

Sélection et décryptage

*Small is beautiful**

Les grandes équipes de recherche développent la connaissance mais ce sont les petites qui jettent les bases des « ruptures » épistémologiques : voilà les conclusions de cet article de *Nature*, conclusions intéressantes que je voudrais discuter avec les lecteurs. Soulignons d'abord que tout le monde a remarqué que le nombre de signataires d'un article a augmenté considérablement au cours du temps ; il n'est pas rare de trouver, dans les « grandes » revues, des articles qui comptent plus d'une dizaine d'auteurs. On a proposé comme raison plausible à cette inflation la surspécialisation des activités scientifiques, l'évolution des stratégies de communication, la complexité croissante des problèmes scientifiques qui requièrent des solutions multidisciplinaires, et bien d'autres causes.

Il faut aussi noter que plusieurs études ont montré que les individus se comportent différemment selon qu'ils appartiennent à des groupes importants ou à de petites entités : dans les grands groupes, les individus auraient tendance à générer moins d'idées, à rejeter plus souvent les points de vue extérieurs, à neutraliser volontiers de façon croisée les différents points de vue internes. En outre, les grands groupes de recherche se focalisent sur des sujets productifs, offrant plus vite des possibilités de publication, alors que les petites équipes, ayant plus à gagner et moins à perdre, ont plus d'aptitudes à innover, à entreprendre des recherches dans des domaines peu explorés et aux résultats incertains. C'est pour apporter des éléments permettant d'asseoir ces données psychosociologiques que les auteurs ont entrepris ce travail.

Ils ont analysé 24 millions d'articles (sans parler de brevets et de logiciels) publiés au cours de la période 1954-2014, pour comparer l'impact de ces articles sur l'avancement des connaissances. Globalement, les « grandes équipes » (celles dont les articles ont beaucoup de signataires) construisent les développements marquants et « populaires », reconnus rapidement par la communauté scientifique ; alors que les « petites équipes » apportent des résultats « disruptifs¹ », reconnus plus tardivement comme ayant marqué le domaine – un changement de paradigme cher à Thomas Kuhn, pourrait-on dire. Il est intéressant de savoir comment ils y sont parvenus.

Tâche bien difficile que de quantifier la « rupture épistémologique » apportée par un article, un brevet ou un logiciel ! La « disruption » D apportée par un article est évaluée par analyse de citations ; les auteurs considèrent qu'un article peut donner lieu à trois types de citations :

- (i) être cité directement ;
- (j) être cité en même temps que ses références ;
- (k) ne pas être cité alors que ses références sont citées.

D est alors égale à la différence $(n_i - n_j)$, divisée par le nombre total $(n_i + n_j + n_k)$ et peut ainsi prendre des valeurs comprises entre -1 (article peu disruptif) et +1 (article très disruptif). La validation de ce calcul est apportée par divers moyens : repérage des articles signés par des Prix Nobel ($D = +0,1$), des revues générales ($D = -0,001$) et des articles cités par ces revues ($D = +0,001$) ; analyse sémantique des mots du titre (par exemple, *change, introduit, avance* par opposition à *endosse, confirme, démontre*), etc.

Ayant ainsi calculé le degré de disruption de millions d'articles, de brevets et de logiciels, les auteurs ont recherché le lien entre ce paramètre et la taille de l'équipe signataire, et ils ont observé que plus l'équipe était importante, moins l'article était disruptif (*figure 1*) et qu'il en est de même pour les brevets et les logiciels. Cela est en faveur de l'hypothèse que les grandes équipes sont plus motivées par le développement de la science et de la technologie « courantes », et que les petites équipes portent plus volontiers leurs efforts sur les problèmes nouveaux. Les articles signés par un seul auteur reçoivent statistiquement autant de citations que ceux signés par cinq auteurs mais ont une probabilité beaucoup plus élevée d'être parmi les 5 % les plus « disruptifs » (*figure 2*). À l'inverse, les articles signés par 10 auteurs reçoivent nettement plus de citations que ceux signés par un à cinq auteurs, mais sont parmi les moins « disruptifs ». Les deux paramètres sont d'ailleurs liés : les articles des petites équipes qui sont le plus cités sont ceux qui sont les plus disruptifs ; et les articles des grandes équipes qui sont le plus cités sont aussi les moins disruptifs (*figure 3*). La décroissance du caractère disruptif en fonction de la taille de l'équipe est surtout marquée pour les articles les plus cités et s'amenuise quand le nombre de citations diminue.

Ce lien entre la taille de l'équipe et le degré de disruption de ces articles est globalement le même pour chacune des décennies composant la période 1954-2004 (*figure 4*) mais le lecteur remarquera au passage que le degré de disruption moyen a singulièrement baissé au cours du temps... L'inflation continue du nombre d'articles publiés en est certainement la cause. Ce lien est également maintenu dans les grandes disciplines de la science, à l'exception de l'informatique et des sciences de l'ingénieur, disciplines dans lesquelles les chercheurs publient plus dans les *proceedings* des congrès que dans les journaux scientifiques. Bien d'autres observations passionnantes sont faites par les auteurs à partir de leurs bases de données, sans que l'on puisse les citer toutes ; par

¹ J'utiliserai cet anglicisme peu harmonieux, à mon corps défendant, tout au long de l'article. Il évoque l'idée de rupture (dont il partage l'étymologie) pour laquelle nous n'avons pas l'adjectif correspondant en français.

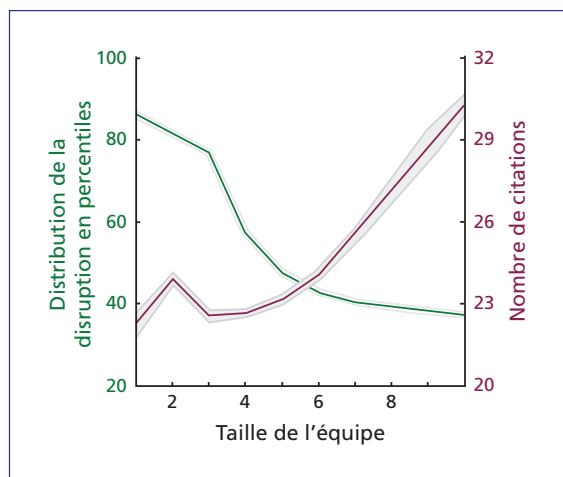


Figure 1. Degré de « disruption » apporté par un article et nombre de citations qu'il a reçues en fonction de la taille de l'équipe signataire de cet article. Les articles publiés par les petites équipes reçoivent moins de citations (23 contre 30 en moyenne) mais se situent dans des percentiles de disruption plus élevés que ceux publiés par les grandes équipes (80-100 contre 20-40).

Figure 1. The level of "disruption" of an article and the number of citations received according to the number of authors. Articles published by small teams get fewer citations (23 vs 30 on average) but are within higher percentiles for disruption than those published by larger teams (80-100 vs 20-40).

exemple, les petites équipes citent des articles plus anciens et moins « populaires » que les grandes équipes (figure 5), ce qui pourrait correspondre à des recherches plus « profondes » ; et leurs travaux sont cités plus tardivement que ceux des grandes équipes.

Je me suis posé beaucoup de questions en lisant cet article, car de nombreux biais sont *a priori* inévitables.

- Parler de « petites » et de « grandes » équipes en fonction du nombre de signataires me paraît abusif. Les « petites » équipes ont souvent besoin de faire appel à d'autres pour utiliser telle ou telle technique dont elles ne disposent pas : cela fait chaque fois quelques signataires de plus² et l'on a l'impression que l'équipe est de grande taille.
- La définition du paramètre *D* mesurant le degré de « disruption » d'un article me paraît discutable. Bien que les auteurs présentent des tests qui à leurs yeux valident ce paramètre, je crains que l'analyse serrée d'une cohorte d'une centaine ou d'un millier de publications prises au hasard dans un champ disciplinaire donné ne montre la fragilité de ce paramètre tel qu'ils l'ont défini ; par ailleurs, ce paramètre

² Exemple personnel : quand nous ne disposons pas, dans mon laboratoire, d'un cytomètre en flux, le voisin nous prêtait le sien et demandait, en contrepartie, que son nom figure sur la publication. Quand notre équipe a grandi et que nous avons pu acquérir un cytomètre, nous avons réalisé les mêmes « manips » sans avoir à ajouter un auteur supplémentaire...

repose sur une analyse de citations (celles qu'inclut l'article et celles qu'il reçoit), et quand on le compare ensuite au nombre de citations reçues, je crains que l'on entre dans un raisonnement circulaire...

