Étude de l'impact des microARNs sur la carcinogenèse des cancers colorectaux instables sur les séquences répétées microsatellites du génome

Nizar El-Murr

▶ **To cite this version:**

Nizar El-Murr. Étude de l'impact des microARNs sur la carcinogenèse des cancers colorectaux instables sur les séquences répétées microsatellites du génome. Interactions cellulaires [q-bio.CB]. Université Pierre et Marie Curie - Paris VI, 2014. Français. NNT : 2014PA066038 . tel-00990895

HAL Id: tel-00990895

https://theses.hal.science/tel-00990895

Submitted on 14 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE PIERRE ET MARIE CURIE

Ecole doctorale Physiologie et Physiopathologie (ED 394)

Centre de Recherche Saint-Antoine UMR_S 938 INSERM/UPMC

Equipe « Instabilité des Microsatellites et Cancers »

Étude de l'Impact des MicroARNs sur la Carcinogenèse des Cancers Colorectaux Instables sur les Séquences Répétées Microsatellites du Génome

Par Nizar EL-MURR

Thèse de doctorat de Génétique et Epigénétique des Cancers

Dirigée par le Docteur Alex DUVAL

Présentée et soutenue publiquement le 19 février 2014

Devant un jury composé de :

Dr Christian GESPACH	DRCE INSERM	Président
Dr Alexandra HENRION CAUDE	CR1 INSERM	Rapporteur
Dr Eric OGIER-DENIS	DR1 INSERM	Rapporteur
Dr Harry SOKOL	MCU-PH	Examinateur
Dr Alex DUVAL	DR1 INSERM	Directeur de thèse
Dr Magali SVRCEK	MCU-PH	Co-encadrante

RESUME

La progression tumorale MSI (Microsatellite Instable) est décrite comme un processus multi-étapes résultant de mutations générées par un processus d'instabilité génétique qui affecte en majorité les motifs répétés en tandem de l'ADN (microsatellites). Ces mutations contribuent à l'oncogenèse lorsqu'elles perturbent la fonction d'oncogènes ou de gènes suppresseurs de tumeurs (microsatellites codants ou non codants avec un rôle fonctionnel, *i.e.* épissage, région promotrice, …). Le trait phénotypique MSI est consécutif à l'inactivation du système de réparation des mésappariements de l'ADN (système MMR). Dans ce travail, je me suis intéressé au rôle des microARNs dans l'oncogenèse MSI. Les microARNs régulent l'expression de nombreux gènes pouvant avoir un rôle clé dans le cancer. J'ai fait l'hypothèse d'un rôle de ces microARNs lors des différentes étapes du processus tumorigénique MSI, en choisissant le cancer colorectal (CCR) comme modèle d'étude :

(1) Les microARNs et l'initiation des CCR de type MSI : Les miRs 155 et 21 sont deux miARNs dont la surexpression est associée à l'inflammation. Ces miRs ciblent en particulier les principales protéines du système MMR. J'ai observé leur surexpression dans la muqueuse non néoplasique et dans la muqueuse transformée de patients souffrant de maladies inflammatoires chroniques de l'intestin (MICI), en particulier chez ceux développant un CCR MSI. De manière intéressante, le laboratoire d'accueil avait démontré que l'émergence de cancers MSI chez ces patients n'était pas consécutive à des mutations affectant les gènes MMR ni à la méthylation des promoteurs de ces gènes. Ces travaux indiquent donc la possibilité d'un mode d'initiation tumorale MSI originale passant par le miR-155 et sa surexpression au niveau de zones de muqueuse colique encore non transformée chez les malades MICI (notion d'effet de champs favorisant l'émergence de clones MMR-déficients).

(2) Les microARNs comme cibles de l'instabilité microsatellitaire dans les CCR : À ce jour, la carcinogenèse colorectale de type MSI est pensée comme un processus dépendant en majorité de la survenue de mutations affectant des répétitions codantes localisées dans des gènes liés au cancer. Notre hypothèse a été que les gènes de miARN contenant de tels microsatellites pourraient également constituer des cibles privilégiées de mutations dans la

cellule tumorale MSI et intervenir dans le processus tumorigénique. Nos résultats montrent que les séquences microsatellites contenues dans les miRs sont généralement peu touchées par MSI, et que celles qui sont fréquemment mutées (miR-1303, miR-567 et miR-1273c) n'impactent pas significativement la biogenèse et l'expression du miR mature. A titre d'exception, le miR-3613 contient une répétition qui est localisée dans le miR mature dont l'instabilité est fréquente dans les CCR et semble modifier sa biogenèse dans la cellule tumorale ; brièvement, ces mutations ne changent pas la séquence *seed* mais l'extrémité 3' du miR-3613-5p mature (notion d'IsomiRs). Les isomiRs produits ont un répertoire de cibles qui pour certaines sont communes à la forme sauvage et pour d'autres spécifiques à chacun des variants. Le rôle de ces « variants » et l'impact qu'ils pourraient avoir dans l'oncogenèse MSI reste à découvrir.

(3) Les protéines impliquées dans la biogenèse des miRs comme cibles de l'instabilité microsatellitaire : un modèle récent propose que la signature miARN spécifique des cancers MSI soit due à des mutations somatiques affectant les répétitions codantes de gènes de la machinerie de biogenèse des miRs *(XPO-5, TARBP2)*. Le criblage que j'ai réalisé sur une large série de lignées cellulaire et de CCR MSI primitifs permet d'infirmer cette hypothèse de travail, les mutations de ces gènes candidats n'étant que très rarement observées dans ces cancers. D'autres mécanismes sont donc à investiguer, ce qui est en perspective dans le laboratoire d'accueil de par la réalisation d'une méta-analyse d'une série de CCR MSI incluant l'étude de leurs caractéristiques Omiques (Mirnome, Génome, Transcriptome, Méthylome, Exome).

Mots-clés : microARNs ; génétique ; épigénétique ; cancer colorectal ; instabilité des microsatellites ; mutations somatiques ; gènes cibles ; initiation tumorale ; progression tumorale ; maladies inflammatoires chroniques intestinales ; défaut de champs ; miR-155 ; miR-21 ; miR-3613.

ABSTRACT

MSI tumor progression (Microsatellite Instability) is depicted as a multistage process that results from mutations generated by a process of genetic instability affecting mostly DNA tandem repeats (known as microsatellites). These mutations contribute to tumorigenesis when they disrupt the function of oncogenes or tumor suppressor genes (coding or non-coding microsatellites with a functional role, *i.e.* splicing, promoter region …). As a phenotypic trait, MSI is the consequence of DNA mismatch repair inactivation (MMR). This work focused on the role microRNAs might play in MSI tumorigenesis. MicroRNAs regulate the expression of numerous genes and are deregulated in cancer. I have hypothesized a role of theses microRNAs during the various stages of the MSI tumorigenic process, choosing colorectal cancers (CRC) as a working model:

(1) MiRNAs and MSI CRC tumor initiation: miRs 155 and 21 overexpression has been associated with inflammation. These miRs target the principal proteins of the MMR system. I observed their overexpression in the non-neoplastic mucosa and transformed mucosa of patients suffering of Inflammatory Bowel Diseases (IBD), particularly in those developing MSI CRCs. Interestingly, our laboratory had demonstrated that the emergence of MSI cancers in these patients was neither the result of inherited mutations affecting MMR genes nor was it due to promoter hypermethylation. These results suggest the possibility for an original mode of MSI tumor initiation through miR-155 overexpression at the level of non-cancerous colic mucosa in IBD patients (a “Field defect” favoring the emergence of MMR deficient clones).

(2) MiRNAs as targets of microsatellite instability in MSI CRCs: up till today, MSI colorectal carcinogenesis has been thought of as a process mainly dependent on the occurrence of mutations affecting coding repeats localized in cancer related genes. Our hypothesis was that miRNA genes, which might contain such sequences, could also constitute preferred targets for mutations in MSI tumor cells and intervene in the MSI tumorigenic process. Our results show that microsatellite sequences contained within microRNA genes are generally rarely affected by MSI, and that those which are frequently mutated (miR-1303, miR-567 and miR-1273c) have no significant impact on the biogenesis and the expression of mature

miRs. Exceptionally, two microsatellite repeats are fully embedded within mature miR-3613 duplex. Their instability is frequent in MSI CRCs and seems to modify miR-3613 biogenesis in tumor cell. Briefly, these mutations do not seem to alter the *seed* sequence but the 3'-end of mature miR-3613-5p (production of IsomiRs). IsomiRs display a signature among which some mRNA targets are common to the wild form while others are specific to each variant. The role of these "variants" and the impact they could have on MSI tumorigenesis is still to uncover.

(3) Proteins implicated in miRNA biogenesis as targets of microsatellite instability: a recent model suggests that the MSI CRC-specific miRNA signature is due to somatic mutations that affect coding repeats of miRNA-associated genes *(XPO-5, TARBP2)*. Screening a large series of MSI CRC cell lines and primary tumors allowed refuting this hypothesis since mutations in these candidate genes are rarely observed in MSI CRCs. Thus, other mechanisms are still to be investigated, which is in prospect in our laboratory by the realization of a meta-analysis of a series of MSI CRCs and the characterization of their "Omics" (Mirnome, Genome, Transcriptome, Methylome, and Exome).

Keywords: microRNA; genetics; epigenetic; colorectal cancer; microsatellites instability; somatic mutations; MSI target genes; tumor initiation; tumor progression; inflammatory bowel diseases; field defect ; miR-155 ; miR-21 ; miR-3613.

TABLE DES MATIERES

RESUME ..... **i**

ABSTRACT ..... **iii**

TABLE DES MATIERES ..... **v**

TABLE DES ILLUSTRATIONS ..... **viii**

LISTE DES TABLEAUX ..... **ix**

LISTE DES ABREVIATIONS ..... **ix**

AVANT-PROPOS ..... **xi**

REVUE BIBLIOGRAPHIQUE **1**

PARTIE I – LES MICROARNS **2**

1. INTRODUCTION – LA DECOUVERTE DES MICROARNS **2**

2. LA BIOGENESE DES MICROARNS **4**

2.1 ORGANISATION GENOMIQUE DES GENES DE MICROARNS 4

2.2 BIOGENESE DES MIARNS : APERÇU GENERAL 6

2.2.1 Le Modèle canonique 6

2.2.2 Les modèles non canoniques de biogenèse des miRs 8

2.3 LES BASES MOLECULAIRES DE LA BIOGENESE DES MIARNS 11

2.3.1 Les RNases III, Drosha et Dicer 11

2.3.2 L'étape catalysée par Drosha 12

2.3.3 Transport nucléocytoplasmique des pre-miRs 13

2.3.4 L'étape catalysée par Dicer 14

3. LES MECANISMES D'ACTION DES MICROARNS **21**

3.1 LE COMPLEXE RISC 21

3.2 LA RECONNAISSANCE DES ARNM CIBLES 26

3.3 LES DIFFERENTS MODES D'ACTION DES MIARNS 30

4. VARIATIONS GENETIQUES : SNP ET MUTATIONS SOMATIQUES **34**

4.1 SNP, *SINGLE-NUCLEOTIDE POLYMORPHISM* DANS LES GENES DE MICROARNS 34

4.2 DES MUTATIONS GERMINALES ET SOMATIQUES DANS LES GENES DE MICROARNS 36

5. MODIFICATIONS POST-TRANSCRIPTIONNELLES DES MIARNS **37**

5.1 MISE EN EVIDENCE 37

5.2 FREQUENCE 40

5.3 BASES MOLECULAIRES 42
5.3.1 Le rôle de Drosha et DICER1 42
5.3.2 Le rôle des exoribonucléases et des nucléotidyltransférases 43
5.3.3 ADAR et editing des microARNs 43
5.4 CONSEQUENCES FONCTIONNELLES 47
5.4.1 Variabilité à l'extrémité 5' : liaison aux Argonautes et changement de cibles 47
5.4.2 Editing : conséquences physiologiques et physiopathologiques 47
5.4.3 Conséquences de l'uridylation et de l'adénylation 50

PARTIE II – MICROARNS ET CANCERS COLORECTAUX MSI 53

6. CANCERS COLORECTAUX INSTABLES SUR LES MICROSATELLITES 53

6.1 GENERALITES SUR LES CANCERS COLORECTAUX 53
6.2 LE SYSTEME MMR ET SA PERTE DANS LES CANCERS COLORECTAUX 57
6.3 GENES CIBLES DE L'INSTABILITE MICROSATELLITAIRE 59
6.4 CARACTERISTIQUES CLINICO-PATHOLOGIQUES DES CANCERS COLORECTAUX DE TYPE MSI 62
6.4.1 Cancers MSI sporadiques et héréditaires 62
6.4.2 Pronostic tumoral et réponse aux traitements de chimiothérapie 63

7. MICROARNS ET DEREGULATION DE LEUR EXPRESSION DANS LES CCRS 64

7.1 INTRODUCTION : MICROARNS ET CANCERS 64
7.2 GENERALITES SUR LES MICROARNS DANS LES CANCERS COLORECTAUX 67

8. MICROARNS ET CANCERS COLORECTAUX MSI 71

8.1 DEREGULATION DES MICROARNS DANS LES CANCERS COLORECTAUX MSI 71
8.2 VARIATIONS GENETIQUES ET MSI DANS LES GENES DE BIOGENESE DES MICROARNS 75
8.3 LES PROTEINES MMR SONT LA CIBLE DE CERTAINS MICROARNS 78

9. CANCER, INFLAMMATION ET MICROARNS 81

9.1 LA NOTION DE DEFAUT DE CHAMP 81
9.2 CANCERS COMPLIQUANT LES MALADIES INFLAMMATOIRES CHRONIQUES INTESTINALES 83
9.2.1 Qu'est-ce que les MICI ? 83
9.2.2 Les MICI peuvent se compliquer par des cancers colorectaux 84
9.2.3 Développement et progression de cancers colorectaux compliquant les MICI 85
9.2.4 Comment l'inflammation peut-elle favoriser l'apparition des CCR MSI ? 87
9.3 INFLAMMATION ET CANCER : LE ROLE DES MICROARNS 91
9.3.1 Inflammation et expression des miARNs 91
9.3.2 Plusieurs microARNs se trouvent dérégulés dans les MICI 93
9.3.3 Rôles multiples des miRs 155 et 21 96

TRAVAUX DE RECHERCHE 98

VUE SYNOPTIQUE DES TRAVAUX DE RECHERCHE 99

DISCUSSION 131

PERSPECTIVES 144

CONCLUSION 149

ANNEXE 152

BIBLIOGRAPHIE 176

TABLE DES ILLUSTRATIONS

FIGURE 1. LOCALISATION GENOMIQUE ET STRUCTURE DES GENES DE MIARNS. 5

FIGURE 2. LA BIOGENESE DES MIARNS..... 10

FIGURE 3. LES DIFFERENTS DOMAINES MOLECULAIRES CONSTITUANT LES RNASES III DROSHA ET DICER. 17

FIGURE 4. SEQUENCES ET MOTIFS STRUCTURAUX NECESSAIRES A LA RECONNAISSANCE ET AU CLIVAGE DU PRI-MIR. 18

FIGURE 5. TRANSPORT NUCLEOCYTOPLASMIQUE DES PRE-MIRS. 19

FIGURE 6. STRUCTURE ET MODE D'ACTION DE DICER. 20

FIGURE 7. STRUCTURE CRISTALLINE D'AGO2 ET MODELISATION DU RLC. 25

FIGURE 8. LES DIFFERENTS TYPES D'INTERACTION ENTRE MIARNS ET ARNM CIBLES. 29

FIGURE 9. REGULATION DES ARNM CIBLES PAR LES MIARNS..... 33

FIGURE 10. REPRESENTATION SCHEMATIQUE DES DIFFERENTES ESPECES D'ISOMIRS PRODUITES A PARTIR D'UN MEME PRECURSEUR. 39

FIGURE 11. DIFFERENTES HYPOTHESES EXPLIQUANT L'HETEROGENEITE DES EXTREMITES DES MIARNS. 45

FIGURE 12. STRUCTURE DES DOMAINES *MID* ET *PAZ* DE AGO2 QUI LIENT LES EXTREMITES 5' ET 3' RESPECTIVEMENT. 45

FIGURE 13. EDITING A EN I DE L'ARN..... 46

FIGURE 14. REGULATION DE LA BIOGENESE ET DE LA FONCTION DES MICROARNS PAR *EDITING*. 52

FIGURE 15. MODELE DE PROGRESSION MULTI-ETAPES DES CANCERS COLORECTAUX. 56

FIGURE 16. LE SYSTEME DE REPARATION DES MESAPPARIEMENTS ET INSTABILITE MICROSATELLITAIRE. 58

FIGURE 17. MODELE DE PROGRESSION DES CANCERS COLORECTAUX DE TYPE MSI. 61

FIGURE 18. LES MICROARNS EN TANT QUE SUPPRESSEURS DE TUMEURS ET ONCOGENES. 66

FIGURE 19. LA DEREGULATION DES MICROARNS DANS LA PROGRESSION TUMORALE MULTI-ETAPES DES CANCERS COLORECTAUX. ... 70

FIGURE 20. LES DIFFERENTES VOIES DE CARCINOGENESE COLORECTALE DE TYPE MSI. 80

FIGURE 21. MISE EN PLACE ET EVOLUTION D'UN DEFAUT DE CHAMP. 82

FIGURE 22. LES MICI SE COMPLIQUENT PARFOIS PAR L'APPARITION D'UN CANCER. 90

FIGURE 23. CONSEQUENCES CLINIQUES DE LA PERSISTANCE D'UN DEFAUT DE CHAMPS APRES RESECTION DE LA TUMEUR PRIMAIRE. 135

FIGURE 24. MODIFICATIONS STRUCTURALES SURVENANT AU SEIN DE L'EPINGLE A CHEVEUX SUITE A MSI. 139

LISTE DES TABLEAUX

TABLEAU 1. BILAN DES DIFFERENTES PUBLICATIONS ETUDIANT L'EXPRESSION DES MIRS DANS LES CCRS DE TYPE MSI. 74

TABLEAU 2. BILAN DES MICROARNS DEREGULES DANS LES MICIS (TISSU ET SANG DES PATIENTS). 94

LISTE DES ABREVIATIONS

#

3' : extrémité 3-OH des acides nucléiques

3p : miR issu de l'extrémité 3' du pre-miR

5' : extrémité 5-P des acides nucléiques

5p : miR issu de l'extrémité 5' du pre-miR

5-FU : 5-fluoro-uracile

A

A : adénine

Å : Angstrom

ADAR : adenosine deaminase acting on RNA

ATP : adénosine triphosphate

ADN : acide désoxyribonucléique

AGO : Argonaute

APC : adenomatous polyposis coli

ARN : acide ribonucléique

ARNm : ARN messager

ARNt : ARN de transfert

C

C : cytosine

CCR : cancer colorectal

CIMP : CpG island methylator phenotype

CIN : chromosomal instability

CpG: C-phosphate-G

CRC : colorectal cancer

D

Da : Dalton

DGCR8 : DiGeorge syndrome critical region 8

DCL1 : Dicer-Like 1

Dcr : Dicer

ds : double strand

E

eIF : eukaryotic initiation factor

G

G : guanine

GTP : guanosine triphosphate

I

I : inosine

IBD : inflammatory bowel disease

L

LLC : leucémie lymphoïde chronique

LOH : loss of heterozygosity

M

miARN / miRNA / miR : microARN

MC : maladie de Crohn

MGMT : O^6-methylguanine-DNA methyltransferase

MICI : maladies inflammatoires chroniques de l'intestin

miR* : brin passager

MLH : MutL homolog

MMR : mismatch Repair

MSH : MutS homolog

MSI : microsatellite instability

MSS : microsatellite stable

N

nt : nucléotide

O

ORF : open reading frame

P

PAZ : domaine Piwi/Argonaute/Zwille

pb : paire de base

PCR : polymerase chain reaction

PIWI : P-element induced wimpy testes

piARN : Piwi-interacting RNA

PMS2 : postmeiotic segregation increased 2

PolII : ARN polymérase II

Pri-miR : primary miRNA

Pre-miR : precursor miRNA

R

Ran : Ras related nuclear protein

RBD : RNA binding domain

RISC : RNA induced silencing complex

RIIID : RNaseIII domain

RCH : rectocolite hémorragique

RLC : RISC loading complex

RNase : ribonucléase

S

siARN : small interfering RNA

shARN : small hairpin RNA

snoARN : small nucleolar ARN

SNP : single nucleotide polymorphism

ss : single strand

T

T : thymine

tasiARN : trans-acting siARN

TGFβ : transforming growth factor β

TGFβR : transforming growth factor β receptor

TNF : tumor necrosis factor

TNRC : trinucleotide repeat containing

TRBP : TAR RNA binding protein

TU : transcription unit

U

U : uracile

UTR : untranslated region

X

XPO-5 : Exportine-5

AVANT-PROPOS

Les microARNs sont des petits ARNs non codants qui régulent post-transcriptionnellement l'expression de nombreux gènes et sont de ce fait impliqués dans pratiquement tous les processus biologiques. Un concept, certes simple, mais récent, qui a poussé les chercheurs à « réorganiser [leur] vision de l'Univers »[*]. L'intérêt que nous portons aux microARNs vient en particulier du rôle qu'ils jouent dans de nombreuses pathologies humaines, y compris les cancers. Durant ma thèse, je me suis intéressé au(x) rôle(s) que peuvent jouer les microARNs dans les cancers colorectaux instables sur les microsatellites (CCR MSI, pour *Microsatellite Instability*). Ces cancers (environ 15% des cancers colorectaux) sont caractérisés par une déficience du système de réparation des mésappariements (en anglais MMR pour *MisMatch Repair*), qui conduit à l'accumulation de mutations (délétions ou insertions de nucléotides) dans des séquences d'ADN répétées du génome, qu'on appelle microsatellites. Ces mutations conduisent, quand elles affectent des séquences codantes, à un décalage du cadre de lecture et donc à une protéine tronquée, non fonctionnelle, potentiellement tumorigénique.

Le premier but de ma thèse a été de voir si, à l'instar des gènes codants des protéines, les gènes de microARNs peuvent également contenir des répétitions qui pourraient être ciblées par l'instabilité microsatellitaire. Dans ce contexte, la notion du décalage du cadre de lecture n'est plus applicable, mais c'est les effets potentiels que cette instabilité pourrait avoir sur la structure, la séquence et donc la biogenèse et la fonction des miRs qui entreraient en jeu. Dans un deuxième temps, je me suis intéressé au CCRs de type MSI qui compliquent les maladies inflammatoires chroniques intestinales (MICI). Le choix de ce modèle de CCRs est relevant puisqu'on connait encore mal les mécanismes moléculaires gouvernant leur initiation, et particulièrement ceux qui expliqueraient la perte du système MMR. Il a été proposé, mais jamais vérifié, qu'une composante épigénétique, dépendante de certains

[*] *"After reading their paper in the autumn of 2000, I had to set aside 10 minutes to stare out the window and reorganize my view of the universe."* Extrait d'un commentaire de Victor Ambros L'article qu'il mentionne, écrit par Gary Ruvkun et ses collaborateurs, a été le premier à montrer que let-7, un microARN du nématode *C. elegans* était parfaitement conservé chez plusieurs espèces animales. **Ambros, V. (2008). "The evolution of our thinking about microRNAs." *Nat Med* 14(10): 1036-1040.**

miRs, interviendrait dans leur mise en place. Me basant sur des études récentes qui décrivent deux microARNs pouvant cibler des protéines du système MMR (i.e. miR-155 et miR-21), mon travail a consisté à étudier sur une collection unique d'échantillons les niveaux d'expression de ces deux miRs dans les cancers MSI et MSS *(Microsatellite Stable)* associées aux MICI ainsi que dans leur microenvironnement tumoral. L'ensemble des résultats obtenus sont détaillés dans la section « Résultats » de ce manuscrit sous forme de deux articles publiés et de deux manuscrits soumis pour publication.

Précédent la partie « Résultats », une introduction se focalisant principalement sur la biologie des microARNs a été rédigée. J'ai essayé dans un premier temps de passer en revue les différentes notions et concepts connus sur la biogenèse et la fonction des microARNs. Dans un deuxième temps, je me suis intéressé aux cancers colorectaux de type MSI et au rôle rapporté des microARNs dans ces tumeurs. Dans cette partie j'ai décrit plus particulièrement les CCRs de type MSI qui compliquent les MICI.

REVUE BIBLIOGRAPHIQUE

1. Introduction – La découverte des microARNs

Les microARNs ont été découverts en 1993 par Victor Ambros, Rosalind Lee et Rhonda Feinbaum **(Lee 1993)**. Ces auteurs étudiaient le rôle du gène *lin-4* dans le développement de *Caenorhabditis elegans*, et plus particulièrement sa capacité à réprimer l'expression de la protéine LIN-14, qui joue un rôle essentiel dans le développement larvaire du nématode. Ils ont alors constaté que le gène *lin-4* ne codait pour aucune protéine, mais générait deux petits ARNs de 61 et 22 nucléotides chacun. De manière intéressante, une complémentarité partielle existait entre ces petits ARNs et la partie 3'UTR de l'ARNm de LIN-14. Cette complémentarité serait à la fois nécessaire et suffisante pour inhiber la traduction de l'ARNm **(Wightman 1991; Wightman 1993)**. Plus tard, l'ARN de 61 nucléotides a été identifié comme étant le précurseur de celui de 22 nucléotides, ce dernier constituant le microARN mature. Ce n'est que 7 ans plus tard, qu'un autre microARN a été découvert. Il s'agit de let-7 *(Lethal-7)*, capable de réprimer l'expression de *lin-41*, *lin-14*, *lin-28*, *lin-42* et *daf-12* lors des transitions entre les différentes phases de développement chez *C. elegans* **(Pasquinelli 2000; Reinhart 2000)**. Par ailleurs, l'étude réalisée par Pasquinelli *et al.* a permis de fournir la preuve que ce processus de régulation ne se limitait pas au vers nématode, mais qu'il était également présent chez différentes espèces animales **(Pasquinelli 2000)**. Depuis, des centaines de microARNs uniques et conservés chez différentes espèces ont été découverts. On en dénombre actuellement 1872 précurseurs chez l'homme (2578 miARNs matures), répertoriés dans miRBase (http://www.mirbase.org/), une base de données regroupant les microARNs de toutes les espèces étudiées **(Griffiths-Jones 2008)**.

Suite à ces observations, plusieurs équipes ont essayé de comprendre la base moléculaire permettant aux microARNs de réguler leurs cibles.

En 1998, une découverte inattendue est venue apporter quelques réponses. Dans un article publié dans la revue *Nature*, Craig C. Mello et Andrew Fire, injectant de l'ARN double brin dans *C. elegans,* ont réussi à réprimer l'expression de nombreux gènes. De manière

intéressante, ni l'injection d'ARNm ni celle d'ARN anti-sens ne leur a permis d'obtenir le même phénotype **(Fire 1998)**. Ils venaient de découvrir l'ARN interférence qui leur a valu le prix Nobel de physiologie et de médecine en 2006. Les petits ARN interférents (siRNA, pour *small interfering RNA*) sont donc de petites molécules d'ARN de 22 nucléotides capables de causer la dégradation des ARNm. Ils sont générés à partir d'ARN double brins provenant de l'appariement de transcrits génomiques sens et anti-sens, ou comme intermédiaires de la réplication virale **(Cogoni 1999; Aravin 2001)**. SiARNs et miARNs sont les deux composantes majeures de l'ARN interférence. Contrairement aux siARNs, les miARNs ont une origine endogène et sont produits à partir de précurseurs ARNs simple brins se repliant en épingle-à-cheveux ou *hairpin* (et qui contiennent généralement plusieurs mésappariements). Si les microARNs et les siARNs diffèrent d'un point de vue structural et de leur origine, ils sont par contre assez proches d'un point de vue moléculaire et fonctionnel. En effet, les deux sont capables de réguler l'expression de gènes en agissant sur les ARNm et utilisent à cet effet la même machinerie cellulaire à quelques différences près, que nous détaillerons dans les prochains chapitres.

2. La biogenèse des microARNs

2.1 Organisation génomique des gènes de microARNs

Juste après leur découverte chez les vertébrés, et l'homme en particulier, il a été proposé que la majorité des miARNs était codée au niveau de certaines régions intergéniques, nécessitant la présence de promoteurs spécifiques pour réguler leur transcription **(Lagos-Quintana 2001)**. Dans ce travail fondamental, Lagos-Quintana *et al.* avaient réussi à identifier par clonage 21 miARNs humains à partir de la lignée cellulaire HeLa et 33 miARNs chez la drosophile. De manière très intéressante, ils avaient constaté qu'un certain nombre de ces miARNs (que ce soit chez l'homme ou chez la drosophile) était organisé en groupe ou *clusters* (Figure 1 A et C). En 2002, Lee *et al.* ont montré que les miARNs en clusters étaient produits sous forme de longs transcrits primaires polycistroniques qui subissaient plusieurs étapes de maturation avant de donner les formes matures des miARNs **(Lee 2002)**. Il a fallu attendre 2004 pour que Rodriguez *et al.,* analysant un nombre plus conséquent de miARNs, montrent que chez l'homme, une grande partie des microARNs (~50%) était plutôt localisée dans des unités de transcription (TU) correspondant à des gènes codants (Figure 1 C et D) ou non codants (Figure 1 A et B). Ils montrent que deux tiers environ des gènes (soit 117 miARNs sur les 232 étudiés) étaient localisés dans des introns, et beaucoup d'entre eux étaient orientés dans le même sens que leurs gènes hôtes (qui, dans environ 80% des cas, étaient codants pour des protéines (Figure 1C)). Certains miARNs, mais moins nombreux (environ 10% de tous les miARNs), étaient également inclus dans des exons codants (Figure 1D) ou non codants (Figure 1B). La transcription de la plupart des miRs introniques ou exoniques serait alors régulée par le promoteur du gène hôte qui les contient **(Rodriguez 2004; Olena 2010)**.

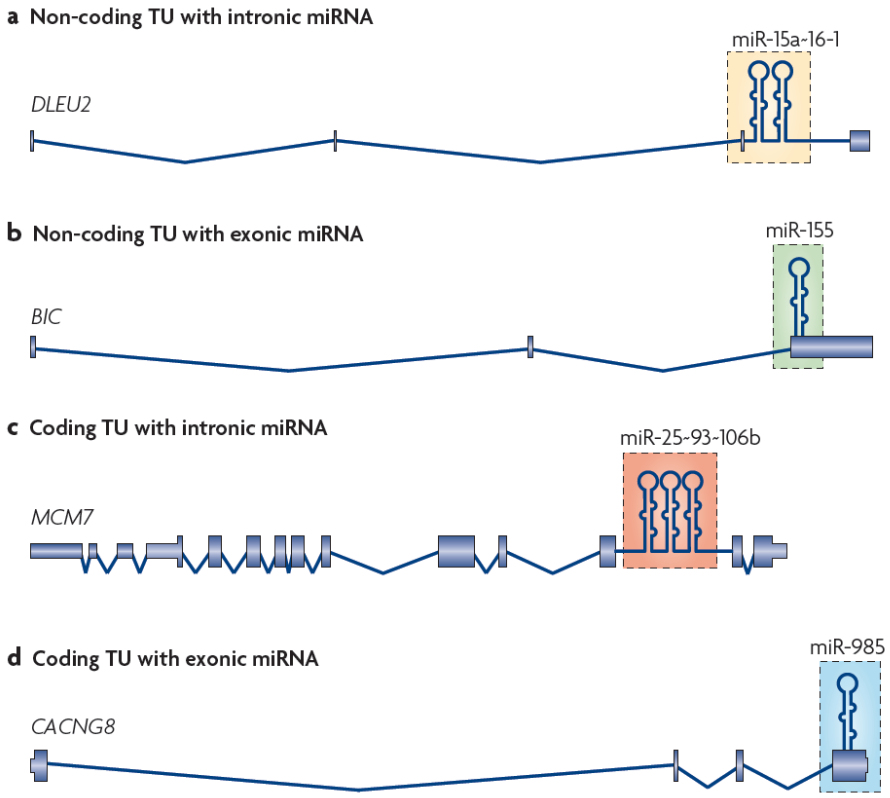


Figure 1. Localisation génomique et structure des gènes de miARNs. Ce schéma, extrait de Kim *et al.*, représente les différentes localisations génomiques des gènes de miARNs. Du moins chez les mammifères, la plupart des gènes de miARNs sont situés dans des unités de transcription (*Transcription Unit*, TU) pouvant être codantes ou non codantes, et sont donc généralement transcrits en même temps que leur gène hôte. Par opposition, les miARNs intergéniques, seuls ou en *cluster*, nécessitent la présence d'un promoteur dédié (non représenté ci-dessus). **A et B.** Dans les TU non codantes, les miARNs peuvent être présents dans des séquences introniques ou exoniques. Par exemple, le cluster miR-15a~16-1 est inclus dans la séquence intronique du transcrit non codant *DLEU2,* et le miR-155 est inclus dans un exon du transcrit non codant *BIC*. **C et D.** De la même façon, les miRs dans les TU codantes peuvent se trouver soit dans des séquences introniques (le cluster miR-25~93~106b renfermé dans un intron de *MCM7*), soit dans des séquences exoniques (miR-987 retrouvé dans le dernier exon du gène *CACNG8*) **(Kim 2009)**.

