**2002**) E2Ls, E2F-like repressors of Arabidopsis that bind to E2F sites in a monomeric form. *J Biol Chem.,* 277:16553-8.

Krause SA, Loupart ML, Vass S, Schoenfelder S, Harrison S, Heck MM. (**2001**) Loss of cell cycle checkpoint control in Drosophila Rfc4 mutants. *Mol Cell Biol.,* 21:5156-68.

Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N. (**2005**) Combinatorial microRNA target predictions. *Nat Genet.,* 37:495-500.

Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, Stoffel M. (**2005**) Silencing of microRNAs in vivo with 'antagomirs'. *Nature,* advance online publication

Kuroki T, Trapasso F, Yendamuri S, Matsuyama A, Alder H, Mori M, Croce CM. (**2003**) Promoter hypermethylation of RASSF1A in esophageal squamous cell carcinoma. *Clin Cancer Res.,* 9:1441-5.

Kurz T, Ozlu N, Rudolf F, O'Rourke SM, Luke B, Hofmann K, Hyman AA, Bowerman B, Peter M. (**2005**) The conserved protein DCN-1/Dcn1p is required for cullin neddylation in C. elegans and S. cerevisiae. *Nature,* 2005 435:1257-61.

Kuwahara Y, Tanabe C, Ikeuchi T, Aoyagi K, Nishigaki M, Sakamoto H, Hoshinaga K, Yoshida T, Sasaki H, Terada M. (**2004**) Alternative mechanisms of gene amplification in human cancers. Genes Chromosomes Cancer. 41:125-32.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. (**2001**) Identification of novel genes coding for small expressed RNAs. *Science,* 294:853-8.

Land H, Parada LF, Weinberg RA. (**1983**) Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature,* 304:596-602.

Lau NC, Lim LP, Weinstein EG, Bartel DP. (**2001**) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science,* 294:858-62.

Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomaki P, Sistonen P, Aaltonen LA, Nystrom-Lahti M, *et al.* (**1993**) Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell,* 75:1215-25.

Lee EY, Cam H, Ziebold U, Rayman JB, Lees JA, Dynlacht BD. (**2002**) E2F4 loss suppresses tumorigenesis in Rb mutant mice. *Cancer Cell,* 2:463-72.

Lee RC, Ambros V. (**2001**) An extensive class of small RNAs in Caenorhabditis elegans. *Science,* 294:862-4.

Lee RC, Feinbaum RL, Ambros V. (**1993**) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell,* 75:843-54.

Lee WH, Murphree AL, Benedict WF. (**1984**) Expression and amplification of the N-myc gene in primary retinoblastoma. *Nature,* 309:458-60.

Lei JX, Liu QY, Sodja C, LeBlanc J, Ribecco-Lutkiewicz M, Smith B, Charlebois C, Walker PR, Sikorska M. (**2005**) S/MAR-binding properties of Sox2 and its involvement in apoptosis of human NT2 neural precursors. *Cell Death Differ.,* 12:1368-77.

Lengauer C**,** Kinzler KW, Vogelstein B. (**1997**) Genetic instability in colorectal cancers. *Nature,* 386:623-7.

Lengauer C, Kinzler KW, Vogelstein B. (**1998**) Genetic instabilities in human cancers. *Nature,* 396:643-9.

Levine AJ. (**1997**) p53, the cellular gatekeeper for growth and division. *Cell,* 88:323-31.

Lewis BP, Burge CB, Bartel DP. (**2005**) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell,* 120:15-20.

Li R, Sonik A, Stindl R, Rasnick D, Duesberg P. (**2000**) Aneuploidy vs. gene mutation hypothesis of cancer: recent study claims mutation but is found to support aneuploidy. *Proc Natl Acad Sci U S A.,* 97:3236-41.

Li R, Rasnick D, Duesberg P. (**2002**) Correspondence re: D. Zimonjic et al., Derivation of human tumor cells in vitro without widespread genomic instability. Cancer Res., 61: 8838-8844, 2001. *Cancer Res.,* 62:6345-8; author reply 6348-9.
Erratum in: *Cancer Res,* 62:7380.

Liu LX, Jiang HC, Liu ZH, Zhu AL, Zhang WH, Wu LF, Zhou J, Wang XQ, Wu M. (**2003**) Expression of cell cycle/growth regulator genes in human hepatocellular carcinoma and adjacent normal liver tissues. *Oncol Rep.,* 10:1771-5.

Lin SC, Liu CJ, Ko SY, Chang HC, Liu TY, Chang KW. (**2005**) Copy number amplification of 3q26-27 oncogenes in microdissected oral squamous cell carcinoma and oral brushed samples from areca chewers. *J Pathol.,* 206:417-22.

Linzer DI, Levine AJ. (**1979**) Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell,* 17:43-52.

Loeb LA. (**1991**) Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.,* 51:3075-9.

Loeb LA, Loeb KR, Anderson JP. (**2003**) Multiple mutations and cancer. *Proc Natl Acad Sci U S A.,* 100:776-81.

Logan N, Delavaine L, Graham A, Reilly C, Wilson J, Brummelkamp TR, Hijmans EM, Bernards R, La Thangue NB. (**2004**) E2F-7: a distinctive E2F family member with an unusual organization of DNA-binding domains. *Oncogene,* 23:5138-50.

Loo LW, Grove DI, Williams EM, Neal CL, Cousens LA, Schubert EL, Holcomb IN, Massa HF, Glogovac J, Li CI, Malone KE, Daling JR, Delrow JJ, Trask BJ, Hsu L, Porter PL. (**2004**) Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. *Cancer Res.,* 64:8541-9.

Loyer P, Trembley JH, Katona R, Kidd VJ, Lahti JM. (**2005**) Role of CDK/cyclin complexes in transcription and RNA splicing. *Cell Signal.,* 17:1033-51.