• Il me semble que le mélange des champs disciplinaires les plus variés est contestable, bien que, là encore, les auteurs affirment qu'en dehors de l'informatique et des sciences de l'ingénieur les mêmes relations entre taille de l'équipe et degré de disruption d'un article subsiste. Il faudrait pouvoir réaliser ce type d'étude sur des champs disciplinaires plus étroits. Bien évidemment, le champ de l'oncologie n'est pas spécifiquement analysé dans cet article. Peut-être que ce qui est vrai pour la physique l'est également pour la biologie fondamentale des cancers, mais la recherche clinique oncologique ne se comporte sans doute pas de la même façon : les « grands » essais de phase III comprennent 20 ou 30 auteurs, chaque participant de l'essai ayant inclus quelques patients dans l'étude se voyant ainsi « récompensé ». Toutefois, de tels articles apportent souvent un changement dans les pratiques, constituent bien une « percée », mais ne sont finalement pas très cités dans la littérature ultérieure : on est à contresens des conclusions de l'article. Bien sûr, on

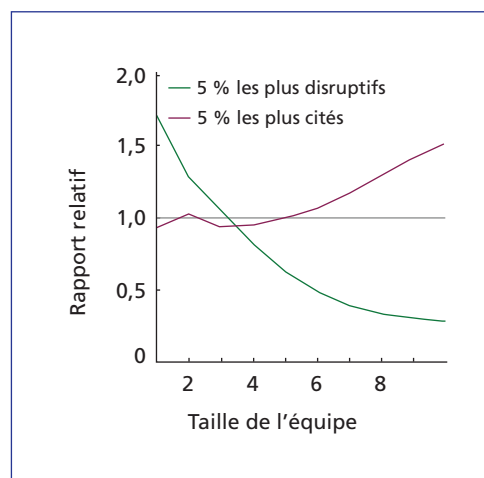


Figure 2. Proportions relatives des articles se trouvant parmi les 5 % les plus « disruptifs » et les 5 % les plus cités en fonction de la taille de l'équipe signataire de ces articles. De un à cinq auteurs, la probabilité que l'article soit dans les 5 % les plus cités ne varie pas, mais elle augmente pour un nombre d'auteurs supérieur à cinq ; la probabilité qu'un article soit parmi les 5 % les plus « disruptifs » décroît continûment en fonction du nombre d'auteurs.

Figure 2. Relative proportion of articles found among the most disruptive 5% and the most cited 5% according to the number of authors. For a team of between 1 and 5 authors, the probability that the article will be among the most cited 5% does not vary but this figure increases for teams with more than five authors. The probability of an article being among the most disruptive 5% decreases steadily according to the number of authors.

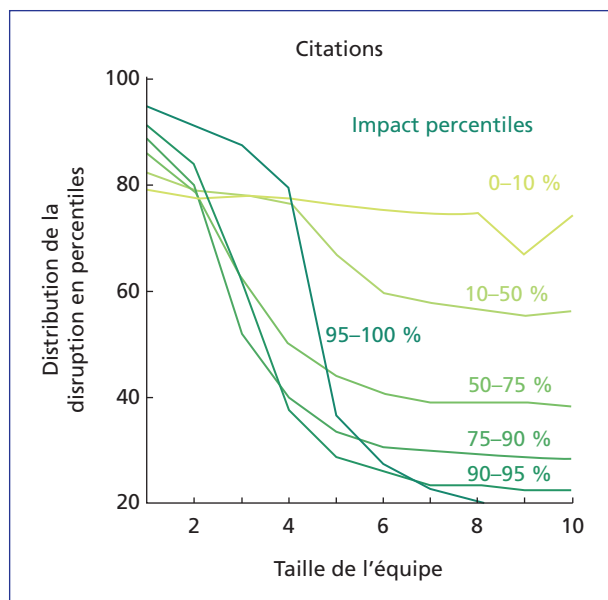


Figure 3. Degré de « disruption » apporté par un article en fonction de la taille de l'équipe signataire de cet article, en liaison avec le nombre de citations reçues par l'article... Plus un article reçoit de citations, moins le lien entre la taille de l'équipe signataire et le degré de disruption de l'article est important : les articles peu cités (percentiles 0-10) ont un degré de disruption identique quelle que soit la taille de l'équipe, alors que les articles très cités (percentiles 95-100) ont un degré de disruption élevé (percentiles 90-100) pour les petites équipes et faible (percentiles 0-20) pour les grandes équipes.

Figure 3. The level of "disruption" of an article according to the number of authors, in association with the number of citations. The more citations an article receives, the less significant the association between the number of authors and the level of disruption. Articles with few citations (percentiles 0-10) have a similar level of disruption regardless of the number of authors, whereas highly-cited articles (percentiles 95-100) have significant levels of disruption (percentiles 90-100) when authored by small teams and low levels of disruption (percentiles 0-20) when authored by larger teams.

pourrait dire qu'une « percée » thérapeutique n'est pas forcément « disruptive »...

- Un autre biais peut être celui qui consiste à pooler tous les journaux scientifiques. Dans notre discipline, nous savons bien qu'un article paru dans *Molecular Cancer Research* a peu de chances d'avoir le même degré de disruption qu'un article paru dans *Cancer Cell* : leurs *impact factors* respectifs sont de 4,6 et de 24. Les deux sont de bons journaux, mais chacun dans sa catégorie ; ils ne jouent pas dans la même cour de récréation... Il y a certainement eu des articles « disruptifs » qui ont été publiés dans le premier, et des articles « routiniers » dans le second, mais ce n'est sans doute pas le cas général ; et pourtant, un survol des tables des matières des numéros parus durant le premier trimestre de l'année 2018 dans les deux journaux montre que le nombre moyen d'auteurs par article de *Molecular Cancer*

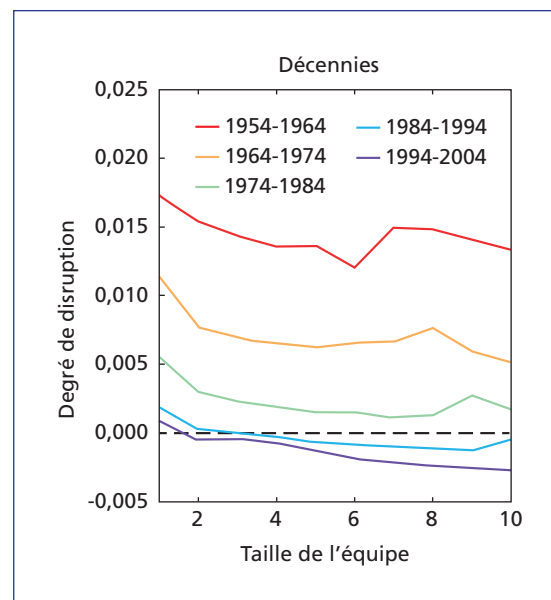


Figure 4. Évolution au cours de cinq décennies de la relation entre le degré de « disruption » apporté par un article et la taille de l'équipe signataire. De façon constante au cours du temps, le degré de disruption diminue avec la taille de l'équipe ; en outre, le degré de disruption, quelle que soit la taille de l'équipe, diminue au fil des décennies.

[Notez que les ordonnées correspondent au paramètre *D* et non à sa distribution en percentiles comme dans les figures précédentes].

Figure 4. Changes over five decades in the relationship between the degree of "disruption" an article causes and the number of authors. Over time, the level of disruption consistently decreases with the number of authors; in addition, the level of disruption, regardless of the number of authors, decreases over decades. [Note that the ordinates correspond to the parameter *D* and not to its distribution among percentiles, as in the preceding diagrams].

Research est de $9,5 \pm 4,4$ alors qu'il est de $20,2 \pm 8,2$ pour *Cancer Cell*, avec une valeur de *p* de $6,6 \times 10^{-11}$, ce qui va à l'encontre des conclusions de l'article.