2.2 Biogenèse des miARNs : aperçu général

2.2.1 Le Modèle canonique

La Figure 2A illustre les différentes étapes régissant la biogenèse ou *processing* des miRs. Seul le modèle dit canonique est représenté. D'autre mécanismes de biogenèse ont été aussi mis en évidence mais sembleraient être moins prépondérants (cf. §2.2.2). La biogenèse des miRs consiste en deux étapes séquentielles, l'une nucléaire et l'autre cytoplasmique, qui conduisent à la production d'un petit ARN de 21-23 nucléotides ayant la capacité de réguler l'expression d'un nombre variable de transcrits. Comme pour les ARNs messagers, les miARNs sont, pour la plupart, produits sous forme de transcrits primaires par l'ARN polymérase II (PolII) : on les appelle pri-miRs (ou miARNs primaires) **(Lee 2004)**. Ces transcrits sont capés, polyadénylés et sont capables de se replier pour donner une structure secondaire en épingle à cheveux ou *hairpin.* Cette structure, caractérisée par une tige d'ARN double brin (dite imparfaite puisque renfermant quelques mésappariements) et une boucle terminale de taille assez variable, est reconnue par la ribonucléase III Drosha et son partenaire DGCR8 *(DiGeorge syndrome critical region 8),* nommé également *Pasha* chez la drosophile et le nématode. Durant cette étape nucléaire, Drosha clive les deux brins du pri-miR et libère une tige-boucle d'environ 60-70 nucléotides nommée pre-miR (ou miARN précurseur) **(Lee 2002; Lee 2003)**. Les microARNs, comme nous venons de le voir, peuvent faire partie de séquences introniques, auquel cas le pri-miR correspondrait en réalité à une partie de l'intron, et l'étape catalysée par Drosha permettant de générer le pre-miR se déroulerait co-transcriptionnellement et au même moment que la réaction d'excision-épissage *(splicing)* (Figure 2B) **(Kim 2007; Morlando 2008; Kim 2009)**. Le *processing* des miARNs exoniques est moins bien connu. Une étude récente du miR-155 a montré que le pre-miR peut être produit à partir de transcrits épissés ou non épissés. L'épissage et le transport cytoplasmique des transcrits épissés constituerait cependant des mécanismes permettant de réguler le taux de miARNs exoniques **(Slezak-Prochazka 2013)**.

Le pre-miR est reconnu et transporté vers le cytoplasme grâce au complexe Exportine-5 couplé à la GTPase Ran **(Yi 2003; Bohnsack 2004; Lund 2004)**. Une fois dans le cytoplasme, le

pre-miR est pris en charge par Dicer, également une RNase III, qui sépare la tige de la boucle et libère un duplex « miR/miR* » ou « miR-5p/miR-3p » d'environ 22 pb. La coupure se fait à deux tours d'hélice de l'extrémité générée par Drosha **(Bernstein 2001; Hutvagner 2001)**. Le duplex se caractérise entre autre par la présence de 2 nucléotides qui dépassent des extrémités 3' (et qu'on appelle *3' overhang*). Cette asymétrie est la résultante directe des coupures effectuées par Drosha et Dicer (cf. §2.3.1). Un seul des deux brins du duplex (miARN guide) est préférentiellement incorporé dans le complexe RISC (pour *RNA-induced silencing complex*), l'autre brin (miR* ou brin passager) est généralement dégradé **(Matranga 2005)**. Actuellement, et suite aux avancées majeures dans les techniques de nouvelles générations et de séquençage à haut débit, il est de plus en plus conseillé d'utiliser l'appellation miR-5p/miR-3p, les deux brins étant généralement exprimés dans les cellules[*].

RISC est un complexe formé par l'assemblage du miR avec plusieurs protéines, dont les plus importantes et les plus étudiées sont les protéines Argonautes (AGO). Il permet au miR mature d'interagir avec son ARNm cible (il s'agit principalement du 3'UTR) et de réprimer son expression. Cette reconnaissance requiert la présence d'une complémentarité parfaite entre les nucléotides 2-7 du miR (qu'on appelle séquence *seed*) et le 3'UTR de l'ARNm. La reconnaissance d'un ARNm par le complexe RISC conduit la plupart du temps au blocage de la traduction de l'ARNm cible (par répression de la traduction suivie éventuellement par le clivage exonucléotidique de l'ARNm) ou au clivage endonucléotidique de l'ARNm (catalysé par AGO2, la seule protéine AGO à posséder une activité catalytique) si la complémentarité entre le miR et l'ARNm est parfait ou quasi-parfaite sur les 22 nucléotides **(He 2004; Kim 2009)**.

Ce modèle de biogenèse est globalement conservé chez toutes les espèces. Cependant des différences existent. Les quelques exemples que nous citerons sont loin d'être exhaustifs, mais ils illustrent bien la grande diversité des mécanismes mis en jeu par différentes espèces pour générer des microARNs. Chez les plantes par exemple, la transition pri-miR → pre-miR et pre-miR → duplex miR/miR* est prise en charge par une seule RNase III jouant à la fois le

[*] *« Mature sequences from all human, mouse, and C. elegans precursors are now designated -5p and -3p, rather than miR/miR* »*. Source : ftp://mirbase.org/pub/mirbase/18/README.

rôle de Drosha et de Dicer : il s'agit de Dicer-Like1 ou DCL1 et son partenaire HYL1 *(Hyponastic leaves 1)*. La structure des précurseurs miARNs chez les plantes est également différente, et nécessite plusieurs clivages par DCL1 avant que le duplex miR/miR* puisse être libéré **(Czech 2011)**. Les plantes possèdent également trois autres protéines DCL nommées DCL 2-4, et qui seraient impliquées dans la biogenèse des siARNs et des tasiARNs *(trans-acting siRNA)* **(Qi 2005)**. Par ailleurs, la drosophile, possède deux enzymes Dicer, nommés Dicer-1 (avec son partenaire Loquacious ou LOQS-PB) et Dicer-2 (avec son partenaire LOQS-PD ou R2D2) qui génèrent les miARNs et les siARNs, respectivement. Chez les mammifères (l'homme en particulier) une seule enzyme se charge à la fois des miARNs et des siARNs **(Czech 2011)**. L'homme possède quatre protéines AGO, nommées AGO 1 à 4, qui n'ont visiblement pas de préférence pour le substrat, qu'il soit miARN ou siARN. La drosophile ne possède que deux protéines AGO (AGO 1 et 2), qui, comme pour Dicer 1 et 2, se chargent respectivement des miARNs et des siARNs (même si l'on trouve plusieurs exceptions à cette règle) **(Czech 2011)**.

2.2.2 Les modèles non canoniques de biogenèse des miRs

Comme nous venons de le voir, certains microARNs se trouvent inclus dans les séquences introniques (Figure 2B). Dans des cas particuliers, la totalité de la séquence intronique, si elle est assez courte et si elle est capable de se replier en épingle à cheveux, constitue un pre-miR : ce type de précurseurs est appelé mirtrons. Le *processing* des mirtrons est indépendant de Drosha, mais nécessite deux étapes supplémentaires : le débranchement du lasso formé durant la réaction d'excision-épissage des messagers et le repliement de l'intron en *hairpin*. Les pre-miRs issus d'un intron rejoignent ensuite la voie classique de biogenèse des miRs **(Okamura 2007)**. Une étude récente dénombre environ 240 mirtrons chez l'homme et 237 mirtrons chez la souris **(Ladewig 2012)**. La biogenèse des mirtrons, si elle est indépendante de Drosha, peut nécessiter parfois la présence d'exonucléases 5'-3' et 3'-5' qui servent à raccourcir le mirtron dans le cas où l'une de ces extrémités est un peu plus longue que l'autre *(trimming)*. Les mirtrons raccourcis à l'extrémité 3' ont été retrouvés uniquement chez la drosophile et sembleraient impliquer le complexe de l'Exosome, alors que les mirtrons raccourcis à l'extrémité 5' ont été annotés uniquement chez les vertébrés sans qu'une exonucléase 5'-3' n'ait été expérimentalement identifiée **(Yang 2011)**. D'autres

mécanismes, indépendants de Drosha, ont également été mis en évidence. Nous citerons les pre-miRs dérivés des snoRNA (*small nucleolar RNA*, petits ARNs nucléolaires régissant les modifications enzymatiques post-transcriptionnelles des ARNs ribosomaux), les miARNs dérivés des shARNs endogènes *(small hairpin RNA)* tels que les miRs 320, 484 et 1980 et les miARNs dérivés des esiARNs *(Endogenous small interfering RNAs)* chez la drosophile. Dans la plupart des cas, l'enzyme qui libère les pre-miRs n'a toujours pas été identifiée **(Miyoshi 2010; Yang 2011)**. Certains miRs peuvent aussi dériver d'ARNs de transferts : l'ARNt de l'isoleucine renferme le miR-1983 et peut adopter, en conformation alternative, une structure en épingle à cheveux pouvant constituer un substrat pour Dicer **(Miyoshi 2010; Yang 2011)**. Par ailleurs, il existe également dans la cellule des voies de biogenèse indépendantes de Dicer. Le pre-miR-451, produit par Drosha, est exporté vers le cytoplasme, où, au lieu d'être pris en charge par Dicer, est incorporé dans AGO2. AGO2 est capable de cliver un des deux brins de la tige du pre-miR et une RNAse cellulaire rogne l'extrémité 3' jusqu'à l'obtention d'un miR-451 mature **(Cheloufi 2010)**.

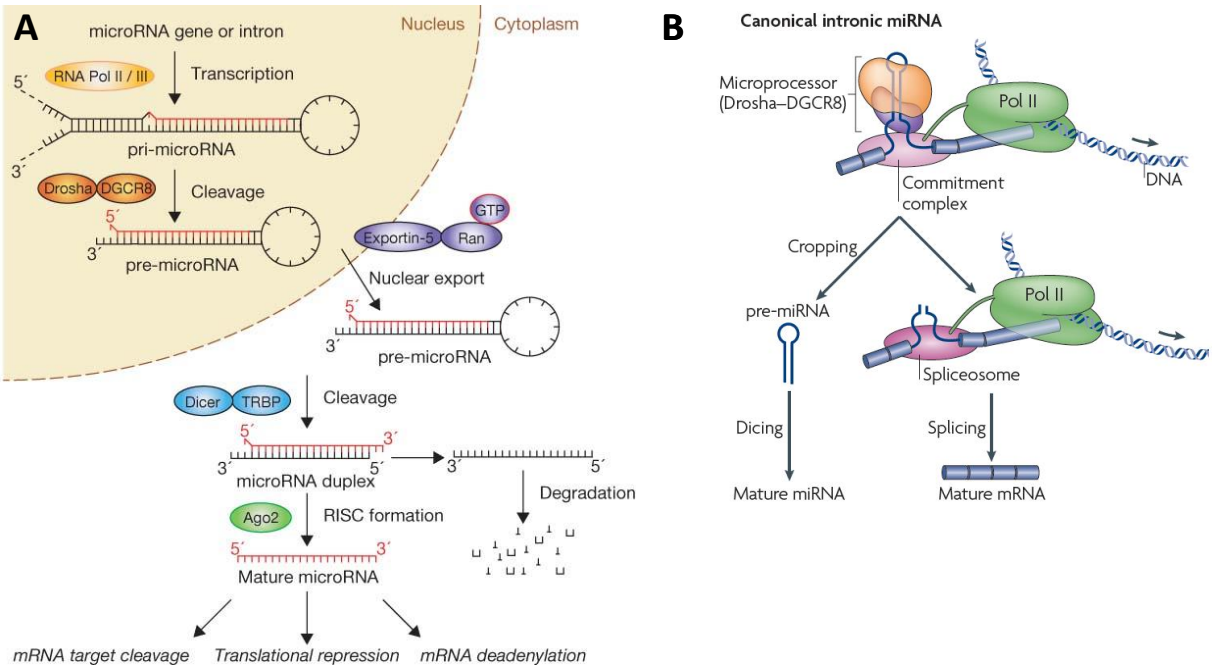


Figure 2. La biogenèse des miARNs. A. Ce schéma représente le modèle canonique de biogenèse des miRs. Les transcrits primaires (ou pri-miRs) sont tout d'abord clivés par la RNase III Drosha *(croping)* pour libérer des tiges-boucles d'environ 70 nucléotides nommées pre-miRs. Les pre-miRs sont ensuite exportés vers le cytoplasme grâce à l'Exportine-5 et sont clivés par Dicer *(Dicing)* pour libérer un duplex miR/miR* de 22 pb environ. Le brin miR* est généralement dégradé et le brin guide est incorporé dans le complexe RISC *(RNA-induced silencing complex),* ce qui permet au miR mature d'interagir avec son ARNm cible (répression de la traduction ou dégradation de l'ARNm). Extrait de Winter *et al.* **(Winter 2009)**. **B.** Quand les microARNs font partie des séquences introniques, leur *processing* se fait co-transcriptionnellement avant le *splicing* (excision-épissage) des introns et implique les mêmes étapes qu'en A. D'après Kim *et al.* **(Kim 2009)**.

2.3 Les bases moléculaires de la biogenèse des miARNs

2.3.1 Les RNases III, Drosha et Dicer

Drosha et Dicer font partie de la famille des endoribonucléases ou RNAses III. Cette famille d'enzymes est capable de cliver les ARNs doubles brins (miARNs, siARNs et autres ARNs doubles brins) et cela en présence d'ions magnésium Mg^{2+}. Le clivage génère une extrémité caractérisée par deux nucléotides qui dépassent de l'extrémité 3' *(3' overhang)* qui est due au positionnement inter ou intramoléculaire des sites catalytiques. Trois classes de RNases III ont jusqu'à présent été découvertes. La classe I, retrouvée exclusivement chez les bactéries et la levure, possède un seul domaine RNase (RIIID) et est active sous forme de dimère. La classe II, dans laquelle figure Drosha, possède deux domaines RNases en tandem (RIIIDa et RIIIDb) qui forment un seul centre catalytique. Finalement la classe III, incluant Dicer, possède en plus des deux domaines RNases, un domaine PAZ *(Piwi/Argonaute/Zwille)* fixant l'extrémité 3' des ARNs doubles brins (Figure 3) **(Tomari 2005; MacRae 2007)**.

Drosha, seule, est incapable de fixer l'ARN double brin et nécessite la présence du cofacteur protéique DGCR8*,* les deux formant alors ce qu'on appelle le complexe Microprocesseur (la spécificité de l'interaction ARN-Microprocesseur sera détaillée par la suite) **(Gregory 2004)**. Les formes les plus simples de Dicer (celle du protozoaire *Giardia intestinalis* par exemple) contiennent uniquement le domaine PAZ et les RNases en tandem. La forme humaine de Dicer contient plusieurs domaines supplémentaires qui régulent la réaction de clivage et jouent un rôle d'échafaudage moléculaire aidant à la fixation de plusieurs cofacteurs protéiques. Il s'agit d'un domaine de fixation au cofacteur TRBP et aux protéines Ago, d'un domaine putatif de liaison à l'ARN double brin nommé DUF283, mais dont la fonction réelle n'est toujours pas connue *(Domain of Unknown Function 283)*, et d'un domaine C-terminal de liaison à l'ARN double brin (Figure 3) **(Kim 2009; Lau 2009)**. La forme humaine contient également un grand domaine N-terminal DExD hélicase-like dont la fonction n'a été qu'en partie élucidée (cf. §2.3.4) **(Lau 2009; Sawh 2012)**.

2.3.2 L'étape catalysée par Drosha

La première étape de maturation des miARNs correspond au clivage du pri-miR afin de libérer l'épingle à cheveux que constitue le pre-miR. Cette étape nucléaire, est catalysée par un complexe protéique formé par la RNase III Drosha et son cofacteur DGCR8, et leur capacité à reconnaître un pri-miR dépend largement de la structure du précurseur **(Kim 2009).** L'une des premières études essayant de répondre à la question de savoir quels sont les déterminants structuraux qui permettent à Drosha et DGCR8 de reconnaître un pri-miR, a révélé l'importance du segment terminal ARN simple brin. Zeng *et al.* ont généré une petite librairie de pri-miRs, variant en taille et en séquence, et ont montré, *in vitro*, qu'indépendamment de la séquence, l'un des prérequis serait la présence de la région simple brin **(Zeng 2005)**. En éliminant cette région ou en introduisant des mutations lui conférant une structure secondaire, ils ont montré que Drosha était incapable de reconnaître et de cliver le pri-miR. En parallèle, la même équipe avait montré que le clivage par Drosha était également dépendant de la présence d'une boucle terminale relativement grande (≥ 10 nucléotides). Le site de clivage serait alors déterminé principalement par la distance (~2 tours d'hélice, soit environ 22 nucléotides) qui le sépare de la boucle terminale **(Zeng 2005)**. Han *et al.*, sans totalement réfuter l'importance d'une grande boucle terminale, soulignent principalement le rôle de la région simple brin **(Han 2006)**. Ils concluent que, si une grande boucle terminale serait dans une certaine mesure avantageuse, le clivage Drosha se ferait plutôt à une distance d'un tour d'hélice (soit environ 11 nucléotides) depuis la jonction entre l'ARN simple brin et la tige. Les auteurs ont également montré que DGCR8 fonctionnerait comme point d'ancrage moléculaire pour Drosha, étant la seule protéine capable d'interagir directement avec les segments simples brins et dans une certaine mesure avec la boucle terminale (Figure 4) **(Han 2006)**.

La structure minimale d'un pri-miR pouvant être reconnue par le complexe microprocesseur consisterait alors en une tige d'environ 33 pb (3 tours d'hélice), une boucle terminale (de taille relativement variable) et un segment basal ARN simple brin **(Kim 2009)**. Cependant, de nombreuses études ont remarqué que tous les ARNs qui se replient de cette façon ne forment pas nécessairement des précurseurs de miARNs. Récemment, Auyeung *et al.* ont

essayé de comprendre quels autres éléments assurent la spécificité de la reconnaissance du pri-miR par le complexe microprocesseur **(Auyeung 2013)**. Les auteurs montrent, tout d'abord, que la plupart des pri-miRs de *Caenorhabditis elegans* (qui ressemblent assez aux pri-miRs humains d'un point de vue structural) n'étaient pas pris en charge par la machinerie de biogenèse, et ne donnaient pas de miRs matures une fois transfectés dans les cellules humaines. Ils ont émis alors l'hypothèse qu'un déterminant « séquence » devait s'ajouter au déterminant « structure » et que le déterminant « séquence » serait spécifique aux espèces. Générant de nombreux variants de pri-miRs humains et séquençant ceux qui restent fonctionnels, ils ont identifié trois principaux déterminants « séquence » conservés chez l'homme (Figure 4B) : un motif CNNC de fixation de la protéine SRp20 (trouvé en aval de la tige, dans la région simple brin), un motif UG basal (au niveau de la jonction tige et ARN simple brin), et un motif GUG apical (au niveau de la boucle terminale). La protéine SRp20, qui d'ailleurs est connue pour être impliquée dans la régulation de l'épissage, le transport des ARNm et l'initiation de la traduction, serait dans ce cas capable d'améliorer la reconnaissance et le *processing* des pri-miRs. Auyeung *et al.* notent en concluant que même si ces motifs sont très conservés dans les pri-miRs humains, tous ne sont pas présents en même temps dans un pri-miR donné. Le modèle actuel qui se dégage met alors en valeur le fait que la reconnaissance des pri-miRs est un phénomène modulaire où chaque pri-miR dépendrait de certains modules (motifs structuraux ou séquences) à des degrés qui varient entre les pri-miRs **(Auyeung 2013).**

2.3.3 Transport nucléocytoplasmique des pre-miRs

L'étape, nucléaire, catalysée par Drosha conduit à la production d'un miARN précurseur ou pre-miR qui doit être transporté dans le cytoplasme pour donner un miR mature. Son transport à travers le complexe du pore nucléaire nécessite la présence de l'Exportine-5 (XPO-5), membre d'une famille conservée de transporteurs nucléaires (incluant également les importines) connus sous le nom de karyopherines (Figure 5) **(Lund 2004)**. L'activité des kariopherines est régulée par la petite GTPase Ran *(Ras-related nuclear protein)*. Dans la cellule, Ran existe sous deux formes : une forme liée au GTP (Ran-GTP) dans le noyau, et une forme liée au GDP (Ran-GDP) dans le cytoplasme. Le gradient RanGTP-RanGDP à travers la

membrane nucléaire est établi par l'action de deux régulateurs: RanGEF *(Ran-GDP-exchange factor)* dans le noyau et RanGAP *(Ran-GTPase-activating protein)* dans le cytoplasme. Dans le noyau, le complexe XPO-5/Ran-GTP lie directement et spécifiquement les pre-miRs produits par Drosha. Après translocation vers le cytoplasme, le GTP est hydrolysé en GDP grâce à l'action de la RanGAP (qui active la fonction GTPase de Ran). XPO-5/Ran-GDP perd son affinité pour le pre-miR qui est alors libéré et pris en charge par Dicer. XPO-5/Ran-GDP transite ensuite vers le noyau à travers le complexe du pore nucléaire, où le GDP est échangé avec du GTP grâce au facteur RanGEF **(Kohler 2007; Guttler 2011)** (Figure 5C).

XPO-5 reconnait et se lie spécifiquement aux pre-miRs. La structure de l'XPO-5 liant un pre-miR a récemment été résolue. Elle montre que la reconnaissance des pre-miRs se fait de manière séquence-indépendante mais structure-dépendante. En effet, XPO-5 forme une sorte de sandwich en forme de U qui lie environ 16 pb du pre-miR. Les deux nucléotides qui dépassent à l'extrémité 3' des pre-miRs jouent dans ce contexte un rôle particulièrement important : ils sont spécifiquement reconnus et insérés dans une poche de XPO-5 chargée positivement *(3' recognition tunnel)* permettant la distinction entre les pre-miRs et leurs précurseurs (Figure 5B) **(Okada 2009; Guttler 2011)**.

2.3.4 L'étape catalysée par Dicer

La deuxième étape catalytique du *processing* des miRs se caractérise par la prise en charge du pre-miR par une autre endoribonucléase multi-domaine, la RNase III Dicer. Dicer et son cofacteur TRBP reconnaissent le pre-miR, clivent la boucle terminale et libèrent un duplex miR/miR*. La majorité des études qui se sont consacrées à l'étude structurale de Dicer, l'ont fait sur une forme simple de l'enzyme (en l'occurrence celle du protozoaire *Giardia intestinalis*, qui fait environ 82 KDa), mais qui n'est pas représentative du Dicer des métazoaires, bien plus large (la forme humaine fait environ 3 fois plus en taille que celle de *Giardia*, soit 219 KDa) **(Lau 2012)**. La première structure cristalline du Dicer de *Giardia* a été obtenue en 2006 (Figure 6A). Cette structure suggère que Dicer fonctionnerait comme une sorte de règle moléculaire, reconnaissant l'extrémité libre d'un ARN double brin (par extension, celle du pre-miR), et clivant l'ARN à une distance fixe de cette extrémité. La

structure a permis de localiser le domaine PAZ capable de lier l'extrémité libre de l'ARN double brin (et plus particulièrement son extrémité 3') et le domaine catalytique formé par les sites ribonucléase III (RIIIDa et b). Une surface plane, chargée positivement *(platform domain)* assure une distance de 65Å (soit 25 pb, taille typique des miARNs de *Giardia*) entre les domaines PAZ et RNases III **(Macrae 2006)**.

Chez l'homme, la taille et la complexité structurale ont jusqu'à présent empêché l'obtention d'une structure cristalline stable de Dicer, mais quelques études ont réussi à avoir une image peu résolutive du complexe. Il s'agirait d'une molécule en forme de L avec une branche longue (de 150 Å environ) et une branche plus courte (de 100 Å environ) (Figure 6B) **(Lau 2009; Wang 2009)**. Récemment, un travail réalisé par Lau *et al.* a permis d'avoir une vision relativement complète du complexe Dicer-TRBP et de son mode d'action (Figure 6B et C). Les auteurs ont réussi à modéliser la structure de Dicer à partir d'un certain nombre d'images obtenues par cryo-microscopie électronique. Ils ont montré qu'un réarrangement des différents domaines par rapport à *Giardia* permettrait d'expliquer pourquoi les microARNs humains sont plus courts. Dans ce modèle, le domaine PAZ serait localisé au niveau apical de l'enzyme. Les domaines hélicases (HEL1, HEL2 et HEL2i), qu'on ne retrouve pas chez *Giardia*, seraient proches du domaine RNAse et formeraient une structure en forme de pince permettant de bien fixer l'ARN double brin (Figure 6B). Le clivage de l'ARN a lieu une fois que l'ARN ait transité à travers le domaine HEL et que son extrémité soit fixée par le domaine PAZ (Figure 6C) **(Lau 2012)**. De manière intéressante, le domaine hélicase posséderait plusieurs propriétés qui diffèrent en fonction du substrat ARN. Chez la drosophile, Welker *et al.* et ensuite Cenik *et al.* ont montré que le domaine hélicase de Dicer-2 serait crucial pour le *processing* des siARNs par Dicer-2. Ces auteurs ont remarqué que le domaine hélicase permettait à Dicer-2 de générer plusieurs siARNs à partir d'un même long ARN double brin. Le rôle du domaine hélicase serait alors de bien fixer l'ARN double brin et d'assurer la processivité de Dicer-2, grâce à l'hydrolyse de l'ATP et à son activité *translocase-like* (Figure 6C, étape 3) **(Cenik 2011; Welker 2011)**. Alternativement, Tsutsumi *et al.* proposent que le domaine hélicase de Dicer-1 (qui chez la drosophile est impliqué dans le *processing* des pre-miARNs), serait capable de reconnaitre spécifiquement et de fixer la partie simple brin de la boucle des pre-miRs. La distance qui sépare le domaine hélicase du domaine PAZ est alors

mesurée, ce qui permettrait à Dicer-1 « d'inspecter rigoureusement » l'authenticité des pre-miRs **(Tsutsumi 2011)**. De manière intéressante, le domaine hélicase de la forme humaine de Dicer ressemble plus à Dicer-1 qu'à Dicer-2. Dans ce contexte, il a été récemment montré que chez l'homme, le domaine hélicase serait également capable de lier la boucle des pre-miRs et d'assurer la spécificité de reconnaissance par Dicer **(Ma 2012)**. Par ailleurs, et de manière totalement inattendue, les mêmes auteurs avaient observé précédemment que, *in vitro*, le domaine hélicase possède un rôle inhibiteur et que son absence (par mutation ou délétion) accélérait considérablement l'activité enzymatique de Dicer et la rendait catalytiquement plus efficace. L'interaction de la forme sauvage de Dicer avec son partenaire TRBP permettait la levée partielle de l'inhibition **(Ma, 2008)**. Il a donc été proposé que le domaine hélicase permettrait de contrôler l'activité endonucléotidique de Dicer, qui dans un environnement riche en ARNs serait préjudiciable à la cellule. La fixation des partenaires de Dicer et le réarrangement moléculaire qui s'ensuit serait alors capable d'assurer la spécificité de reconnaissance des miARNs par le complexe Dicer **(Ma 2012; Nicholson 2012)**. Finalement, nous citerons une étude réalisée par Park *et al.* ayant révélé l'existence d'une poche supplémentaire au niveau de Dicer capable de lier le phosphate à l'extrémité 5' des miARNs. Cette poche se situe à proximité de la poche 3' du domaine PAZ qui, jusqu'à présent, était connu pour pouvoir fixer préférentiellement l'extrémité 3' des pre-miRs. Les auteurs montrent en même temps que l'existence de cette poche est cruciale pour la biogenèse des miARNs ; la mesure de la distance entre le domaine PAZ et les sites catalytique se faisant principalement à partir de l'extrémité 5'. Dicer-1 chez la drosophile possède ce motif, mais pas Dicer de *Giardia* **(Park 2011)**.

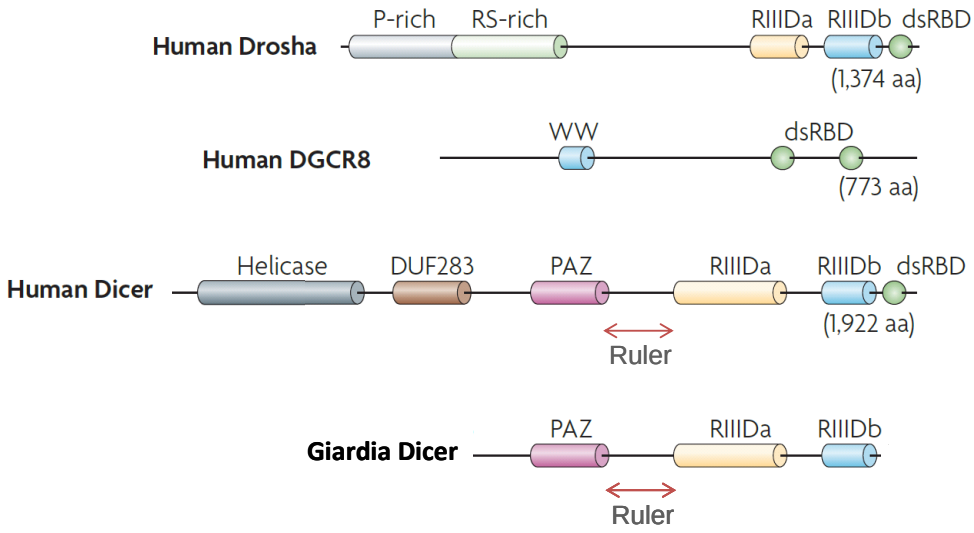


Figure 3. Les différents domaines moléculaires constituant les RNases III Drosha et Dicer. Sur ce schéma sont alignées les formes humaines de Drosha (ainsi que son cofacteur DGCR8) et Dicer. A titre indicatif, est également représentée la structure protéique plus simple du Dicer de *Giardia*. *P-rich*, domaine riche en proline dont la fonction n'est toujours pas connue mais qui pourrait interagir avec le domaine WW de DGCR8. *RS-rich*, région riche en résidus arginine (R) et sérine (S), dont la fonction est également inconnue mais qui s'avère être importante pour l'activité de Drosha. *RIIIDa* et *RIIIDb*, domaines RNase III en tandem. *dsRBD*, domaine de fixation à l'ARN double brin. *WW*, domaine de DGCR8 connu pour son interaction spécifique avec des séquences riches en proline, dont probablement la région riche en proline de Drosha. *DUF283*, domaine dont la fonction n'est toujours pas connue *(domain of unknown function 238). PAZ*, domaine de liaison à l'extrémité 3' des ARNs. *Ruler*, domaine connecteur, fixant la distance entre le domaine PAZ et le centre catalytique de Dicer **(MacRae 2007; Kim 2009; Lau 2009)**.

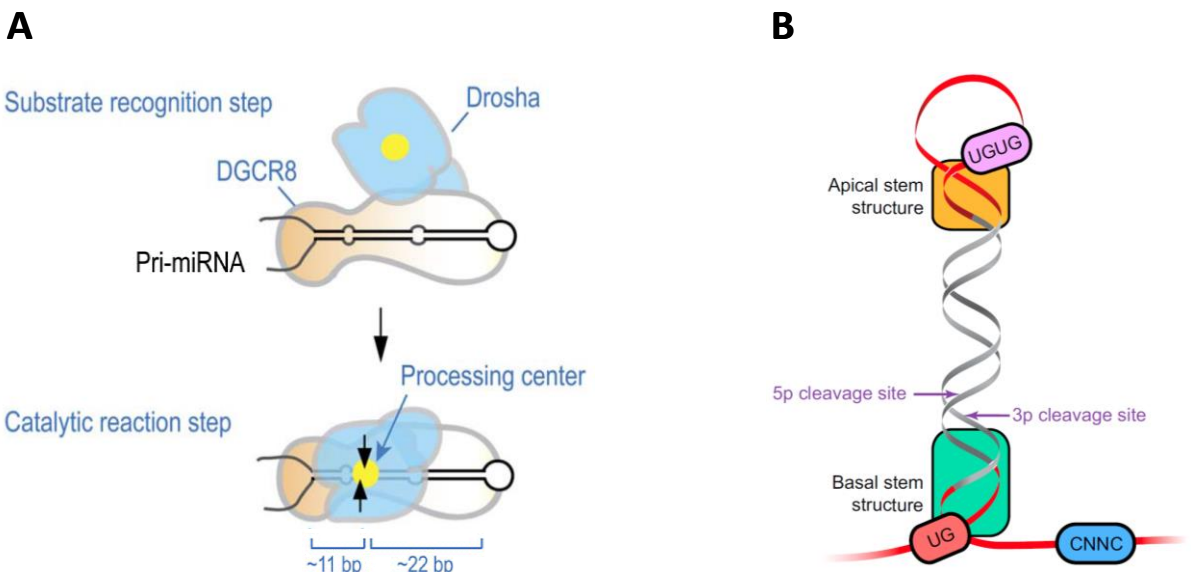


Figure 4. Séquences et motifs structuraux nécessaires à la reconnaissance et au clivage du pri-miR. A. Ce schéma, extrait de Han *et al.*, montre les deux étapes qui caractérisent la réaction de clivage du pri-miR par le complexe microprocesseur. La première étape *(substrate recognition step)* est régie par DGCR8 qui se fixe avec une grande affinité au niveau de la jonction entre l'ARN simple brin et la tige. Cette interaction permet à Drosha de se localiser à un tour d'hélice de la jonction (soit environ 11 nucléotides) et de cliver le pri-miR en pre-miR *(catalytic reaction step)*. Han *et al.* soulignent le fait que la boucle terminal interagirait également avec DGCR8 et serait capable d'améliorer l'efficacité du clivage, mais ne constitue pas en soi un élément essentiel au clivage **(Han 2006)**. **B.** Ce schéma, extrait d'Auyeung *et al.*, résume les principaux déterminants de séquence nécessaires à la réaction de clivage du pri-miR. Il s'agit des motifs CNNC, UG et UGUG. Il souligne également l'importance de la complémentarité des deux brins au niveau basal (rectangle vert), et dans une certaine mesure au niveau apical (rectangle orange) **(Auyeung 2013)**.