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. (**2005**) MicroRNA expression profiles classify human cancers. *Nature,* 435:834-8;

Lyons TE, Salih M, Tuana BS. (**2005**) The Activating E2Fs Mediate the Transcriptional Regulation of the Human E2F6 Repressor. *Am J Physiol Cell Physiol.,* advance online publication.

Mailand N, Lukas C, Kaiser BK, Jackson PK, Bartek J, Lukas. (**2002**) Deregulated human Cdc14A phosphatase disrupts centrosome separation and chromosome segregation. *Nat Cell Biol.,* 4:317-22.

Mairal A, Terrier P, Chibon F, Sastre X, Lecesne A, Aurias A. (**1999**) Loss of chromosome 13 is the most frequent genomic imbalance in malignant fibrous histiocytomas. A comparative genomic hybridization analysis of a series of 30 cases. *Cancer Genet Cytogenet.,* 111:134-8.

Mairal A, Terrier P, Chibon F, Sastre X, Lecesne A, Aurias A. (**2000**) Loss of chromosome 13 is the most frequent genomic imbalance in malignant fibrous histiocytomas. A comparative genomic hybridization analysis of a series of 30 cases. *Cancer Genet Cytogenet*., 111:134-8.

Maiti B, Li J, de Bruin A, Gordon F, Timmers C, Opavsky R, Patil K, Tuttle J, Cleghorn W, Leone G. (**2005**) Cloning and characterization of mouse E2F8, a novel mammalian E2F family member capable of blocking cellular proliferation. *J Biol Chem.,* 280:18211-20.

Mariconti L, Pellegrini B, Cantoni R, Stevens R, Bergounioux C, Cella R, Albani D. (**2002**) The E2F family of transcription factors from Arabidopsis thaliana. Novel and conserved components of the retinoblastoma/E2F pathway in plants. *J Biol Chem.,* 277:9911-9.

Martin GS. (**2004**) The road to Src. *Oncogene,* 23:7910-7.

Mashberg A. (**1977**) Erythroplasia vs. leukoplasia in the diagnosis of early asymptomatic oral squamous carcinoma. *N Engl J Med.,* 297:109-10.

Massion PP, Kuo WL, Stokoe D, Olshen AB, Treseler PA, Chin K, Chen C, Polikoff D, Jain AN, Pinkel D, Albertson DG, Jablons DM, Gray JW. (**2002**) Genomic copy number analysis of non-small cell lung cancer using array comparative genomic hybridization: implications of the phosphatidylinositol 3-kinase pathway.
Cancer Res., 62:3636-40.

Mayer ML**,** Gygi SP, Aebersold R, Hieter P. (**2001**) Identification of RFC (Ctf18p, Ctf8p, Dcc1p): an alternative RFC complex required for sister chromatid cohesion in S. cerevisiae. *Mol Cell*, 7:959-70.

Melo J, Toczyski D. (**2002**) A unified view of the DNA-damage checkpoint. *Curr Opin Cell Biol.,* 14:237-45.

Meuwissen R, Berns A. (**2005**) Mouse models for human lung cancer. *Genes Dev.,*

Michael MZ, O' Connor SM, van Holst Pellekaan NG, Young GP, James RJ. (**2003**) Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol Cancer Res.,* 1:882-91.

Michor F, Iwasa Y, Rajagopalan H, Lengauer C, Nowak MA. (**2004**) Linear model of colon cancer initiation. *Cell Cycl*e, 3:358-62.

Miska EA. (**2005**) How microRNAs control cell division, differentiation and death. *Curr Opin Genet Dev.,* 15:563-8.

Mitelman F. (**1974**) The Rous sarcoma virus story: Cytogenetics of tumors induced by RSV. *In Chromosomes and Cancer. J German Editor, Wyley, New York 1974. 675-693.*

Miyakis S, Liloglou T, Kearney S, Xinarianos G, Spandidos DA, Field JK. (**2003**) Absence of mutations in the VHL gene but frequent loss of heterozygosity at 3p25-26 in non-small cell lung carcinomas. *Lung Cancer,* 39:273-7.

Momand J, Zambetti GP, Olson DC, George D, Levine AJ. (**1992**) The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell,* 69:1237-45.

Monni O, Barlund M, Mousses S, Kononen J, Sauter G, Heiskanen M, Paavola P, Avela K, Chen Y, Bittner ML, Kallioniemi A. (**2001**) Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc Natl Acad Sci U S A.,* 98:5711-6.

Mowat M, Cheng A, Kimura N, Bernstein A, Benchimol S. (**1985**) Rearrangements of the cellular p53 gene in erythroleukaemic cells transformed by Friend virus. *Nature,* 314:633-6.

Mu D, Chen L, Zhang X, See LH, Koch CM, Yen C, Tong JJ, Spiegel L, Nguyen KC, Servoss A, Peng Y, Pei L, Marks JR, Lowe S, Hoey T, Jan LY, McCombie WR, Wigler MH, Powers S. (**2003**) Genomic amplification and oncogenic properties of the KCNK9 potassium channel gene. *Cancer Cell,* 3:297-302.

Muller H, Bracken AP, Vernell R, Moroni MC, Christians F, Grassilli E, Prosperini E, Vigo E, Oliner JD, Helin K. (**2001**) E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes Dev.,* 15:267-85.

Narayan S, Roy D. (**2003**) Role of APC and DNA mismatch repair genes in the development of colorectal cancers. *Mol Cancer,* 2:41.

Nigg EA. (**2002**) Centrosome aberrations: cause or consequence of cancer progression? *Nat Rev Cancer,* 2:815-25.