Sans pouvoir mener d'analyse quantitative exhaustive, je me suis amusé à repérer les trois articles les plus « disruptifs » dans le domaine de la biologie des cancers, ceux dont je souligne l'importance à mes étudiants de master. À tout seigneur tout honneur : l'article princeps sur le concept d'oncogène [1] a été signé par quatre auteurs ; celui du groupe de Robert Weinberg montrant que la transformation maligne *in vitro* requiert au moins trois altérations moléculaires [2] a été signé par six auteurs ; et celui du groupe de Bert Vogelstein montrant (entre autres choses) que le nombre de mutations de « gènes de cancer » est très élevé dans les cancers du sein et du côlon [3] a été signé par 29 auteurs. Ces trois articles ne représentent bien sûr qu'une petite partie de tous ceux qui ont contribué de

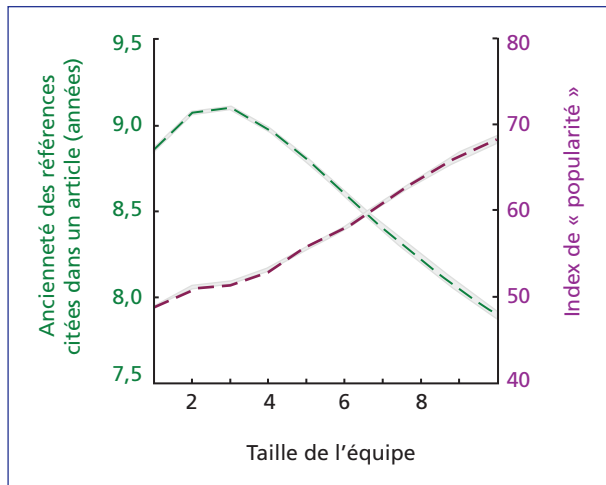


Figure 5. Ancienneté des références citées dans un article et index de « popularité » de cet article (correspondant à la rapidité avec laquelle il est cité) en fonction de la taille de l'équipe. Les petites équipes citent des articles plus anciens que les grandes équipes (neuf ans contre huit ans en moyenne), et ces articles sont cités plus tardivement que ceux des grandes équipes (index de 50 contre 70 en moyenne).
Figure 5. The timing of citation of references in articles and the popularity index of these articles (corresponding to the speed with which they were cited) according to the number of authors. Small teams cite older articles compared to larger teams (9 years vs 8 years on average), and these articles are cited later than those by larger teams (popularity index of 50 vs 70 on average).

façon majeure à notre compréhension du cancer. Le premier a reçu 1 514 citations, le second 2 572 et le troisième 3 420. Là encore, cet exemple ne va que très partiellement dans le sens des conclusions de l'article. Cela dit, je ne prétends pas que l'analyse de trois articles a la même valeur statistique que l'analyse de 24 millions d'articles ! Je me propose de faire une enquête auprès d'un panel de collègues afin d'identifier un corpus plus important d'articles « disruptifs » dans notre champ disciplinaire et de vérifier si les associations mises en évidence par les auteurs de cet article sont validées.

Les modalités de financement de la recherche en France (et sans doute ailleurs...) vont exactement à l'encontre de ce qui permettrait de faire émerger des avancées « pointues » (pour ne pas reprendre encore une fois le terme « disruptif »), au moins dans le domaine du cancer. L'Institut national du cancer (INCa) privilégie lourdement les « grandes équipes », les recherches collaboratives associant le plus possible d'équipes, les croisements multidisciplinaires. J'ai fait partie ces dernières années du comité chargé d'évaluer les projets de recherche translationnelle soumis à l'INCa et je peux dire que ce ne sont pas les projets pouvant conduire à des découvertes « disruptives » qui sont financés (je n'ai pas fait mieux que mes collègues dans cette instance, et l'INCa a certainement développé les meilleures méthodes d'évaluation que j'ai vues à l'œuvre au cours de ma carrière). Ce sont les projets « réalistes », d'une bonne « faisabilité », s'appuyant sur des travaux antérieurs et venant appuyer, confirmer, démontrer (*endorse, confirm, demonstrate*) et non pas changer, innover, promouvoir (*change, introduce, advance*), qui ont le plus de chances d'être financés (ce qui ne veut pas dire qu'ils ne sont pas, en définitive, les meilleurs... dans le cadre imposé).

Je dirai pour conclure qu'il est heureux de voir paraître d'excellents articles de sociologie de la science, une discipline peu représentée en France, où a sévi trop longtemps un post-modernisme bien peu scientifique, reposant sur des expériences individuelles contestables...

* Wu L, Wang D, Evans JA. Large teams develop and small teams disrupt science and technology. *Nature* 2019 ; 566 : 378-82. doi : 10.1038/s41586-019-0941-9.

RÉFÉRENCES

1. Stehelin D, Varmus HE, Bishop JM, Vogt PK. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* 1976 ; 260 : 170-3.
2. Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, Weinberg RA. Creation of human tumour cells with defined genetic elements. *Nature* 1999 ; 400 : 464-8.
3. Sjöblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006 ; 314 : 268-74.

Jacques Robert

Déterminants hétérozygotes de l'instabilité génomique*

L'instabilité génomique est le déterminant principal, sinon, unique, du phénotype mutateur des cellules tumorales. Il n'est pas définitivement établi qu'elle constitue le *primum movens*, la cause originelle, de l'acquisition des mutations qui caractérisent la cellule cancéreuse, mais j'en suis, à titre personnel, profondément convaincu. Elle est liée le plus souvent à la perte de l'activité de protéines responsables de la détection et de la réparation des lésions de l'ADN (mutations génétiques délétères, diminution épigénétique d'expression des

gènes codant ces protéines, etc.). La nécessité que soit perdu, de façon homozygote, le « travail » des deux allèles pour y parvenir est une objection fréquente au rôle primordial que peut jouer l'instabilité génomique dans l'oncogenèse. Cet article vient nous rappeler qu'il existe plusieurs possibilités de mise en jeu des gènes suppresseurs de tumeurs (GST) en dehors du *double hit* postulé par Knudson il y a près de 50 ans [1], tout particulièrement l'haplo-insuffisance et la dominance négative ; cela signifie que des altérations hétérozygotes de certains gènes

peuvent conduire à l'instabilité génomique, donc au phénotype mutateur, de la cellule cancéreuse.

Les auteurs se sont adressés à la levure *Saccharomyces cerevisiae* pour étudier les mutations hétérozygotes susceptibles de contribuer à l'instabilité génomique et d'identifier des mutations correspondantes dans le génome humain. Rappelons qu'il existe deux types de GST : ceux qui interfèrent avec la prolifération cellulaire, comme *RB1* ou *PTEN* ; et ceux qui gouvernent la reconnaissance et la réparation des lésions de l'ADN (en d'autres termes, la stabilité du génome), comme *TP53*, *BRCA1* ou *MSH1*. On appelait ces deux types de gènes les *gatekeepers* et les *caretakers*, respectivement. Les auteurs ont dans un premier temps sélectionné les clones diploïdes de levure qui survivent et prolifèrent après exposition séquentielle à trois molécules qui inhibent la croissance cellulaire : le 5-fluoroanthranilate, l' α -aminoadipate et la canavanine. Cela leur permet d'éliminer les cellules mutantes au niveau de GST du premier type pour ne sélectionner que les cellules ayant une mutation impliquant l'instabilité génomique.

Sur 59 clones étudiés, l'instabilité génomique est due dans 41 clones à une mutation hétérozygote unique, dans 16 clones à plusieurs mutations hétérozygotes, et dans deux clones à une mutation homozygote. Un séquençage d'exome permet alors d'identifier les gènes mutés ; et le *knock-out* de ces gènes dans la souche originale par ingénierie génétique permet de s'assurer de leur implication dans l'instabilité génomique. Un total de 92 gènes a été ainsi identifié ; 57 ont des orthologues dans le génome humain et 10 d'entre eux sont connus pour avoir un rôle dans l'oncogenèse (tableau 1). Afin de savoir si les autres gènes (ceux qui ne sont pas connus pour avoir un rôle dans l'oncogenèse) étaient capables, lorsqu'ils étaient mutés, d'être responsables d'instabilité génomique, les auteurs ont inactivé de façon mono-allélique les homologues humains de six d'entre eux dans une lignée cellulaire humaine diploïde : l'inactivation de cinq de ces gènes est effectivement responsable d'une augmentation du taux de mutations des cellules humaines (tableau 2), d'un facteur compris entre 4 et 440. Ces gènes peuvent bien évidemment être impliqués dans l'oncogenèse.