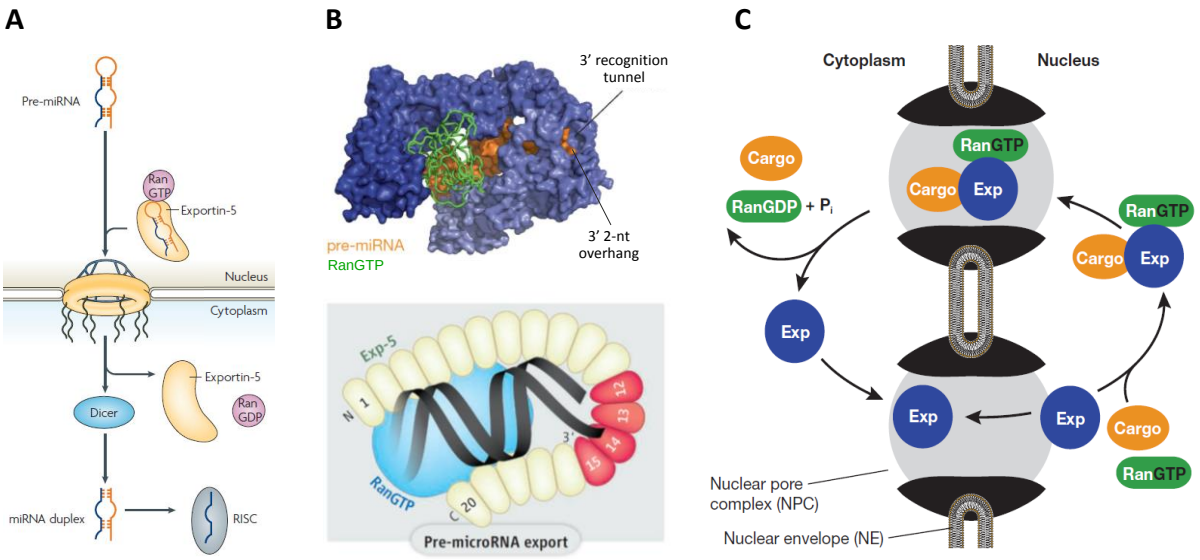


Figure 5. Transport nucléocytoplasmique des pre-miRs. Le transport des pre-miRs du noyau vers le cytoplasme requiert la présence de l'Exportine-5 et de son partenaire, la GTPase Ran. **A.** Couplage moléculaire des différentes étapes du *processing* des miRs. Adapté de Köhler *et al.* **(Kohler 2007)**. **B.** Structure cristalline de complexe XPO-5/Ran-GTP/premiR. La GTPase interagit avec plusieurs domaines HEAT (domaines protéiques consérvés retrouvés dans certaines protéines cytoplasmiques telles que la Huntingtin, l'EF3 *(elongation factor 3)*, la PP2A *(protein phosphatase 2A)*, et la kinase TOR1 chez la levure) et permet de maintenir la forme en U de XPO-5. Une poche (domaines HEAT 12-15) fixe spécifiquement les deux nucléotides qui dépassent de l'extrémité 3' des pre-miRs et assurent en même temps la spécificité de l'interaction et la protection des extrémités de l'action des exoribonucléases. D'après Gutter *et al.* **(Guttler 2011)** et Stewart M. **(Stewart 2009)**. **C.** Le cycle de Ran. Ran-GTP favorise la liaison entre l'Exportine et sa molécule cargo (dans ce cas, il s'agit des pre-miRs). Ran-GDP a par contre peu d'affinité avec les pre-miRs qui sont alors relachés dans le cytoplasme. Le gradient GTP-GDP à travers la membrane nucléaire est maintenu grace à l'action des cofacteurs protéiques, Ran-GAP dans le cytoplasme, et Ran-GEF dans le noyau (ne figurent pas sur le schéma). Adapté de Güttler *et al.* **(Guttler 2011)**.

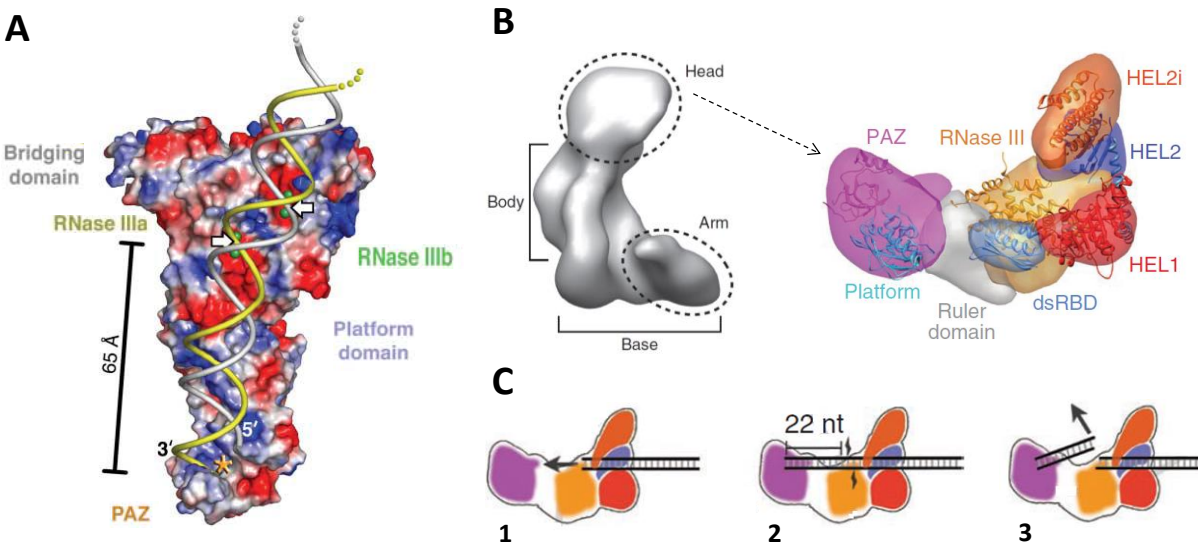


Figure 6. Structure et mode d'action de Dicer. A. Structure cristalline du Dicer de *Giardia*. Le domaine PAZ assure la liaison avec l'extrémité 3' de l'ARN double brin (*), la plateforme *(platform domain)* permet à Dicer de mesurer une distance de 65 Å, les motifs RNases (petites flèches) permettent de cliver l'ARN double brin, et le pont *(bridging domain)* permet de lier les deux domaines RNases. D'après MacRae *et al.* **(Macrae 2006)**. **B.** Structure 3D du Dicer humain. Sur la représentation de gauche figurent les principaux composants de Dicer avec sa forme en « L » caractéristique. La représentation de droite, détaille chaque domaine formant le complexe. Il s'agit principalement du domaine PAZ (fixation de l'extrémité 3' libre de l'ARN double brin), d'un domaine hélicase HEL (assurant la processivité de la réaction de clivage), d'un domaine de fixation à l'ARN (dsRBD), du domaine RNase III, et d'un domaine connecteur (*ruler domain*, dont la longueur et l'angle permettent de mesurer une distance d'environ 22 bp entre les domaines PAZ et RNAses III). **C.** Mode d'action de Dicer. (1) L'ARN double brin passe à travers les domaines hélicases (orange, rouge et bleu). (2) L'extrémité libre est reconnue par le domaine PAZ (violet). Une fois positionné, l'ARN double brin est clivé par le domaine RNase III (jaune) à ~22 pb du domaine PAZ. (3) Le duplex miR/miR* est libéré et pris en charge par les protéines Argonautes. **B** et **C**, d'après Lau *et al.* **(Lau 2012)**.

3. Les mécanismes d'action des microARNs

3.1 Le complexe RISC

La réaction catalysée par Dicer libère un duplex miR/miR* qui est pris en charge par les protéines Argonautes. Les protéines Argonautes forment avec Dicer le complexe RISC et ne sont capables de fixer qu'un seul brin (le miR ou miARN guide qui va cibler les ARNm), l'autre brin (miR*) étant généralement dégradé. Ce processus est connu sous le nom de *RISC loading*. Les miRs étant continuellement sous pression sélective pour cibler un set particulier de gènes, un changement du brin chargé dans AGO conduirait à des effets indésirables, voire néfastes pour la cellule **(Czech 2011)**. Deux études parues en 2003 ont montré que le chargement du brin fonctionnel dans le complexe RISC dépendait des propriétés thermodynamiques du duplex miR/miR* et plus particulièrement au niveau de ses extrémités 5'. Le brin qui possède l'extrémité la moins stable est celui qui est préférentiellement incorporé dans le complexe RISC **(Khvorova 2003; Schwarz 2003)**.

Caractéristiques des protéines Argonautes. La régulation post-transcriptionnelle se fait à travers les protéines Argonautes, qui sont guidées par les petits ARNs vers les ARNm cibles. La famille des protéines Argonautes est subdivisée en deux sous-catégories : celle des protéines AGO et celles des protéines PIWI *(P-element induced wimpy testis)*. Les protéines AGO, ubiquitaires, se lient aux siARNs et miARNs (dont la taille varie entre 21 et 23 nucléotides), alors que les protéines PIWI se lient aux piARNs *(Piwi-interacting RNA,* qui sont généralement plus grands et font entre 23 et 30 nucléotides). Brièvement, les piARNs et leurs protéines associées sont principalement exprimées dans les lignées germinales et assurent la stabilité du génome en inhibant l'activité des éléments transposables. Chez l'homme il existe 4 protéines Argonautes (AGO 1-4) alors que la drosophile, par exemple, n'en possède que deux (AGO 1-2) **(Cenik 2011; Czech 2011)**.

Les protéines AGO (et également PIWI) possèdent 3 domaines protéiques conservés (Figure 7A). Le domaine amino-terminal PAZ contient une cavité hydrophobe qui reconnait particulièrement les deux nucléotides dépassant de l'extrémité 3' du duplex. Ce domaine

fixe également l'extrémité 3' du petit ARN guide. Vient ensuite le domaine MID, qui se lie préférentiellement au phosphate à l'extrémité 5'. Cette liaison, forte, permet de recycler le miR ou le siARN afin qu'il puisse cibler successivement plusieurs ARNm. Finalement, vient le domaine carboxy-terminal PIWI. Ce domaine ressemble structuralement à l'endoribonucléase RNase H (enzyme qui clive l'ARN dans les hétéroduplex ARN/ADN) et se caractérise par la présence de trois résidus négativement chargés et conservés (Aspartate-Aspartate-Glutamate ou motif DDE) assurant l'activité catalytique d'AGO (le site de clivage se situant entre les nucléotides 10 et 11 du brin guide). Chez l'homme, seule AGO2 est capable de cliver les ARNm et est donc nommée *« Slicer »* **(Cenik 2011)**. L'activité catalytique d'AGO2 permet également de cliver le brin passager du duplex (miR*) et facilite sa dégradation par diverses exonucléases. Cette activité dépend de la complémentarité parfaite ou quasi parfaite entre les deux brins **(Matranga 2005; Rand 2005)**. Pour AGO 1, 3 et 4 la dissociation entre le miR et le miR* est indépendante de toute activité catalytique. Le brin passager est libéré par débobinage ou *unwinding*, une réaction rendue possible par la présence de un ou plusieurs mésappariements aux extrémités du duplex **(Czech 2011)**. L'interaction entre les protéines AGO et les petits ARNs se fait par le biais de leurs squelette sucre-phosphate, les bases sont donc libres d'interagir avec les ARNm cibles **(Ender 2010)**.

Le tri des microARNs et association aux Argonautes. Chez la drosophile, il a été démontré que généralement les miARNs (produits par Dcr-1) s'associaient avec AGO1 et les siARNs (produits par Dcr-2) s'associaient avec AGO2 **(Czech 2011).** Ce modèle a cependant été rapidement révoqué et l'image qui se dégage à présent souligne le fait que différents éléments structuraux régissent l'association des petits ARNs avec les protéines AGO. Plus particulièrement, il a été montré que les duplex libérés par Dcr-1 pouvaient s'associer avec AGO1 mais aussi avec AGO2 ; et que si globalement les miARNs s'associaient avec AGO1 et les siARNs avec AGO2, cela ne dépendait pas du fait qu'ils soient produits par Dcr-1 ou Dcr-2 **(Forstemann 2007; Tomari 2007)**. Kawamata *et al.* ont établi les déterminants structuraux derrière ces associations. Les duplex siARNs, du fait de la parfaite complémentarité entre les deux brins, sont systématiquement dirigés vers AGO2. L'existence d'un mésappariement central dans le duplex miR/miR* dirigerait le duplex vers AGO1, et défavoriserait en même temps la liaison avec AGO2 en bloquant son activité catalytique. Par contre les duplex

miR/miR* qui ne possèdent pas le mésappariement central mais des mésappariements aux extrémités sont incorporés dans AGO2. L'existence à la fois du mésappariement central et de un ou plusieurs mésappariements aux extrémités du duplex seraient capables de promouvoir la formation d'un complexe RISC mature et fonctionnel **(Kawamata 2009)**. D'une façon moins importante, le premier nucléotide en position 5' du miR joue également un rôle dans l'association avec les protéines AGO. Chez la drosophile, AGO1 se lie préférentiellement avec les miRs commençant par un U, alors que AGO2 préfère les miRs commençant par un C **(Czech 2009; Ghildiyal 2010; Czech 2011)**. Chez l'homme, jusqu'à présent, une association préférentielle d'un duplex avec l'une des protéines AGO n'a pas encore été établie **(Czech 2011; Dueck 2012)**.

Le couplage moléculaire des complexes RISC et Dicer. En 2005, Gregory *et al.* ont observé, après immunoprécipitation du complexe RISC chez l'homme, que Dicer et son cofacteur TRBP co-purifiaient avec AGO2 et le duplex miR/miR*. Ce complexe moléculaire a été appelé *RISC loading complex* (RLC). Les auteurs ont alors proposé un modèle où la production du duplex par Dicer et son chargement dans AGO, même si d'un point de vue temporel ils se faisaient de manière séquentielle, se dérouleraient en réalité au niveau d'un même complexe. Le RLC constituerait alors un état de transition entre le *processing* des pre-miRs (pre-RISC caractérisé par l'association d'un duplex miR/miR* avec les protéines AGOs) et le complexe RISC mature (seul le brin guide reste associé aux AGOs) et permettrait l'incorporation de duplex juste après sa production par Dicer **(Gregory 2005)**. A l'époque, les bases moléculaires de cette association étaient encore inconnues, et le rôle, voire même l'existence du RLC chez l'homme a été contesté **(Yoda 2010)**.

Une étude structurale réalisée par Wang *et al.* a montré que d'un point de vue moléculaire ce complexe peut théoriquement exister chez l'homme. Les auteurs ont en effet réussi à modéliser en même temps Dicer et AGO2, et observent que la forme en L de Dicer (en gris sur la figure) serait adaptée pour recevoir AGO2 (en jaune sur la figure) et pour interagir avec elle (Figure 7B) **(Wang 2009)**. Récemment, un travail réalisé par Noland *et al.* a permis de comprendre – du moins en partie – le rôle et les implications d'une association entre Dicer et les protéines AGO (plus particulièrement AGO2). Le RLC ne constituerait pas un

complexe moléculairement figé : son assemblage et son désassemblage constitueraient des étapes clés déterminant le choix du brin guide et la dégradation du brin passager. Les auteurs ont démontré qu'une fois produit, le duplex d'ARN est libéré dans le milieu interstitiel et subissait ensuite un repositionnement au sein de Dicer. Ce repositionnement ne serait pas aléatoire et permettrait d'orienter le duplex en fonction de ses propriétés thermodynamiques. C'est en se basant sur cette orientation que le choix du brin guide va être pris par la protéine AGO. Selon ce modèle, Dicer serait capable lui-même de détecter l'asymétrie thermodynamique du duplex, et possèderait deux sites de liaison à l'ARN : le premier pour le *processing* des pre-miRs et le deuxième pour la détection de l'asymétrie thermodynamique du duplex **(Noland 2011)**. De manière intéressante, la présentation du duplex à AGO serait facilitée par d'autres protéines connues pour interagir avec le RLC telles que Hsc70 *(heat shock cognate 70)* et Hsp90 *(heat shock protein 90),* et nécessiterait l'hydrolyse de l'ATP. Cette hydrolyse induirait des changements structuraux au niveau des protéines AGO et leur permettrait d'accepter plus facilement le duplex **(Czech 2011)**. Même si AGO2 chez la drosophile est structuralement différente de AGO1 et des AGOs 1-4 chez l'homme, Tomari *et al.*, l'avaient associée à ce même mécanisme. Leurs résultats montrent que R2D2, un partenaire de Dcr-2 serait capable de lier l'extrémité la plus stable, et Dcr-2 l'extrémité la moins stable du duplex, les deux constituant des « capteurs » d'asymétrie thermodynamique **(Tomari 2004).**

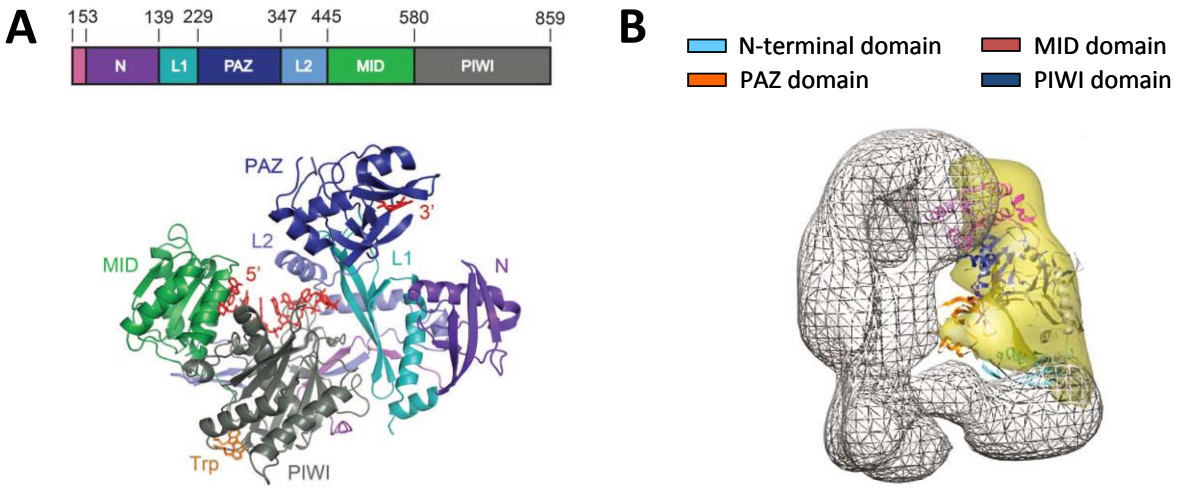


Figure 7. Structure cristalline d'AGO2 et modélisation du RLC. A. Structure cristalline de la forme humaine d'AGO2. Les extrémités 5' et 3' d'un miARN sont représenté en rouge. D'après Schirle *et al.* **(Schirle 2012). B.** Modélisation de l'association entre AGO2 et Dicer pour former le *RISC loading complex* ou RLC. Le domaine hélicase de Dicer constituerait une base sur laquelle repose AGO2. Le duplex miR/miR* se faufilerait à l'intérieur de l'espace vide qui se trouve à l'interface des deux protéines. En gris, Dicer et en jaune, AGO2. D'après Wang *et al.* **(Wang 2009)**.

3.2 La reconnaissance des ARNm cibles

Il est actuellement établi que chez les animaux, la grande majorité des miARNs forment des duplex imparfaits avec les 3'UTRs de leur ARNm cibles **(Bartel 2009; Brodersen 2009; Pasquinelli 2012)**. Cette interaction nécessite cependant la présence d'une complémentarité parfaite sur les nucléotides 2-7 du miR, région connue sous le nom de séquence *seed*. En accord avec cette affirmation, plusieurs études ont été publiées montrant que la partie 5' des miARNs serait conservée, et qu'une substitution de nucléotides perturbant l'appariement avec la séquence *seed* aurait pour conséquence une perte de l'activité régulatrice des miARNs **(Lai 2002; Lim 2003; Doench 2004)**. C'est d'ailleurs en se basant sur cette conservation de la séquence *seed* et de la partie complémentaire au sein de l'ARNm que plusieurs algorithmes tels que TargetScan (http://www.targetscan.org/) et PicTar (http://pictar.mdc-berlin.de/) réussissent à prédire les cibles d'un miARN donné : si, au niveau de l'ARNm, certains sites sont conservés, il est fort probable qu'ils soient sous pression sélective et qu'ils aient une fonction biologique. Ces algorithmes, combinant données expérimentales, génomiques, voire même parfois thermodynamiques, essayent d'en ressortir des règles qui expliqueraient la reconnaissance miARN-ARNm. Cependant, l'hétérogénéité des cibles et la multitude de facteurs cellulaires (protéiques et structuraux) qui entrent en jeu, limitent considérablement la mise en place de modèles prédictifs fiables. Le critère « conservation de la séquence *seed* » par exemple, ne détermine pas en soit la reconnaissance des cibles : la cellule n'ayant pas le moyen distinguer entre une séquence conservée et une autre non conservée. Il constitue cependant un outil très utile permettant de sélectionner les candidats les plus probables parmi une liste de cibles pouvant inclure des milliers de gènes dont la plupart constitueraient des faux-positifs **(Bartel 2009; Friedman 2009)**. Le contexte structural et moléculaire du 3'UTR joue également un rôle très important. Le site de reconnaissance doit : (1) se situer au moins à 15 nucléotides du codon stop ; (2) être proche du codon stop ou de la queue polyA, pour les longs 3'UTRs ; (3) se trouver dans une région riche en AU, ou tout autre élément structural facilitant l'accessibilité à l'ARNm ; (4) coexister avec des sites de reconnaissance d'autres miARNs pour une plus grande efficacité de régulation **(Grimson 2007; Bartel 2009)**.

La Figure 8 résume les différents types d'interactions qui peuvent exister entre un miARN et son ARNm cible **(Bartel 2009)**. Les exemples donnés sont ceux qui ressortent, *in silico*, en utilisant le logiciel TargetScan pour prédire les cibles d'un miARN donné **(Friedman 2009)**. Les cases A à C représentent les modes d'interaction dits canoniques, qui à eux seuls représentent environ 60% du total des ARNm ciblés par un miARN-type (case H). Il s'agit dans le premier cas, d'une complémentarité parfaite sur les 6 nucléotides de la séquence *seed* et une adénine en position 1 de l'ARNm (7mer-A1 site). Dans le second, la séquence *seed* s'étend au 8ème nucléotide du miARN (7mer-m8 site), alors que le troisième cas combine les deux premiers c'est-à-dire, une complémentarité parfaite sur les nucléotides 2-8 du miARN et un A en position 1 de l'ARNm (8mer site). Dans ces 3 cas, le nucléotide en position 1 du miARN n'interagit pas avec l'ARNm, et se trouve la plupart du temps face à une adénine. Ce mésappariement a tout d'abord été suggéré par une étude génomique chez les vertébrés, qui montre la surreprésentation des adénosines flanquant les sites conservées complémentaires aux séquences *seed* **(Lewis 2005)**. Il a été ensuite confirmé par plusieurs études protéomiques montrant qu'un appariement A-U entre le nucléotide en position 1 du miARN et l'ARNm réduisait l'efficacité de la régulation **(Nielsen 2007; Baek 2008)**.

La Figure 8 représente également ce qu'on appelle des sites de reconnaissances dits marginaux *(marginal sites)*. Il s'agit dans le premier cas d'une complémentarité parfaite sur les nucléotides 2-7 du miARN (6mer site) en l'absence d'un A en position 1 de l'ARNm (case D) et dans le deuxième cas, d'un décalage de la séquence *seed* (offset 6mer site) qui se retrouve aux positions 3-8 du miARNs (case E). Ces sites de reconnaissance sont globalement peu efficaces et dans beaucoup de cas, leur conservation apparait plus liée au hasard qu'à une pression de sélection **(Bartel 2009)**.

L'importance de la région 3' du miARN. Plusieurs études réalisées chez la drosophile et les mammifères montrent que, d'un point de vue conservation des séquences, la région 3' du miARN, très peu conservée, semblerait ne pas jouer un rôle majeur dans la reconnaissance des cibles **(Brennecke 2005; Lewis 2005)**. Cependant, il reste envisageable qu'une complémentarité entre la partie 3' du miR et l'ARNm puisse complémenter l'appariement avec la séquence *seed*, renforçant la spécificité et l'efficacité de l'interaction avec la cible

(Bartel 2009). Une étude très intéressante réalisée en 2007 par Grimson *et al.* basée sur une approche à la fois expérimentale et computationnelle, a permis de mettre en évidence certaines caractéristiques permettant d'améliorer l'efficacité de reconnaissance de la cible : il s'agit notamment, et en synergie avec la complémentarité avec la séquence *seed*, d'une complémentarité parfaite sur les nucléotides 13-16 (et parfois 12-17) du miARN (en orange, case F) **(Grimson 2007)**. Ces sites complémentaires coexistent, d'une façon plus ou moins importante, avec les sites canoniques et marginaux (barres oranges, case H). Ils sont certes très spécifiques, mais sont par contre plus rares (sites atypiques) **(Bartel 2009; Friedman 2009)**. De manière très intéressante, dans certains cas peu nombreux, cette complémentarité permet de compenser des mésappariements survenant au niveau de la séquence *seed* (case G). Le miR let-7 avec sa cible lin-41 et le miR-196 avec sa cible Hoxb8 font partie de cette catégorie. Dans le cas du miR-196, la complémentarité en 3' est tellement importante que le duplex miR-196/Hoxb8 est dirigé vers la protéine AGO2 qui catalyse le clivage de l'ARNm **(Yekta 2004; Bartel 2009)**.

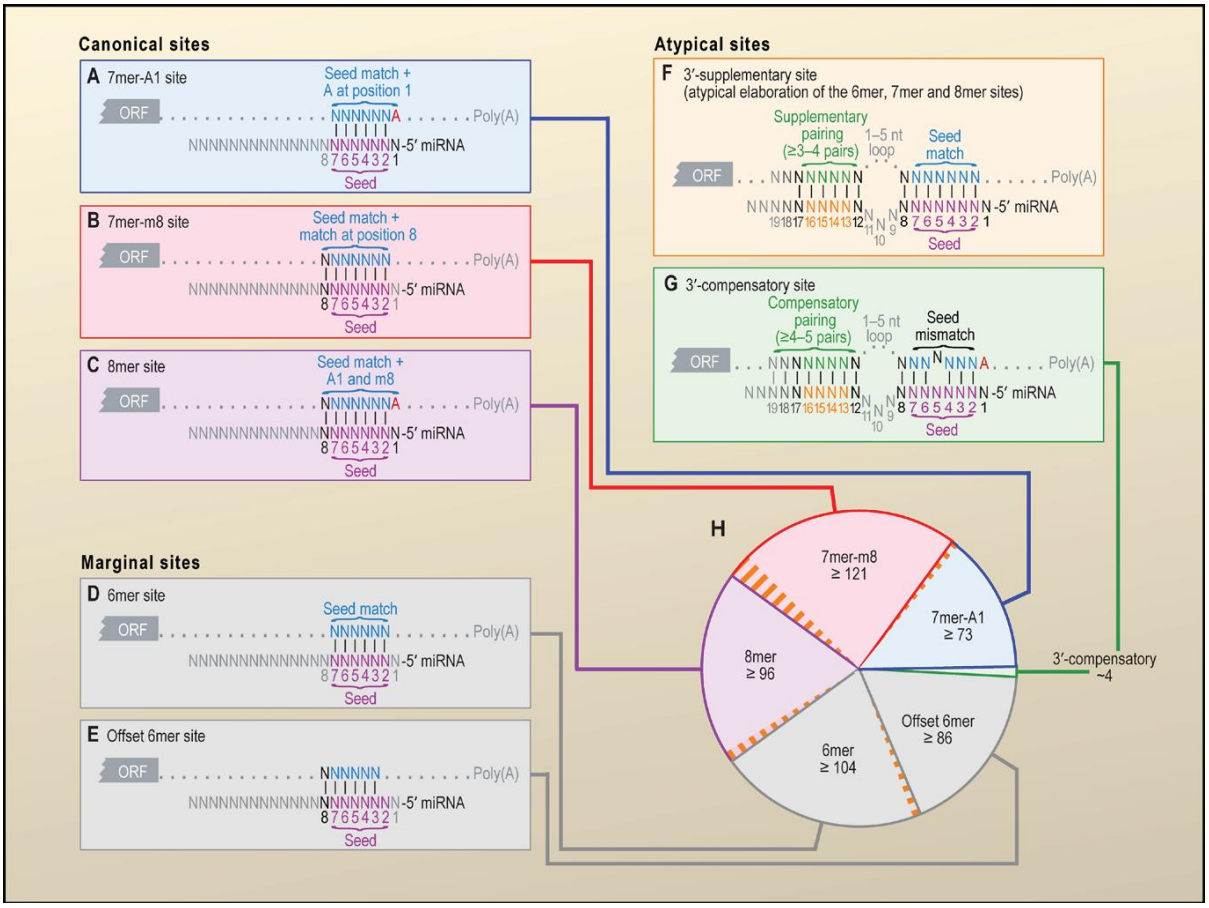


Figure 8. Les différents types d'interaction entre miARNs et ARNm cibles. Ce schéma illustre les différentes configurations possibles obtenues pour un miARN-type. **A-C.** Les sites de reconnaissances dits « canoniques ». Les tirets verticaux illustrent la complémentarité parfaite entre la séquence *seed* et le 3'UTR de l'ARNm. **D-E.** Les sites de reconnaissances « marginaux », beaucoup moins spécifiques et efficaces. **F-G.** Les sites de reconnaissances « atypiques », nécessitant une complémentarité entre la partie 3' du miARN et l'ARNm. Dans certains cas peu nombreux (case G), cette complémentarité vient compenser un mésappariement au niveau de la séquence *seed*. **H.** Cette case illustre la proportion de chaque type d'interaction, en moyenne sur 87 miARNs appartenant à des familles conservées. Adaptée de Friedman *et al.* **(Friedman 2009)**. Les petites barres en oranges dans chaque tranche du diagramme représentent la proportion de sites nécessitant une complémentarité avec la région 3' du miARN. ORF, cadre de lecture *(Open Reading Frame)*. Extrait de Bartel D.P. **(Bartel 2009)**.

3.3 Les différents modes d'action des miARNs

Comment un miARN peut-il réguler l'expression de sa cible ? Répondre à cette question nécessite préalablement une compréhension de la structure des ARNs messagers et des différentes protéines avec lesquelles ils interagissent. Un ARNm se caractérise en particulier par sa coiffe (7-methylguanosine en position 5' terminale) et sa queue Poly(A). Ces deux éléments coopèrent pour promouvoir la traduction de l'ARNm. La coiffe en 5' est reconnue par le facteur d'initiation de la traduction eIF4F *(eukaryotic initiation factor 4F)*. eIF4F est un complexe protéique formé entre autres, par les facteurs eIF4E (liaison à la coiffe), eIF4G (échafaudage moléculaire permettant de recruter la petite sous-unité 40S du ribosome) et eIF4A (une ARN hélicase). La queue poly(A) est recouverte par une protéine spécifique, la PABP *(poly(A) binding protein)*, qui elle-même interagit avec eIF4G ce qui conduit à une circularisation de l'ARNm (Figure 9, structure centrale). Ceci permet le rapprochement du 5'UTR et du 3'UTR et augmente l'efficacité de la traduction en facilitant le recrutement de plusieurs facteurs protéiques et le turnover des sous-unités ribosomiques. La traduction des ARNm procède selon trois étapes : (1) l'initiation, qui correspond au recrutement des différents facteurs eIF, leur phosphorylation et l'assemblage de la petite sous unité ribosomique ; (2) l'élongation, qui correspond au recrutement des différents facteurs d'élongation (eIF, *eukaryotic initiation factor*), leur phosphorylation, l'assemblage de la grande sous-unité ribosomique et à la production de la protéine et (3) la terminaison de la traduction quand le ribosome arrive au codon « stop » **(Filipowicz 2008; Huntzinger 2011)**.

De nos jours plusieurs arguments suggèrent différents mécanismes permettant aux miARNs de réguler l'expression de leurs ARNm cibles. Ces mécanismes varient entre déstabilisation de l'ARNm, répression de sa traduction et même parfois activation de la traduction (Figure 9) **(Huntzinger 2011; Pasquinelli 2012)**.

Dégradation de l'ARNm. Comme nous venons de le voir, une complémentarité parfaite entre le miARN et sa cible est responsable du recrutement de la protéine AGO2 et du clivage endonucléotidique de l'ARNm et de sa dégradation (Figure 9a). Ce mécanisme, qu'on retrouve surtout chez les végétaux, reste assez rare chez les animaux **(Huntzinger 2011;**

Pasquinelli 2012). De nombreuses études ont néanmoins montré qu'une régulation par les miARNs se manifestait la plupart du temps par une déstabilisation de la cible et cela indépendamment de la présence d'une complémentarité parfaite (Figure 9b). Une étude très intéressante, combinant des techniques de séquençage profond, de puces d'expressions et de *« ribosome profiling »*, montre qu'à l'échelle du génome une diminution du niveau d'expression des ARNm est responsable à elle seule de plus de 84% de la baisse de l'expression protéique **(Guo 2010)**. D'autres études sont venues corroborer ces résultats, montrant que la déstabilisation des ARNm cibles serait une conséquence générale de la régulation par les miARNs **(Huntzinger 2011)**. Moléculairement, il s'agit d'une interaction entre les protéines Argonautes et les protéines GW182 *(Glycine-tryptophan repeats containing proteins)* également connues sous le nom de TNRC6A, TNRC6B et TNRC6C chez les vertébrés, qui permet le recrutement du complexe de déadénylation CAF1-CCR4-NOT et l'élimination de la queue poly(A). La perte de la queue poly(A) sensibilise les ARNm : la coiffe est éliminée par le complexe DCP2 et ses partenaires, et les ARNm décoiffés sont finalement dégradés par XRN1, une exonculéase 5'-3'. Ces différents complexes protéiques sont organisés dans le cytoplasme au sein de structures appelées P-bodies, ou GW-bodies d'après les protéines GW182 qui les constituent **(Filipowicz 2008; Huntzinger 2011)**.

Répression de la traduction. Les premières études qui ont été réalisées sur les microARNs et leurs modes d'action ont montré que la régulation de l'expression des protéines par les miARNs se fait avec peu ou pas d'effet sur l'abondance des ARNm **(Olsen 1999)**. Nous savons bien évidemment que ceci est loin d'être généralisable sur tous les miARNs. L'étude réalisée par Olsen *et al.* a par contre permis de montrer, qu'en absence de dégradation des ARNm, alors que la production de protéines est inhibée, les ARNm de LIN-14 et LIN-28 restent attachés au polysomes, suggérant que la répression a lieu une fois que la traduction a été initiée (Figure 9d) **(Olsen 1999)**. L'inhibition de la production des protéines serait alors due soit à une dégradation du polypeptide durant la traduction, soit à une dissociation prématurée des deux sous-unités ribosomiques **(Huntzinger 2011; Pasquinelli 2012)**. En dépit de ces données, d'autres résultats semblent indiquer que les microARNs sont capables d'inhiber l'étape d'initiation de la traduction (Figure 9c). Dans une étude publiée en 2005, Pillai *et al.* montrent que les ARNm qui ne dépendent pas de la coiffe pour leur traduction

n'étaient pas sujets à la répression par les miARNs, laissant suggérer que le complexe RISC interagirait directement avec la coiffe pour inhiber la traduction. Ils montrent également qu'en présence de miARNs, les ARNm cibles ne co-sédimentent pas avec la fraction contenant les polysomes sur gradient de saccharose mais avec les fractions plus légères appauvries en ribonucléoprotéines **(Pillai 2005)**.