Nigro JM, Baker SJ, Preisinger AC, Jessup JM, Hostetter R, Cleary K, Bigner SH, Davidson N, Baylin S, Devilee P, *et al.* (**1989**) Mutations in the p53 gene occur in diverse human tumour types. *Nature,* 342:705-8.

Nowell PC, Hungerford DA. (**1960**) A minute chromosome in human chronic granulocytic leukaemia. *Science*, 132: 1497

Nowell PC (**1976**) The clonal evolution of tumor cell populations. *Science,* 194 : 23-8.

Nowell PC, Rowley JD, Knudson AG Jr. (**1998**) Cancer genetics, cytogenetics--defining the enemy within. *Nat Med*,..4 :1107-11.

Nunberg JH, Kaufman RJ, Schimke RT, Urlaub G, Chasin LA. (**1978**) Amplified dihydrofolate reductase genes are localized to a homogeneously staining region of a single chromosome in a methotrexate-resistant Chinese hamster ovary cell line. *Proc Natl Acad Sci U S A.,* 75:5553-6.

Oberley MJ, Inman DR, Farnham PJ. (**2003**) E2F6 negatively regulates BRCA1 in human cancer cells without methylation of histone H3 on lysine 9. *J Biol Chem.,* 278:42466-76.

O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. (**2005**) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature,* 435:839-43.

Oeggerli M, Tomovska S, Schraml P, Calvano-Forte D, Schafroth S, Simon R, Gasser T, Mihatsch MJ, Sauter G. (**2004**) E2F3 amplification and overexpression is associated with invasive tumor growth and rapid tumor cell proliferation in urinary bladder cancer. *Oncogene, 23*:5616-23.

Okada H, Mak TW. (**2004**) Pathways of apoptotic and non-apoptotic death in tumour cells. *Nat Rev Cancer,* 4:592-603.

Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y. (**2002**) A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science,* 296:1132-6

Oliner JD, Kinzler KW, Meltzer PS, George DL, Vogelstein B. (**1992**) Amplification of a gene encoding a p53-associated protein in human sarcomas. *Nature,* 358:80-3.

Orntoft TF, Thykjaer T, Waldman FM, Wolf H, Celis JE. (**2002**) Genome-wide study of gene copy numbers, transcripts, and protein levels in pairs of non-invasive and invasive human transitional cell carcinomas. *Mol Cell Proteomics,* 1(1):37-45.

Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, Yoshida Y, Seto M. (**2004**) Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma.
Cancer Res., 64:3087-95.

Parada LF, Land H, Weinberg RA, Wolf D, Rotter V. (**1984**) Cooperation between gene encoding p53 tumour antigen and ras in cellular transformation. *Nature,* 312:649-51.

Parsons R, Li GM, Longley MJ, Fang WH, Papadopoulos N, Jen J, de la Chapelle A, Kinzler KW, Vogelstein B, Modrich P. (**1993**) Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell,* 75:1227-36.

Parsons DW, Wang TL, Samuels Y, Bardelli A, Cummins JM, DeLong L, Silliman N, Ptak J, Szabo S, Willson JK, Markowitz S, Kinzler KW, Vogelstein B, Lengauer C, Velculescu VE. (**2005**) Colorectal cancer: mutations in a signalling pathway. *Nature,* 436:792.

Patel JH, Loboda AP, Showe MK, Showe LC, McMahon SB. (**2004**) Analysis of genomic targets reveals complex functions of MYC. *Nat Rev Cancer.,* 4:562-8.

Pekarsky Y, Palamarchuk A, Huebner K, Croce CM. (**2002**) FHIT as tumor suppressor: mechanisms and therapeutic opportunities. *Cancer Biol Ther.,* 1:232-6.

Pihan G, Doxsey SJ. (**2003**) Mutations and aneuploidy: co-conspirators in cancer? *Cancer Cell*, 4:89-94.

Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W. (**2005**) Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science,* 309:1573-6.

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. (**1998**) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet.,* 20:207-11.

Pinkel D, Albertson DG. (**2005**) Comparative genomic hybridization. *Annu Rev Genomics Hum Genet.,* 6:331-54.

Platzer P, Upender MB, Wilson K, Willis J, Lutterbaugh J, Nosrati A, Willson JK, Mack D, Ried T, Markowitz S. (**2002**) Silence of chromosomal amplifications in colon cancer. *Cancer Res.,* 62:1134-8.

Pohlers M, Truss M, Frede U, Scholz A, Strehle M, Kuban RJ, Hoffmann B, Morkel M, Birchmeier C, Hagemeier C. (**2005**) A role for E2F6 in the restriction of male-germ-cell-specific gene expression. *Curr Biol.,* 15:1051-7.

Polanowska J, Le Cam L, Orsetti B, Valles H, Fabbrizio E, Fajas L, Taviaux S, Theillet C, Sardet C. (**2000**) Human E2F5 gene is oncogenic in primary rodent cells and is amplified in human breast tumors. *Genes Chromosomes Cancer.,* 28:126-30.

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. (**1999**) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet.,* 23:41-6.

Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, Brown PO. (**2002**) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A.,* 99:12963-8.

Prevot D, Decimo D, Herbreteau CH, Roux F, Garin J, Darlix JL, Ohlmann T. (**2003**) Characterization of a novel RNA-binding region of eIF4GI critical for ribosomal scanning. *EMBO J.,* 22:1909-21.

Rajagopalan H, Nowak MA, Vogelstein B, Lengauer C. (**2003**) The significance of unstable chromosomes in colorectal cancer. *Nat Rev Cancer,* 3:695-701.

Rajagopalan H, Jallepalli PV, Rago C, Velculescu VE, Kinzler KW, Vogelstein B, Lengauer C. (**2004**) Inactivation of hCDC4 can cause chromosomal instability. *Nature,* 428:77-81.