Parmi les 92 gènes dont l'inactivation conduit à l'instabilité génomique chez la levure figurent 21 gènes impliqués dans le « métabolisme » de l'ADN et des chromosomes, et les autres ont divers rôles dans la vie cellulaire, ce qui suggère que les voies auxquelles appartiennent les protéines qu'ils codent pourraient, elles aussi, jouer un rôle dans le métabolisme de l'ADN et des chromosomes. Les auteurs identifient ainsi un ensemble de nouveaux gènes pouvant être impliqués dans l'oncogenèse. Il est à noter que ce sont des mutations mono-alléliques (hétérozygotes) qui induisent ainsi l'instabilité génomique ; la survenue d'un seul événement mutationnel peut être alors rendue responsable de l'instabilité génomique et, partant, du phénotype mutateur des cellules ayant subi cette mutation, alors que nos connaissances de l'instabilité génomique privilégiaient essentiellement le mécanisme du *double hit* de Knudson.

Tableau 1. Liste des dix gènes de levure, ayant un homologue dans le génome humain jouant un rôle dans l'oncogenèse, dont l'inactivation conduit à l'instabilité génomique.

Tableau 1. List of the 10 yeast genes with a homologue in the human genome that play a role in oncogenesis and lead to genomic instability once inactivated.

Gène de levure	Gène humain	Fonction
<i>HSC82</i>	<i>HSP90AB1</i>	Protéine chaperonne
<i>MPE1</i>	<i>RBBP6</i>	Réplication de l'ADN
<i>MSH2</i>	<i>MSH2</i>	Réparation des mésappariements
<i>PAC10</i>	<i>VBP1</i>	Protéine chaperonne
<i>RAD1</i>	<i>ERCC4</i>	Endonucléase
<i>SCC2</i>	<i>NIPBL</i>	Cohésion de l'ADN
<i>SNF2</i>	<i>SMARCA4</i>	Remodelage de la chromatine
<i>TEL1</i>	<i>ATM</i>	Réparation de l'ADN
<i>TOR1</i>	<i>MTOR</i>	Survie et prolifération
<i>UFD4</i>	<i>TRIP12</i>	Réparation de l'ADN

Tableau 2. Liste de cinq gènes de levure, ayant un homologue dans le génome humain non antérieurement connu pour jouer un rôle dans l'oncogenèse, et dont l'inactivation mono-allélique conduit à l'instabilité génomique.

Tableau 2. List of five yeast genes with a human homologue, not previously known to be play a role in oncogenesis, that lead to genomic instability upon mono-allelic inactivation.

Gène de levure	Gène humain	Fonction
<i>MUS81</i>	<i>MUS81</i>	Recombinaison
<i>YAK1</i>	<i>DYRK1B</i>	Métabolisme
<i>DDL1</i>	<i>DDHD1</i>	Mitochondries
<i>TAO3</i>	<i>FRYL</i>	Cytosquelette
<i>BNI1</i>	<i>GRID2IP</i>	Cytosquelette

Cet article apporte donc des arguments pour faire de l'instabilité génomique un *primum movens* dans la naissance des cancers : une seule mutation peut induire le phénotype mutateur permettant la survenue de nombreuses mutations, parmi lesquelles pourront être sélectionnées celles qui confèrent un avantage prolifératif aux cellules.

* Coelho MC, Pinto RM, Murray AW. Heterozygous mutations cause genetic instability in a yeast model of cancer evolution. *Nature* 2019 ; 566 : 275-278. doi : 10.1038/s41586-019-0887-y.

RÉFÉRENCE

1. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 1971 ; 68 : 820-3.

Jacques Robert

Rôle du polymorphisme P72R du gène *TP53* sur le risque de survenue des cancers mammaires*

On connaît beaucoup de choses sur les mutations du gène *TP53* et leur rôle prépondérant dans l'oncogénèse et le développement des cancers [1], mais le polymorphisme P72R du gène est finalement bien moins connu. Il s'agit d'un polymorphisme commun (fréquence allélique = 0,45) dont les effets ont été surtout étudiés *in vitro*, avec des résultats contradictoires : à la suite d'un endommagement de l'ADN, le variant 72P favoriserait l'arrêt du cycle cellulaire, l'entrée en sénescence et la réparation de l'ADN alors que le variant 72R faciliterait l'entrée des cellules en apoptose. Mais le rôle de ce polymorphisme dans l'oncogénèse et le risque de survenue d'un cancer reste incertain. Les auteurs de cet article ont étudié cette question *in vivo*, sur deux modèles distincts de souris auxquelles ils ont fait exprimer de façon homozygote chacun des deux variants de la protéine p53 humaine.

Le premier modèle est celui de tumeurs mammaires de souris induites par le 7,12-diméthylbenz(a)anthracène (DMBA), le second celui de tumeurs mammaires survenant chez des souris hémizygotiques pour une mutation du gène *ErbB2*. Les deux modèles donnent des résultats concordants : dans le premier modèle, les tumeurs apparaissent 21 jours après traitement chez les souris 72R et 71 jours après traitement chez les souris 72P (figure 1) ; 50 % des souris 72P développent une tumeur, contre 100 % des souris 72R. Notons qu'en absence de traitement aucune tumeur n'apparaît, ce qui montre que le polymorphisme n'a pas d'effet spontané sur l'oncogénèse, à l'inverse des mutations délétères. Dans le second modèle, l'apparition des premières tumeurs est également plus rapide chez les

souris 72R (122 jours) que chez les souris 72P (173 jours) et la latence moyenne d'apparition des tumeurs est décalée de façon similaire (226 jours contre 250) (figure 2). Le taux de croissance tumorale est respectivement de $37,8 \pm 5,1 \text{ mm}^3$ par jour chez les souris 72R et de $22,8 \pm 5,0 \text{ mm}^3$ par jour chez les souris 72P.

Quel mécanisme peut expliquer ces différences ? Les auteurs, en remontant vers les fonctions de p53, ont recherché, dans les glandes mammaires, des différences dans l'expression d'une série de gènes cibles de p53. Aucune différence n'est observée pour les gènes codant les protéines Puma, Noxa ou Bax, alors que des différences significatives sont mises en évidence pour ceux codant p21 et p16 (figure 3), les glandes mammaires des souris 72R exprimant de façon plus élevée ces deux gènes suppresseurs de tumeurs que les glandes mammaires des souris 72P. Le résultat est une plus faible phosphorylation de la protéine RB1 chez les souris 72R, donc une moindre activation des facteurs de transcription de la famille E2F. Ces résultats peuvent sembler paradoxaux, puisque les souris 72R ont une croissance tumorale mammaire plus rapide que les souris 72P.

Les auteurs résolvent cette contradiction en s'intéressant au phénotype associant la sénescence, liée à l'arrêt du cycle cellulaire (signe habituel d'inhibition de l'oncogénèse), à l'acquisition d'un phénotype inflammatoire protumoral. Effectivement, une série de marqueurs d'inflammation, dont l'expression est parfois sous la dépendance de p53, sont surexprimés dans les glandes mammaires des souris 72R : la chimiokine Ccl2, l'interleukine Il6, le facteur de transcription RelA (un des

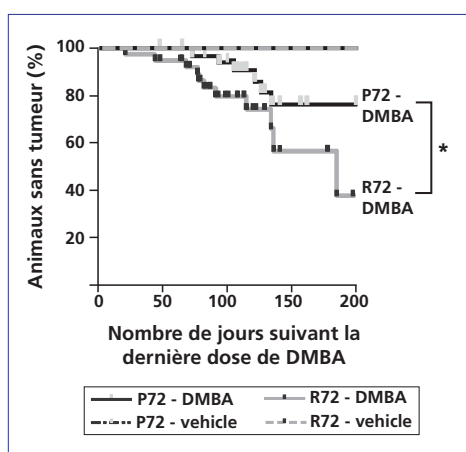


Figure 1. Apparition des tumeurs mammaires chez des souris traitées par un agent cancérogène, le diméthylbenz(a)anthracène (DMBA), selon que les souris expriment une protéine p53 « sauvage » (72P) ou « variante » (72R).

Figure 1. The occurrence of mammary tumours in mice treated with a carcinogenic agent, DMBA, in wild-type (72P) and variant (72R) p53-expressing mice.