Activation de la traduction. Les microARNs sont également capables de stimuler la traduction de certains gènes, même si cela reste encore assez rare (Figure 9e) **(Pasquinelli 2012)**. Pendant un arrêt du cycle cellulaire, les séquences riches en A-U (ARE, *AU-rich elements*) de l'ARNm de TNFα sont capables de recruter diverses protéines Argonautes et autres facteurs associés au miARNs. Dans ce contexte, le miR-369-3 serait capable d'activer la traduction de TNFα **(Vasudevan 2007)**. Les auteurs ont également constaté que d'autres miRs tel que let-7, connus pour réprimer la traduction, sont capables d'activer la traduction suite à un arrêt du cycle cellulaire **(Vasudevan 2007)**. Un cas assez intéressant a été décrit par Ørom *et al.* Il s'agit particulièrement d'une interaction entre le miR-10a et le 5'UTR des ARNm de diverses protéines ribosomiques capables d'activer leur traduction **(Orom 2008)**.

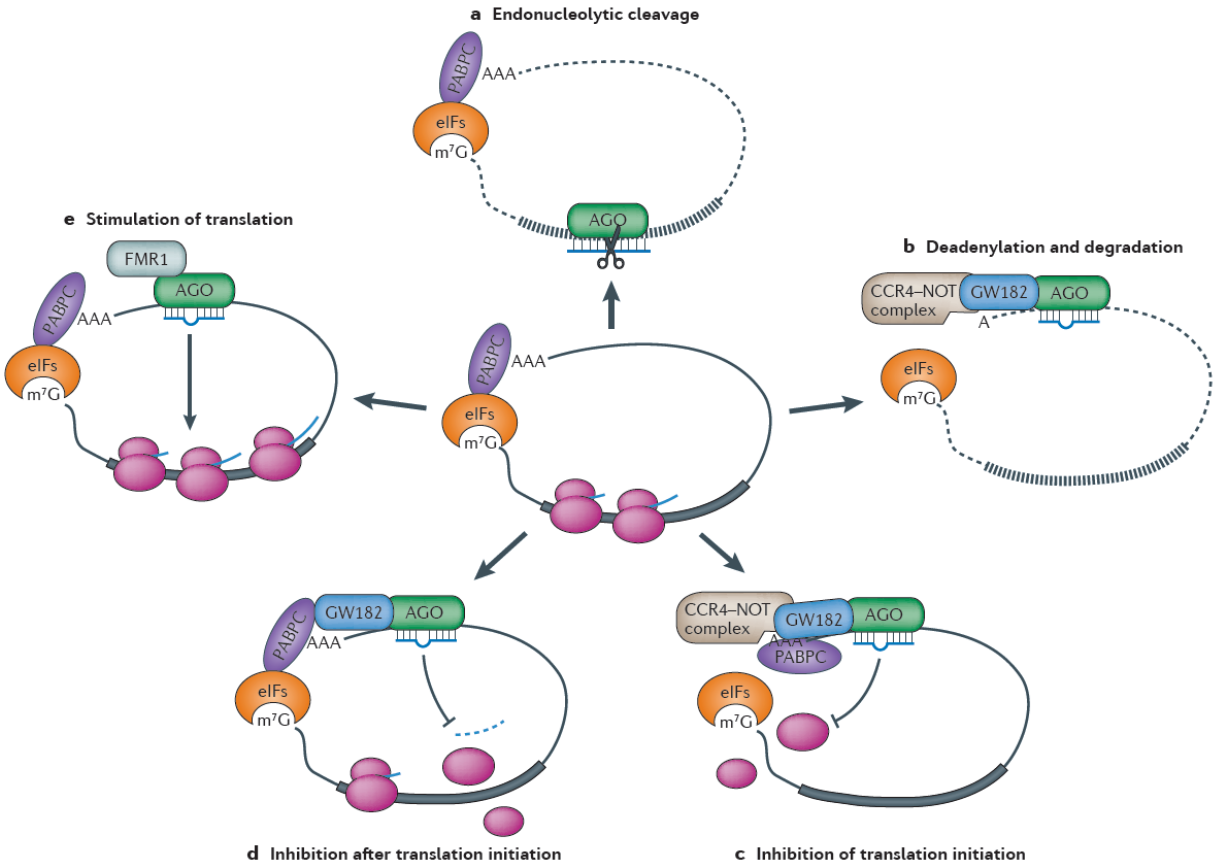


Figure 9. Régulation des ARNm cibles par les miARNs. a. Clivage endonucléotidique de l'ARNm par AGO2. **b.** Déadénylation et dégradation de l'ARNm. **c.** Inhibition de l'initiation de la traduction. **d.** Inhibition de l'étape d'élongation. **e.** Activation de la traduction. La partie centrale représente la circularisation des ARNm. En rose, les ribosomes ; en orange, les complexes d'initiation eIF et en violet, les protéines de liaison à la queue poly(A) PABPC. Extrait de *Pasquinelli A.M.* **(Pasquinelli 2012)**.

4. Variations génétiques : SNP et mutations somatiques

4.1 SNP, *single-nucleotide polymorphism* dans les gènes de microARNs

De manière générale, des variations de séquences dans les gènes de microARNs (que ce soit le miR mature ou ses précurseurs) ont potentiellement pour conséquences un effet sur la biogenèse et/ou la fonction des miRs matures **(Ryan 2010; Slaby 2012)**. Les SNP et leurs conséquences fonctionnelles peuvent être classés en trois catégories : (1) des polymorphismes en cis ou trans du promoteur du gène de miARN, ou du gène dans lequel se trouve le pri-miR pouvant influer sur les niveaux de transcription ; (2) des variations dans la séquence même du pri- ou du pré-miARN (miR-SNP) affectant la biogenèse du miR mature et (3) des variations dans la séquence du miR mature (que ce soit au niveau de la séquence *seed* ou de la partie 3' du miR) ayant des conséquence sur sa capacité à réguler une cibler et/ou sa stabilité. Nous pouvons également inclure les polymorphismes qui affectent les 3'UTR des ARNm (miR-TS-SNP, TS pour *target site*) et qui peuvent moduler, créer ou éliminer des sites de fixation des miARNs, et que nous discuterons très brièvement dans ce chapitre.

Une étude *in silico* réalisée par Saunders *et al.*, a eu pour but d'évaluer à quelle fréquence les séquences de pre-miR étaient polymorphes et cela par rapport aux régions flanquantes en 5' et 3' (environ 100 nucléotides de part et d'autre du pre-miR) **(Saunders 2007)**. Sur les 474 pre-miRs étudiés, environs 90% ne présentaient aucun polymorphisme, et pour ceux qui étaient polymorphes, la région *seed* n'était pratiquement pas affectée. Ils conclurent alors qu'une forte pression de sélection négative *(purifying selection)* agissait afin de limiter les conséquences potentiellement délétères que pourrait avoir un changement dans la séquence des miRs matures. Dans une étude plus récente, Quach *et al.*, séquençant 92 régions contenant des microARNs (soit 117 miRs matures) chez 91 individus de différentes populations, ont également montré qu'une pression sélective, particulièrement au niveau du miR mature, limitait considérablement le polymorphisme de ces gènes **(Quach 2009)**.

Malgré ces données, plusieurs exemples dans la littérature montrent qu'il existe bel et bien un polymorphisme (quoique relativement limité) au niveau des gènes de microARNs **(Ryan**

2010; Slaby 2012). Les premières études fonctionnelles qui ont été réalisées dans ce domaine semblent indiquer que le polymorphisme n'a aucun effet sur la biogenèse des miRs. Particulièrement, une étude réalisée par Diederichs *et al.,* qui montre que des SNP au niveau des gènes de microARNs (la majorité étant au niveau du pri-miR), même s'ils entrainent la plupart du temps des changements structuraux assez visibles, n'auraient aucun effet sur l'expression des miRs matures **(Diederichs 2006)**. Par opposition, Duan *et al.* ont identifié un SNP localisé en position 8 du miR-125a mature et qui, de manière inattendue, bloquait la biogenèse du miR et plus particulièrement, le passage pri-miR vers pre-miR **(Duan 2007)**. Nous citerons également une étude réalisée par Sun *et al.*, où les auteurs montrent qu'en plus d'un effet négatif sur la biogenèse et l'expression des miRs matures, les SNP peuvent, dans certains cas, améliorer l'efficacité du processing (cas de la transition G → A au sein du pri-miR-510) **(Sun 2009)**. De manière très intéressante, les auteurs montrent également que les SNP peuvent parfois altérer les sites de coupures de Drosha ou Dicer. Une conversion T → G au niveau du premier nucléotide du miR-934-5p conduit d'une part à l'apparition d'une espèce 5p plus grande et en moindre quantité, et d'autre part à la production en grande quantité de miR-934-3p. Il s'agit en effet d'une inversion, suite au polymorphisme, entre le brin guide (5p) et le brin passager (3p) **(Sun 2009)**.

De manière assez remarquable, plusieurs SNP touchant les gènes de microARNs ont été associés à diverses pathologies humaines, y compris les cancers **(Ryan 2010; Slaby 2012)**. En 2008, Jazdzewski *et al.* ont identifié un polymorphisme au niveau du pre-miR-146a qui conduisait à une diminution de l'expression du miR mature et prédisposait au carcinome papillaire de la thyroïde **(Jazdzewski 2008)**. Le miR-146a ciblait, entre autres, des gènes appartenant aux voies de signalisations des cytokines et des récepteurs de type Toll, fréquemment surexprimés dans ce type de cancers. Quelques mois plus tard, Hu *et al.* étudiant un polymorphisme au niveau du miR-196a2 ont observé qu'il était associé à un mauvais pronostic chez les patients homozygotes possédant ce SNP et atteints de cancer bronchique non à petites cellules *(non-small cell lung carcinoma)* **(Hu 2008)**. De façon plus générale, une étude récente de 149 SNP chez 6 espèces animales a permis de localiser ces polymorphismes au niveau de divers Locus de caractères quantitatifs (QTL, *quantitative trait loci*), de sites fragiles de chromosomes et de sites de susceptibilité au cancer, et leur a

attribué un rôle potentiel dans le contrôle génétique de différents caractères complexes **(Zorc 2012)**. En ce qui concerne les cancers colorectaux, une étude réalisée par Lee *et al.* sur 436 patients, a permis d'identifier un polymorphisme au niveau du pre-miR-492. Ce SNP serait alors associé à une mauvaise survie sans progression (PFS, *progression free survival*) chez les patients hétérozygotes et homozygotes possédant ce SNP, mais n'aurait aucun effet sur la survie globale des patients **(Lee 2010)**. Toujours dans les cancers colorectaux, nous citerons un polymorphisme qui se trouve dans un site complémentaire au miR let-7 (lcs6, *let-7 complementary site*) au niveau du 3'UTR du gène *KRAS*. Il est en fait connu que des mutations du proto-oncogène *KRAS* sont associées à une résistance aux thérapies basées sur des anticorps monoclonaux ciblant le récepteur EGF. Zhang *et al.* ont donc montré que pour 130 patients atteints de cancer colorectal métastatique et qui sont sauvages pour *KRAS*, la présence du SNP au niveau de lcs6, qui conduit à une forte expression de KRAS, serait associée à une meilleure réponse au traitement monoclonal (cetuximab) **(Zhang 2011)**. Ce même polymorphisme a d'ailleurs été également associé à un plus grand risque de développement de cancer bronchique non à petites cellules, et à un mauvais pronostic (survie globale réduite) chez les patients atteints de HNSCC *(head and neck squamous cell carcinoma)* **(Chin 2008; Christensen 2009)**.

4.2 Des mutations germinales et somatiques dans les gènes de microARNs

Peu de mutations germinales ou somatiques ont été identifiées dans les gènes de microARNs. L'exemple le plus connu et le plus important reste une mutation germinale des miR-16-1–miR-15a **(Calin 2005)**. Dans les leucémies lymphoïdes chroniques (LLC), une délétion fréquente affecte le chromosome 13q13.4 dans lequel se trouvent ces deux microARNs suppresseurs de tumeurs (ils agissent sur BCL2 et répriment son activité anti-apoptotique). Calin *et al.* ont montré que la perte de l'expression des miR-16-1 et miR-15a serait également due à une mutation constitutionnelle C→T au niveau de leur précurseur (pri-miR) et que l'on retrouve uniquement chez les patients atteint de LLC. Cette mutation serait associée, la plupart du temps, à une délétion du second allèle normal, rappelant le modèle de Knudson d'inactivation des gènes suppresseurs de tumeurs **(Calin 2005)**.

5. Modifications post-transcriptionnelles des miARNs

5.1 Mise en évidence

L'avènement des technologies de séquençage à haut débit et particulièrement celles du RNA-seq a permis le rapide essor de l'identification et de la caractérisation des microARNs. Ces technologies ont révélé la diversité de ces petits ARNs et de part ce fait, soulignent la difficulté d'établir des règles pouvant s'appliquer à l'ensemble des miRs. À grande échelle, l'un des premiers exemples rapportant des variations de séquences dans les miRs matures date de 2006. Ruby *et al.* ont séquencé environ 440 000 petits ARNs de *Caenorhabditis elegans*. Ils remarquent la grande hétérogénéité des microARNs et confirment ce que certains auteurs avaient déjà observé par *Northern blot* **(Chen 2004)** : à partir d'un même précurseur plusieurs microARNs de tailles ou séquences différentes sont capables d'être produits. Ils jugent cependant cette hétérogénéité comme insignifiante physiologiquement et ne considèrent, pour un miR donné, que ces deux séquences majoritaires (celle du miR sur un brin, et celle du miR* sur l'autre) **(Ruby 2006)**. Chez les mammifères, Landgraf *et al.*, clonant et séquençant plus de 330 000 petits ARNs non codants, constatent qu'environ 20% des séquences clonées ne s'alignaient pas parfaitement sur le génome. Les auteurs, n'excluant pas le fait que ces variations puissent être artéfactuelles et liées à des erreurs de PCR ou de séquençage, font en même temps le constat que certains des changements ne sont pas aléatoires : ils dépendent du nucléotide et de sa localisation au sein du miR mature et sont parfois spécifiques des tissus en question. Il s'agissait surtout de substitutions de nucléotides qu'ils ont attribuées à des modifications enzymatiques d'un A en I (conversion de l'adénosine en inosine) et d'ajouts ou de délétions de nucléotides en 3' et 5' du miR. Même si un nombre conséquent des séquences majoritaires qu'ils ont identifiées ne correspondaient pas pour les mêmes miRs aux séquences déposées dans miRBase, les auteurs proposent eux aussi de ne garder qu'une seule séquence de référence pour un miR donné **(Landgraf 2007)**. Le terme *isomiR* n'est apparu qu'un peu plus tard. Morin *et al.*, utilisant également des technologies de nouvelle génération aboutissent à la même conclusion : les microARNs présentent fréquemment des variations par rapport à la séquence de référence (en l'occurrence celle de miRBase), de multiples variants de taille et de séquence ou *isomiRs* étant produits à partir d'un même précurseur et même parfois

exprimés à des taux assez élevés **(Morin 2008)**. Ce ne fut donc qu'une question de temps avant que les études ne se succèdent, appuyant ces constats et suggérant que ces isoformes auraient potentiellement un rôle physiologique. Ils soulignent finalement l'importance d'une reclassification des isomiRs vis-à-vis de nombreuses pathologies humaines **(Ryan 2010; Neilsen 2012)**.

De nos jours, les isomiRs sont groupés en trois catégories (Figure 10) : (1) IsomiRs avec des **variations à l'extrémité 5'**, (2) IsomiRs avec des **variations à l'extrémité 3'** et (3) IsomiRs avec des modifications internes ou **substitutions de nucléotides** connues sous le nom de *miRNA editing* **(Neilsen 2012).**

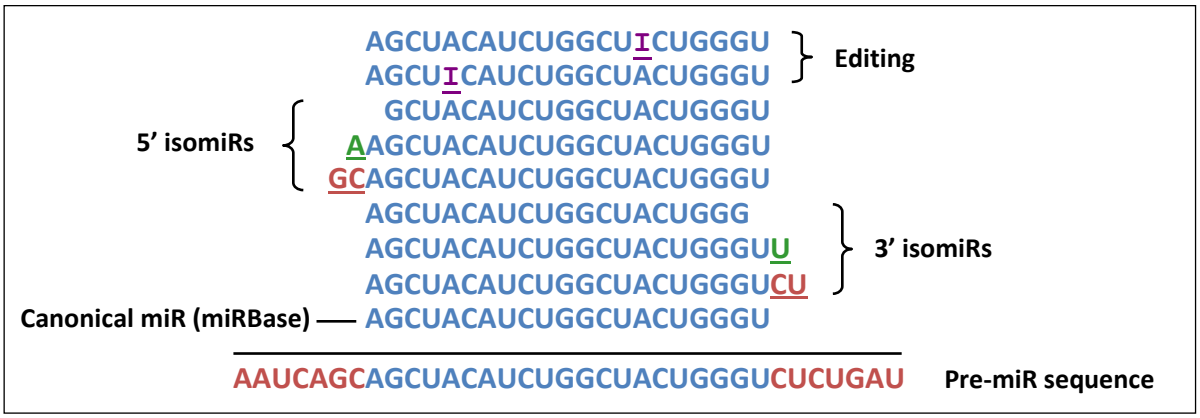


Figure 10. Représentation schématique des différentes espèces d'isomiRs produites à partir d'un même précurseur. Ce schéma, adapté de Neilsen *et al.* **(Neilsen 2012)**, illustre les différents types d'isoformes qui sont généralement obtenues par séquençage profond. En bleu, la séquence mature de référence (celle répertoriée dans miRBase); en violet les modifications de A en I (*editing*). Les ajouts de nucléotides peuvent dépendre (rouge) ou pas (vert) de la séquence matrice du précurseur.

5.2 Fréquence

Il est généralement admis qu'à partir d'un précurseur en épingle à cheveux, une seule espèce dominante de miARN mature est produite **(Bartel 2009)**. Ce constat vient du fait que la plupart des études, sans vraiment remettre en question la présence d'isomiRs, regroupaient en une seule espèce fonctionnelle tous les isoformes partageant la même extrémité 5' **(Landgraf 2007; Morin 2008)**. Cela semble à premier abord assez acceptable dans la mesure où c'est la séquence *seed* qui guide le miARN vers son messager cible. Si l'on se limite à ce raisonnement, seules des modifications en 5' sont intéressantes d'un point de vue fonctionnel puisqu'elles seraient capables d'affecter la reconnaissance miARN/ARNm (en modifiant la séquence *seed*), et permettraient éventuellement au miR mature de cibler d'autres transcrits. En même temps, on pourrait s'attendre à une contre-sélection de ces formes qui, en ciblant un autre répertoire d'ARNm, auraient potentiellement des conséquences néfastes pour la cellule **(Lewis 2003; Ruby 2006)**. De la même façon, la désamination de l'adénine et sa conversion en inosine peut aussi considérablement affecter la capacité d'un miR à reconnaître ses cibles, du fait que l'inosine s'apparie avec la cytidine. Globalement on s'attendrait donc à ce que les modifications au niveau 5' soient peu fréquentes à l'échelle du génome et qu'inversement, les modifications en 3' soient les plus fréquemment représentées. Dans ce contexte, Chiang *et al.* montrent qu'en accord avec l'hypothèse d'une pression sélective, il y globalement peu d'hétérogénéité en 5' – chez la souris, ces variants constituant 8% de toutes les séquences obtenues. Cependant ces isomiRs peuvent être parfois fortement exprimés et devraient donc d'un point de vue fonctionnel être considérés comme espèces moléculaires à part entière **(Chiang 2010)**. Concernant l'*editing,* nous savons qu'il est plus rare chez la souris que chez l'homme, chez qui le taux peut atteindre 16-20% des séquences pri-miR **(Kawahara 2008)**, voire même plus pour les miRs matures **(Reid 2008)**. Cependant, et pour différentes raisons techniques, plusieurs auteurs se sont interrogés sur la réalité de ce taux assez élevé au sein des miRs matures. De Hoon *et al.* suggèrent que comme dans les expériences de biopuces, les courtes séquences d'ARN obtenues par séquençage profond peuvent incorrectement s'aligner sur certains loci, et cela d'autant plus qu'elles risquent d'avoir plusieurs séquences similaires dispersées dans le génome. Cette *cross-hybridization* pourrait donc expliquer la surreprésentation des évènements d'*editing* rapportée dans certaines études. Elaborant un algorithme permettant

d'éviter ces artefacts, ils constatent alors que si l'on peut trouver des modifications au niveau des précurseurs, la plupart du temps ces modifications ne touchent pas la séquence *seed* voire même pas le miR mature **(de Hoon 2010)**. Ceci est loin de sous-entendre un impact mineur de l'*editing* sur les microARNs. En effet, si l'*editing* semble être assez peu répandu au sein des miRs matures en général, il reste toujours capable d'affecter individuellement certains miARNs et de manière plus large de jouer un rôle dans *processing* des précurseurs **(Kawahara 2008; Chiang 2010; Neilsen 2012)**. Finalement, et comme prévu, ce sont les évènements à l'extrémité 3' qui sont les plus fréquents. Il s'agit, comme l'illustre la Figure 10, d'extrémités plus ou moins raccourcies et d'ajouts de nucléotides (des A et des U représentant respectivement ~50% et ~25% de l'ensemble des ajouts) **(Ruby 2006; Chiang 2010; Wyman 2011)**.

5.3 Bases moléculaires

Dès l'instant où ils sont transcrits sous forme de pri-miRs jusqu'à la reconnaissance de leurs cibles, les microARNs sont couplés à des protéines. C'est en étudiant ces interactions que l'on pourra comprendre la base moléculaire qui explique pourquoi tant de variants sont produits à partir d'un même précurseur. Ce paragraphe essayera donc de revoir les principales avancées dans ce domaine.

5.3.1 Le rôle de Drosha et DICER1

En partie, les variations aux extrémités 5' et 3' peuvent s'expliquer par ce qu'on appelle des « *templated sequence modifications* ». Comme leur nom l'indique, ce sont des modifications qui dépendent de la séquence matrice du précurseur, et les IsomiRs produits ne diffèrent que par leur taille. Certains auteurs suggèrent que cette hétérogénéité serait due à des « erreurs » liées à la flexibilité structurale de Dicer, et dans une moindre mesure, à celle de Drosha (Figure 11A) **(Starega-Roslan 2011)**. Ce modèle implique que les extrémités 5' des miARNs issus du brin 3p sont aussi hétérogènes que les extrémités 3' des miARNs issus du brin 5p. Or, ceci n'est pas le cas : que le miR mature provienne de bras 5p ou 3p de pre-miR, son extrémité 5' est toujours moins variable que son extrémité 3' (Figure 11B, et §1.1) **(Seitz 2008)**. Un autre mécanisme doit donc être proposé pour expliquer cette différence ; et ce mécanisme se baserait non pas sur les propriétés catalytiques de Drosha et Dicer (qui génèrent de toutes façon des extrémités hétérogènes), mais sur la structure des protéines Argonautes et leur affinités avec les microARNs (Figure 11C) **(Seitz 2008; Frank 2010)**. En effet, plusieurs études montrent qu'un déséquilibre existe en faveur de la présence d'un U ou d'un A en position 1 du miR mature chez l'homme, et de la présence d'un U en position 1 chez la drosophile **(Ghildiyal 2008; Czech 2009; Hu 2009)**. Se basant sur ce constat, Seitz *et al.*, proposent que, du moins chez la drosophile, ce biais serait dû en grande partie à la liaison avec les protéines argonautes qui imposeraient une sorte de « goulot d'étranglement » permettant d'homogénéiser les extrémités 5' **(Seitz 2008; Seitz 2011)**. Ceci est confirmé par une étude structurale approfondie qui montre que chez l'homme, AGO2 se lie avec une affinité 30 fois plus importante aux bases U et A qu'aux bases C et G, et qu'elle

serait donc capable de discriminer entre les différents microARNs en fonction du nucléotide qu'ils portent en position 1 (Figure 11C et Figure 12B) **(Frank 2010)**.

5.3.2 Le rôle des exoribonucléases et des nucléotidyltransférases

Comme nous venons de le voir, c'est la liaison avec les protéines Argonautes qui explique pourquoi les extrémités 5' sont peu variables. De manière très intéressante, c'est cette même liaison qui explique l'hétérogénéité des extrémités 3' **(Neilsen 2012)**. La structure protéique de la forme humaine d'AGO2 récemment publiée révèle les différences structurales qui caractérisent les groupements protéiques liant les extrémités 5' et 3' **(Elkayam 2012; Schirle 2012)**. La Figure 12A montre le large réseau d'interactions qui maintient en place et protège l'extrémité 5' dans le domaine *Mid* d'AGO2. Quant à l'extrémité 3', elle dépasse du domaine *PAZ*, et est donc vulnérable à l'attaque des exoribonucléases ou à l'ajout de nucléotides par des nucléotidyltransférases (Figure 12B). Parmi les exoribonucléases qui raccourcissent les miRs en 3', nous pouvons citer Nibbler chez la drosophile et le complexe de l'exosome et la nucléase QID chez certains champignons**.** Les nucléotidyltransférases que l'on retrouve chez l'homme possèdent des activités uridyl transférases et/ou adénine transférases, en accord avec l'idée que les modifications en 3' correspondent le plus souvent à des ajouts de U et de A **(Neilsen 2012)**.

5.3.3 ADAR et editing des microARNs

L'*editing* de l'ARN est très répandu au sein du génome et ne se limite pas aux ARNs non codants. En tant que mécanisme épigénétique, il est considéré comme créateur de diversité à la fois transcriptomique et protéomique. L'*editing* de A en I est la principale forme d'édition chez les mammifères, chez qui l'on retrouve également de l'*editing* du C en U. Il est catalysé par des enzymes de la famille des adénosines désaminases qui agissent sur l'ARN (ADAR pour *Adenosine Deaminase Acting on RNA*) et convertissent post-transcriptionnellement les adénosines en inosines, ces dernières étant reconnues par la machinerie cellulaire comme étant des guanosines (Figure 13). Ces enzymes (il en existe 3

chez les mammifères), exprimés dans beaucoup de tissus (ADAR1 et ADAR2) et particulièrement dans le cerveau (ADAR3), reconnaissent spécifiquement les ARNs double brin (et donc forcément les précurseur pri- et pre-miRs) **(Nishikura 2006; O'Connell 2006)**.

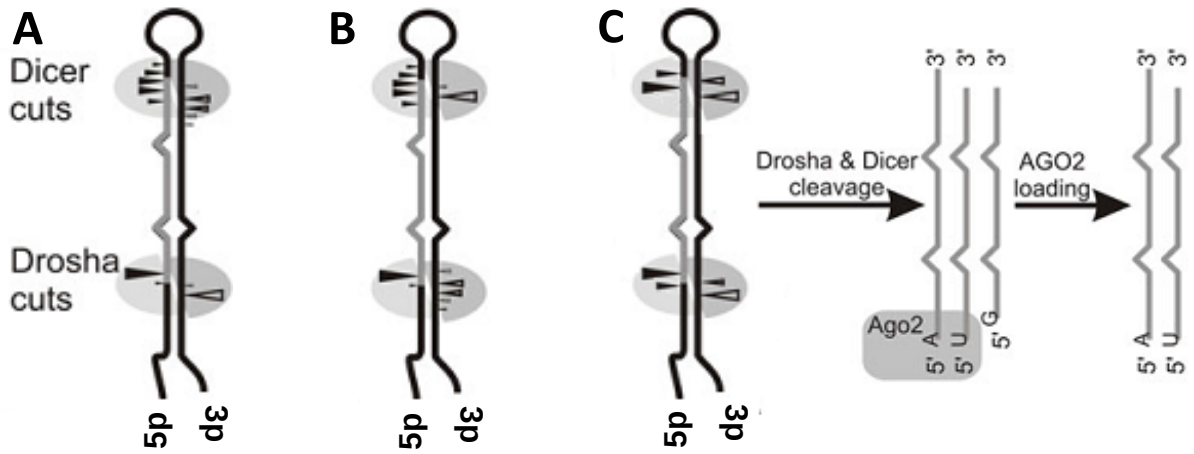


Figure 11. Différentes hypothèses expliquant l'hétérogénéité des extrémités des miARNs. A. Les coupures par Drosha et Dicer sont les principales sources d'hétérogénéité des miARNs, les extrémités générées par Drosha étant moins hétérogènes que celles générées par Dicer. **B.** Quel que soit le brin en question, l'extrémité 5' est toujours moins hétérogène que l'extrémité 3'. Cette hypothèse implique qu'intrinsèquement, le clivage par Drosha ou par Dicer serait plus précis en 5' qu'en 3' du miR, ce qui jusqu'à présent n'a pas été démontré **C.** Une troisième hypothèse stipule que Drosha et Dicer génèrent tous deux des extrémités variables, mais que l'interaction avec les protéines Argonautes limite l'hétérogénéité des miRs en 5'. Adapté de Starega-Roslan *et al.* **(Starega-Roslan 2011)**.

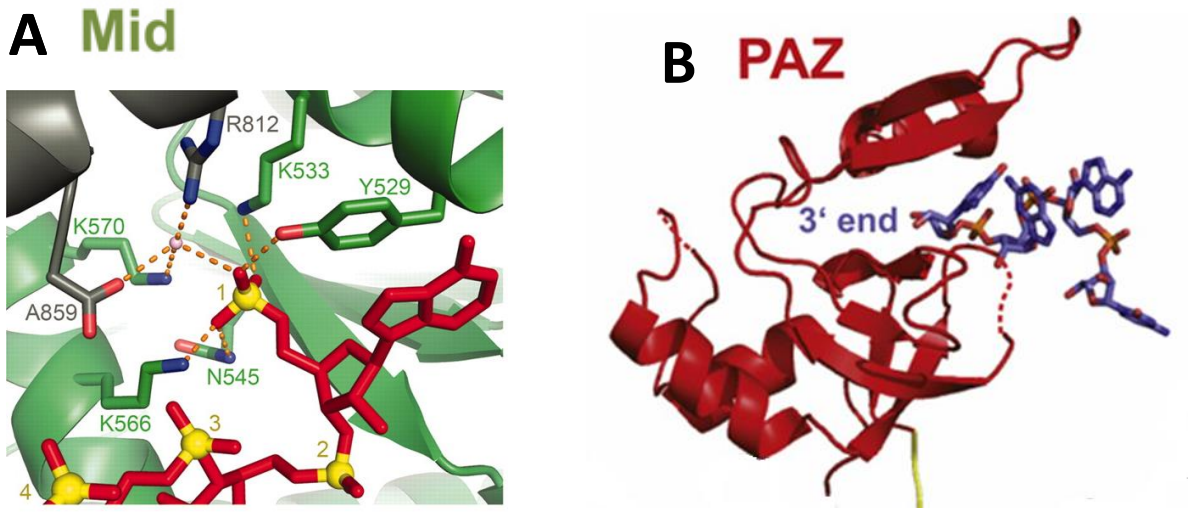


Figure 12. Structure des domaines *Mid* et *PAZ* de AGO2 qui lient les extrémités 5' et 3' respectivement. A. Plusieurs résidus du domaine *Mid* (en vert) interagissent directement avec les bases de l'extrémité 5' du miR (en rouge). D'après Schirle *et al.* **(Schirle 2012)**. **B.** L'extrémité 3' du miR (en bleu) dépasse du domaine *PAZ* (en rouge), et serait plus sujette à l'action des exoribonucléases et des nucléotidyl-transférases. D'après Elkayam *et al.* **(Elkayam 2012)**.

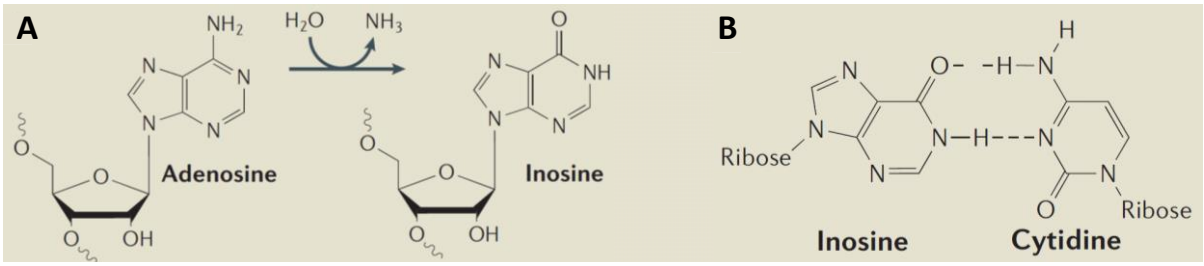


Figure 13. Editing A en I de l'ARN. A. L'inosine est obtenue par désamination hydrolytique de l'adénosine par l'enzyme ADAR. **B.** Contrairement à l'adénosine, l'inosine s'apparie avec la cytidine comme le ferait une guanosine. D'après Nishikura K. **(Nishikura 2006)**.