Rakha EA, Pinder SE, Paish EC, Robertson JF, Ellis IO. (**2004**) Expression of E2F-4 in invasive breast carcinomas is associated with poor prognosis. *J Pathol.,* 203:754-61.

Ralhan R, Chakravarti N, Kaur J, Sharma C, Kumar A, Mathur M, Bahadur S, Shukla NK, Deo SV. (**2005**) Clinical significance of altered expression of retinoid receptors in oral precancerous and cancerous lesions: Relationship with cell cycle regulators. *Int J Cancer.,* advance online publication.

Rangarajan A, Weinberg RA. (**2003**) Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nat Rev Cancer,* 3:952-9.

Rangarajan A, Hong SJ, Gifford A, Weinberg RA. (**2004**) Species- and cell type-specific requirements for cellular transformation. *Cancer Cell,* 6:171-83.

Raphael BJ, Volik S, Collins C, Pevzner PA. (**2003**) Reconstructing tumor genome architectures. *Bioinformatics,* 19 Suppl 2:II162-II171.

Redon R, Muller D, Caulee K, Wanherdrick K, Abecassis J, du Manoir S. (**2001**) A simple specific pattern of chromosomal aberrations at early stages of head and neck squamous cell carcinomas: PIK3CA but not p63 gene as a likely target of 3q26-qter gains. *Cancer Res.,* 61:4122-9.

Redon R, Hussenet T, Bour G, Caulee K, Jost B, Muller D, Abecassis J, du Manoir S. (**2002**) Amplicon mapping and transcriptional analysis pinpoint cyclin L as a candidate oncogene in head and neck cancer.
Cancer Res., 62:6211-7.

Redon R, Rio M, Gregory SG, Cooper RA, Fiegler H, Sanlaville D, Banerjee R, Scott C, Carr P, Langford C, Cormier-Daire V, Munnich A, Carter NP, Colleaux L. (**2005**) Tiling path resolution mapping of constitutional 1p36 deletions by array-CGH: contiguous gene deletion or "deletion with positional effect" syndrome? *J Med Genet.,* 42:166-71.

Reichel R, Kovesdi I, Nevins JR. (**1987**) Developmental control of a promoter-specific factor that is also regulated by the E1A gene product. *Cell,* 48:501-6.

Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, Young RA, Dynlacht BD. (**2002**) E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev.,* 16:245-56.

Ren S, Rollins BJ. (**2004**) Cyclin C/cdk3 promotes Rb-dependent G0 exit. *Cell,* 117:239-51.

Rhodes DR, Chinnaiyan AM. (**2005**) Integrative analysis of the cancer transcriptome. *Nat Genet.,* 37 Suppl:S31-7.

Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. (**2004**) Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci U S A.;* 101:9309-14.

Rhodes DR, Chinnaiyan AM. (**2004**) Bioinformatics strategies for translating genome-wide expression analyses into clinically useful cancer markers. *Ann N Y Acad Sci.,* 1020:32-40.

Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Barrette TR, Ghosh D, Chinnaiyan AM. (**2005**) Mining for regulatory programs in the cancer transcriptome. *Nat Genet.,* 37:579-83.

Ried T, Heselmeyer-Haddad K, Blegen H, Schrock E, Auer G. (**1999**) Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation. *Genes Chromosomes Cancer,* 25:195-204.

Roerig P, Nessling M, Radlwimmer B, Joos S, Wrobel G, Schwaenen C, Reifenberger G, Lichter P. (**2005**) Molecular classification of human gliomas using matrix-based comparative genomic hybridization. *Int J Cancer,* 117:95-103;

Rowley JD. (**1973**) A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*, 243:290-3.

Rowley JD. (**1984**) Biological implications of consistent chromosome rearrangements in leukemia and lymphoma. *Cancer Res.,* 44:3159-68.

Roussel M, Saule S, Lagrou C, Rommens C, Beug H, Graf T, Stehelin D. (**1979**) Three new types of viral oncogene of cellular origin specific for haematopoietic cell transformation. *Nature,* 281:452-5.

Rubio-Moscardo F, Climent J, Siebert R, Piris MA, Martin-Subero JI, Nielander I, Garcia-Conde J, Dyer MJ, Terol MJ, Pinkel D, Martinez-Climent JA. (**2005**) Mantle-cell lymphoma genotypes identified with CGH to BAC microarrays define a leukemic subgroup of disease and predict patient outcome. *Blood,* 105:4445-54.

Rudkin GT, Stollar BD. (**1977**) High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence. *Nature,* 265:472-3.

Ruley HE. (**1983**) Adenovirus early region 1A enables viral and cellular transforming genes to transform primary cells in culture. *Nature,* 304:602-6.

Saavedra HI, Maiti B, Timmers C, Altura R, Tokuyama Y, Fukasawa K, Leone G. (**2003**) Inactivation of E2F3 results in centrosome amplification. *Cancer Cell,* 3:333-46.

Sabatier L, Ricoul M, Pottier G, Murnane JP. (**2005**)The loss of a single telomere can result in instability of multiple chromosomes in a human tumor cell line. *Mol Cancer Res.,* 3:139-50.

Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, Yan H, Gazdar A, Powell SM, Riggins GJ, Willson JK, Markowitz S, Kinzler KW, Vogelstein B, Velculescu VE. (**2004**) High frequency of mutations of the PIK3CA gene in human cancers. *Science,* 304:554.

Samuels Y, Diaz LA Jr, Schmidt-Kittler O, Cummins JM, Delong L, Cheong I, Rago C, Huso DL, Lengauer C, Kinzler KW, Vogelstein B, Velculescu VE. (**2005**) Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell,* 7:561-73.

Scheijen B, Bronk M, van der Meer T, De Jong D, Bernards R. (**2004**) High incidence of thymic epithelial tumors in E2F2 transgenic mice*. J Biol Chem.,* 279:10476-83.