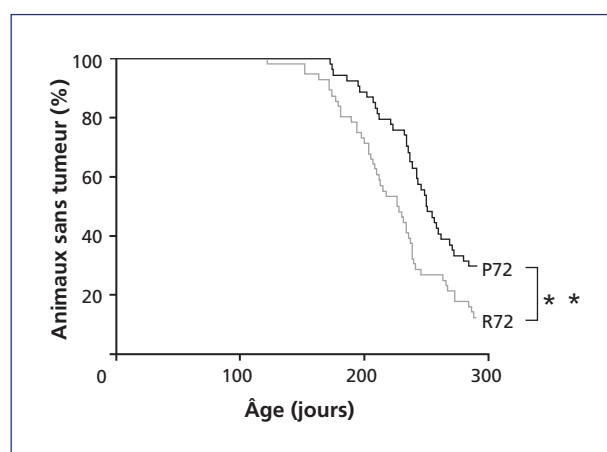


Figure 2. Apparition des tumeurs mammaires chez des souris exprimant une forme mutante oncogénique du gène *ErbB2*, selon que les souris expriment une protéine p53 « sauvage » (72P) ou « variante » (72R).

Figure 2. The occurrence of mammary tumours in mice expressing an oncogenic mutant form of the *ErbB2* gene in wild-type (72P) and variant (72R) p53-expressing mice.

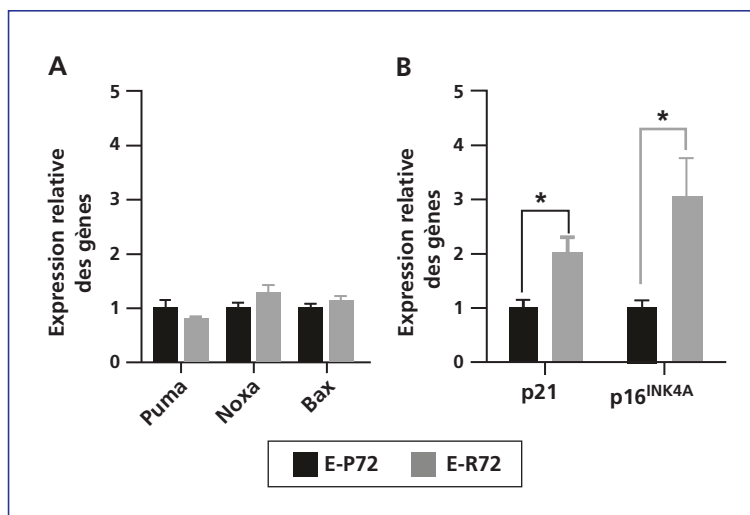


Figure 3. Expression relative de gènes codant (A) des protéines pro-apoptotiques (Puma, Noxa, Bax) ; (B) des protéines impliquées dans l'arrêt du cycle cellulaire (p21, p16) dans les glandes mammaires de souris exprimant une forme mutante oncogénique du gène *ErbB2*, selon que les souris expriment une protéine p53 « sauvage » (72P) ou « variante » (72R).
Figure 3. Relative expression of genes encoding (A) pro-apoptotic proteins (Puma, Noxa, Bax), and (B) proteins involved in cell cycle arrest (p21, p16) in mammary glands of mice expressing an oncogenic mutant form of the *ErbB2* gene in wild-type (72P) and variant (72R) p53-expressing mice.

composants du complexe Nfkb), la cytokine Tnf α , la métalloprotéinase Mmp9, etc. En outre, la présence de macrophages associés aux tumeurs, la densité vasculaire et d'autres signaux témoignent de l'état inflammatoire majeur des tumeurs mammaires des souris 72R. Il faut chercher l'origine de ces différences dans le fait que l'arginine de p53 72R donne à la protéine une plus grande affinité pour les *response elements* des gènes codant ces protéines inflammatoires, comme le montrent les expériences d'enrichissement réalisées par

immuno-précipitation de la chromatine (ChIP) suivie de réaction de polymérisation en chaîne (PCR) quantitative (*figure 4*).

Cet article est important parce qu'il apporte une pierre originale à la solution du problème de fond des variants génétiques difficilement interprétables. Classiquement, on attribue les différences fonctionnelles entre variants à une différence d'activité de la protéine ou de son expression. C'est certes le cas dans de nombreux exemples : les variants des régions codantes des gènes de

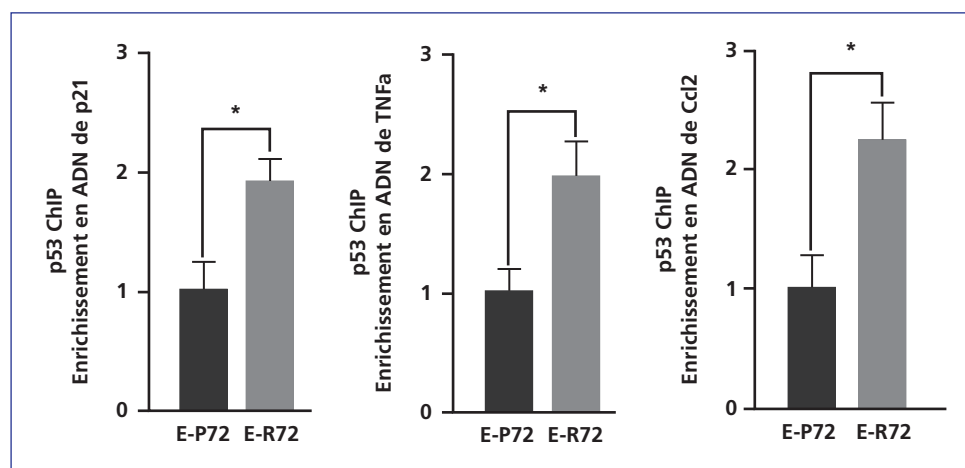


Figure 4. Résultat des expériences d'immuno-précipitation de la chromatine par p53 pour les *response elements* des promoteurs des gènes codant les protéines p21, Tnf α et Ccl2.

Figure 4. Immunoprecipitation analysis of chromatin/p53 binding to the *response elements* within the promoters of the genes encoding p21, Tnf α and Ccl2 proteins.

réparation de l'ADN aboutissent à des protéines de moindre activité expliquant le risque accru de tumeurs ; les variants des promoteurs de certains gènes comme *MDM2* aboutissent à une augmentation ou à une diminution de l'expression des protéines codées. Mais quand la variation polymorphique n'a pas de conséquence apparente sur l'activité ou l'expression de la protéine, on reste souvent dans l'incertitude : il faut aller chercher au niveau des interactions ADN-protéine, en

particulier (mais pas seulement) lorsque la protéine est un facteur de transcription, comme p53.

* Gunaratna RT, Santos A, Luo L, et al. Dynamic role of the codon 72 p53 single-nucleotide polymorphism in mammary tumorigenesis in a humanized mouse model. *Oncogene* 2019. [epub ahead of print]. doi : 10.1038/s41388-018-0630-4.

Jacques Robert

Peut-on comparer quantitativement l'effet oncogénique des mutations pilotes rencontrées dans une tumeur ?*

Comment déterminer l'importance relative des altérations génomiques qui confèrent un avantage sélectif aux cellules cancéreuses ? C'est un problème récurrent pour qui s'intéresse à l'oncogénèse... Comment distinguer les mutations « pilotes » (*driver*) des mutations « en passant » (*passenger*)³ ? Dans un des grands travaux pionniers sur le séquençage de l'ADN tumoral [1], le groupe de Vogelstein avait rencontré ce problème et tenté de définir des « gènes de cancer » à l'aide d'un algorithme un peu subjectif. Les approches se sont considérablement affinées depuis cette époque. Les auteurs de cet article souhaitent apporter des éléments de quantification permettant de mesurer l'ampleur des effets de telle ou telle mutation, et pas seulement la probabilité que cette mutation soit ou non oncogénique. Ils proposent pour cela d'évaluer l'intensité de la sélection des variants somatiques dans des lignées de cellules cancéreuses, en tant que mesure de l'avantage reproductif conféré par les mutations dans les tissus somatiques. Une telle évaluation quantitative revêt une grande importance en oncologie : l'identification des mutations pilotes d'une tumeur donnée ou d'un type tumoral particulier peut apporter une aide à la décision thérapeutique en médecine de précision « personnalisée », à la conception des essais cliniques, au développement thérapeutique et à la priorisation de la recherche fondamentale.