5.4 Conséquences fonctionnelles

5.4.1 Variabilité à l'extrémité 5' : liaison aux Argonautes et changement de cibles

Il a été proposé que la pression sélective exercée par le domaine *Mid* des protéines AGO permette d'éviter la grande variabilité aux extrémités 5' des miRs. Cependant, même si les modifications en 5' sont peu représentées par rapport à celles qui ont lieu en 3', elles restent toujours détectables avec potentiellement des conséquences fonctionnelles assez importantes. Azuma-Mukai *et al.* montrent qu'un variant du miR-142-5p possède deux nucléotides supplémentaires en 5' (+2C). *In vitro,* le miR et son variant seraient incorporés dans AGO2, et chacun d'eux aurait une spécificité différente par rapport à une cible artificielle **(Azuma-Mukai 2008)**. Pour ce même miR, Burroughs *et al.* confirment son association préférentielle avec AGO2 (81% d'affinité contre 21% pour AGO1). Par contre, ils montrent que le variant s'associe différemment avec les protéines argonautes (79% pour AGO1 et 19% pour AGO2). AGO1 et AGO2 possédant *in vivo* des activités différentes, seul AGO2 étant capable de cliver les ARNm, les auteurs proposent que les spécificités différentes entre le miR-142-5p et son variant soient en même temps liées au changement dans la séquence *seed* et à leurs affinités différentes avec les protéines argonautes. Ils montrent que ceci n'est pas valable uniquement pour le miR-142-5p, mais également pour d'autres miRs matures (miR-140-3p, miR-16, miR-29a) **(Burroughs 2011)**.

5.4.2 Editing : conséquences physiologiques et physiopathologiques

Une étude récente montre que chez la souris le niveau d'*editing* de certains miARNs augmente au cours du développement du cerveau ; ceci a poussé les auteurs à supposer que l'*editing* des microARNs jouerait un rôle physiologique important et ajouterait alors une couche supplémentaire de complexité aux différentes voies de régulation post-transcriptionnelles à certains moments clés **(Ekdahl 2012)**.

Comment ? D'une certaine façon, le peu de modifications que l'on retrouve dans la séquence des miRs matures par rapport au pri-miRs soutient l'idée que chez les animaux, la

principale fonction biologique de l'*editing* vis-à-vis des microARNs, si elle existe, serait la régulation de leur biogenèse plutôt que la modulation de leur capacité à réguler les messagers cibles **(de Hoon 2010)**. Cependant nous retrouvons dans la littérature certains exemples très intéressants infirmant cette hypothèse et soulignant le grand spectre d'action de l'*editing* et la diversité des conséquences fonctionnelles qui en découlent, que ce soit d'un point de vue physiologique ou physiopathologique.

Yang *et al.* montrent que l'*editing* ne serait pas aléatoire mais constituerait plutôt un phénomène régulé et spécifique de certaines séquences. Ils montrent que la conversion en inosine de certains résidus adénosine du pri-miR-142 conduit à l'inhibition de son *processing* par Drosha. Le pri-miR-142 édité est en effet reconnu par Tudor-SN, une ribonucléase qui fait partie du complexe effecteur RISC et qui reconnaît particulièrement les inosines des ARNs double brin (Figure 14A) **(Yang 2006)**. La même équipe a également montré que l'*editing* du pri-miR-151 n'a aucun effet sur l'étape catalysée par Drosha, mais qu'il interfère par contre dans la reconnaissance du pre-miR par le complexe protéique Dicer-TRPB. Ceci conduit à l'accumulation dans la cellule de pre-miR-151 édités qui seront éventuellement dégradés sans donner des miRs matures (Figure 14B) **(Kawahara 2007)**. Par extension, nous pouvons comprendre pourquoi tellement peu d'*editing* est retrouvé dans les séquences des miRs matures : leur rareté serait due à une dégradation rapide des précurseurs **(O'Connell 2006; Yang 2006)**. Finalement, de manière originale et par opposition à l'idée que l'*editing* aurait simplement pour conséquence l'inhibition de la biogenèse des miRs matures, les mêmes auteurs rapportent l'expression de plusieurs isoformes éditées des miRs de la famille miR-376. Un des sites d'*editing* se trouve localisé en plein milieu de la séquence *seed* et confère aux isomiRs leur capacité à cibler un ensemble spécifique de transcrits qui diffère de ceux ciblés par le miR non édité (Figure 14C) **(Kawahara 2007)**.

D'un point de vue oncogenèse, la relation qui existe entre *editin*g et développement tumoral est à établir. Plusieurs points restent à adresser comme par exemple savoir comment ce mécanisme est régulé et qu'est-ce qui assure sa spécificité vis-à-vis de certaines séquences **(Dominissini 2011; Gommans 2012)**. Une étude très intéressante réalisée par Paz *et al.* montre que globalement plusieurs tumeurs se caractérisent par une diminution du taux

d'*editing* par rapport au tissu sain. Les auteurs qualifient cette diminution de *« hypoediting »*, et l'attribuent à une sous-expression des enzymes ADAR **(Paz 2007)**. Partant de ce constat, Choudhury *et al.*, étudiant eux aussi le cluster miR-376, montrent une diminution de la fréquence d'*editing* de ce cluster dans les gliomes. Ils concluent leur étude en montrant que l'accumulation des formes non éditées dans les tumeurs serait à l'origine de leur agressivité et d'une augmentation de leurs capacités invasives. Ce phénotype étant largement lié au fait que les isoformes éditées et non éditées ciblent différents transcrits **(Choudhury 2012; Dominissini 2012)**.

Ce qui rend l'étude de l'*editing* et de ses conséquences fonctionnelles assez intéressante c'est d'une part sa spécificité vis-à-vis de certains transcrits en particulier mais d'autre part, et en même temps, son activité globale au sein du génome. L'*editing* peut virtuellement toucher n'importe quel ARN double brin, que ce soit les microARNs, les long ARNs non codants, les ARNs de transfert ou les ARNs messagers **(Peng 2012)**. Si l'on s'intéresse particulièrement à l'effet qu'il peut avoir sur la biogenèse et la fonction des miRs matures, il ne suffirait pas d'aller rechercher les sites d'*editing* au sein des miRs matures ou de leurs précurseurs, mais il faudrait voir si l'*editing* peut, au niveau des ARNm, créer ou perturber des sites de reconnaissance des miARNs. Une étude réalisée par Borchert *et al.* a essayé de répondre à cette question et cela en dehors d'un contexte tumoral. Les auteurs montrent que parmi les 12723 évènements d'*editing* qu'ils ont étudiés au niveau des 3'UTRs, 3000 ont une complémentarité parfaite avec des séquences *seed*, et que beaucoup de ces nouveau sites peuvent spécifiquement être reconnus par des microARNs **(Borchert 2009)**. Il serait donc envisageable qu'une perturbation de l'*editing* au cours de la tumorigenèse puisse activer des oncogènes et inversement inactiver des gènes suppresseurs de tumeurs. Cela en introduisant d'une part des mutations dans les séquences codantes (puisque les inosines sont reconnues comme des guanosines) et d'autre part, au niveau des 3'UTRs en affectant la spécificité de reconnaissance miARN/ARNm **(Dominissini 2011)**.

5.4.3 Conséquences de l'uridylation et de l'adénylation

Comme nous l'avons précédemment décrit, les modifications qu'on retrouve aux extrémités 3' des miRs peuvent être « *templated* » ou « *non-templated* ». Les *« templated »* sont dues à des erreurs commises par Drosha et Dicer au cours de la biogenèse des miRs et les « *non-templated »* sous-entendent le fait qu'un mécanisme particulier se charge d'ajouter des nucléotides (principalement des A et des U), et cela indépendamment de la séquence du miR mature.

Il a été montré que l'uridylation joue un rôle régulateur très important dans la biogenèse des miRs, que cela soit chez les plantes ou chez les mammifères **(Kim 2010)**. La biogenèse des miRs de la famille let-7 en est un parfait exemple. Cette famille, dont les membres sont parmi les plus anciens et les plus conservés des microARNs, régule l'activité de plusieurs processus physiologiques (cycle cellulaire, développement, différentiation des cellules souches, etc. …) et de ce fait est impliquée dans de nombreuses pathologies humaines y compris les cancers **(Thornton 2012)**. De manière intéressante, la grande variabilité de séquence qui caractérise cette famille (du moins au niveau pri- et pre-miR) se trouve contrebalancée par des motifs structuraux conservés communs à tous ses membres **(Nam 2011)**. Ces motifs, nécessaires à la régulation de la biogenèse des miRs, permettent l'interaction avec la protéine Lin28, facteur crucial impliqué dans l'auto-renouvèlement des cellules souches embryonnaires. Dans ces cellules, Lin28B serait en fait capable de lier les pri-miR let-7 et d'inhiber l'étape catalysée par Drosha (cela par un mécanisme encore inconnu) et Lin28A serait capable de lier les pre-miR let-7 et d'induire leur oligo-uridylation en recrutant la nucléotidyltransférase TUT4 (*Terminal UridylylTransferase,* connue également sous le nom de ZCCHC11). L'ajout de la queue poly-U en 3' – qui par ailleurs est connue pour être associée à la dégradation exonucléotidique des ARNm – empêcherait la reconnaissance et le *processing* du pre-let-7 par Dicer et faciliterait donc sa dégradation. Dans les cellules souches embryonnaires non différenciées, Lin28 est fortement exprimé et fait partie d'un réseau moléculaire assurant la totipotence cellulaire. Au cours du développement, le niveau d'expression de Lin28 diminue permettant l'accumulation des

miRs de la famille let-7. Cette accumulation conduit à l'arrêt de l'auto-renouvèlement des cellules souches et favorise la différenciation cellulaire **(Kim 2010; Thornton 2012)**.

L'oligo-uridylation interfère donc avec la capacité de Dicer à reconnaître les pre-miR. Rappelons que cette capacité est largement dépendante du domaine protéique PAZ qui reconnaît spécifiquement les **deux nucléotides** en 3' *overhang* du pre-miR (voir précédemment, §2.3.1). Très récemment, une étude a identifié une classe un peu particulière de pre-miRs – qu'ils appellent *« group II pre-miRNA »* – dont les membres possèdent **un seul nucléotide** en 3' *overhang* au lieu d'en avoir deux. Comme pour la queue poly-U, la présence d'un nucléotide en 3' *overhang* empêche le pre-miR de se fixer au domaine PAZ. De manière complètement inattendue, et indépendamment de Lin28, ces pre-miRs subissent systématiquement une uridylation en 3' (ajout d'un seul U) qui restaure leur capacité à être reconnus par Dicer. Même si la mono-uridylation n'a été rapportée que pour les membres de la famille let-7, certains auteurs laissent suggérer qu'elle ne se limiterait pas à cette famille, mais constituerait plutôt une composante clé de la machinerie de biogenèse des microARNs **(Heo 2012; Scott 2013).**

Qu'en est-il du rôle de l'adénylation ? Chez les mammifères, il a été en fait démontré que l'adénylation du miR-122 dans les cellules hépatocytaires permet sa stabilisation **(Katoh 2009)**. Comme pour l'uridylation, nous nous attendons aussi à une dualité dans le rôle de l'adénylation. En effet, une étude réalisée par Burroughs *et al.*, souligne le fait que l'adénylation, d'un point de vue global, si elle est assez prépondérante, n'affecte pas la stabilité des microARNs, mais réduit leur capacité à interagir avec les protéines Argonautes. Les auteurs notent également que l'adénylation se fait sur le miR mature et non pas sur ses précurseurs comme il a été décrit pour l'uridylation **(Burroughs 2010; Burroughs 2011)**. Que ce soit en stabilisant le miR ou en impactant sa capacité à interagir avec les protéines Argonautes, les bases moléculaires qui expliquent l'une ou l'autre de ces hypothèses restent encore inconnues **(Kim 2010)**.

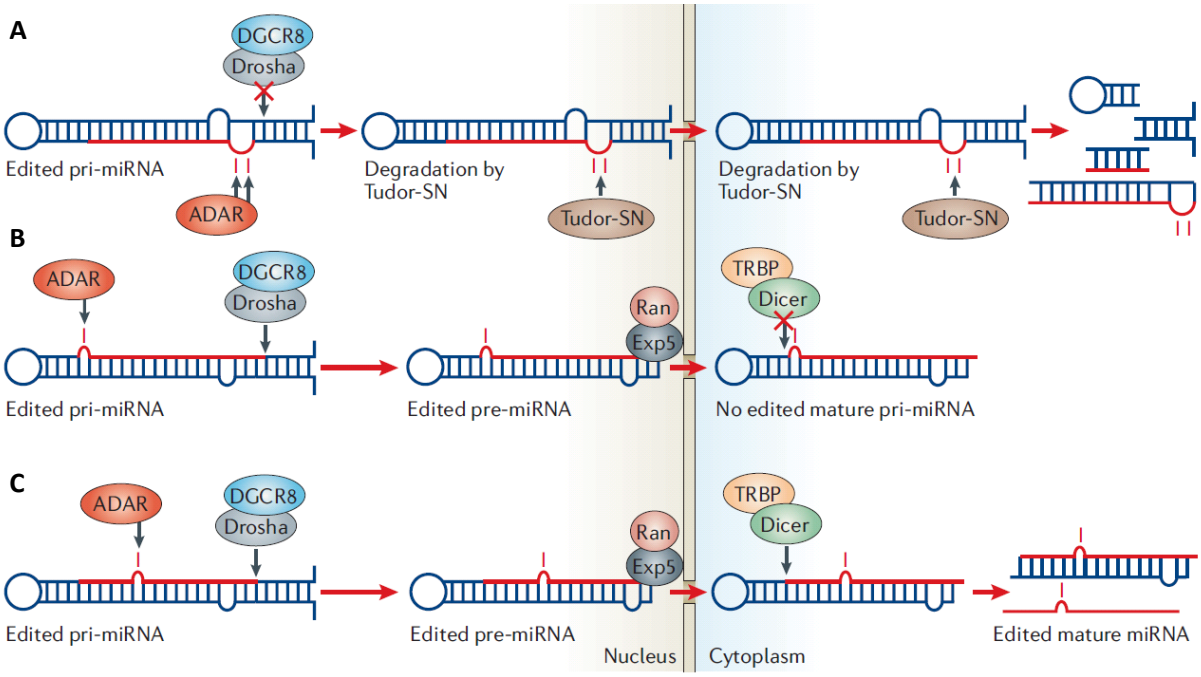


Figure 14. Régulation de la biogenèse et de la fonction des microARNs par *Editing*. Ce schéma représente les effets connus jusqu'à présent de l'*editing* sur les miARNs. **A.** L'*editing* empêche l'interaction avec Drosha et conduit à la dégradation du pri-miR par la nucléase Tudor-SN qui reconnaît spécifiquement les inosines (cas du pri-miR-142). **B.** De la même façon, l'*editing* peut bloquer l'étape catalysée par Dicer (cas du pre-miR-151). **C.** L'*editing* n'interfère pas dans la biogenèse du miR mais module sa spécificité vis-à-vis de ses cibles (cas du cluster miR-376). D'après Nishikura K. **(Nishikura 2006).**

6. Cancers colorectaux instables sur les microsatellites

6.1 Généralités sur les cancers colorectaux

Selon les chiffres de l'Institut National du Cancer (INCA)[*] et en partenariat avec la Ligue Nationale contre le Cancer (LNCC), les cancers colorectaux se situent, tous sexes confondus, au troisième rang des cancers les plus fréquents en France (deuxième cancer chez la femme et troisième chez l'homme). Il s'agit essentiellement d'un cancer de l'âge mûr, avec près de 85 % des cas survenant après 65 ans. En 2011, le nombre estimé de nouveaux cas par année était d'environ 40 500. D'après les estimations, et toujours selon l'INCA, le nombre de cancers colorectaux devrait augmenter dans les prochaines années pour atteindre les 45 000 nouveaux cas annuels en 2020 en France. Le taux de mortalité (standardisé monde) en 2011 atteignait 13,8 pour 100 000 hommes et 8,2 pour 100 000 femmes. Aux Etats-Unis, le *National Cancer Institute* estime à 1 150 000 le nombre de personnes vivant aux USA avec un cancer du côlon ou de rectum en 2013[†]. Le nombre de nouveaux cas diagnostiqués par an est quant à lui de 142 820 en 2013 (soit environ 8.6% de tous les cancers).

Les cancers colorectaux, comme leur nom l'indique, sont des cancers qui affectent les épithéliums de côlon et du rectum. Il s'agit en fait d'une monocouche de cellules épithéliales présentant plusieurs invaginations (qu'on appelle cryptes) permettant d'augmenter la surface d'absorption (les deux rôles principaux du côlon étant l'absorption de l'eau et l'élimination des résidus alimentaires). A la base de chaque invagination, 4 à 6 cellules souches donnent naissance aux 3 types cellulaires majoritaires qui composent la crypte : les cellules neuroépithéliales, les cellules caliciformes muco-sécrétantes et les cellules absorbantes de type entérocytaire. Le taux de prolifération cellulaire à la base de la crypte est compensé par une perte progressive des cellules différenciées au niveau de l'apex. Une

[*] http://www.e-cancer.fr/cancerinfo/les-cancers/cancers-du-colon/quelques-chiffres-sur-les-cancers-colorectaux (mise à jour du 27.01.2012)

[†] http://www.cancer.gov/cancertopics/types/colon-and-rectal

perturbation de cet équilibre résulte en l'accumulation anormale de cellules filles provenant d'une seule cellule progénitrice que l'organisme n'arrive pas à éliminer. Un épithélium hyperplasique est un épithélium caractérisé par un développement anormal des cellules sans que l'architecture globale de la crypte ou la morphologie des cellules ne soient altérés (Figure 15, *hyperplastic epithelium*). L'une des premières manifestations d'une transformation cellulaire se caractérise par l'apparition d'un polype adénomateux (Figure 15, *adenomas*). Il s'agit d'un bourgeonnement – non cancéreux au début – de l'épithélium colique présentant une organisation intercellulaire (plusieurs couches de cellules) et intracellulaire (noyaux cellulaires plus grands et mal positionnés) aberrantes. On parle alors de dysplasie, qui augmente en grade avec la croissance du polype. L'acquisition de la capacité à envahir les tissus adjacents caractérise l'étape finale de la transformation de la muqueuse colique (Figure 15, *carcinoma/invasion and metastasis*). Les adénomes se transforment en adénocarcinomes caractérisés par leur agressivité et leur capacité à faire des métastases (vers les ganglions lymphatiques, la surface péritonéale ou le foie) **(Vogelstein 2002; Weinberg 2007)**.

En 1990, dans un célèbre article publié dans la revue *Cell*, Fearon et Vogelstein ont établi, en termes moléculaires, la première séquence de transformation d'un épithélium colique sain en adénocarcinome invasif. C'est un processus qui requiert l'accumulation progressive et par étapes de plusieurs mutations touchant des gènes clés dans la cellule (Figure 15) **(Fearon 1990)**.

Le principal gène impliqué dans l'initiation des cancers colorectaux est le gène *APC (Adenomatous Polyposis Coli)*. La perte d'*APC* (un gène suppresseur de tumeurs) conduit à une activation constitutionnelle de la voie de signalisation WNT/β-catenine et à une profération anarchique et continue des cellules coliques (la forme sauvage de l'APC forme un complexe protéique qui dégrade la β-catenine et l'empêche donc de transloquer vers le noyau pour activer entre autre les gènes de prolifération cellulaire). Le polype qui résulte de cette mutation (en réalité il s'agit de la prolifération clonale d'une seule cellule, du moins au départ), est susceptible d'acquérir d'autres mutations. Une mutation activatrice dans le gène *KRAS* (un oncogène) conduit à une augmentation de la prolifération cellulaire (KRAS est une

enzyme impliquée dans les voies de signalisation en réponse à divers facteurs de croissance dont l'EGF, *epidermal growth factor*). D'autres mutations s'ensuivent, créant ainsi plusieurs sous-populations clonales. Il s'agit en particulier de gènes suppresseurs de tumeurs localisés sur le chromosome 18q (entre autres DCC, *deleted in colorectal cancer*, SMAD 2 et 4, impliqués dans la voie de signalisation TGF-β, …) et le suppresseur de tumeur p53 qui conduit finalement à la formation d'un adénocarcinome. Les cellules cancéreuses présentent alors une forte instabilité caractérisée par des gains et des pertes de fragments chromosomiques qui se rajoutent à ces différentes mutations et confère aux cellules la capacité d'envahir les tissus adjacents. De manière assez intéressante, ce modèle donne une grande importance à la survenue séquentielle des différentes mutations. Par exemple, si une mutation de *KRAS* survient avant celle d'*APC*, les cellules évoluent très rarement en cancer. De plus, toutes les tumeurs coliques ne présentent pas systématiquement des mutations dans tous les gènes précédemment cités, indiquant que d'autres gènes seraient également impliqués de manière non négligeable dans la carcinogenèse colique **(Fearon 1990; Vogelstein 2002)**.

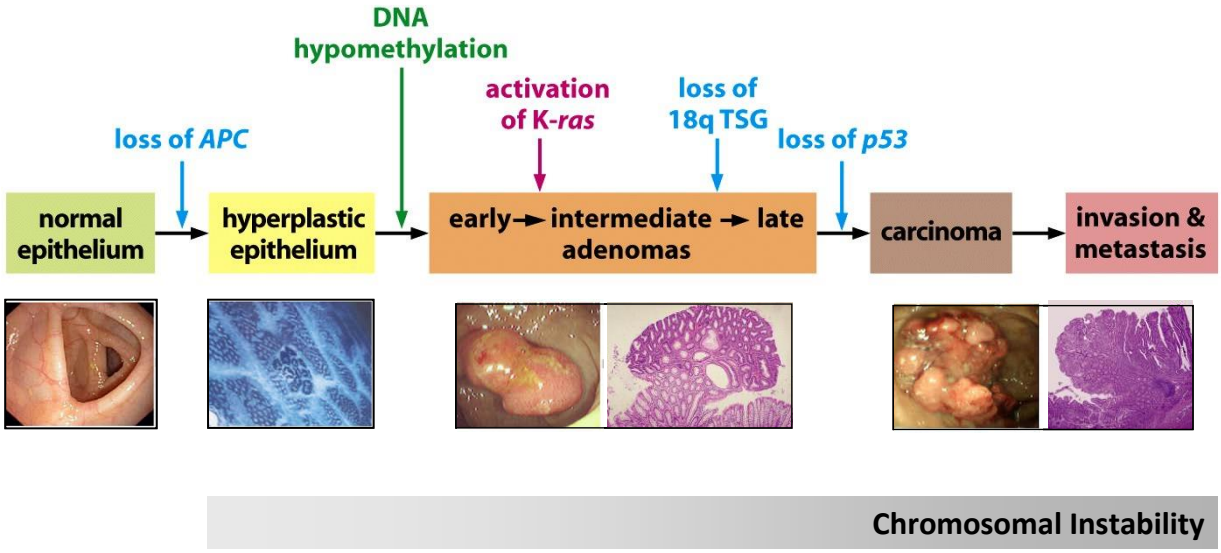


Figure 15. Modèle de progression multi-étapes des cancers colorectaux. Cette figure illustre les différentes étapes de la progression tumorale des cancers colorectaux selon le modèle de Fearon et Vogelstein. La partie du haut représente les principales altérations moléculaires qui accompagnent les modifications morphologiques et anatomiques (partie du bas) **(Fodde 2001; Weinberg 2007; Pino 2010)**.

6.2 Le système MMR et sa perte dans les cancers colorectaux

Nous savons actuellement que cette voie de carcinogenèse colique correspond principalement aux cancers dits de type MSS, caractérisés par une instabilité chromosomique (ou CIN, pour *Chromosomal Instability*). Le terme MSS *(Microsatellite Stable)* est utilisé pour marquer leur opposition par rapport aux cancers colorectaux instables sur les microsatellites (MSI, pour *Microsatellite Instability*). Les microsatellites sont des séquences d'ADN formées par une répétition continue de motifs composés de 1 à 6 nucléotides et dispersées dans le tout le génome. En 1993, trois équipe différentes ont montré qu'entre 12 et 15% des cancers colorectaux présentaient un profil mutationnel bien distinct : des mutations somatiques (insertions ou délétions de nucléotides) altérant la taille des séquences microsatellites (Figure 16B) **(Aaltonen 1993; Ionov 1993; Peltomaki 1993; Thibodeau 1993)**. Thibodeau *et al.*, ont particulièrement montré que le phénotype MSI était inversement corrélé à la perte d'hétérozygotie des chromosomes 5q, 17p et 18q (associés à l'instabilité chromosomique), et corrélait positivement avec une meilleure survie des patients **(Thibodeau 1993)**. Très rapidement, ce phénotype (appelé également hypermutateur) a été associé à une perte fonctionnelle du système de réparation des mésappariements de l'ADN (MMR, pour *Mismatch Repair*) (Figure 16) **(Fishel 1993; Leach 1993)**.

Les protéines qui forment le système MMR sont : MSH2, MSH3, MSH6, MLH1, MLH3, PMS1 et PMS2. MSH2 et MSH6 forment le complexe MutSα capable de reconnaitre les simples mésappariement de bases (incorporation du mauvais nucléotide) et les petites boucles liées à l'insertion ou la délétion de quelques nucléotides (généralement 1 ou 2 nucléotides). MSH2 et MSH3 forment le complexe MutSβ capable de reconnaitre les grandes boucles d'insertions ou de délétions (de 2 à 16 nucléotides). Après l'étape de reconnaissance, MutSα ou MutSβ recrutent le complexe MutLα (hétéro-dimère formé par MLH1 et PMS2). MutLα forme alors un anneau coulissant qui scanne l'ADN à la recherche d'entailles *(nicks)* ou de sites hypo-méthylés caractérisant le brin néosynthétisé. La partie du brin néosynthétisé est alors excisé, l'ADN polymérase de nouveau recrutée et un nouveau brin est alors synthétisé (Figure 16A) **(Weinberg 2007; Grady 2008; Hamelin 2008; Boland 2010)**.

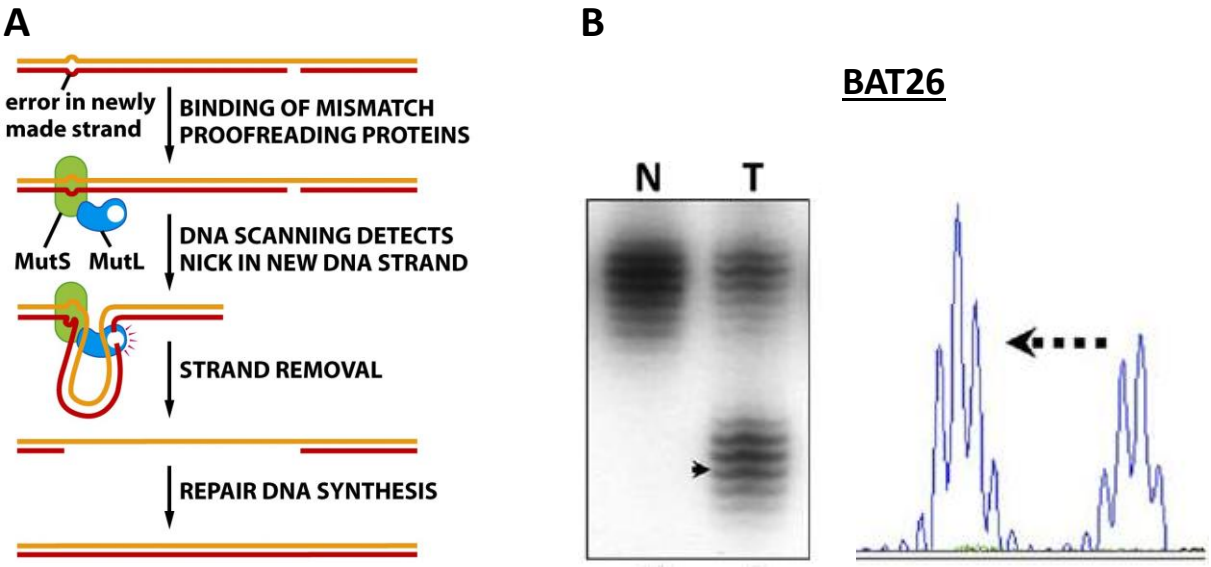


Figure 16. Le système de réparation des mésappariements et instabilité microsatellitaire. A. Le système MMR prend en charge les « glissements » (*stuttering* ou *strand slippage*) des ADN polymérases qui conduisent la plupart du temps à des délétions de nucléotides au niveau des microsatellites et donc à la formation d'un boucle *(bulge)* au sein de l'ADN. **B.** Un défaut du système MMR se manifeste par l'accumulation de ces mutations, on parle alors d'instabilité microsatellitaire ou MSI. MSI est détectable par autoradiographie après électrophorèse des produits PCR (gauche), ou par électrophorèse capillaire avec des amorces fluorescentes (droite). Le marqueur utilisé dans cette figure est BAT26 (répétition mononucléotidique), un des 5 marqueurs recommandés par le *« National Cancer Institut »* pour la détection de MSI. Les délétions se manifestent par l'apparition de bandes supplémentaires (flèches). N, tissu normal ; T, Tissu tumoral. **A**, d'après Robert A. Weinberg, The Biology of Cancer **(Weinberg 2007)**. **B**, d'après Boland *et al.* **(Boland 2010)**.

6.3 Gènes cibles de l'instabilité microsatellitaire

Le processus d'instabilité microsatellitaire n'est pas oncogénique en soit, mais c'est l'accumulation de nombreuses mutations au niveau de gènes cibles d'instabilité qui semble être responsable du processus tumoral. L'immense majorité des séquences répétées se trouve localisée dans des régions intergéniques. Il existe cependant des séquences répétées localisée dans les exons de certains gènes. Des mutations (insertions ou délétions de nucléotides) aux niveaux de ces séquences entraînent un décalage dans le cadre de lecture (ORF, *Open Reading Frame*) et donc l'apparition d'une protéine tronquée, le plus souvent non fonctionnelle. S'il s'agit de protéines impliquées dans des fonctions telles que la prolifération cellulaire, la réponse aux facteurs de croissance ou l'apoptose, elles risquent alors de jouer un rôle dans le développement ou la progression tumorale MSI **(Duval 2002; Hamelin 2008)**. Le premier gène identifié comme cible de l'instabilité microsatellitaire est le récepteur du TGFβ (TGFβR2). Markowitz *et al.* ont montré que le gène *TGFβR2* ne s'exprimait pas dans les lignées cellulaires et xénogreffes MSI et que cela était dû à la présence dans sa partie codante d'une séquence microsatellite très fréquemment mutées dans un contexte MSI **(Markowitz 1995)**. TGFβ et son récepteur TGFβR2 agissent dans ce cas comme des suppresseurs de tumeurs, et leur perte de fonction dans un contexte MSI favorise la prolifération des cellules épithéliales. D'autres gènes ont également été montrés comme étant la cible de MSI. Il s'agit notamment de *BAX*, *IGF2R*, *TCF4, RAD50*, *CHK1*, *PTEN*, *MSH6*, *MLH3*, ainsi que d'autres gènes impliqués dans différentes voies de signalisation comme la transduction du signal, l'apoptose et l'inflammation, la régulation de la transcription, la signalisation des dommages de l'ADN et leur réparation (Figure 17) **(Duval 2002; Hamelin 2008)**.

Outre l'instabilité microsatellitaire, on retrouve dans les cancers colorectaux de type MSI des mutations somatiques (conséquence de l'instabilité nucléotidique générale caractérisant ces tumeurs) qui les distinguent des cancers de type MSS. Par exemple, une mutation activatrice de *BRAF* (un gène qui code pour une serine/thréonine kinase dans la voie de signalisation MAP/ERK) se retrouve presque exclusivement dans les cancers MSI et est mutuellement exclusive des mutations *KRAS* (qu'on retrouve majoritairement dans les cancers MSS et très

peu dans les tumeurs MSI). De manière assez intrigante, la mutation *BRAF* se retrouve uniquement chez les patients atteint de cancers MSI sporadiques et non héréditaires (voir ci-dessous) et serait donc un moyen – quoique peu fiable – de distinguer entre ces deux entités (car la fréquence de la mutation n'atteint pas 100% dans les tumeurs sporadiques). BRAF et KRAS appartenant à la même voie de signalisation, les conséquences de leur mutation se ressemblent donc assez entre les cancers MSS et MSI. Un autre exemple est celui du gène *APC* (fréquemment muté dans les cancers MSS). La grande majorité des cancers MSI possèdent des taux d'expression normaux de la protéine APC. Dans ce cas, une mutation de la β-catenine la rend incapable d'être reconnue par l'APC et a donc les mêmes conséquences qu'une perte de fonction de l'APC **(Hamelin 2008; Boland 2010; Vilar 2010).** En outre, l'équipe d'accueil a rapporté des mutations fréquentes de *TCF-4 (T cell transcription factor-4)* dans les cancers MSI ; ce gène, qui code pour un facteur de transcription, est l'effecteur principal de la voie Wnt et les mutations empêchent l'inactivation du mutant par CtBP (C-terminal-binding protein) **(Duval 1999; Duval 2000; Cuilliere-Dartigues 2006)**.

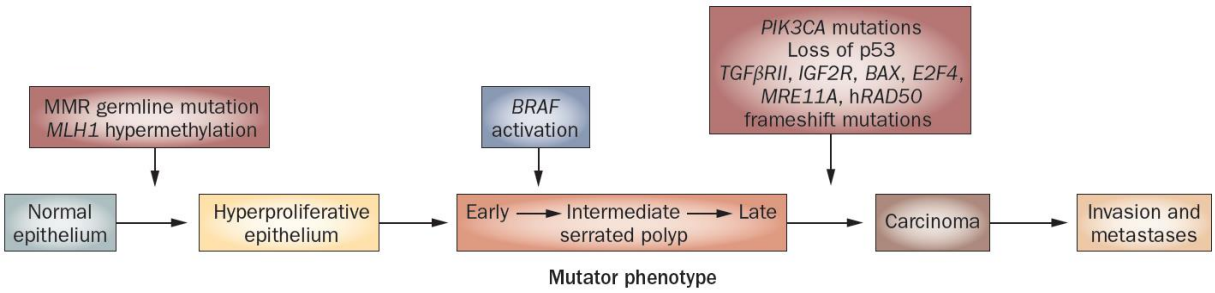


Figure 17. Modèle de progression des cancers colorectaux de type MSI. L'évènement principal dans la mise en place des cancers colorectaux de type MSI est la perte de fonction d'une ou de plusieurs protéines du système MMR. Cette perte a pour conséquence une instabilité génétique qui se manifeste surtout au niveau des séquences microsatellites et est oncogénique quand elle touche des gènes impliqués dans des fonctions cellulaires clés. Extrait de Vilar *et al.* **(Vilar 2010)**.