Schena M, Shalon D, Davis RW, Brown PO. (**1995**) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science,* 270:467-70.

Schmittgen TD, Jiang J, Liu Q, Yang L. (**2004**) A high-throughput method to monitor the expression of microRNA precursors. *Nucleic Acids Res.,* 32:e43.

Schrock E, du Manoir S, Veldman T, Schoell B, Wienberg J, Ferguson-Smith MA, Ning Y, Ledbetter DH, Bar-Am I, Soenksen D, Garini Y, Ried T. (**1996**) Multicolor spectral karyotyping of human chromosomes. *Science,* 273:494-7.

Schwab M, Alitalo K, Klempnauer KH, Varmus HE, Bishop JM, Gilbert F, Brodeur G, Goldstein M, Trent J. (**1983**) Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature,* 305:245-8.

Schwab M, Varmus HE, Bishop JM. (**1985**) Human N-myc gene contributes to neoplastic transformation of mammalian cells in culture. *Nature,* 316:160-2.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. (**2004**)
Large-scale copy number polymorphism in the human genome. *Science,* 305:525-8.

Segal E, Friedman N, Kaminski N, Regev A, Koller D. (**2005**) From signatures to models: understanding cancer using microarrays. *Nat Genet.,* 37 Suppl:S38-45.

Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE. (**2005**) Segmental duplications and copy-number variation in the human genome. Am J Hum Genet., 77:78-88.

Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H, Firth H, Sanlaville D, Winter R, Colleaux L, Bobrow M, Carter NP. (**2004**) Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J Med Genet.,* 41:241-8.

Sheiness D, Bister K, Moscovici C, Fanshier L, Gonda T, Bishop JM. (**1980**) Avian retroviruses that cause carcinoma and leukemia: identification of nucleotide sequences associated with pathogenicity. *J Virol.,* 33:962-8.

Shih IM, Zhou W, Goodman SN, Lengauer C, Kinzler KW, Vogelstein B. (**2001**) Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. *Cancer Res.,* 61:818-22.

Shimizu N, Kanda T, Wahl GM. (**1996**) Selective capture of acentric fragments by micronuclei provides a rapid method for purifying extrachromosomally amplified DNA. *Nat Genet.,* 12:65-71.

Shimizu N, Miura Y, Sakamoto Y, Tsutsui K (**2001**) Plasmids with a mammalian replication origin and a matrix attachment region initiate the event similar to gene amplification. *Cancer Res.,* 61:6987-90.

Sieber OM, Heinimann K, Tomlinson IP. (**2003**) Genomic instability--the engine of tumorigenesis? *Nat Rev Cancer*, 3:701-8.

Singh B, Stoffel A, Gogineni S, Poluri A, Pfister DG, Shaha AR, Pathak A, Bosl G, Cordon-Cardo C, Shah JP, Rao PH. (**2002a**) Amplification of the 3q26.3 locus is associated with progression to invasive cancer and is a negative prognostic factor in head and neck squamous cell carcinomas. *Am J Pathol.,* 161:365-71.

Singh B, Reddy PG, Goberdhan A, Walsh C, Dao S, Ngai I, Chou TC, O-Charoenrat P, Levine AJ, Rao PH, Stoffel A. (**2002b**) p53 regulates cell survival by inhibiting PIK3CA in squamous cell carcinomas. *Genes Dev.,* 16:984-93.

Singhal S, Vachani A, Antin-Ozerkis D, Kaiser LR, Albelda SM. (**2005**) Prognostic implications of cell cycle, apoptosis, and angiogenesis biomarkers in non-small cell lung cancer: a review. *Clin Cancer Res.,* 11:3974-86.

Singh P, Wong SH, Hong W. (**1994**) Overexpression of E2F-1 in rat embryo fibroblasts leads to neoplastic transformation. *EMBO J.,* 13:3329-38.

Smith S, Hwang JY, Banerjee S, Majeed A, Gupta A, Myung K. (**2004**) Mutator genes for suppression of gross chromosomal rearrangements identified by a genome-wide screening in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A.,* 101:9039-44.

Smits VA, Medema RH. (**2001**) Checking out the G(2)/M transition. *Biochim Biophys Acta.,* 1519:1-12.

Snijders AM, Schmidt BL, Fridlyand J, Dekker N, Pinkel D, Jordan RC, Albertson DG. (2005) Rare amplicons implicate frequent deregulation of cell fate specification pathways in oral squamous cell carcinoma. *Oncogene,* 24:4232-42.

Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P. (**1997**)
Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer, 20:399-407.

Song MS, Song SJ, Ayad NG, Chang JS, Lee JH, Hong HK, Lee H, Choi N, Kim J, Kim H, Kim JW, Choi EJ, Kirschner MW, Lim DS. (**2004**) The tumour suppressor RASSF1A regulates mitosis by inhibiting the APC-Cdc20 complex. *Nat Cell Biol.,* 6:129-37.

Speicher MR, Gwyn Ballard S, Ward DC. (**1996**) Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat Genet.,* 12:368-75.

Speicher MR, Carter NP. (**2005**) The new cytogenetics: blurring the boundaries with molecular biology. *Nat Rev Genet.,* 6:782-92.

Stegmeier F, Visintin R, Amon A. (**2002**) Separase, polo kinase, the kinetochore protein Slk19, and Spo12 function in a network that controls Cdc14 localization during early anaphase. *Cell,* 108:207-20.

Stegmeier F, Amon A (2004) Closing mitosis: the functions of the Cdc14 phosphatase and its regulation. *Annu Rev Genet.,* 38:203-32.

Stehelin D, Varmus HE, Bishop JM, Vogt PK. (**1976**) DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature,*260:170-3.