Classiquement, l'identification de mutations oncogéniques parmi le pool immense de mutations présentes en raison du phénotype mutateur de la cellule cancéreuse est réalisée en quantifiant les mutations surreprésentées dans une tumeur par rapport au tissu normal ; la disproportion des mutations présentes dans un gène donné est considérée comme la preuve du rôle causal de ce gène dans l'oncogénèse [2-5]. La simple constatation de cette disproportion et le calcul de la probabilité qu'une mutation donnée soit oncogénique ont longtemps suffi aux chercheurs ; une quantification plus précise permettrait de reconnaître le rôle pilote de certaines mutations dans les cancers et d'en tirer des conséquences pratiques.

Un calcul de p évalue un seuil de probabilité mais ne mesure pas l'ampleur d'un effet ; il établit une corrélation mais sûrement pas une relation causale. Les « faux positifs » sont nombreux : le gène *TNN* code une protéine structurale du muscle strié ; il est très long, il est répliqué tardivement lors de la phase S, et il est inaccessible à la réparation transcriptionnelle des tissus non musculaires ; en conséquence, il subit un taux de mutations élevé et se retrouve fréquemment muté dans les carcinomes – mais il n'est pas oncogénique. Ce type de situation peut être décelé en tenant compte, lors du calcul de p , de la longueur du gène et de son taux de mutations de base.

Comme les cancers se développent selon un modèle évolutionnaire, les auteurs de cet article ont décidé de quantifier le rôle oncogénique des mutations en mesurant leur effet sélectif sur la lignée cancéreuse ; au lieu de mesurer un taux global de mutations, ils cherchent à identifier uniquement les mutations qui sont sélectionnées en raison de leur effet sur la survie et la prolifération cellulaires. Il en résulte que le taux d'apparition de variants nucléotidiques dans une tumeur en absence de sélection est évaluable par la mesure du taux de mutations silencieuses au niveau de chaque gène. Le nombre de substitutions nucléotidiques d'un site (au-dessus de la ligne de base des mutations silencieuses) représente l'intensité exercée par la sélection sur cette mutation dans la population cellulaire. L'intensité de sélection peut ainsi quantifier l'avantage prolifératif conféré aux variants et permet de comparer l'importance relative des mutations pilotes. Les auteurs ont utilisé les données génomiques du TCGA (*The Cancer Genome Atlas*) et d'autres bases de données comme la CCLE (*Cancer Cell Line Encyclopedia*). L'intensité de sélection est évaluée site par site en divisant le taux de substitutions observé pour ce site par le taux de substitutions attendu en absence de sélection ; ce dernier est évalué par le taux de mutations synonymes survenant sur ce site. Le principe est finalement très simple mais sa mise en œuvre requiert des moyens informatiques importants.

Un exemple de ce qui peut être obtenu est représenté sur la *figure 1* qui présente les 25 mutations ayant l'intensité de sélection la plus élevée dans les adénocarcinomes et les

³ Je me refuse à utiliser le terme de mutations « passagères », ce mot ayant aussi le sens de « transitoire » en français.

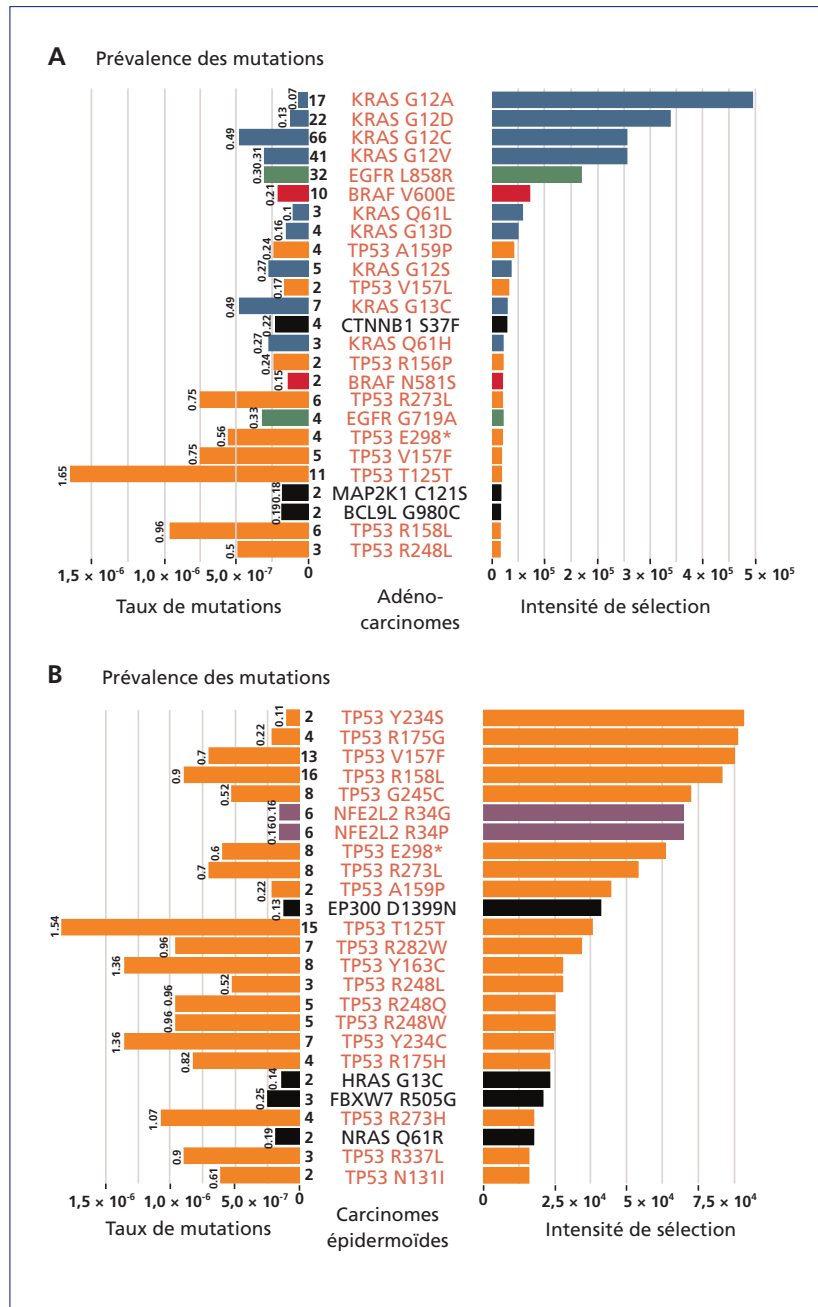


Figure 1. Liste des 25 mutations ayant l'intensité de sélection la plus élevée dans des collections de 675 adénocarcinomes (A) et 600 carcinomes épidermoïdes (B) du poumon. Les gènes et leurs variants sont indiqués au centre des figures ; les intensités de sélection sont classées par ordre décroissant, et représentées par des barres dans la partie droite des figures ; le nombre de tumeurs présentant la mutation (prévalence des variants) est indiqué à gauche des noms des gènes ; le taux de substitutions de chaque site est représenté par des barres dans la partie gauche des figures et la valeur de ce taux est indiquée à gauche de ces barres. L'intensité de sélection est proportionnelle au quotient de la prévalence des variants par le taux de substitutions du site.

Figure 1. List of the 25 mutations based on the highest level of selection in (A) lung adenocarcinomas (675 samples) and (B) squamous cell carcinomas (600 samples). The genes and their variants are indicated in the middle of the figures. The strength of the selection is ranked in decreasing order and presented as bars on the right side of the figures. The number of tumours harbouring the mutation (variant prevalence) is indicated to the left of the gene names. The substitution rate for each site is represented by bars on the left side of the figures and the value of this rate is indicated to the left of these bars. The strength of selection is proportional to the quotient of prevalence of variants according to the substitution rate.

carcinomes épidermoïdes du poumon. On peut voir que l'importance relative des mutations de *TP53*, *KRAS*, *EGFR*, *BRAF* et autres oncogènes varie considérablement entre les deux types de cancers : ce sont les mutations du codon 12 de *KRAS* qui sont les plus déterminantes pour les adénocarcinomes et les diverses mutations délétères de *TP53* qui le sont pour les carcinomes épidermoïdes. Dans un type tumoral donné (adénocarcinomes du poumon), l'intensité de sélection est clairement découplée de la fréquence des mutations : la mutation L858R de *EGFR* est presque deux fois plus fréquente que la mutation G12A de *KRAS*. Pourtant, l'ampleur de l'effet de cette dernière est très supérieure à la première, suggérant un rôle plus déterminant dans l'oncogenèse puisqu'elle confère un avantage sélectif plus important aux tumeurs qui la contiennent. Par ailleurs, une même mutation peut avoir un rôle sélectif plus important dans un type tumoral que dans un autre ; c'est le cas de la mutation V600E de *BRAF*, dont l'intensité de sélection est beaucoup plus élevée dans les mélanomes que dans les adénocarcinomes pulmonaires.