6.4 Caractéristiques clinico-pathologiques des cancers colorectaux de type MSI

6.4.1 Cancers MSI sporadiques et héréditaires

Les études réalisées sur les cancers colorectaux ont montré qu'il existe deux mécanismes d'apparition des tumeurs MSI. Le premier est héréditaire : des mutations germinales dans les gènes *MLH1*, *MSH2*, et plus rarement *MSH6* et *PMS2* sont responsable de la prédisposition génétique aux cancers connue sous le nom du syndrome de Lynch. Le syndrome de Lynch (du nom de Henry T. Lynch qui fut le premier à caractériser ce syndrome en 1966) est une prédisposition héréditaire au cancer de transmission autosomique dominante qui touche environ 2 à 5% des patients atteints de cancers colorectaux **(Hampel 2005)**. Les patients héritent d'un allèle muté, la deuxième mutation ou perte d'hétérozygotie survenant plus tard de manière sporadique et à un âge assez jeune (généralement avant 50 ans). Le spectre clinique du syndrome de Lynch est assez large, pouvant inclure outre les cancers colorectaux, certains cancers de l'endomètre, des voies urinaires, de l'intestin grêle, de l'estomac, des voies biliaires, du pancréas, des ovaires, etc. … Le second mécanisme, qui est impliqué dans la majorité des cas de cancers MSI, est épigénétique et consiste en une hyperméthylation du promoteur de *MLH1*. Ces cancers sont sporadiques (pas de prédisposition familiale). Les raisons sous-jacentes à cette hyperméthylation sont multiples et incluent les épisodes d'inflammation de l'épithélium colique, l'utilisation de certaines drogues en clinique, etc. La méthylation du promoteur de *MLH1* fait partie quant à elle d'un phénotype plus large de méthylation anormale de plusieurs gènes qui inclut également certaines tumeurs MSS : le phénotype CIMP *(CpG Island Methylator Phenotype)* dont l'existence en tant qu'entité clinique reste encore controversée **(Hamelin 2008)**. Il existe également, mais de manière très rare, des cas d'hyperméthylation héritables. Il s'agit d'une mutation germinale caractérisée par la délétion du dernier exon du gène *EPCAM,* un gène qui précède *MSH2*. La délétion de l'exon conduit à un transcrit de fusion formé par *EPCAM* et *MSH2* et à la méthylation du promoteur de *MSH2* **(Ligtenberg 2009)**.

6.4.2 Pronostic tumoral et réponse aux traitements de chimiothérapie

L'importance de déterminer le statut d'instabilité vient des propriétés clinico-pathologiques des tumeurs MSI qui les différencient des tumeurs MSS : leur pronostic et leur réponses aux traitements. En outre, le dépistage du phénotype MSI permet d'aider à la détection du Syndrome de Lynch chez un patient. Gryfe *et al.* ont été les premiers à montrer sur une cohorte de 607 patients de CCR MSS et MSI que les tumeurs MSI auraient un meilleur pronostic que les tumeurs MSS et qu'elles seraient moins sujettes à former des métastases dans les ganglions lymphatiques ou à distance **(Gryfe 2000)**. Plus récemment, une méta-analyse réalisée par Popat *et al.*, regroupant les données de 32 études différentes (soit au total 7642 patients, dont 1277 CCRs MSI) a confirmé de manière significative l'avantage pronostique de MSI **(Popat 2005)**.

L'impact du phénotype MSI sur la réponse au 5-fluoro-uracile (5FU) a fait l'objet de plusieurs études avec des résultats assez contradictoires (le 5-FU agit principalement sur la synthèse d'ADN en bloquant l'activité de la thimidylate synthase, conduisant à un arrêt du cycle cellulaire et à l'apoptose). Il semblerait que ce traitement ne bénéficie pas aux patients ayant une tumeur MSI (stades II ou III), et pourrait même être délétère pour ces patients **(Ribic 2003; Hamelin 2008)**. Inversement, les tumeurs de patients atteints de cancer colorectal métastatique sont plus susceptibles et répondent favorablement à un traitement par Irinotecan, un inhibiteur de topo-isomérase I (et donc l'inhibition à la fois de la réplication de l'ADN et de la transcription), si leur tumeur était de phénotype MSI **(Fallik 2003; Hamelin 2008)**. Il existe encore très peu de données pour comprendre les raisons qui sont à l'origine du meilleur pronostic des cancers MSI et de leur réponse différente à la chimiothérapie. Dans ce champ de recherche, les travaux de l'équipe d'accueil ont rapporté des premières pistes intéressantes, relativement à la découverte de mutations dues à MSI qui affectent la chaperonne HSP110 dans ces tumeurs **(Dorard 2011)**.

7. MicroARNs et dérégulation de leur expression dans les CCRs

7.1 Introduction : microARNs et cancers

De nombreuses études ont montré que les miARN sont exprimés de manière aberrante dans divers cancers, laissant suggérer que ces petites molécules pourraient jouer un rôle d'oncogène ou de gène suppresseur de tumeurs (Figure 18) **(Esquela-Kerscher 2006; Zhang 2007)**. L'une des premières études explorant cette thématique a été réalisée en 2002 par le laboratoire de Carlo Croce. Calin *et al.* ont en effet montré que les miR-15 et miR-16 étaient situés au niveau du chromosome 13q14, une région fréquemment perdue dans les leucémies lymphoïdes chroniques, et que dans environ 68% des cas, ces deux miRs étaient significativement sous-exprimés **(Calin, 2002)**. Cette perte d'expression, jouant probablement un rôle pro-tumorigénique, laisse supposer que les deux miRs seraient en réalité des suppresseurs de tumeurs et régulerait l'expression d'oncogènes. Nous savons actuellement que les miR-15 et miR-16 régulent entre autres l'expression du facteur anti-apoptotique BCL2 **(Fabbri 2011)**. Deux ans plus tard, la même équipe a montré qu'environ 50% des gènes de miARN (98 miRs sur les 186 étudiés) sont localisés dans des régions génomiques associées au cancer ou au sein de sites fragiles de chromosomes et qu'ils sont sous-exprimés dans la plupart des tumeurs analysées **(Calin 2004)**. De manière plus approfondie, Lu *et al.*, étudiant les niveaux d'expression de 217 microARNs dans 334 échantillons humains, ont mis en évidence une signature miARN spécifique à chaque type de cancer étudié. Cette signature, qui distinguerait les cancers des tissus sains, constituerait un outil de classification surpassant les ARNm et essentiel à la compréhension des mécanismes moléculaires de tumorigenèse, et aurait des implications à la fois diagnostiques et thérapeutiques **(Lu 2005)**.

Depuis, des centaines de microARNs ont été caractérisés comme étant dérégulés dans divers tissus tumoraux **(Esquela-Kerscher 2006; Croce 2009; Lujambio 2012)**. Nous citerons par exemple une sous-expression de certains membres de la famille let-7 capables de cibler la protéine oncogénique RAS et qui serait associée à un mauvais pronostic chez les patients atteints de cancer du poumon mais également dans d'autres types de cancers (dans ce cas, il

s'agit donc de suppresseurs de tumeurs) **(Takamizawa 2004; Johnson 2005)**. Nous citerons également l'exemple du *cluster* polycistronique *mir-17~92*, situé sur le chromosome 13q31 (région connue pour être fréquemment amplifiée dans plusieurs types de lymphomes). Il a été montré que l'expression de ce *cluster* était induite par le proto-oncogène c-Myc : les deux (i.e. le *cluster* et c-Myc) agiraient alors afin de promouvoir la division cellulaire et accélérer le développement tumoral, particulièrement dans les lymphomes B **(He 2005)**. Les miRs du cluster cibleraient entre autres des molécules de régulation du cycle cellulaire comme PTEN, CDKN1A, etc… (il s'agit donc plutôt d'oncomiRs) **(Lujambio 2012)**.

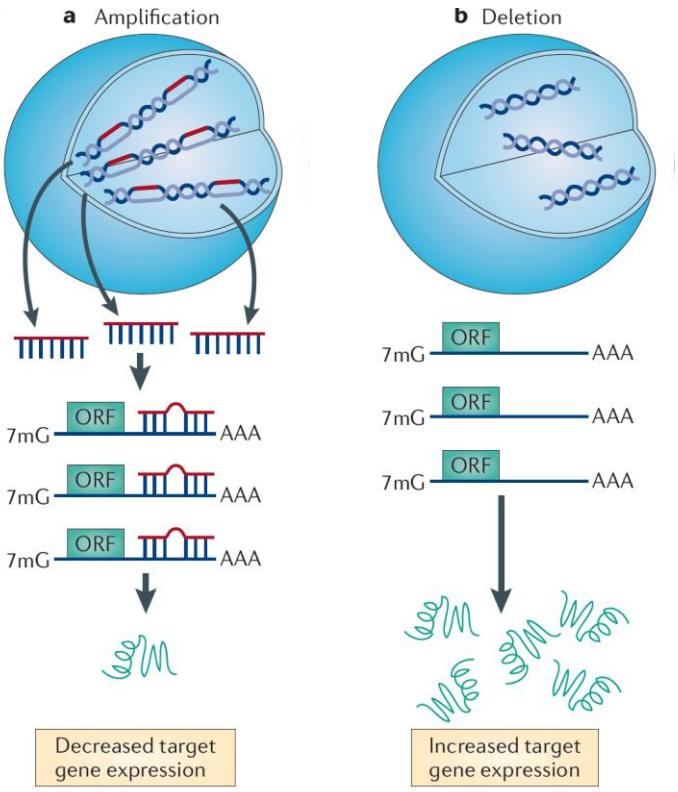


Figure 18. Les microARNs en tant que suppresseurs de tumeurs et oncogènes. Cette figure illustre les différents aspects de la dérégulation de l'expression des miRs. **a.** Une amplification génomique entraîne la surexpression de certains miRs et une sous-expression de leurs gènes cibles. Si les gènes cibles en question régulent la progression dans le cycle cellulaire ou sont pro-apoptotiques par exemple, ces miRs sont alors considérés comme oncomiRs. **b.** Inversement, des délétions chromosomiques conduisent à la sous-expression de certains miRs et s'accompagnent par une surexpression des cibles. Si les gènes cibles sont oncogéniques, les miRs sont alors considérés comme suppresseurs de tumeurs **(Kasinski 2011)**.

7.2 Généralités sur les microARNs dans les cancers colorectaux

Sans vouloir en faire une revue exhaustive, nous décrirons dans ce paragraphe les principales découvertes concernant des miARNs particuliers et leur rôle dans le développement et la progression des cancers colorectaux.

La première étude dans le cadre de cette thématique a été réalisée en 2003 par Michael *et al.* En effet, ces auteurs ont remarqué une sous-expression des miR-143 et miR-145 dans les tumeurs coliques **(Michael 2003)**. De manière intéressante, les niveaux d'expression des précurseurs restent inchangés, soulignant donc une régulation post-transcriptionnelle de la biogenèse de ces miRs dans les cellules tumorales. La sous-expression des miRs 143 et 145 serait capable d'accélérer la croissance et la prolifération des cellules tumorales. Le miR-145 ciblerait entre autres *l'Insulin Receptor Substrate 1* (IRS-1, une molécule jouant un rôle important à la fois dans des voies métaboliques et de croissance cellulaire), le proto-oncogène c-Myc, le proto-oncogène YES1 *(Yamaguchi sarcoma viral oncogen homolog 1)* et STAT1 *(Signal Transducers and Activators of Transcription 1),* tous jouant un rôle dans la carcinogenèse colorectale **(Schetter 2011; Goel 2012)**. Le miR-143 ciblerait entre autres KRAS, DNMT3A (DNA (cytosine-5)-methyltransferase 3A, une enzyme impliquée dans la méthylation des ilots CpG) et ERK5 (ou MAPK7*, Mitogen-activated protein kinase 7*) **(Schetter 2011; Goel 2012)**. D'autres microARNs sont également impliqués dans la carcinogenèse colorectale. Il a été rapporté en particulier la surexpression du miR-21 (que nous détaillerons dans le §9.3.3) et des miRs du cluster mir-17~92 (ciblant le facteur de transcription E2F1 et donc régulant la prolifération cellulaire) et une sous-expression des membres de la famille let-7 **(Schetter 2011; Goel 2012)**. A grande échelle, une première signature miARN (mirnome) caractérisant les cancers colorectaux a été établie. L'étude, réalisée par Cummins *et al.* a permis d'identifier 133 nouveaux miRs caractéristiques des tumeurs colorectales parmi lesquels figurent entre autres les miRs 143 et 145 (en comparaison au tissu sain) **(Cummins, 2006)**. Plus récemment une étude réalisée par le *Cancer Genome Atlas Network* sur 224 tumeurs colorectales et muqueuses saines, a combiné des données de séquençage profond, d'études de variations du nombre de copies d'ADN, de profils de méthylation des promoteurs et d'expression d'ARNm et de miARNs. En

ce qui concerne les microARNs, les auteurs ont remarqué qu'il existerait une association entre l'agressivité des tumeurs et une altération de l'expression de plusieurs miARN **(Cancer.Genome.Atlas.Network 2012)**.

Au début de ce chapitre, nous avons brièvement discuté du modèle de progression multi-étapes des cancers colorectaux établi en 1990 par Fearon et Vogelstein. De manière remarquable, la dérégulation de l'expression de plusieurs miRs intervient dans chaque étape de la transition adénome-carcinome-métastase **(Goel 2012)**. La Figure 19 illustre les différentes étapes de cette transition. Les principales voies concernées et les miRs qui les régulent y sont également représentés. En bref, il s'agit des principales voies de signalisation régissant la prolifération cellulaire, l'apoptose, le contrôle du cycle cellulaire, la différenciation, la migration et l'invasion, … **(Schetter 2011; Goel 2012)**.

MicroARNs et initiation des cancers colorectaux. Comme exemple nous citerons les miR-135a et 135b. En effet, il a été montré que leur expression serait augmentée dans les adénomes et des adénocarcinomes du côlon. Ces miRs ciblent particulièrement le gène APC de la voie de signalisation WNT/β-catenine (gène principal dans la Polypose Adénomateuse Familiale, PAF) (Figure 19). APC est fréquemment muté ou perdu dans les adénomes du côlon, et donc la surexpression des miRs 135a et 135b constituerait un événement précoce dans la transformation cellulaire de l'épithélium colique **(Nagel 2008)**.

MicroARNs et p53. Le facteur de transcription p53 est le gène suppresseur de tumeurs le plus connu et le plus étudié. Très succinctement, il est activé suite à des dommages à l'ADN et agit afin d'arrêter la division cellulaire le temps de réparer ces dommages. P53 est dérégulé dans environ 51-74% des cancers colorectaux **(Faber 2009)**. Il fait également partie d'une boucle de rétroaction avec les membres de la famille miR-34. Chang *et al.* ont montré dans un modèle cellulaire que p53 serait capable d'induire l'expression des miR-34a/b/c, et qu'en partie, les effets de p53 seraient médiés par ces miRs (qui, comme p53, seraient des suppresseurs de tumeurs). Une perte de p53 conduirait alors à une sous-expression des miR-34 et une surexpression de leurs cibles (qui incluent, entre autres, CDK4/6, la cycline E2, E2F5, …) **(Chang 2007)**.

MicroARNs et agressivité des cancers colorectaux. Les protéines ZEB1 et ZEB2 *(zinc finger E-box binding homeobox 1/2)* sont impliquées dans l'initiation de la transition épithélio-mésenchymateuse (EMT), étape clé de la formation des métastases, et leur expression est régulée par les membres de la famille miR-200 **(Mongroo 2010)**. De manière intéressante, ZEB1 et ZEB2 sont eux-mêmes capable de réprimer l'expression des miR-200. Une sous-expression des miR-200 conduit donc à la surexpression des facteurs ZEB1 et ZEB2 et corrélerait donc avec une plus grande agressivité des cellules tumorales **(Faber 2009; Schetter 2011)**.

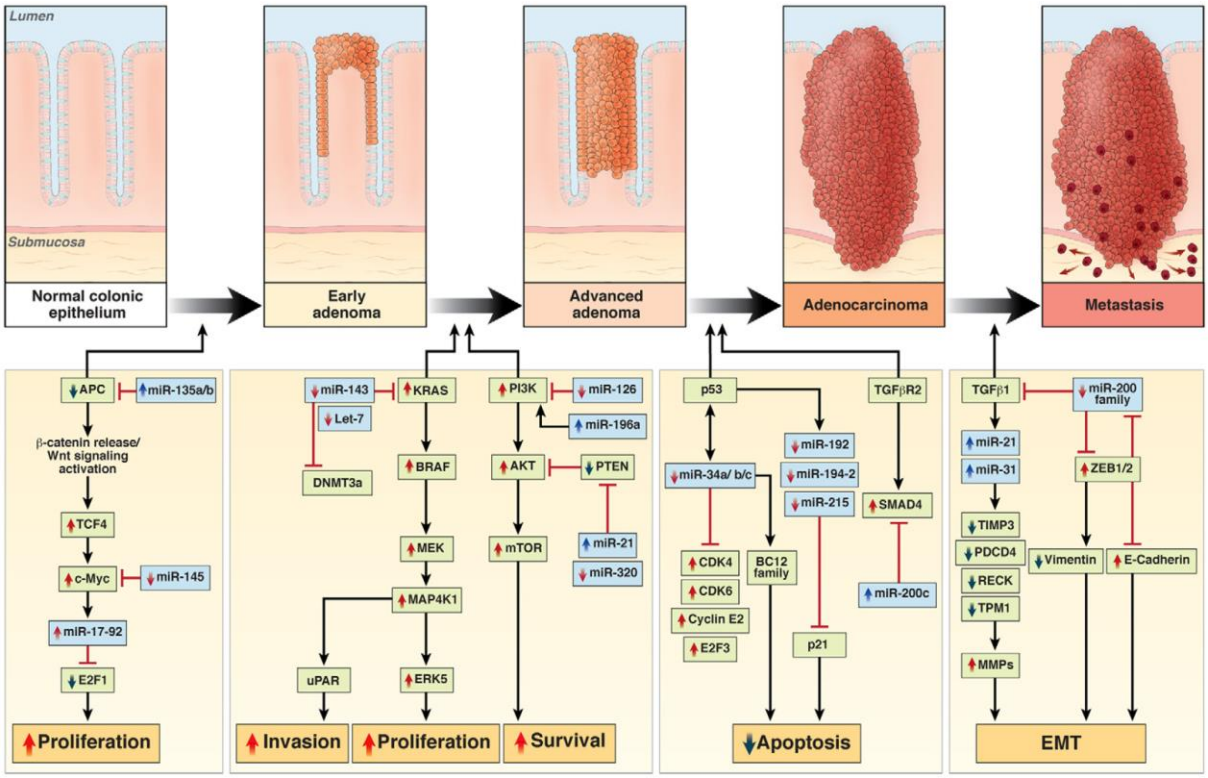


Figure 19. La dérégulation des microARNs dans la progression tumorale multi-étapes des cancers colorectaux. Sont représentés les microARNs dérégulés (cases bleues), leurs principales cibles (cases vertes) et les principales voies concernées (cases jaunes). Les flèches indiquent une surexpression ou une sous-expression des microARNs ou de leurs cibles. Extrait de Goel *et al.* **(Goel 2012)**.

8. MicroARNs et cancers colorectaux MSI

Un impact direct de la dérégulation des miARNs sur les cancers a été établi durant ces dernières années ; la notion d'« OncomiRs » (miARN oncogènes ou suppresseurs de tumeurs) vient témoigner du rôle crucial qu'ils peuvent jouer dans l'initiation et la progression tumorale **(Esquela-Kerscher 2006)**. Ce chapitre se consacrera à l'étude de la relation directe et réciproque qui existe entre les miARNs et les cancers colorectaux de type MSI. Ainsi les miARNs constituent : (1) de par leur dérégulation, des bio-marqueurs pouvant distinguer les cancers MSI des cancers MSS et de la muqueuse colique normale, et (2) de par leur capacité à cibler directement certaines protéines du système MMR, une nouvelle voie épigénétique de tumorigenèse MSI. Inversement, l'expression des miARNs peut être affectée par l'instabilité microsatellitaire qui toucherait certaines des protéines impliquées dans leur *processing* et impacterait donc leur biogenèse. Toutes ces voies contribuant – ainsi que les autres voies canoniques – à la carcinogenèse colique de type MSI et conférant à ces cancers leurs caractéristiques cliniques, pathologiques et moléculaires.

8.1 Dérégulation des microARNs dans les cancers colorectaux MSI

Plusieurs études ont montré la capacité de certains miARNs dérégulés à être des marqueurs des cancers MSI ou des cancers MSS, voire au sein même des cancers MSI, à séparer les cancers sporadiques des cancers héréditaires survenant dans le cadre du syndrome de Lynch **(Vilar 2011; Hrasovec 2012; Yamamoto 2012)**. Je reprendrai donc par la suite les principales études réalisées dans ce domaine et discuterai de l'importance et des difficultés qui accompagnent la mise en place d'une signature miARN dans les cancers colorectaux de type MSI. Une vue d'ensemble de toutes ces publications ainsi que les principales conclusions de chacune d'elles est présentée dans le Tableau 1.

Lanza *et al.* furent les premiers à montrer l'existence d'un profil associant les données d'expression de ARNm et de miARN capable de distinguer les cancers MSI des cancers MSS **(Lanza 2007)**. Leur travail a permis d'identifier un ensemble de 27 gènes, incluant 8 gènes de miARN, différentiellement exprimés entre ces deux types de CCRs. Parmi les 8 gènes de

miARN, on retrouve, sous-exprimés dans MSI versus MSS, 7 membres du cluster miR-17-92 (*miR-17-5p*, *miR-20*, *miR-25*, *miR-92-1*, *miR-92-2*, *miR-93-1* et *miR-106a*). De manière intéressante, cette famille de miARN est particulièrement connue et étudiée pour son rôle oncogénique qu'elle peut avoir dans de nombreux cancers **(Olive 2010)**. La sous-expression de ce cluster dans les CCR MSI pourrait en partie expliquer leur moindre agressivité et donc un meilleur pronostic comparé aux CCR MSS **(Lanza 2007)**.

Schepeler *et al.* **(Schepeler 2008)**, sans pour autant axer leur investigation sur les cancers MSI et s'intéressant particulièrement aux CCRs de stade II (CCRs MSI et CCRs MSS avec ou sans invasion ganglionnaire métastatique), ont également identifié des signatures miARN caractéristiques des cancers MSI. Leurs travaux montrent que les tumeurs MSI sont clairement distinctes des tumeurs MSS, et possèdent un profil miARN les rapprochant plus des muqueuses normales. Elles restent cependant assez distinctes du tissu normal à cause d'une expression aberrante de plusieurs miRs, incluant le miR-145. En dehors d'un contexte MSI, ce travail a surtout permis d'identifier plusieurs miARN associés aux récidives des cancers MSS de stade II, laissant sous-entendre une application pronostique voire thérapeutique de ces miRs.

Sarver *et al.* **(Sarver 2009)** se sont consacrés à l'étude des cancers MSI sporadiques hyperméthylés sur le promoteur de *MLH1,* ainsi qu'aux CCRs MMR proficients (stade II à IV). Ils identifièrent ainsi 6 miARNs (*miR-552*, *miR-592*, *miR-181-c*, *miR-196-b*, *miR-31*, *miR-625*) dont la dérégulation serait capable de distinguer les CCRs MSI sporadiques des CCRs MSS. De manière très intéressante, ils constatèrent que les tumeurs MSI peuvent facilement se regrouper (d'un point de vue miARN) avec certaines tumeurs MSS, à condition de partager un même phénotype CIMP sous-jacent. Ceci laisse suggérer qu'une composante épigénétique (méthylation de l'ADN et phénotype CIMP) serait à l'origine de la dérégulation des miARN.

Enfin, Balaguer *et al.* **(Balaguer 2011)** ont pu établir pour la première fois une signature miARN permettant de distinguer les cancers du côlon MSI héréditaires des cancers MSI sporadiques ; un phénotype CIMP pouvant expliquer cette différence. Leur travail a permis en outre, de montrer que les tumeurs de patients suspectés d'être atteints du syndrome de

Lynch, mais chez lesquels aucune mutation germinale n'a été retrouvée (*Lynch-like*), présentent un profil d'expression miARN similaire à celui des patients Lynch avérés. Les auteurs ont alors émis l'hypothèse d'une base moléculaire commune existant entre ces deux entités clinico-pathologiques. Les cas *Lynch-like* seraient donc en réalité des vrais cas Lynch qui, pour certaines limitations techniques, n'ont pas été correctement diagnostiqués. Cette hypothèse sera discutée plus en détail par la suite (§0).

Tableau 1. Bilan des différentes publications étudiant l'expression des miRs dans les CCRs de type MSI.

Références	Nombre de cas					Techniques utilisées (nombre de sondes miRs)	Principaux résultats
	CCR MSI		CCR MSS	Muqueuse normale	Total		
	Lynch	Sporadiques					
Lanza et al. (2007)	16		23	0	39	Microarrays (N=230) qRT-PCR Northern	Analyse combinée de l'expression des miARN et des ARNm. Identification de 8 miARN différentiellement exprimés entre MSS et MSI, parmi lesquels figurent certains membres du cluster miR-17-92
Schepeler et al. (2008)	49 (37 MSS et 12 MSI) patients de stade II (30 sans récidive et 19 avec récidive à distance)			10	59	miARN microarrays (N=315)	MSI clairement distincts des MSS, et possèdent un profil miARN les rapprochant plus des muqueuses normales.
Sarver et al. (2009)	0	12	68 (stage II : 28, stage III : 24, stage IV : 15)	28	108	miARN microarrays (N=735) et qRT-PCR	Identification de 6 miARNs dont la dérégulation serait capable de distinguer les CCRs MSI sporadiques des CCRs MSS.
Earle et al. (2010)	22 MSI-H et 8 MSI-L		25	55 appariées	110	qRT-PCR (N=24)	Expression des miRs dans les CCRs associée au statut MSI (syndrome de Lynch, MSI-L)
Slattery et al. (2011)	70 tumeurs sélectionnées en fonction de leur localisation (40 coliques et 30 rectales)			20 coliques et 10 rectales	100	miARN microarrays (N=866)	Profils d'expression spécifiques à la localisation des tumeurs et à leurs altérations (CIMP+, MSI+, *KRAS2*, *TP53*)
Balaguer et al. (2011)	35	33	19	20	107	miARN microarrays (N=743) et qRT-PCR	Profils d'expression pouvant distinguer entre les CCRs MSI compliquant un Lynch, les CCR MSI sporadiques et les CCR MSS sporadiques

8.2 Variations génétiques et MSI dans les gènes de biogenèse des microARNs

Comme nous venons de le voir, une différence d'expression des miARNs existe entre les différentes entités clinico-pathologiques des cancers colorectaux. Certains auteurs attribuent cette différence à une dérégulation au niveau génétique, d'autres à un niveau épigénétique. Très peu cependant sont allés explorer l'étiologie d'une telle dérégulation **(Hrasovec 2012; Yamamoto 2012)**.

L'équipe de Manel Esteller a étudié plus en détails cette problématique. Leur hypothèse est que certains gènes codant des protéines impliquées dans la biogenèse des miARNs, s'ils contiennent des séquences répétées, peuvent eux aussi être mutés suite à la perte de fonction du système MMR, et cela comme toute autre « vraie cible » MSI. Le premier gène cible qu'ils ont identifié est *TARBP2* (*TAR RNA-binding protein 2*) qui code pour TRBP, un partenaire fonctionnel essentiel qui interagit avec DICER1 et la stabilise **(Melo 2009)**. Dans les lignées cellulaires et tumeurs primaires MSI, une instabilité au niveau des séquences répétées de ce gène conduit à la perte de fonction de la protéine tronquée ; une perte de fonction qui a pour conséquence la déstabilisation de DICER1, l'inhibition du *processing* pre-miR en duplex miR/miR*, et finalement une diminution de l'expression d'un certain nombre de miRs matures. Le fait que les mutations sont la plupart du temps sous forme hétérozygote expliquerait pourquoi tous les miRs ne sont pas affectés de la même façon. Parmi les microARNs sous-exprimés, les auteurs ont identifié plusieurs miRs connus pour avoir un rôle anti-tumorigénique (miARNs suppresseurs de tumeurs tels que let-7f, miR-205, miR-26a, miR-125a et miR-125b). En parallèle, l'expression ectopique de TRBP dans les lignées cellulaires déficientes, à part sa capacité à restaurer la biogenèse des microARNs, a conduit à une baisse du potentiel tumorigénique des lignées en question (viabilité cellulaire, formation de colonies, formation de tumeurs chez la souris), ouvrant ainsi la voie vers le développement de nouvelles stratégies thérapeutiques éventuelles pour le traitement des cancers MSI **(Melo 2009)**.

Le gène *XPO-5* codant l'Exportine-5, protéine qui transporte les pre-miRs du noyau vers le cytoplasme, est également ciblé par l'instabilité microsatellitaire. En dehors du contexte

colorectal, des SNP ont été retrouvés dans les gènes *XPO-5* et *RAN* (partenaire essentiel de *XPO-5*), et seraient associés à un risque élevé de développer un cancer de l'œsophage. Ces SNP, localisés dans la partie 3'UTR, impacteraient la stabilité des ARNm sans pour autant entrainer des changements majeurs de la structure protéique **(Ye 2008)**. XPO-5 est également sous-exprimé dans d'autres types de cancers tels que les cancers du poumon de stade 1 ou les carcinomes bronchiolo-alvéolaire **(Chiosea 2007; Ryan 2010)**. Le gène de l'Exportine-5 contient dans sa partie codante une séquence microsatellite parfois mutée de manière hétérozygote dans les lignées et tumeurs primaires MSI **(Melo 2010)**. Cette mutation conduit, suite à un décalage du cadre de lecture, à la production d'une protéine tronquée, incapable de reconnaître ou de se lier aux pre-miRs (due à la perte de son extrémité C-terminale). Le complexe ternaire pre-miARN/XPO-5/RAN-GTP ne pouvant pas s'assembler, conduit à l'accumulation des précurseurs dans le noyau et à une baisse de l'expression de nombreux miRs matures. Comme pour TRBP, les auteurs ont observés que cette mutation confère aux lignées cellulaires étudiées un phénotype tumorigénique agressif, et qu'à l'inverse, le sauvetage du mutant, en restaurant l'expression de plusieurs miARNs, se traduit par une baisse du potentiel tumorigénique **(Melo 2010)**.

De manière très intéressante, l'effet tumorigénique observé suite à la perte de fonction de certaines protéines impliquées dans la biogenèse des miRs a déjà été observé dans un contexte indépendant de MSI. Des *knockdown* de Drosha, DGCR8 et de DICER1 ont tous pour conséquences une plus grande vitesse de prolifération cellulaire et une augmentation des capacités invasives des cellules tumorales **(Kumar 2007)**. Les auteurs ont observé de manière concomitante à la perte d'expression de certains miRs, une surexpression de plusieurs oncogènes, en particulier K-Ras et c-Myc. Beaucoup d'études se sont consacrées à *DICER1*, qui se trouve particulièrement sous-exprimé dans plusieurs de types de cancers (cancers du sein, du rein, de l'ovaire etc. …) **(Kumar 2007; Merritt 2008; Lambertz 2010)**. *In vitro* et *in vivo*, la délétion d'une seule copie de *DICER1* se manifeste par un plus grand potentiel tumorigénique (prolifération cellulaire, formation de tumeurs chez la souris). Une perte bi-allélique par contre entrave la viabilité cellulaire et est donc contre sélectionnée **(Kumar 2007; Lambertz 2010)**. Cette observation fait de *DICER1* un suppresseur de tumeurs bien particulier, qui fonctionne en haploinsuffisance contrairement aux autres gènes

suppresseurs de tumeurs pour lesquels seule l'inactivation des deux allèles (la deuxième inactivation se faisant généralement par perte d'hétérozygotie) fournit à la cellule tumorale un avantage sélectif (en dehors des mutations dominantes négatives). Nous remarquerons que *TARBP2* et *XPO-5* sont aussi mutés de manière hétérozygote **(Melo 2009; Melo 2010)**, et qu'ils contribueraient donc de la même manière que DICER1 à la formation et au développement des tumeurs MSI (*hoploinsufficent tumor supressors*) **(Grosshans 2010)**. En ce qui concerne l'expression de DICER1 dans les cancers colorectaux MSI, Melo *et al.*, n'ayant pas trouvé des mutations ou des pertes génomiques pouvant expliquer la baisse de son expression, concluent que cette baisse serait principalement liée à la perte de fonction de TARBP2 **(Melo 2009)**. Ils sont rejoints dans leur observation par Faggad *et al.*, qui montrent que dans les cancers du côlon, la sous-expression de DICER1 est associée à un stade avancé de la maladie et corrèle avec une survie réduite des patients **(Faggad 2012)**. Inversement, d'autres auteurs rapportent que c'est la surexpression de DICER1 qui serait associée à un mauvais pronostic chez les patients atteints de cancers colorectaux **(Faber 2011; Papachristou 2011; Stratmann 2011)**. Ces résultats contradictoires soulignent certes le caractère pléiotrope de DICER1, mais montrent cependant qu'on est encore loin d'avoir une vision complète de son rôle dans un contexte à la fois normal et physiopathologique **(Faber 2011).**

La dérégulation de l'expression des miRs est caractéristique des cancers colorectaux, y compris dans les cancers MSI. Comme nous venons de le voir, cette dérégulation pourrait être expliquée entre autre par une perte de fonction de certaines protéines impliquées dans la biogenèse des miRs, qui serait associée à l'instabilité microsatellitaire. À partir de ces observations, un nouveau phénotype caractérisant les CCRs MSI a été proposé : MMMP pour *« Mutated MiRNA Machinery Phenotype »* **(Davis-Dusenbery 2010)**. Ce que cela implique d'un point de vue classification des cancers MSI et d'un point de vue clinique reste toujours à déterminer.