Stephens P, Edkins S, Davies H, Greenman C, Cox C, Hunter C, Bignell G, Teague J, Smith R, Stevens C, O'Meara S, Parker A, Tarpey P, Avis T, Barthorpe A, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Edwards K, Forbes S, Gorton M, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jones D, Kosmidou V, Laman R, Lugg R, Menzies A, Perry J, Petty R, Raine K, Shepherd R, Small A, Solomon H, Stephens Y, Tofts C, Varian J, Webb A, West S, Widaa S, Yates A, Brasseur F, Cooper CS, Flanagan AM, Green A, Knowles M, Leung SY, Looijenga LH, Malkowicz B, Pierotti MA, Teh B, Yuen ST, Nicholson AG, Lakhani S, Easton DF, Weber BL, Stratton MR, Futreal PA, Wooster R. (**2005**) A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat Genet.,* 37:590-2.

Stevaux O, Dyson NJ. (**2002**) A revised picture of the E2F transcriptional network and RB function. *Curr Opin Cell Biol.,* 14:684-91.

Sticht C, Hofele C, Flechtenmacher C, Bosch FX, Freier K, Lichter P, Joos S. (**2005**) Amplification of Cyclin L1 is associated with lymph node metastases in head and neck squamous cell carcinoma (HNSCC). *Br J Cancer,* 92:770-4.

Storlazzi CT, Fioretos T, Paulsson K, Strombeck B, Lassen C, Ahlgren T, Juliusson G, Mitelman F, Rocchi M, Johansson B. (**2004**) Identification of a commonly amplified 4.3 Mb region with overexpression of C8FW, but not MYC in MYC-containing double minutes in myeloid malignancies. *Hum Mol Genet.,* 13:1479-85.

Storre J, Schafer A, Reichert N, Barbero JL, Hauser S, Eilers M, Gaubatz S. (**2005**) Silencing of the meiotic genes SMC1beta and STAG3 in somatic cells by E2F6. *J Biol Chem.,* advance online publication.

Strand M, Prolla TA, Liskay RM, Petes TD. (**1993**) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature,* 365:274-6.
Erratum in: *Nature,* 368:569.

Struski S, Doco-Fenzy M, Cornillet-Lefebvre P. (**2002**) Compilation of published comparative genomic Hybridization studies. *Cancer Genet Cytogenet.,* 135:63-90.

Sudbo J, Kildal W, Risberg B, Koppang HS, Danielsen HE, Reith A. (**2001**) DNA content as a prognostic marker in patients with oral leukoplakia. *N Engl J Med.,* 344:1270-8.

Sudbo J, Lippman SM, Lee JJ, Mao L, Kildal W, Sudbo A, Sagen S, Bryne M, El-Naggar A, Risberg B, Evensen JF, Reith A. (**2004**) The influence of resection and aneuploidy on mortality in oral leukoplakia. *N Engl J Med.,*350:1405-13.

Suh MR, Lee Y, Kim JY, Kim SK, Moon SH, Lee JY, Cha KY, Chung HM, Yoon HS, Moon SY, Kim VN, Kim KS (**2004**) Human embryonic stem cells express a unique set of microRNAs. *Dev Biol.,* 270:488-98.

Sullivan M, Uhlmann F. (**2003**) A non-proteolytic function of separase links the onset of anaphase to mitotic exit. *Nat Cell Biol.,* 5:249-54.

Talbot SG, O-charoenrat P, Sarkaria IS, Ghossein R, Reddy P, Ngai I, Cordeiro CN, Wong RJ, Kris MG, Rusch VW, Singh B. (**2004**) Squamous cell carcinoma related oncogene regulates angiogenesis through vascular endothelial growth factor-A. *Ann Surg Onco*l., 11:530-4.

Tagawa H, Karnan S, Suzuki R, Matsuo K, Zhang X, Ota A, Morishima Y, Nakamura S, Seto M. (**2005a**)
Genome-wide array-based CGH for mantle cell lymphoma: identification of homozygous deletions of the proapoptotic gene BIM. *Oncogene,* 24:1348-58.

Tagawa H, Suguro M, Tsuzuki S, Matsuo K, Karnan S, Ohshima K, Okamoto M, Morishima Y, Nakamura S, Seto M. (**2005b**) Comparison of genome profiles for identification of distinct subgroups of diffuse large B-cell lymphoma. *Blood,* 106:1770-7.

Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, Harano T, Yatabe Y, Nagino M, Nimura Y, Mitsudomi T, Takahashi T. (**2004**) Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res.,* 64:3753-6.

Tam W, Dahlberg JE. (2005) miR-155/BIC as an oncogenic microRNA. Genes Chromosomes Cancer advance online publication.

Tang R, Changchien CR, Wu MC, Fan CW, Liu KW, Chen JS, Chien HT, Hsieh LL. (**2004**) Colorectal cancer without high microsatellite instability and chromosomal instability--an alternative genetic pathway to human colorectal cancer. *Carcinogenesis,* 25:841-6.

Tao Y, Kassatly RF, Cress WD, Horowitz JM. (**1997**) Subunit composition determines E2F DNA-binding site specificity. *Mol Cell Biol.,* 17:6994-7007.

Telenius H, Pelmear AH, Tunnacliffe A, Carter NP, Behmel A, Ferguson-Smith MA, Nordenskjold M, Pfragner R, Ponder BA. (**1992**) Cytogenetic analysis by chromosome painting using DOP-PCR amplified flow-sorted chromosomes. *Genes Chromosomes Cancer,* 4:257-63.

Tighe A, Johnson VL, Taylor SS. (**2004**) Truncating APC mutations have dominant effects on proliferation, spindle checkpoint control, survival and chromosome stability. *J Cell Sci.,* 117:6339-53

Thiagalingam S, Laken S, Willson JK, Markowitz SD, Kinzler KW, Vogelstein B, Lengauer C. (**2001**) Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proc Natl Acad Sci U S A.*, 98:2698-702.