Je laisse le lecteur découvrir dans l'article quelles sont les mutations conférant un avantage sélectif majeur à chacun des 22 types de cancers étudiés, et surtout les différences qui existent entre des cancers de type (d'étiologie ?) différent dans le même organe (cancers des voies aérodigestives supérieures positives ou négatives

pour HPV, cancers du sein exprimant ou non les récepteurs hormonaux). Cette approche quantitative remarquable apporte un regard précis sur les mécanismes différentiels de l'oncogenèse selon l'organe et l'histologie, et cet article me paraît être l'un des plus « disruptifs »⁴ que j'ai lus depuis longtemps... et il est signé par trois auteurs seulement.

* Cannataro VL, Gaffney SG, Townsend JP. Effect sizes of somatic mutations in cancer. *J Natl Cancer Inst* 2018 ; 110 : 1171-7.

RÉFÉRENCES

1. Sjöblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006 ; 314 : 268-74.
2. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014 ; 505 : 495-501.
3. Martincorena I, Raine KM, Gerstung M, et al. Universal patterns of selection in cancer and somatic tissues. *Cell* 2017 ; 171 : 1029-41.
4. Armenia J, Wankowicz SAM, Liu D, et al. The long tail of oncogenic drivers in prostate cancer. *Nat Genet* 2018 ; 50 : 645-51.
5. Kan Z, Jaiswal BS, Stinson J, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* 2010 ; 466 : 869-73.

Jacques Robert

Nouvelles mutations oncogéniques d'*ERBB2* (HER2) dans les cancers*

Parmi nos 58 récepteurs à activité tyrosine kinase (RTK), nombreux sont ceux à pouvoir être impliqués dans l'oncogenèse. Le récepteur HER2, codé par le gène *ERBB2*, est sans doute, avec le récepteur EGFR, le plus étudié en raison de son implication dans plusieurs cancers majeurs (sein et estomac notamment) et celui pour lequel on dispose du plus grand nombre d'inhibiteurs, anticorps ou petites molécules inhibitrices de son activité catalytique (ITK). L'activation oncogénique la plus fréquente d'HER2 est l'amplification du gène *ERBB2*, à l'origine d'une surexpression du récepteur, qui survient dans près de 20 % des cancers du sein. Il existe également des mutations oncogéniques bien connues dans le domaine tyrosine kinase, comme dans la plupart des RTK. Mais les autres mutations que l'on rencontre dans ce gène restent mal connues quant à leur rôle oncogénique potentiel et quant à la sensibilité des récepteurs mutés aux médicaments anti-HER2.

Les auteurs ont recherché les mutations d'*ERBB2* dans plus de 100 000 tumeurs provenant de 54 types de cancers (et de 400 sous-types). Rappelons que le caractère potentiellement oncogénique d'*ERBB2* avait été découvert dans un modèle de glioblastome de rat induit par la nitrosoéthylurée (en abrégé, neu), et était dû à une mutation présente dans le domaine transmembranaire du récepteur,

correspondant chez l'homme à la position V659E. La plupart des 3 851 mutations observées (3,5 %) siègent dans le domaine extracellulaire et le domaine kinase (figure 1). Toutefois, les auteurs se sont intéressés aux mutations les moins connues, présentes dans les domaines transmembranaire (TMD, 27 acides aminés) et juxtamembranaire (JMD, 39 acides aminés) du récepteur. La mutation la plus fréquente du TMD (G660D) concerne 16 cas et la plus fréquente du JMD (R678Q/P/W) concerne 192 cas.

L'étude du récepteur par résonance magnétique nucléaire a permis de mettre en évidence une zone polaire sur une des faces de la partie N-terminale du TMD, et les mutations les plus fréquentes du TMD (V659E et G660D/R) amplifient ce caractère polaire. Les auteurs ont alors étudié, après mutagenèse dirigée, l'effet de toutes les mutations possibles des 27 résidus du TMD et d'une partie du JMD générant ainsi un ensemble de 836 mutants. Ces mutants ont servi à infecter des cellules BaF3 en présence ou non d'*ERBB2* sauvage, afin de mimer les mutations humaines hétérozygotes et homozygotes respectivement. L'enrichissement au cours du temps en cellules portant telle ou telle mutation permettait d'évaluer le rôle oncogénique de chaque mutation.

⁴ Voir la première note de « Sélection et Décryptage » de ce numéro.

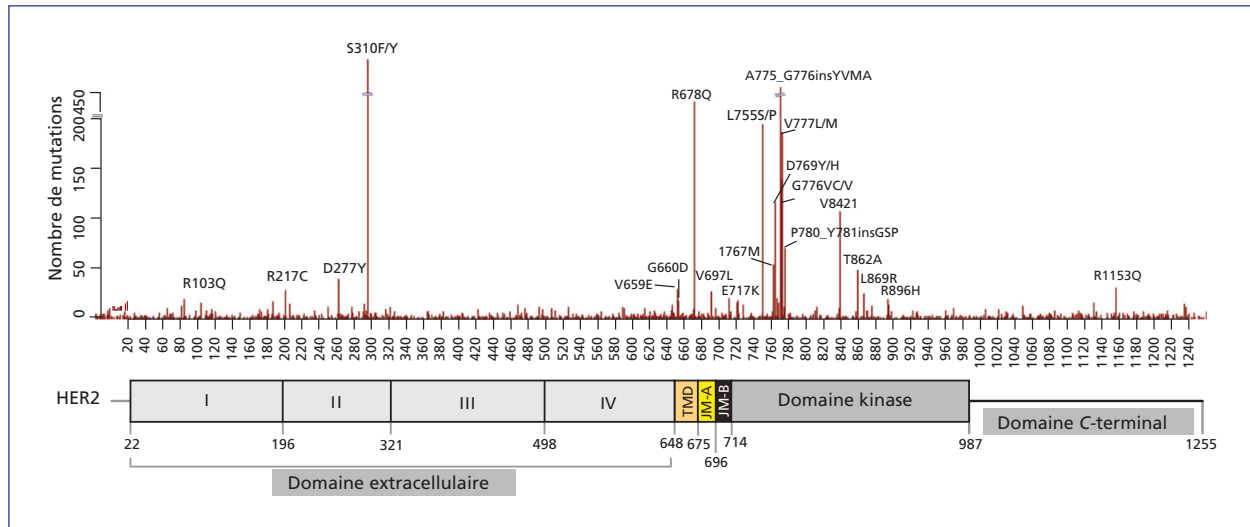


Figure 1. Distribution des mutations du gène *ERBB2* rencontrées dans 111 176 tumeurs de patients atteints de 54 types de cancers différents. On remarque le *hot spot* F310S/Y du domaine extracellulaire ainsi que ceux, assez classiques, présents dans le domaine kinase. Les mutations du TMD et du JMD sont moins fréquentes, à l'exception de la mutation R678Q, présente dans 0,17 % des échantillons.

Figure 1. Distribution of *ERBB2* gene mutations found in 111,176 tumours from patients with 54 different types of cancer. The F310S/Y hot spot is present in the extracellular domain and classic hot spots are present in the kinase domain. TMD and JMD mutations are less frequent, with the exception of R678Q mutation which is present in 0.17% of samples.