8.3 Les protéines MMR sont la cible de certains microARNs

La relation entre miRs et instabilité microsatellitaire se complexifie par l'observation que certains miRs ciblent directement les 3'UTR des protéines MMR, et qu'inversement le système MMR interviendrait directement dans la biogenèse de certains miRs. L'exemple du miR-422a qui cible *MLH1*, est particulièrement intéressant dans ce contexte. Tous deux feraient partie d'une boucle de rétroaction ou le miR-422a régulerait le niveau d'expression de MLH1 et à son tour MLH1 (au sein du complexe MutLα) stimulerait la conversion du pri-miR-422a en pre-miR-422a **(Mao 2012)**. Dans une étude précédente, les mêmes auteurs avaient réussi à détecter une délétion de 3 nucléotides dans la région 3'UTR de MLH1 sur une rechute de patient atteint de leucémie myéloïde aiguë et qui était de phénotype MSI **(Mao 2008)**. Cette mutation, pour une raison qui à l'époque était inconnue des auteurs, déstabilisait l'ARN messager et conduisait à la diminution du taux d'expression de MLH1. L'apparition d'un phénotype hypermutateur, sans qu'une base génétique ou épigénétique bien définies ne lui soient associées, a parfois été observé (cancers MSI associées aux maladies inflammatoires chroniques de l'intestin (MICI) et lymphomes MSI chez les patients immunodéprimés) **(Svrcek 2007; Borie 2009)**. Il a été estimé que ces patients constituent environ 5% des patients MSI **(Hampel 2005)**. Il serait donc tout à fait envisageable qu'une composante microARN entre en jeu, qui serait d'une certaine manière responsable de la déficience du système MMR et permettant ainsi la mise en place et le développement des cancers MSI dans certains cas minoritaires.

Deux publications majeures sont venues appuyer ce concept. Le laboratoire de Carlo Croce a en effet identifié deux microARNs capables de cibler les différentes composantes du système MMR. Le miR-155, surexprimé dans les cancers colorectaux et particulièrement dans les CCRs MSI, serait capable de cibler directement hMSH2, hMSH6 et hMLH1, réduisant leur taux d'expression et induisant un phénotype hypermutateur. Une surexpression du miR-155 a été également notée dans plusieurs tumeurs colorectales instables mais chez lesquelles les auteurs n'avaient pas réussi à détecter ni mutations ni méthylation des gènes MMR **(Valeri 2010)**. Ces derniers ressemblent aux tumeurs que Balaguer *et al.* avaient qualifié de *Lynch-like* et qui possèdent une signature miR quasi-identique aux vrais patients Lynch (§8.1)

(Balaguer 2011). Le miR-21, surexprimé dans les cancers colorectaux, serait aussi capable de cibler hMSH2 et hMSH6. Sa surexpression dans les CCRs MSI corrèlerait en même temps avec une diminution du nombre de cellules apoptotiques retrouvées dans ces tumeurs et caractéristique d'une résistance au 5-fluorouracil (5-FU) **(Valeri 2010)**.

La modulation du système MMR par les miR-155 et 21 peut donc être considérée comme nouvelle voie épigénétique de tumorigenèse colorectale de type MSI **(Hrasovec 2012).** A ce modèle, nous ajouterons également l'instabilité au niveau des gènes impliqués dans la biogenèse des miRs matures et la dérégulation de l'expression de nombreux miRs caractéristique des cancers MSI. Toutes contribuant à la mise en place, au développement et à l'agressivité des tumeurs MSI **(Figure 20)**.

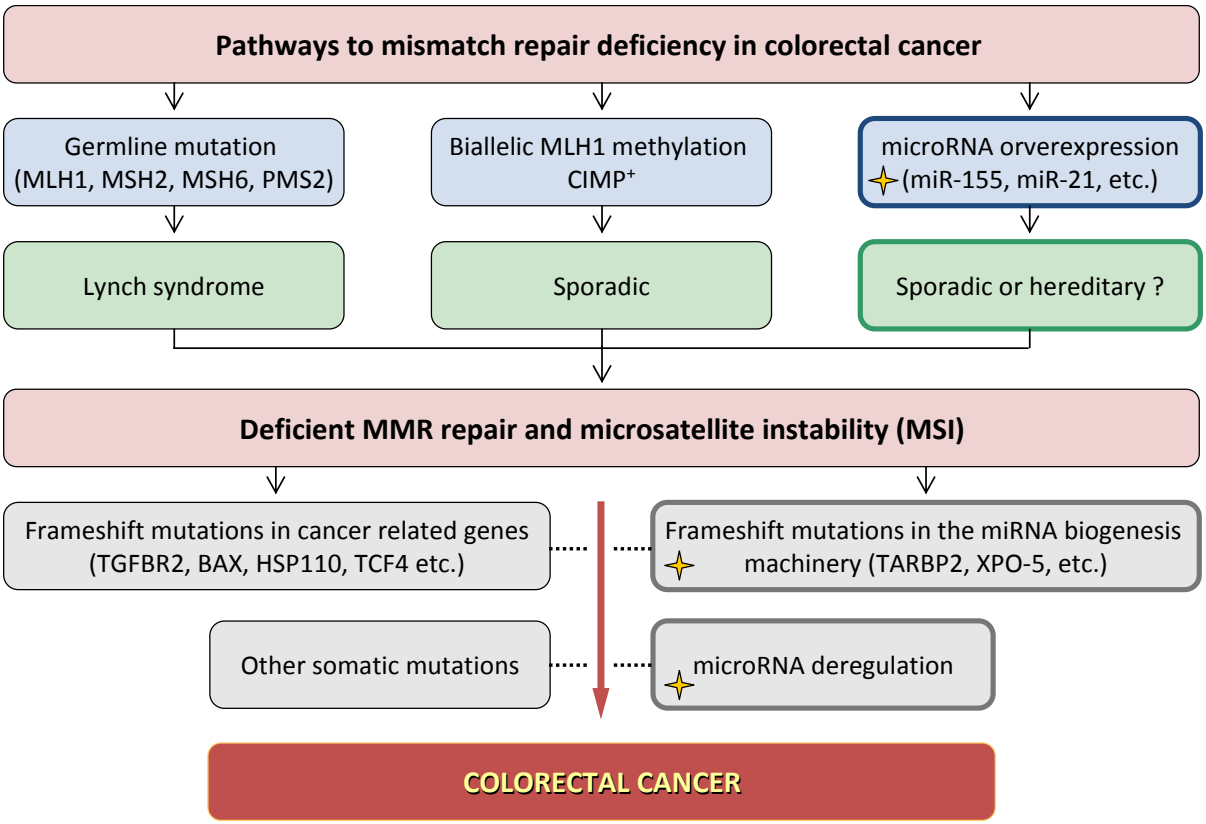


Figure 20. Les différentes voies de carcinogenèse colorectale de type MSI. Ce schéma, adapté de Sinicrope *et al* **(Sinicrope 2012)** reprend les principales voies moléculaires conduisant aux cancers colorectaux MSI. A part les deux voix classiques (prédisposition à développer des cancers MSI dans le cadre du syndrome de Lynch, et cancers sporadiques suite à la méthylation du promoteur de *MLH1*), est également représentée la voie microARNs. Il s'agit d'une part d'une surexpression de certains miRs ciblant les protéines du système MMR, et d'autre part, de mutations dues à MSI dans les gènes impliqués dans la biogenèse des miRs et qui seraient en partie responsables de la dérégulation des miRs dans les cellules tumorales. Cette dérégulation, couplée à d'autres mutations somatiques, donne aux cancers MSI leurs propriétés moléculaires et clinico-pathologiques. Les étapes supplémentaires qui ne figurent pas dans le modèle classique de la tumorigenèse MSI sont marquées par une étoile.

9. Cancer, inflammation et microARNs

9.1 La notion de défaut de champ

Le développement d'un cancer est un processus temporel qui requiert l'accumulation par les cellules de nombreuses altérations moléculaires. Ce processus graduel est connu sous le nom de progression tumorale multi-étapes **(Fearon 1990)**. Son expansion et son agressivité dépendent de la capacité des différentes populations cellulaires à s'adapter aux conditions environnementales qui les entourent, et cela par l'accumulation d'altérations génétiques ou épigénétique successives. Les altérations confèrent aux cellules un avantage sélectif et permettent au tissu néoplasique d'envahir petit à petit la muqueuse normale adjacente et éventuellement métastaser **(Fearon 1990; Yates 2012)**. La notion de défaut de champ (*« Field defect »* ou *« Field cancerization »* en anglais) sous-entend la présence et le développement au sein d'un tissu à apparence normale de cellules génétiquement ou épigénétiquement favorisées à former un cancer (Figure 21). Temporellement, il se limite aux étapes antérieures à la progression tumorale proprement dite, avant l'apparition du cancer, au moment où une cellule non transformée acquiert les premières altérations génétiques lui permettant de former une descendance clonale de cellules filles génétiquement altérées *(patch)* (en jaune et vert sur la Figure 21) **(Braakhuis 2003; Rivenbark 2012)**. La première étude mentionnant le terme de « défaut de champ » date de 1953 pour décrire les multiples cancers aérodigestifs développés par les fumeurs **(Slaughter 1953; Braakhuis 2003)**. Les patients en question peuvent développer des tumeurs indépendantes allant de la muqueuse labiale jusqu'au parenchyme pulmonaire ; autrement dit, tout endroit ayant été en contact avec la fumée de cigarette. Depuis, la notion de défaut de champ s'est généralisée pour inclure d'autres types de cancers et d'autres types de situations à risque (cancer colorectal, cancer du sein, de la prostate, etc. …). Comme nous allons le voir, l'inflammation via certaines anomalies moléculaires qui lui sont associées a été rapportée comme une situation pourvoyeuse de défaut de champs.

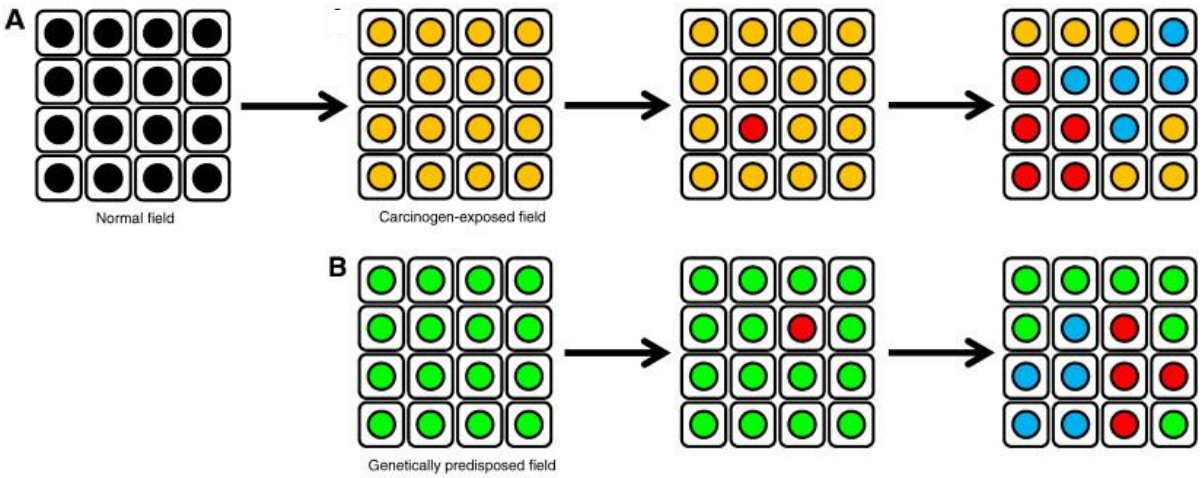


Figure 21. Mise en place et évolution d'un défaut de champ. Cette notion sous-entend le fait qu'un champ de cellules à apparence normale (*field* sur le schéma) possède initialement des altérations favorisant l'émergence de cellules cancéreuses. **A.** En l'absence de prédisposition au cancer, un tissu normal (noir) peut se trouver exposé à différentes situations potentiellement carcinogènes (fumée de cigarettes, exposition aux rayonnements UV, agents méthylants, etc…). Parmi les cellules exposées certaines peuvent accumuler des altérations génétiques et épigénétiques supplémentaires (rouge et bleu) augmentant leur risque de se transformer en cellules tumorales. **B.** Dans le cas d'une prédisposition génétique à développer des cancers, toutes les cellules de l'organisme possèdent la même altération moléculaire (vert) qui en s'ajoutant à d'autres (rouge et bleu) permettent au cancer de se développer (notion de défaut de champs constitutive). D'après Rivenbark *et al* **(Rivenbark 2012)**.

9.2 Cancers compliquant les maladies inflammatoires chroniques intestinales

9.2.1 Qu'est-ce que les MICI ?

Les maladies inflammatoires chroniques intestinales (MICI ou IBD en anglais pour *inflammatory bowel diseases*) sont des maladies idiopathiques, qui comprennent un certain nombre de troubles inflammatoires chroniques et récurrents affectant l'épithélium du tube digestif. La pathogenèse des MICI reste encore relativement peu connue, mais plusieurs études montrent qu'elle résulterait à la fois de facteurs génétiques et environnementaux principalement liés à la réaction immunitaire du système digestif à certains régimes alimentaires et à des agents infectieux **(Baumgart 2007)**.

Les principaux types de MICI sont la **maladie de Crohn** (MC, pouvant concerner la totalité du tube digestif, mais le plus souvent l'iléon terminal, le caecum, la zone péri-anale et le côlon) et la **rectocolite hémorragique** (RCH, limitée aux régions du rectum et parfois du côlon). Dans environ 10% des cas, ces deux formes sont indiscernables (ces cas ne concernent que le côlon et sont qualifiés de **colite indéterminée** ou *indeterminate colitis*) **(Podolsky 2002)**. Bien que ces deux formes apparaissent généralement chez des personnes génétiquement prédisposées et cela suite à la dérégulation de leur réponse immunitaire, la MC et la RCH sont génétiquement, moléculairement et immunologiquement distinctes, et impliquent différents mécanismes physiopathologiques (facteurs de risques génétiques, signatures protéiques, profils de cellules immunitaires et de cytokines sécrétées, etc.) **(Baumgart 2007)**. L'incidence des MICIs en Amérique du Nord et en Europe varierait entre 10 et 200 cas diagnostiqués pour 100 000 personnes (avec l'incidence de la MC augmentant au cours des quatre dernières décennies). Les pays les plus à risque seraient les plus industrialisés **(Bouma 2003)**. Le traitement des MICIs repose principalement sur l'utilisation continue de composés anti-inflammatoires tels que ceux dérivés du 5-ASA (acide 5-aminosalicylique), des corticostéroïdes, des immunosuppresseurs comme l'azathioprine, des anti-métabolites comme le méthotrexate, ou alternativement, la résection chirurgicale des principaux segments affectés et en cas de gravité extrême, la colectomie **(Bouma 2003)**.

Les MICI (très rarement fatales, mais limitant considérablement la qualité de vie des patients), peuvent se compliquer par l'apparition de cancers. Les processus physiologiques et physiopathologiques qui aboutissent à la transformation cellulaire et à l'apparition d'une tumeur ont été largement étudiés, et dans plusieurs cas, ces études semblent pointer vers l'hypothèse que l'inflammation pourrait constituer un défaut de champ duquel peuvent émerger les cellules tumorales. Dans ce qui suit, nous essayerons de reprendre les principales avancées dans ce domaine, en essayant de nous focaliser principalement sur les microARNs et le rôle qu'ils peuvent jouer dans la mise en place, le développement et la progression des tumeurs compliquant les MICIs (et plus particulièrement celles de type MSI).

9.2.2 Les MICI peuvent se compliquer par des cancers colorectaux

Parmi tous les cancers colorectaux, seulement 1% surviennent chez des patients atteints de MICI. Cependant, même à cette fréquence, les patients atteints de MICI représentent l'un des groupes les plus à risque de développer des CCRs. En parallèle, de nombreuses études soulignent qu'une augmentation du risque serait associée à l'âge, la durée, l'étendue de l'inflammation, et la gravité de la maladie (dans une revue récente, Dyson *et al.* explorent en détail tous ces aspects **(Dyson 2012)**). Dans ce qui suit, nous citerons quelques-unes des principales études et méta-analyses qui font référence dans le domaine. Ces travaux essayent de donner une estimation du risque de développer un cancer chez les patients MICI et quels en sont les principaux facteurs de risque ; sachant que jusqu'à présent, l'ampleur du risque reste encore assez controversée **(Zisman 2008)**.

Sur une population suédoise de 3117 patients atteint de RCH, Ekbom *et al.* ont détecté 91 patients ayant développé un CCR (~3%). Ils calculent un risque de développement de cancer dans cette population d'environ 5,7 avec un intervalle de confiance à 95% compris entre 4,6 et 7 (au-delà du risque de développer un CCR dans la population générale). En même temps, ils font le constat que l'âge au moment du diagnostic de la RCH (le risque d'un CCR par exemple chez les patients diagnostiqués avant l'âge de 15 ans est de 40%) et l'étendue de l'inflammation (le risque est de 30% pour les patients chez qui l'inflammation affecte la totalité du gros intestin, une situation connue sous le nom de pancolite) constituent des

facteurs de risque forts et indépendants pour le développement de cancers **(Ekbom 1990)**. Plus tard, une méta-analyse réalisée par Eaden *et al.*, combinant les résultats de 116 publications (soit 54 478 patients parmi lesquels 1698 patients ayant développé au moins un cancer), estime à 3,7% la probabilité de développer un cancer chez les patients atteints de RCH. En séparant les patients en classes, ils calculent une probabilité de 2% de développer un CCR 10 ans après le diagnostic ; 8% après 20 ans ; et 18% après 30 ans **(Eaden 2001)**. En 2004, Rutter *et al.* établissent que la sévérité de l'inflammation chez les patients RCH (échelle établie par colonoscopie et analyse histologique), constitue un facteur de risque supplémentaire non négligeable, à prendre en compte dans le suivi clinique des patients **(Rutter 2004)**.

L'estimation du risque de développement de CCR dans le cadre de la maladie de Crohn est par contre bien plus difficile et présente plusieurs limitations méthodologiques liées à l'hétérogénéité de la maladie (puisqu'elle ne se limite pas seulement au côlon) **(Zisman 2008)**. Tenant compte de ce fait, Ekbom *et al.* estiment sur une cohorte de 1655 patients (parmi lesquels 12 cancers colorectaux diagnostiqués) que quand l'inflammation est localisée uniquement au niveau de l'iléon terminal, le risque de développer un CCR ne dépasse pas celui de la population générale (risque relatif égal à 1). Quand l'inflammation est localisée au niveau de l'iléon terminal et affecte en même temps certaines parties du côlon, le risque s'élève à 3,2. Finalement, quand seul le côlon est touché par l'inflammation, le risque de développer un CCR s'élève à 5,6 **(Ekbom 1990)**. Dans une méta-analyse incluant environ 12 000 patients, Canavan *et al.* estiment que globalement, chez les patients atteints de MC et quelle que soit sa localisation, le risque de développer un CCR est d'environ 2,5 avec un intervalle de confiance à 95% compris entre 1,3 et 4,7. En tenant compte du paramètre « localisation de l'inflammation », ils retrouvent approximativement les mêmes risques qu'Ekbom *et al.* avaient trouvés **(Canavan 2006)**.

9.2.3 Développement et progression de cancers colorectaux compliquant les MICI

Comme pour les cancers colorectaux sporadiques, le développement des cancers associés aux MICI (et qu'on appellera par la suite MICI-CCRs) se fait suivant un processus multi-

étapes ; le passage d'une étape à la suivante se produisant suite à l'accumulation de mutations supplémentaires dans des gènes suppresseurs de tumeurs ou des oncogènes (Figure 22). Il est à noter cependant, qu'il n'est pas rare qu'un cancer survienne sans nécessairement passer par toutes les étapes **(Itzkowitz 2004; Ullman 2011)**. Dans la Figure 22, le terme dysplasie caractérise un état de prolifération des cellules épithéliales clairement défini comme étant anormal. La dysplasie peut-être de haut ou de bas grade selon le degré de prolifération des cellules mais reste par contre confinée à la membrane basale, et les cellules n'envahissent pas le tissu sous-jacent **(Zisman 2008)**. De manière très intéressante, et contrairement aux cancers sporadiques, les MICI-CCRs sont multifocaux, c'est-à-dire se développent généralement à partir de nombreux foyers dispersés au niveau de la muqueuse colique inflammatoire, ce qui reflèterait potentiellement un large défaut de champs **(Itzkowitz 2004; Ullman 2011)**.

Les mêmes évènements moléculaires qui régissent le développement et la progression des cancers sporadiques se retrouvent dans les MICI-CCRs (Figure 22). Willenbucher *et al.* ont montré que l'instabilité génomique, qu'elle soit à l'échelle chromosomique ou des microsatellites, faisait partie des premiers évènements détectés et cela même avant l'apparition d'une dysplasie. Leurs travaux ont également montré que, comme pour les CCRs sporadiques, 85% des MICI-CCRs étaient de type MSS et 15% de type MSI **(Willenbucher 1999)**. Par contre, l'ordre d'apparition et la fréquence de certains évènements diffèrent entre les MICI-CCRs et les CCRs sporadiques. La perte de l'APC par exemple serait moins fréquente et constituerait un évènement tardif, très rarement rencontré dans les dysplasies de bas grade **(Aust 2002; Ullman 2011).** Inversement, celle de p53 serait un évènement très précoce, qu'on retrouverait même dans la muqueuse inflammatoire de patients n'ayant pas encore développé de cancers ; une autre allusion au défaut de champ qui visiblement serait visiblement assez marqué dans les MICI-CCRs **(Hussain 2000)**.

L'hyperméthylation des promoteurs de plusieurs gènes suppresseurs de tumeurs, tel que *ER (Estrogen Receptor), MYOD (Myogenic differentiation) et CDKN2A* (codant la protéine $p16^{Ink4A}$), a été détectée dans la muqueuse néoplasique de patients atteints de RCH (chez qui une dysplasie de haut grade ou un cancer ont été trouvés). De manière très intéressante,

chez ces mêmes patients, l'hyperméthylation a aussi été retrouvée dans l'épithélium colique d'apparence normale et non dysplasique, laissant suggérer qu'une méthylation anormale des promoteurs constituerait un évènement précoce (défaut de champ) dans la tumorigenèse colique des patients atteints de RCH **(Issa 2001)**. A plus grande échelle, Koizumi *et al.*, travaillant sur 14 muqueuses coliques non cancéreuses de patients atteints de RCH, ont ressorti un profil de méthylation rapprochant ces muqueuses de certaines lignées cellulaires cancéreuses (en l'occurrence, la lignée CW-2). Leur travail leur a permis en outre de détecter, au sein des muqueuses non dysplasiques, une hyperméthylation de plusieurs gènes codant des facteurs de transcription et de remodelage de la chromatine, de nombreuses protéines de signalisation cellulaire généralement associées aux cancers (voie WNT par exemple), et des protéines de réparation de l'ADN **(Koizumi 2012)**.

9.2.4 Comment l'inflammation peut-elle favoriser l'apparition des CCR MSI ?

Les travaux de Willenbucher *et al.* ont permis de montrer que certains MICI-CCRs, et particulièrement ceux compliquant la RCH, étaient de phénotype MSI **(Willenbucher 1999)**. Inversement, le phénotype MSI serait beaucoup plus rare dans le cadre de la MC, suggérant que seuls certains types d'inflammation chronique seraient à l'origine de l'instabilité microsatellitaire **(Noffsinger 2000; Ullman 2011)**.

Les mécanismes qui pourraient expliquer l'apparition d'un tel phénotype dans un contexte inflammatoire sont encore très peu connus. Plusieurs études montrent que les MICI-CCRs de type MSI constitueraient une classe à part, se distinguant à la fois des cancers MSI sporadiques et des cancers MSI héréditaires (Syndrome de Lynch). Schulmann *et al.* ont, en effet, montré que l'instabilité au niveau de certains microsatellites codants, tel que ceux qu'on retrouve par exemple dans *TGFBR2* et *ACVR2,* était beaucoup moins fréquente dans les MICI-CCRs de type MSI que dans les CCRs MSI sporadiques. Par ailleurs, sur les 124 tissus néoplasiques inclus dans leur étude, 19 étaient de phénotype MSI, et aucun patient parmi eux n'était atteint de la MC **(Schulmann 2005)**. Les travaux de l'équipe d'accueil, sans remettre en cause la différence entre MICI-CCRs de type MSI et les CCRs MSI sporadiques, l'attribuent plutôt à la présentation clinique et moléculaire de ces cancers. Les MICI-CCRs de

type MSI (17 patients sur les 205 inclus dans l'étude, soit environ 12%) se présentent en fait de manière assez hétérogène, ayant perdu n'importe quelle protéine du système MMR, indépendamment du statut de méthylation du promoteur de MLH1 (spécifique des tumeurs sporadiques). En même temps, leurs résultats laissent suggérer que les MICI-CCRs de type MSI, même s'ils surviennent de manière plus précoce que les cancers sporadiques, ne sont pas non plus des cancers héréditaires, puisque la plupart des patients inclus dans l'étude ne faisaient pas partie de familles Lynch **(Svrcek 2007)**.

L'ensemble des travaux de Schulmann et de Svrcek, même s'il réfute l'hypothèse que les MICI-CCRs soient de même origine que les CCRs sporadiques, laisse suggérer qu'une composante épigénétique serait capable d'expliquer – du moins en partie – l'apparition d'un cancer de type MSI dans un contexte inflammatoire. Comme nous venons de le voir, l'hyperméthylation des promoteurs, que l'on retrouve fréquemment associée aux cancers compliquant les MICIs, serait une conséquence directe d'un état inflammatoire chronique. Dans ce contexte, Fleisher *et al.* ont remarqué que pour 6 (46%) des 13 lésions MSI (cancéreuses ou dysplasiques) prélevées chez des patients MICIs, la perte de la protéine MLH1 serait due à une hyperméthylation de son promoteur. Ils trouvent en même temps que 15% des lésions MSS étaient aussi méthylés sur le promoteur de *hMLH1*, laissant suggérer que les couples d'amorces utilisées couvraient des régions non spécifiques et que donc l'impact de la méthylation du promoteur de *hMLH1* serait en réalité beaucoup plus faible **(Fleisher 2000; Capel 2007; Svrcek 2007)**.

Au vu de ces résultats, se pose alors la question de savoir comment peuvent apparaître les cancers MSI associées au MICIs sans qu'il y ait prédisposition génétique ou hyperméthylation du promoteur de *hMLH1*. Les mécanismes moléculaires qui entrent en jeu restent encore à découvrir, mais les quelques études qui essayent de répondre à cette question associent la perte des protéines du système MMR au stress oxydant. De manière très intéressante, Gasche *et al.* et ensuite Chang *et al.*, en exposant des cellules MMR proficientes à des niveaux élevés mais non cytotoxiques de H_2O_2, ont remarqué qu'un stress oxydant s'accompagnait par l'inactivation du système MMR et l'apparition d'un phénotype MSI (les complexes hMutSα, hMutSβ et hMutLα subissant des dommages oxydatifs). Ils concluent en

émettant l'hypothèse que le « stress oxydant » pourrait expliquer l'apparition de cancers MSI dans un contexte inflammatoire **(Gasche 2001; Chang 2002)**.

Par ailleurs, l'hyperméthylation anormale de certains îlots CpG a été retrouvée dans de nombreuses muqueuses d'apparence saine et non dysplasiques mais qui par contre sont adjacentes à des lésions cancéreuses (muqueuses coliques, gastrique, bronchique, vésicale, …) **(Giovannucci 2005)**. Dans le cadre des cancers du côlon, Shen *et al* ont été les premiers à mettre en évidence une déficience de la O^6-methylguanine-DNA methyltransferase (MGMT) dans la muqueuse saine de patients atteints de CCRs due à une hyperméthylation de son promoteur. La MGMT est une protéine de réparation de l'ADN qui prend en charge les O^6-alkyl-guanines générées par divers mutagènes alkylants. Sa déficience est assez fréquente dans les cancers colorectaux (~50% des cas) et dans les muqueuses saines adjacentes. Au cours de la réplication, les résidus O^6-alkyl-guanines sont lus comme des adénines et s'apparient avec les bases thymines, ce qui génère des mutations (conversion des G O^6-méthylés en A). Il a donc été rapporté que pendant la tumorigenèse colorectale, la déficience de la MGMT précède et est fortement associée à des mutations de *KRAS* **(Shen 2005; Chai 2009)**. Dans une étude ultérieure, Svrcek *et al* montrent qu'un défaut de champ de la MGMT favoriserait l'émergence de clones cellulaires MMR-déficients. Ils partent du fait que généralement les mésappariements générés par les O^6-alkyl-guanines sont reconnues par le système MMR qui, incapable de les réparer, pousse la cellule à entrer en apoptose. Des clones MMR déficients auraient alors un avantage sélectif (la tolérance à la méthylation) puisque leur système MMR défectueux est incapable de détecter ces mésappariements. Un champ cellulaire déficient en MGMT constituerait alors une lésion pré-néoplasique précédant les lésions cancéreuses et évoluant potentiellement en cancer colorectal de phénotype MSI. De manière très intéressante, cette même étude a également permis de mettre en évidence un défaut de champ de la MGMT dans la muqueuse non transformée de patients MICI ayant ou non un cancer **(Svrcek 2010)**.

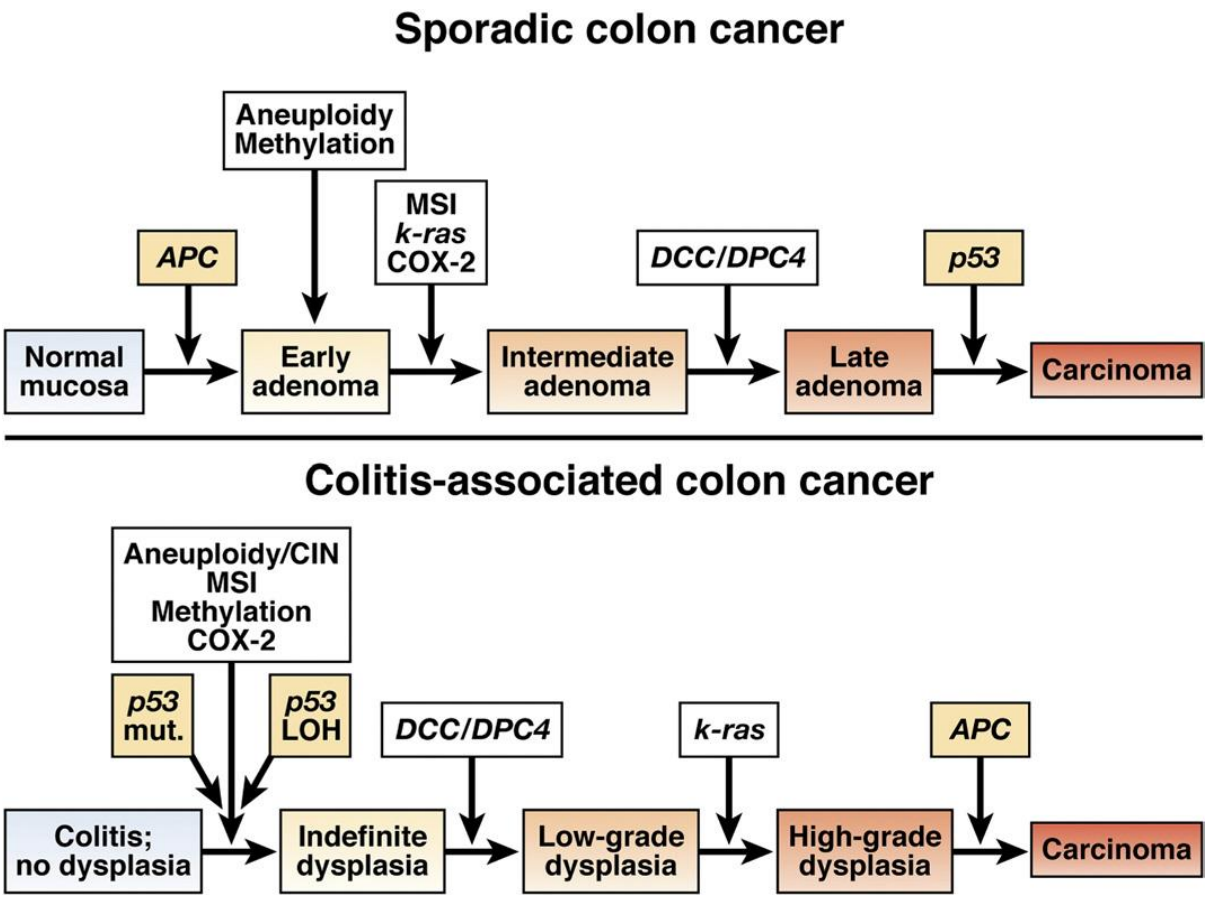


Figure 22. Les MICI se compliquent parfois par l'apparition d'un cancer. Le développement tumoral des cancers associés aux MICI (panel du bas) se fait, comme pour les cancers sporadiques (panel du haut) suivant un mécanisme multi-étapes. Plusieurs voies mutées sont communes à la fois aux cancers associés aux MICI et aux cancers sporadiques (inactivation de l'APC et de P53, activation de l'oncogène K-ras, perte d'hétérozygotie (LOH), instabilité chromosomique (CIN), instabilité des microsatellites (MSI), etc.). La fréquence et l'ordre d'apparition de ces altérations diffèrent par contre en fonction du contexte clinico-pathologique. Tiré d'Ullman *et al.* **(Ullman 2011)**.

9.3 Inflammation et cancer : le rôle des microARNs

9.3.1 Inflammation et expression des miARNs

L'inflammation chronique est un processus conduisant au remodelage moléculaire à la fois des cellules immunitaires et des cellules épithéliales, et se matérialise globalement par la sécrétion continue de différentes cytokines (par les cellules immunitaires) et la dérégulation qui s'en suit de l'expression de nombreuses protéines (dans les cellules épithéliales). L'expression des protéines est modulée par le stress inflammatoire et par les différents stimuli que la cellule reçoit, et leur dérégulation conduit éventuellement à la transformation cellulaire. Leur niveau d'expression dans un contexte inflammatoire et tumoral dépend de plusieurs paramètres (méthylation, action des cytokines, stress oxydant, …), et de manière toute aussi importante, de leur régulation post-transcriptionnelle par des microARNs. Les microARNs sont en effet dérégulés dans de très nombreuses pathologies humaines, y compris l'inflammation. Ils répondent aux différents stimuli et seraient eux aussi sujets à des inactivations épigénétiques, des amplifications ou des délétions chromosomiques, ... **(Chiba 2012; Contreras 2012; Iborra 2012)**.