Torres-Rosell J, Machin F, Aragon L. (**2005**) Cdc14 and the temporal coordination between mitotic exit and chromosome segregation. *Cell Cycle,* 4:109-12.

Trautmann S, McCollum D. (**2002**) Cell cycle: new functions for Cdc14 family phosphatases. *Curr Biol.*, 2002 12:R733-5.

Trimarchi JM, Fairchild B, Wen J, Lees JA. (**2001**) The E2F6 transcription factor is a component of the mammalian Bmi1-containing polycomb complex. *Proc Natl Acad Sci U S A.,* 98:1519-24.

Tyzzer E. E. (**1916**) Tumor immunity. *J. Cancer Res.*, 1: 125–156.

Thibodeau SN, Bren G, Schaid D. (**1993**) Microsatellite instability in cancer of the proximal colon. *Science,* 260:816-9.

Tijo JH, Levan A. (**1956**) The chromosome number of man. *Hereditas*. 42: 1-6.

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. (**2005**) Fine-scale structural variation of the human genome. *Nat Genet.,* 37:727-32.

Uematsu K, He B, You L, Xu Z, McCormick F, Jablons DM. (**2003**) Activation of the Wnt pathway in non small cell lung cancer: evidence of dishevelled overexpression. *Oncogene,* 22:7218-21.

van Duin M, van Marion R, Watson JE, Paris PL, Lapuk A, Brown N, Oseroff VV, Albertson DG, Pinkel D, de Jong P, Nacheva EP, Dinjens W, van Dekken H, Collins C. (**2005**) Construction and application of a full-coverage, high-resolution, human chromosome 8q genomic microarray for comparative genomic hybridization. *Cytometry A.,* 63:10-9.

Varmus HE. (**1990**) Nobel lecture. Retroviruses and oncogenes. I. *Biosci Rep.,* 10:413-30.

Vennstrom B, Bishop JM. (**1982a**) Isolation and characterization of chicken DNA homologous to the two putative oncogenes of avian erythroblastosis virus. Cell, 28:135-43.

Vennstrom B, Sheiness D, Zabielski J, Bishop JM. (**1982b**) Isolation and characterization of c-myc, a cellular homolog of the oncogene (v-myc) of avian myelocytomatosis virus strain 29. *J Virol.,* 42:773-9.

Virtaneva K, Wright FA, Tanner SM, Yuan B, Lemon WJ, Caligiuri MA, Bloomfield CD, de La Chapelle A, Krahe R. (**2001**) Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc Natl Acad Sci U S A.,* 98:1124-9.

Vogelstein B, Kinzler KW. (**2004**) Cancer genes and the pathways they control. *Nat Med.,* 10:789-99

Volik S, Zhao S, Chin K, Brebner JH, Herndon DR, Tao Q, Kowbel D, Huang G, Lapuk A, Kuo WL, Magrane G, De Jong P, Gray JW, Collins C. (**2003**) End-sequence profiling: sequence-based analysis of aberrant genomes. *Proc Natl Acad Sci U S A.,* 100:7696-701.

Von Hansemann D. (**1890**) Ueber asymmetrische Zelltheilung in epithel Krebsen und deren biologische Bedeutung. *Virschows Arch Pathol Anat*, 119: 299-326.
[litterall translation: On asymmetric mitoses in epithelial cancers and their biological significance].

Vos MD, Martinez A, Elam C, Dallol A, Taylor BJ, Latif F, Clark GJ. (**2004**) A role for the RASSF1A tumor suppressor in the regulation of tubulin polymerization and genomic stability. *Cancer Res.,* 64:4244-50.

Waber PG, Lee NK, Nisen PD. (**1996**) Frequent allelic loss at chromosome arm 3p is distinct from genetic alterations of the Von-Hippel Lindau tumor suppressor gene in head and neck cancer. *Oncogene,* 12:365-9.

Wang Z, Cummins JM, Shen D, Cahill DP, Jallepalli PV, Wang TL, Parsons DW, Traverso G, Awad M, Silliman N, Ptak J, Szabo S, Willson JK, Markowitz SD, Goldberg ML, Karess R, Kinzler KW, Vogelstein B, Velculescu VE, Lengauer C. (**2004a**) Three classes of genes mutated in colorectal cancers with chromosomal instability. *Cancer Res.,* 64:2998-3001.

Wang Z, Shen D, Parsons DW, Bardelli A, Sager J, Szabo S, Ptak J, Silliman N, Peters BA, van der Heijden MS, Parmigiani G, Yan H, Wang TL, Riggins G, Powell SM, Willson JK, Markowitz S, Kinzler KW, Vogelstein B, Velculescu VE. (**2004b**) Mutational analysis of the tyrosine phosphatome in colorectal cancers. *Science,* 304:1164-6.

Weinberg RA. (**1996**) E2F and cell proliferation: a world turned upside down. *Cell,* 85:457-9.

Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ. (**2002**) Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev.,* 16:235-44.

Weir B, Zhao X, Meyerson M. (**2004**) Somatic alterations in the human cancer genome. *Cancer Cell*, 6:433-8.

Weiss MM, Kuipers EJ, Postma C, Snijders AM, Siccama I, Pinkel D, Westerga J, Meuwissen SG, Albertson DG, Meijer GA. (**2003**) Genomic profiling of gastric cancer predicts lymph node status and survival. *Oncogene,* 22:1872-9.

Weiss MM, Kuipers EJ, Postma C, Snijders AM, Pinkel D, Meuwissen SG, Albertson D, Meijer GA. (**2004**)
Genomic alterations in primary gastric adenocarcinomas correlate with clinicopathological characteristics and survival. *Cell Oncol.*. 26:307-17.

Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, Botstein D. (**2002**) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell,* 13:1977-2000.

Winge Ö. (**1930**) Zytologishes Untersuchungen über die Natur Maligner Tumoren. II. Teerkarzinome bein Mäusen. *Z. Zellforsch. Mikrosk. Anat.* 10, 683–735.

Wistuba II, Behrens C, Virmani AK, Mele G, Milchgrub S, Girard L, Fondon JW 3rd, Garner HR, McKay B, Latif F, Lerman MI, Lam S, Gazdar AF, Minna JD. (**2000**) High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. *Cancer Res.,* 60:1949-60.

Wittwer CT, Ririe KM, Andrew RV, David DA, Gundry RA, Balis UJ. (**1997**) The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques,* 22:176-81.

Wolf D, Rotter V. (**1984**) Inactivation of p53 gene expression by an insertion of Moloney murine leukemia virus-like DNA sequences. *Mol Cell Biol.,* 4:1402-10.

Wolf D, Rotter V. (**1985**) Major deletions in the gene encoding the p53 tumor antigen cause lack of p53 expression in HL-60 cells. *Proc Natl Acad Sci U S A.,* 82:790-4.

Wolf M, Mousses S, Hautaniemi S, Karhu R, Huusko P, Allinen M, Elkahloun A, Monni O, Chen Y, Kallioniemi A, Kallioniemi OP. (**2004**) High-resolution analysis of gene copy number alterations in human prostate cancer using CGH on cDNA microarrays: impact of copy number on gene expression. *Neoplasia,* 6:240-7.

Woo RA, Poon RY. (**2003**) Cyclin-dependent kinases and S phase control in mammalian cells. *Cell Cycle,* 2:316-24.

Wu L, Timmers C, Maiti B, Saavedra HI, Sang L, Chong GT, Nuckolls F, Giangrande P, Wright FA, Field SJ, Greenberg ME, Orkin S, Nevins JR, Robinson ML, Leone G. (**2001**) The E2F1-3 transcription factors are essential for cellular proliferation. *Nature,* 414:457-62.

Wu G, Xing M, Mambo E, Huang X, Liu J, Guo Z, Chatterjee A, Goldenberg D, Gollin SM, Sukumar S, Trink B, Sidransky D. (**2005**) Somatic mutation and gain of copy number of PIK3CA in human breast cancer. Breast *Cancer Res.,* 7:R609-16.

Xu G, Livingston DM, Krek W. (**1995**) Multiple members of the E2F transcription factor family are the products of oncogenes. *Proc Natl Acad Sci U S A.,* 92:1357-61.

Yancopoulos GD, Nisen PD, Tesfaye A, Kohl NE, Goldfarb MP, Alt FW. (**1985**) N-myc can cooperate with ras to transform normal cells in culture. *Proc Natl Acad Sci U S A.,* 82:5455-9.

Yasui K, Arii S, Zhao C, Imoto I, Ueda M, Nagai H, Emi M, Inazawa J. (**2002**) TFDP1, CUL4A, and CDC16 identified as targets for amplification at 13q34 in hepatocellular carcinomas. *Hepatology,* 35:1476-84.

Yasui K, Okamoto H, Arii S, Inazawa J. (**2003**) Association of over-expressed TFDP1 with progression of hepatocellular carcinomas. *J Hum Genet.,* 48:609-13.

Yen CC, Chen YJ, Pan CC, Lu KH, Chen PC, Hsia JY, Chen JT, Wu YC, Hsu WH, Wang LS, Huang MH, Huang BS, Hu CP, Chen PM, Lin CH. (**2005**) Copy number changes of target genes in chromosome 3q25.3-qter of esophageal squamous cell carcinoma: TP63 is amplified in early carcinogenesis but down-regulated as disease progressed. *World J Gastroenterol.,* 11:1267-72.

Yokoi S, Yasui K, Iizasa T, Imoto I, Fujisawa T, Inazawa J. (**2003**) TERC identified as a probable target within the 3q26 amplicon that is detected frequently in non-small cell lung cancers. *Clin Cancer Res.,* 9:4705-13.

Zanesi N, Mancini R, Sevignani C, Vecchione A, Kaou M, Valtieri M, Calin GA, Pekarsky Y, Gnarra JR, Croce CM, Huebner K. (**2005**) Lung cancer susceptibility in Fhit-deficient mice is increased by Vhl haploinsufficiency. *Cancer Res.,* 65:6576-82.

Zhang H, Lawrence CW. (**2005**) The error-free component of the RAD6/RAD18 DNA damage tolerance pathway of budding yeast employs sister-strand recombination. *Proc Natl Acad Sci U S A.,* 102:15954-9.

Zhang Y, Chellappan SP. (**1995**) Cloning and characterization of human DP2, a novel dimerization partner of E2F. *Oncogene,* 10:2085-93.

Zhao H, Kim Y, Wang P, Lapointe J, Tibshirani R, Pollack JR, Brooks JD. (**2005**) Genome-wide characterization of gene expression variations and DNA copy number changes in prostate cancer cell lines. *Prostate,* 63:187-97.

Zhou X, Cole SW, Hu S, Wong DT. (**2004**) Detection of DNA copy number abnormality by microarray expression analysis. *Hum Genet.,* 114:464-7.

Zhou X, Rao NP, Cole SW, Mok SC, Chen Z, Wong DT. (**2005**) Progress in concurrent analysis of loss of heterozygosity and comparative genomic hybridization utilizing high density single nucleotide polymorphism arrays.Cancer Genet Cytogenet., 159:53-7.

Zimonjic D, Brooks MW, Popescu N, Weinberg RA, Hahn WC. (**2001**) Derivation of human tumor cells in vitro without widespread genomic instability. *Cancer Res. ,* 61:8838-44.