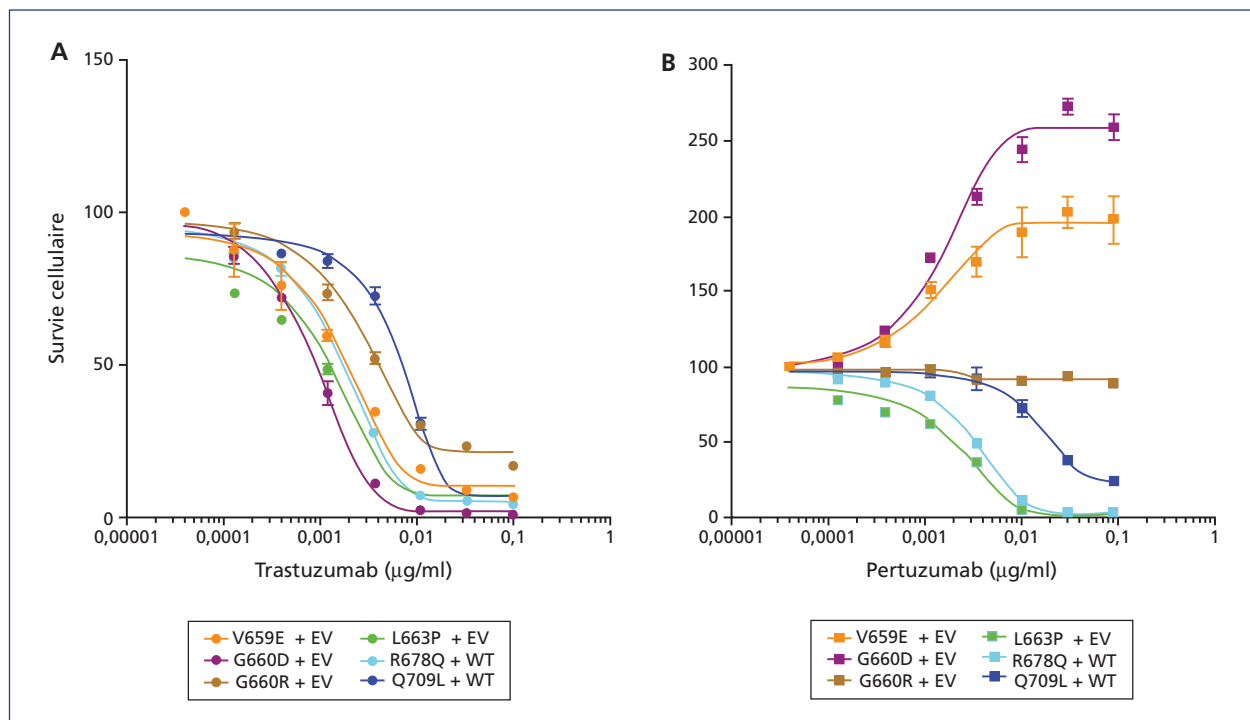


Figure 2. Sensibilité au trastuzumab (A) et au pertuzumab (B) des cellules exprimant le gène *ERBB2* porteur de diverses mutations, soit à l'état hétérozygote (+WT), soit à l'état homozygote (+EV, c'est-à-dire vecteur vide). Le trastuzumab est actif sur les mutations homozygotes et hétérozygotes, indifféremment, alors que le pertuzumab n'est actif que sur les mutations hétérozygotes, stimulant même la prolifération des cellules portant la mutation à l'état homozygote.

Figure 2. Sensitivity of cells expressing the *ERBB2* gene, either heterozygously (+WT) or homozygously (+EV, for empty vector), to trastuzumab (A) and pertuzumab (B). Trastuzumab is active in cells with both homozygous and heterozygous mutations, whereas pertuzumab is active in cells with only heterozygous mutations, and even stimulates the proliferation of cells with homozygous mutations.

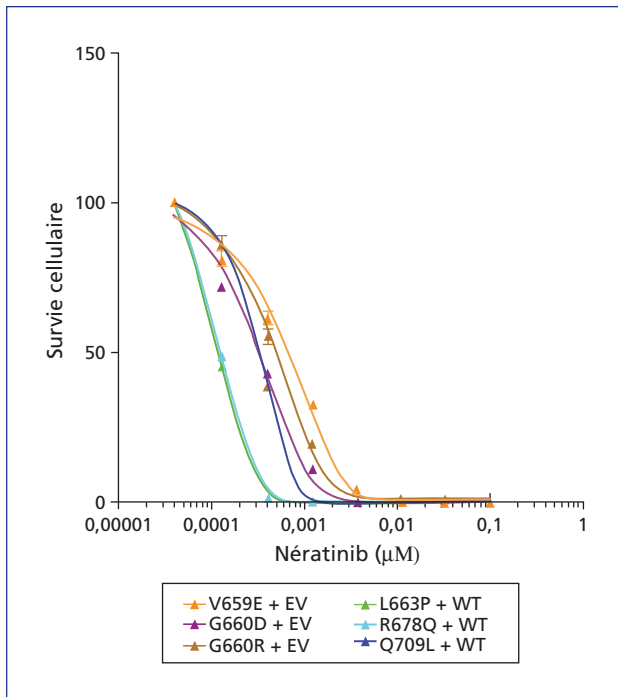


Figure 3. Sensibilité au nératinib des cellules exprimant le gène *ERBB2* porteur de diverses mutations, soit à l'état hétérozygote (+WT), soit à l'état homozygote (+EV, c'est-à-dire vecteur vide). Toutes les mutations sont sensibles au médicament.

Figure 3. Sensitivity of cells expressing the *ERBB2* gene with various mutations, either heterozygously (+WT) or homozygously (+EV, for empty vector), to neratinib. All mutations are sensitive to the drug.

Ce sont précisément les mutations rencontrées dans les échantillons cliniques (plus quelques autres) qui sont portées par les cellules ayant la plus forte prolifération, et ce, en présence ou en absence du gène *ERBB2* sauvage, les mutations G660D et V659E étant les plus représentées. À l'inverse, la plupart des mutations rencontrées dans les échantillons cliniques sont associées à une stimulation de la prolifération, en présence ou en absence du récepteur sauvage selon les cas. Je ne décrirai pas ici l'analyse conformationnelle des protéines mutantes, qui a permis de comprendre le mécanisme par lequel ces mutations,

situées hors du domaine kinase, sont susceptibles d'activer le récepteur et de rendre les cellules « autosuffisantes en facteurs de croissance ». Ce travail de biologie structurale est certainement la partie la plus originale et la plus difficile de l'expérimentation. Disons pour faire simple que ce mécanisme implique les interactions entre monomères au niveau de leur interface permettant de constituer le dimère actif, et/ou la stabilisation de la conformation active du récepteur.

Plus intéressante pour nous est d'abord la confirmation que le domaine kinase est indispensable au fonctionnement du récepteur pour assurer survie et prolifération, mais surtout que les différents outils disponibles pour inhiber le récepteur, anticorps anti-HER2 (trastuzumab et pertuzumab) et les ITK anti-HER2 (lapatinib, nératinib, afatinib) sont capables de bloquer l'activité des récepteurs mutants au niveau du TMD et du JMD, et permettent des réponses thérapeutiques majeures. Je ne donnerai que quelques exemples des nombreuses courbes de cytotoxicité des cellules BaF3 présentées dans l'article ([figures 2 et 3](#)) : elles se passent de commentaires. Une confirmation de l'efficacité du trastuzumab a été apportée *in vivo* sur des tumeurs murines obtenues par greffe de cellules BaF3 G660D. Enfin, une observation clinique intéressante est rapportée dans l'article : les membres d'une famille indienne (un frère, une sœur et leur cousine) ont présenté un adénocarcinome pulmonaire à l'âge de 53, 47 et 41 ans ; la mutation germinale G660D d'*ERBB2* a été découverte chez ces patients. Ils avaient été traités sans succès par pémétréxed et carboplatine, puis erlotinib. L'afatinib en 4^e ligne a permis une réponse majeure et surtout une survie de plus de 15 mois. On peut en conclure que les patients présentant, au niveau germinale et sans doute aussi au niveau somatique, une des mutations d'*ERBB2* au niveau du TMD ou du JMD, peuvent être traités par ITK ; ce ne peut guère être mis en œuvre que lors de séquençages d'exome systématiques, dans le cadre de programmes dédiés... ce qui limite l'importance pratique du travail présenté dans cet article !

* Pahuja KB, Nguyen TT, Jaiswal BS, et al. Actionable activating oncogenic *ERBB2/HER2* transmembrane and juxtamembrane domain mutations. *Cancer Cell* 2018 ; 34 : 792-806.

Jacques Robert