Plusieurs facteurs protéiques impliqués dans la biogenèse des miRs ont été retrouvés dérégulés dans un contexte inflammatoire. Bien que cette dérégulation laisse sous-entendre en parallèle une dérégulation au niveau de l'expression des miRs, une relation directe entre inflammation et biogenèse des miRs reste encore à établir **(Contreras 2012)**. Nous citerons :

1. **Les protéines SMAD** : Ces protéines font partie des voies de signalisation TGF-β *(transforming growth factor β)* et BMP *(bone morphogenetic protein)*. TGF- β est une cytokine pléiotropique intervenant dans de nombreux processus cellulaires et possédant des activités antiprolifératives, immunosuppressives et anti-inflammatoires **(Tili 2013)**. En même temps certaines protéines SMAD font partie, avec l'ARN hélicase p68, du complexe endoribonucléotidique Drosha **(Davis 2008)**. Il est donc suggéré que l'activation continue des voies de signalisation TGF- β pourrait avoir des conséquences sur la biogenèse des miRs **(Contreras 2012)**.

2. **<u>Les enzymes ADAR</u>** : Ce sont des enzymes capable de modifier la séquence des miRs matures ou d'affecter leur biogenèse (cf. §5.3.3). Elles seraient surexprimées dans les tissus inflammatoires et participerait à la réponse inflammatoire et immunitaire **(Yang 2003)**. Elles joueraient également un rôle protecteur contre les effets délétères (induction de l'apoptose) de l'activation des voies de signalisation induites par les interférons (glycoprotéines de la famille des cytokines associées à de nombreux processus pathologiques tels que l'inflammation chronique, les maladies auto-immunes et les cancers) **(Hartner 2009)**.
3. **<u>La RNase III Dicer</u>** : son expression serait directement inhibée par les radicaux libres oxygénés et par l'oncogène *RAS*. Par contre, l'interféron-γ a été montré comme pouvant induire l'expression de Dicer **(Wiesen 2009)**.

La persistance des épisodes inflammatoires conduit à la transformation cellulaire et à l'apparition d'un cancer. L'exemple de let-7 illustre parfaitement le double rôle que jouent les microARNs dans l'inflammation et la transformation cellulaire. Le facteur de transcription NF-κB est l'un des principaux médiateurs de la réponse immunitaire. Sa surexpression est synonyme de prolifération cellulaire et de diminution du taux apoptotique. Dans une étude récente réalisée par Iliopoulos *et al.* sur des cellules mammaires, les auteurs ont montré que l'activation de NF-κB aurait pour conséquence une surexpression de la protéine Lin28 (un régulateur négatif de l'expression des miRs de la famille let-7, cf. §5.4.3). La diminution de l'expression de let-7 – qui par ailleurs ciblerait les messagers de l'interleukine 6 et de plusieurs oncogènes tels que *RAS* – conduirait à la levée de leur inhibition. IL6 et RAS activent eux-mêmes NF-κB, potentialisant ses effets, et IL6 active l'activateur de transcription STAT3 *(signal transducer and activator of transcription 3)*, ce qui a pour conséquence une prolifération cellulaire accrue et le développement de cancers. Let-7, NF-κB, IL6, RAS et Lin28 feraient donc partie d'une boucle de rétrocontrôle positif aboutissant à l'augmentation du potentiel transformant des cellules et à l'apparition d'un cancer **(Iliopoulos 2009)**.

9.3.2 Plusieurs microARNs se trouvent dérégulés dans les MICI

Comme il a été proposé pour plusieurs pathologies et comme pour les CCRs MSI, les miRs peuvent constituer des biomarqueurs et leurs différences d'expression pourrait être utilisées à des fins diagnostiques. Dans ce contexte, plusieurs miRs ont été trouvés dérégulés dans les MICIs **(Dalal 2010; Coskun 2012).** Dans une première étude réalisé par Wu *et al.*, les auteurs ont ressorti une signature de 11 miRs différentiellement exprimés qui serait spécifique des patients atteints de RCH : sous-expression des miRs 192, 375 et 422b, et surexpression des miRs 16, 21, 23a, 24, 29a, 126, 195 et let-7f par rapport aux muqueuses saines **(Wu 2008)**. Plusieurs autres études ont ensuite identifié d'autres miARNs, les associant ou pas à certains aspects fonctionnels et pathologiques. Le Tableau 2 reprend l'ensemble de ces travaux et rapporte les principaux miRs qui ont été trouvés dérégulés dans les MICIs. Ce tableau répertorie également, dans un but diagnostique, tous les microARNs circulant ayant été associés aux MICIs (sang périphérique et sérum). De manière très intéressante, plusieurs des miRs qui figurent dans ce tableau sont déjà connus pour jouer un rôle dans le développement de nombreux cancers (membres de la famille miR-30, let-7f, miR-21, miR-155, les miRs 143 et 145 suppresseurs de tumeurs, …) **(Coskun 2012)**.

Tableau 2. Bilan des microARNs dérégulés dans les MICIs (tissu et sang des patients). En gras, les miARNs surexprimés. *RCH*, rectocolite hémorragique. *MC*, maladie de Crohn. *ISH*, hybridation *in situ*. D'après Coskun *et al.* **(Coskun 2012).**

miARNs dérégulés	Type d'échantillon	Population (*n*)	Technique utilisée	Référence
miRs-192, 375, 422b, **16, 21, 23a, 24, 29a, 126, 195,** and **let-7f**	Colon sigmoïde	RCH active (*n* = 15) *vs* contrôles normaux (*n* = 15)	Microarray and qRT-PCR	Wu *et al.* **(Wu 2008)**
miR-**21** and miR-**155**	Colon sigmoïde	RCH active (*n* = 12) *vs* contrôles normaux (*n* = 12)	Microarray and qRT-PCR	Takagi *et al.* **(Takagi 2010)**
miRs-19b, 629, **23b, 106a**, and **191**	Colon sigmoïde	MC active (*n* = 5) *vs* contrôles normaux (*n* = 13)	Microarray and qRT-PCR	Wu *et al.* **(Wu 2010)**
miRs-**16, 21, 223,** and **594**	Iléon terminal	MC active (*n* = 6) *vs* contrôles normaux (*n* = 13)	Microarray and qRT-PCR	Wu *et al.* **(Wu 2010)**
miRs-188-5p, 215, 320a, 346, **7, 31, 135b, 223, 29a, 29b, 126*, 127-3p,** and **324-3p**	Côlon	RCH active (*n* = 8) *vs* contrôles normaux (*n* = 8)	qRT-PCR	Fasseu *et al.* **(Fasseu 2010)**
miRs-188-5p, 215, 320a, 346, **196a, 29a, 29b, 126*, 127-3p,** and **324-3p**	Côlon	RCH inactive (*n* = 8) *vs* contrôles normaux (*n* = 8)	qRT-PCR	Fasseu *et al.* **(Fasseu 2010)**
miRs-**9, 126, 130a, 181c, 375, 26a, 29b, 30b, 34c-5p, 126*, 127-3p, 133b, 155, 196a, 324-3p, 21, 22, 29c, 31, 106a, 146a, 146b-3p,** and **150**	Côlon	MC active (*n* = 8) *vs* contrôles normaux (*n* = 8)	qRT-PCR	Fasseu *et al.* **(Fasseu 2010)**
miRs-**9*, 30a*, 30c, 223, 26a, 29b, 30b, 34c-5p, 126*, 127-3p, 133b, 155, 196a, 324-3p, 21, 22, 29c, 31, 106a, 146a, 146b-3p**, and **150**	Côlon	MC inactive (*n* = 8) *vs* contrôles normaux (*n* = 8)	qRT-PCR	Fasseu *et al.* **(Fasseu 2010)**
miRs-150, 196b, 199a-3p, 199b-5p, 223, and 320a	Côlon	RCH inactive (*n* = 8) *vs* MC inactive (*n* = 8)	qRT-PCR	Fasseu *et al.* **(Fasseu 2010)**
miR-7	Côlon	MC active (*n* = 8) *vs* contrôles normaux (*n* = 6)	qRT-PCR	Nguyen *et al.* **(Nguyen 2010)**
miR-**150**	Côlon	RCH active (*n* = 5) *vs* contrôles normaux (*n* = 4)	qRT-PCR	Bian *et al.* **(Bian 2011)**

miR-**196**	Côlon	MC active (*n* = 83) *vs* contrôles normaux (*n* = 67)	qRT-PCR and ISH	Brest *et al.* **(Brest 2011)**
miR-143 and miR-145	Côlon	RCH active (*n* = 8) *vs* contrôles normaux (*n* = 8)	qRT-PCR	Pekow *et al.* **(Pekow 2012)**
miRs-149*, miRplus-F1065, **199a-5p, 362-3p, 340*, 532-3p,** and **miRplus-E1271**	Sang périphérique	MC active (*n* = 14) *vs* contrôles normaux (*n* = 13)	Microarray and qRT-PCR	Wu *et al.* **(Wu 2011)**
miR-149* and miR-**340***	Sang périphérique	MC inactive (*n* = 5) *vs* contrôles normaux (*n* = 13)	Microarray and qRT-PCR	Wu *et al.* **(Wu 2011)**
miRs-505***, 28-5p, 151-5p, 103-2*, 199a-5p, 340*, 362-3p, 532-3p,** and **miRplus-E1271**	Sang périphérique	RCH active (*n* = 13) *vs* contrôles normaux (*n* = 13)	Microarray and qRT-PCR	Wu *et al.* **(Wu 2011)**
miRs-505*, **28-5p, 103-2*, 149*, 151-5p, 340*, 532-3p,** and **miRplus-E1153**	Sang périphérique	RCH active (*n* = 10) *vs* MC active (*n* = 14)	Microarray and qRT-PCR	Wu *et al.* **(Wu 2011)**
miRs-**195, 16, 93, 140, 30e, 20a, 106a, 192, 21, 484,** and **let-7b**	Sérum	MC active (*n* = 46) *vs* contrôles normaux (*n* = 32)	LDA qRT-PCR	Zahm *et al.* **(Zahm 2011)**
miRs-**16, 23a, 29a, 106a, 107, 126, 191, 199a-5p, 200c, 362-3p,** and **532-3p**	Sang périphérique	MC active (*n* = 128) *vs* contrôles normaux (*n* = 162)	qRT-PCR	Paraskevi *et al.* **(Paraskevi 2012)**
miRs-**16, 21, 28-5p, 151-5p, 155,** and **199a-5p**	Sang périphérique	RCH active (*n* = 88) *vs* contrôles normaux (*n* = 162)	qRT-PCR	Paraskevi *et al.* **(Paraskevi 2012)**
miRs-**188-5p, 422a, 378, 500, 501-5p, 769-5p,** and **874**	Sang périphérique	RCH (*n* = 20) *vs* contrôles normaux (*n* = 20)	qRT-PCR	Duttagupta *et al.* **(Duttagupta 2012)**

9.3.3 Rôles multiples des miRs 155 et 21

Nous nous intéressons particulièrement aux miRs 155 et 21. Ces miRs, comme le montre le Tableau 2, ont fréquemment été retrouvés surexprimés dans les tissus inflammatoires des patients MICIs (en gras et en rouge). Comme nous l'avons vu dans le §8.3, ils seraient capables de cibler les ARNm des protéines du système MMR (hMSH2, hMSH6, hMLH1 pour le miR-155, et hMSH2 pour le miR-21) et d'induire l'apparition d'un phénotype MSI **(Valeri 2010; Valeri 2010)**.

Le miR-155 est généralement considéré comme un microARN oncogène, son expression étant élevée dans de nombreux cancers comme par exemple le cancer du sein, du côlon, de la thyroïde, de l'estomac et plusieurs leucémies. De plus, dans les cancers des poumons et du pancréas, sa surexpression serait associée à un mauvais pronostic **(Tili 2013)**. Produit à partir du transcrit non codant *BIC* (fortement exprimé dans les lymphocytes B et T activés ainsi que dans les macrophages), il serait impliqué dans la maturation et la différenciation des lignées hématopoïétiques et jouerait un rôle clé au cours de la réponse immunitaire : des souris *knockout* pour le miR-155 sont immunodéficientes et sont caractérisées par un développement anormale des cellules T et B lymphocytaires ainsi que des cellules dendritiques **(Rodriguez 2007)**. De part ce rôle, la dérégulation de l'expression du miR-155 contribuerait au développement des maladies auto-immunes et à l'apparition de nombreux syndromes myéloprolifératifs **(Tili 2009)**. Dans un contexte inflammatoire, O'Connell *et al.* ont montré que l'activation des macrophages se traduisait par une surexpression du miR-155, qui serait médiée par l'interféron-β, le facteur de transcription NF-kB et les voies de signalisation TNF-α et TLR (*Toll-like receptors*, récepteurs membranaires impliqués dans la reconnaissance des pathogènes) **(O'Connell 2007)**. Plus récemment et toujours dans un contexte inflammatoire, Tili *et al.* ont montré qu'une surexpression du miR-155 serait associée à un fort taux de mutations. De manière intéressante, ce taux élevé ne semblait pas être exclusivement liée à la perte des protéines MMR. Les auteurs ont alors proposé l'hypothèse que le miR-155 serait capable de cibler – et donc d'inhiber l'expression – de nombreux gènes de réparation de l'ADN et de contrôle du cycle cellulaire. Ces résultats mettent en lumière l'idée qu'une surexpression du miR-155 pourrait se traduire par

l'apparition d'un cancer, et cela suite à l'accumulation de nombreuses mutations et à l'incapacité de la cellule à empêcher son entrée en mitose **(Tili 2011)**.

Le miR-21 est l'un des microARN les plus fortement exprimés dans les tissus coliques de patients atteints de rectocolite hémorragique (Tableau 2). Comme le miR-155, il a été retrouvé surexprimé dans de nombreux cancers (cancer des poumons, hepatocarcinomes, mélanomes, leucémies, …) et serait presque invariablement associé à des formes tumorales agressives et à un mauvais pronostic chez les patients. Stimulé par des signaux inflammatoires, il serait capable de cibler plusieurs gènes suppresseurs de tumeurs **(Tili 2013)**. Asangani *et al.* ont montré que le miR-21 serait capable de cibler PDCD4 *(programmed cell death protein 4*, inhibiteur de la transformation cellulaire et de la progression tumorale). Dans les cancers colorectaux, une surexpression du miR-21 serait associée à une augmentation des propriétés invasives des cellules tumorales et un fort taux de métastases **(Asangani 2008)**. Le miR-21 serait également capable de cibler PTEN *(phosphatase and tensin homolog)* un autre suppresseur de tumeur, et de moduler la voie de signalisation TGF-β favorisant ainsi la prolifération et la survie cellulaire **(Tili 2013)**.

TRAVAUX DE RECHERCHE

VUE SYNOPTIQUE DES TRAVAUX DE RECHERCHE

L'initiation des cancers colorectaux de types MSI se fait suite à la perte de fonction – par mutation ou hyperméthylation de promoteur – de l'une des protéines du système de réparation des mésappariements. Cette perte engendre une instabilité génétique qui s'observe au niveau des séquences d'ADN répétées (microsatellites) dispersée dans le génome. L'instabilité microsatellitaire est oncogénique lorsqu'elle impacte des séquences répétées présentes dans les exons ou les introns de nombreux gènes tels que *TGFβR2*, *TCF4*, *HSP110*, *BAX*, *MRE11*, etc... L'ensemble de ces « gènes cibles » et leurs mutations sont à l'origine des propriétés moléculaires et clinico-pathologiques des tumeurs MSI. Nous séparerons par la suite en deux étapes ce type de carcinogenèse : (1) l'étape d'initiation, une étape pré-tumorale qui conduit à la perte de fonction du système MMR, et (2) l'étape de développement ou de progression tumorale et qui se manifeste par des mutations somatiques au niveau des différents gènes cibles d'instabilité. Mon travail a consisté à investiguer le rôle des microARNs dans ces deux étapes. Tout d'abord en tant que mécanisme moléculaire permettant la mise en place du phénotype MSI (de par la capacité de certains miRs à réguler l'expression des gènes MMR), et ensuite en tant que « gènes cibles » potentiels de l'instabilité microsatellitaire, pouvant jouer un rôle dans la tumorigenèse colique de type MSI.

A. <u>Initiation des tumeurs MSI</u> : Surexpression des microARNs 155 et 21 dans les CCRs MSI compliquant les MICI (Svrcek* & El-Murr* *et al*., *Carcinogenesis*, 2013)

Les MICI peuvent se compliquer par l'émergence d'un cancer colorectal. Rarement, ce cancer est de type MSI. Les travaux de l'équipe d'accueil ont montré que ces cancers formaient une classe à part, n'étant que rarement hyperméthylés sur le promoteur de *MLH1* sans que les individus soient génétiquement prédisposés au cancer (Syndrome de Lynch, lié à l'inactivation de l'un des gènes du système MMR sans hyperméthylation de leurs promoteurs). Les mécanismes moléculaires pouvant expliquer l'émergence du phénotype MSI dans les cancers associés aux MICI restent donc à investiguer. Notre hypothèse est qu'une composante alternative interviendrait, basée sur la dérégulation de l'expression de certains microARNs et permettant de réguler négativement le système MMR. Dans ce

contexte, deux études récentes ont montré que certains microARNs (i.e. miR-155 et miR-21) seraient capables de cibler MSH2, MSH6 et MLH1 et donc moduler le système de réparation des mésappariements **(Valeri 2010; Valeri 2010)**. De manière intéressante, ces deux miRs sont particulièrement connus pour être surexprimés dans un contexte tumoral et/ou d'inflammation chronique. Notre étude a consisté à quantifier par qRT-PCR l'expression de ces deux miRs dans une collection unique d'échantillons provenant de patients MICI ou non MICI ayant ou non développé un CCR. Il s'agit particulièrement de couples tumeurs/muqueuses saines appariées de patients MICI-CCR (MSI=10 et MSS=10) ou non MICI-CCR (MSI=32 et MSS=30), de 18 muqueuses coliques de patients opérés pour diverticulite et de 20 muqueuses coliques de malades ayant une MICI sans cancer. Globalement, nos résultats montrent une surexpression des deux miRs dans un contexte inflammatoire (MICI sans cancer) et de manière plus importante encore dans un contexte tumoral (MICI CCR). Comparées aux muqueuses saines (provenant de malades opérés pour diverticulite), les muqueuses non transformées des patients MICI-CCR présentaient elles aussi des niveaux d'expression élevés. Cette dernière observation laisse suggérer un rôle de ces miRs dans les étapes pré-tumorales permettant de favoriser à terme l'émergence de CCR MSI. Ce phénomène est compatible avec une notion importante en cancérologie qui est celle du défaut de champ (*« Field defect »* ou *« Field cancerization »*). Particulièrement pour le miR-155, sa surexpression a tendance à être spécifiquement associée aux cancers MSI, alors que la surexpression du miR-21 serait un événement commun aux tumeurs MSS et MSI. Nos résultats suggèrent donc que la surexpression de certains miRs (particulièrement le miR-155) pourrait constituer un événement sous-tendant l'émergence d'un phénotype MSI dans les CCR compliquant les MICI, en accord avec la fonction de ces miRs (cibler les protéines du système MMR). Ils laissent également suggérer qu'à l'avenir, la détection et le suivi du défaut de champ des miRs 155 et 21 puissent devenir un outil pour prévenir la survenue des CCRs MSI associés aux MICI. De plus, la surexpression du miR-155 pourrait constituer une composante supplémentaire s'ajoutant au défaut de champ de la MGMT **(Svrcek 2010)** et permettant de comprendre comment certains cancers compliquant les MICI peuvent être de type MSI.

B. Progression et développement tumoral MSI :

Identification de gènes de miARNs directement ciblés par un processus d'instabilité microsatellitaire dans les CCR MSI (El-Murr *et al.*, *PLoS ONE*, 2013)

L'impact mutagène de l'instabilité microsatellitaire sur les gènes de miARNs (conséquence sur la biogenèse et l'expression des miRs matures) qui à l'instar des gènes codants peuvent être activement impliqués dans des processus de transformation, n'a pas encore été rapporté. Identifier des séquences répétées instables dans les gènes de miRNA pourrait donc conduire à la découverte de miRs ayant possiblement un rôle clinico-pathologique important dans l'oncogenèse MSI. Pour répondre à cette problématique, nous avons réalisé une analyse des microsatellites mono- et di-nucléotidiques contenus dans les gènes de miARNs et de leur instabilité dans les CCR MSI. Une recherche *in silic*o de répétitions de type microsatellitaire (≥ 7 unités répétées) a été réalisée sur les 940 précurseurs épingles-à-cheveux de miARNs (source : miRbase V15). De ces 940 précurseurs, un petit nombre se dégagent comme ayant des répétitions mono-nucléotidiques (24/940, 2.5%) alors que très peu ont des répétitions di-nucléotidiques (3/940, 0.3%). L'instabilité a été étudiée dans des lignées cellulaires et tumeurs primaires coliques (MSI comparativement à MSS) ; et des mutations liées à MSI ont été retrouvées à des fréquences assez variables dans 16 des 24 gènes testés (fréquences allant de 2,5% à 100% en fonction de la taille de la répétition). Utilisant la méthode statistique du maximum de vraisemblance, les gènes de miARN possédant des répétitions mono-nucléotidiques ont été séparés en deux groupes selon leur taux de mutabilité dans les CCR MSI : ceux qui sont peu ou pas mutés d'une part (18/21) et ceux très fréquemment altérés d'autre part (3/21). Ces deux groupes peuvent être définis selon leur tendance à présenter des mutations sélectionnées ou non au cours de la tumorigenèse MSI. Des 3 cibles qui se dégagent de cette étude, *hsa-mir-1303*, *hsa-mir-1273c* et *hsa-mir-567*, seul le miR-1303 est exprimé dans les tissus colorectaux. Il mute à des fréquences très élevées (>80%) et d'une manière biallélique, mais son niveau d'expression ne semble pas être affecté par l'instabilité. Dans l'ensemble, cette étude montre que les séquences répétées dans les gènes de miARN sont relativement rares et dans la majorité préservées des mutations liées à MSI. Les microARNs formeraient alors un classe de gènes naturellement préservés des séquences microsatellites (et donc des conséquences

potentiellement néfastes qu'elles peuvent avoir). De plus, les rares fois où ces séquences sont, à la fois, présentes et mutées dans les gènes de miARN, les mutations n'affectent pas la biogenèse et l'expression du miR mature. Ceci laisse alors suggérer que, contrairement aux gènes codant, l'instabilité dans les séquences répétées des gènes de miARNs ne constituerait pas une voie majeure dans la tumorigenèse colique de type MSI. En dehors d'un contexte MSI, notre étude laisse sous-entendre que la présence des séquences répétées au sein des précurseurs miARNs est « tolérée » du moment qu'elle n'affecte par la structure de l'épingle-à-cheveux (souplesse structurale de ces séquences qui peuvent accumuler des mutations). Un point que nous reprendrons plus en détail dans la partie « discussion » de ce manuscrit.

L'instabilité microsatellitaire altère la biogenèse du miR-3613-5p et conduit à l'apparition de *variants* spécifiquement associées aux CCRs de type MSI (El-Murr *et al.*, *in prep.*)

Dans la continuité de l'étude réalisée sur la version 15 de miRBase, et malgré ces résultats préliminaires, une étude plus approfondies a été réalisée sur un miARN candidat, qui semble constituer une exception de par la localisation de la séquence répétée. Il s'agit du gène *hsa-miR-3613*. Ce gène, rapporté pour la première fois en 2010 (miRBase V16) dans le tissu cervical normal et tumoral, est situé sur le chromosome 13q14.2. Il contient deux répétitions de 10 nucléotides chacune, incluses dans leur totalité dans la séquence des miRs matures produit à partir de l'épingle à cheveux : le miR-3613-5p (poly(U) en position 9-18 du miR), et le miR-3613-3p (poly(A) position 3-12 du miR). *hsa-miR-3613* est fréquemment muté dans les lignées cellulaires (86%) et les tumeurs primaires MSI (en moyenne 66%). Seule le miR-3613-5p est exprimé de manière significative dans la muqueuse colique. Contrairement aux gènes précédemment identifiés, l'étude de l'expression du miR-3613-5p par qRT-PCR dans les lignées cellulaires et les tumeurs primaires MSI et MSS montre que l'instabilité se traduit cette fois-ci par une diminution de l'expression du miR dans les lignées et tumeurs instables comparativement aux lignées et tumeurs stables sur ce gène. L'instabilité au niveau du mir-3613 génère en outre la production de miRs mutants. Des analyses à la fois par Northern Blot (constructions plasmidiques reproduisant le spectre de délétions retrouvées au niveau de la séquence microsatellite dans les CCR MSI) et par séquençage profond (analyse d'une série de CCR MSI et MSS) mettent en évidence que les mutations conduisent à une

diminution de l'expression du miR mature et à la production de plusieurs isoformes aberrantes (IsomiRs). Ces isoformes possèdent la même séquence *seed* que le miR-3613-5p sauvage (nucléotides 2 à 8) mais ont une extrémité 3' qui varie (l'endroit où est présente, entre autre, la répétition). Il s'agit d'ajouts de nucléotides qui compensent les délétions au niveau de la séquence microsatellite. La transfection d'une lignée cellulaire MSI non mutée sur ce gène par des *mimics* de la forme sauvage et des mutantes et l'analyse par puces à ADN montrent que malgré leur faible taux d'expression, les variants semblent avoir une signature transcriptomique qui pourrait se distinguer de celle de la forme sauvage (répertoire de gènes cibles différents). Ce que ces résultats impliquent d'un point de vue fonctionnel vis-à-vis de la tumorigenèse colique de type MSI reste à déterminer. Ils montrent néanmoins pour la première fois l'existence de mutations somatiques dans la séquence mature d'un microARN capables d'altérer sa biogenèse. La question, à savoir si cette mutation est un évènement isolé ou si elle concerne d'autres séquences de miRs non identifiées et des implications que cela peut avoir d'un point de vue moléculaire, clinique et thérapeutique, reste encore ouverte.

Evaluation de la fréquence d'instabilité des gènes impliqués dans la biogenèse des miRs et la régulation de leur fonction (El-Murr *et al.*, *in prep.*)

Comme tout autre gène, ceux codants pour les protéines impliquées dans la biogenèse des miRs et la régulation de leur fonction sont susceptibles de contenir dans leurs exons des séquences répétées de type microsatellitaire. Des mutations dans ces séquences, et la perte de fonction qui s'en suit, peuvent donc expliquer – du moins en partie – la signature miRs qui différencie les cancers colorectaux de type MSI des autres types de cancers colorectaux **(Vilar 2011; Hrasovec 2012; Yamamoto 2012)**. L'étude s'est focalisée sur 11 gènes contenant dans leurs exons une ou plusieurs séquences mononucléotidiques répétées d'au moins 7 nucléotides. Parmi ces gènes, dont toutes les répétions sont comprises entre 7 et 8 unités répétées, figurent *XPO-5*, codant l'Exportine-5 (export des pre-miRs du noyau vers le cytoplasme), *AGO1* et *AGO2* (composants principaux du complexe effecteur RISC), *TARBP2* (partenaire de Dicer) et *TNRC6A* (formation des P-bodies). L'analyse dans des lignées cellulaires et tumeurs primaires MSI a révélé des fréquences de mutations assez faibles pour la plupart des gènes. Le test statistique du maximum de vraisemblance réalisé ces gènes

ainsi que sur d'autres gènes ayant des répétitions de taille identique, a permis de séparer en deux classes ces gènes en fonction de leur fréquence de mutation. Tous les gènes impliqués dans la biogenèse des miARNs ont été alors classés parmi ceux dont les mutations, de faible fréquence, ne seraient pas sélectionnées dans l'oncogenèse MSI.

Quelques études se sont précédemment intéressées à *XPO-5* et *TARBP2* avec des résultats assez contradictoires (allant de 2,4% à 26,3% pour *XPO-5* et 2% à 43,3% pour *TARBP2*) **(Melo 2009; Garre 2010; Kim 2010; Melo 2010)**. Nos résultats pour ces deux gènes (9% pour *XPO-5* et 6% pour *TARBP2* dans les tumeurs primaires MSI) tranchent en faveur des faibles fréquences et laissent sous-entendre un rôle mineur de ces mutations dans la tumorigenèse colique de type MSI. Les gènes les plus fréquemment mutés dans notre cohorte de patients MSI sont *AGO2* (10,3%), *DDX17* (13,4%) et *FXR2* (25,4%). Mise à part *AGO2*, le rôle de DDX17 (une ARN hélicase partenaire de Drosha, également connue sous le nom de p72) et FXR2 (*fragile X mental retardation, autosomal homolog 2*) dans la biogenèse des miRs reste encore assez peu étudié. Les conséquences de leur mutation dans les cancers de type MSI restent donc à déterminer.

Article 1

Overexpression of microRNAs-155 and 21 targeting mismatch repair proteins in inflammatory bowel diseases

Svrcek M*, El-Murr N*, Wanherdrick K, Dumont S, Beaugerie L, Cosnes J, Colombel J-F, Tiret E, Fléjou J-F, Lesuffleur T, Duval A. (2013) *Carcinogenesis* 34 (4): 828-834.

Carcinogenesis vol.34 no.4 pp.828–834, 2013
doi:10.1093/carcin/bgs408
Advance Access publication January 3, 2013

Overexpression of microRNAs-155 and 21 targeting mismatch repair proteins in inflammatory bowel diseases

Magali Svrcek[1,2,3,*,†], Nizar El-Murr[1,2,†], Kristell Wanherdrick[1,2], Sylvie Dumont[2], Laurent Beaugerie[2,4], Jacques Cosnes[2,4], Jean-Frédéric Colombel[5,6], Emmanuel Tiret[2,7], Jean-François Fléjou[1,2,3,8], Thécla Lesuffleur[1,2] and Alex Duval[1,2]

[1]INSERM, UMR S 938 – Centre de Recherche Saint-Antoine, Equipe ''Instabilité des Microsatellites et Cancers'' – Equipe Labellisée Ligue Contre le Cancer, Paris F-75012, France, [2]UPMC – Université Pierre et Marie Curie – Paris 6, Paris F-75005, France, [3]Service d'Anatomie et Cytologie Pathologiques and [4]Service de Gastroentérologie et Nutrition, AP-HP, Hôpitaux Universitaires Est Parisien, Hôpital Saint-Antoine, Paris F-75012, France, [5]Service d'Hépato-gastroentérologie, Centre Hospitalier Régional Universitaire de Lille, Lille F-59037, France, [6]INSERM UMR 995, Equipe ''Maladies inflammatoires digestives: physiopathologie et développement de cibles thérapeutiques'', Université Lille Nord de France, Lille F-59044, France, [7]Service de Chirurgie Viscérale, AP-HP, Hôpitaux Universitaires Est Parisien, Hôpital Saint-Antoine, Paris F-75012, France, and [8]Tumorothèque Cancer Est, AP-HP, Hôpitaux Universitaires Est Parisien, Hôpital Saint-Antoine, Paris F-75012, France.

*To whom correspondence should be addressed. Tel: +33 1 49 28 21 79; Fax: +33 1 49 28 28 78;
Email: magali.svrcek@sat.aphp.fr
Correspondence may also be addressed to Nizar El-Murr.
Tel: +33 1 49 28 46 09; Fax: +33 1 49 28 66 81;
Email: nizar.el-murr@inserm.fr
Correspondence may also be addressed to Alex Duval. Tel: +33 1 49 28 66 80; Fax: +33 1 49 28 66 81;
Email: alex.duval@inserm.fr

Microsatellite instability (MSI) due to mismatch repair (MMR) deficiency is reported in 5–10% of colorectal cancers (CRCs) complicating inflammatory bowel diseases (IBD). The molecular mechanisms underlying MMR deficiency may be different in IBD CRCs, and in sporadic and hereditary MSI tumors. Here, we hypothesize that overexpression of miR-155 and miR-21, two inflammation-related microRNAs that target core MMR proteins, may constitute a pre-neoplastic event for the development of MSI IBD CRCs. We studied miR-155 and miR-21 expression using real-time quantitative PCR in MSI (*n* = 10) and microsatellite stable (*n* = 10) IBD CRCs, and in MSI (*n* = 32) and microsatellite stable (*n* = 30) non-IBD CRCs. We also screened colonic samples from IBD patients without cancer (*n* = 18) and used healthy colonic mucosa as controls (*n* = 20). MiR-155 and miR-21 appeared significantly overexpressed not only in the colonic mucosa of IBD subjects without CRC but also in neoplastic tissues of IBD patients compared with non-IBD controls (*P* < 0.001). Importantly, in patients with IBD CRCs, miR-155 and miR-21 overexpression extended to the distant non-neoplastic mucosa (*P* < 0.001). Ratios of expressions in tumors versus matched distant mucosa revealed a nearly significant association between miR-155 overexpression and MSI in IBDs (*P* = 0.057). These results show a strong deregulation of both MMR-targeting microRNAs in IBD subjects with or without cancer. MiR-155 overexpression being particularly associated to MSI IBD CRCs and extending to distant non-neoplastic mucosa, strongly suggests that a pre-neoplastic miR-155 field defect may promote MSI-driven transformation of the colonic mucosa. The detection and monitoring of miR-155 field defect may, therefore, have implications for the prevention and treatment of MSI IBD CRCs.

Abbreviations: CD, Crohn's disease; CRC, colorectal cancer; FC, fold changes; FFPE, formalin-fixed, paraffin-embedded; HNPCC, hereditary non-polyposis CRC; IBD, inflammatory bowel diseases; miR, microRNA; MMR, mismatch repair; MSI, microsatellite instability; MSS, microsatellite stable; UC, ulcerative colitis.

† These authors contributed equally to this work.

Introduction

An increased risk of developing colorectal cancers (CRCs) is described in patients with inflammatory bowel diseases (IBD), that is, ulcerative colitis (UC) and Crohn's disease (CD) (1–3). In both UC and CD, the risk of developing cancer has been reported to be strongly associated with duration, extent and severity of colonic mucosal inflammation (4,5). To date, the exact mechanisms underlying the emergence of CRCs in IBD patients are poorly characterized. We and researchers have recently reported the existence of microsatellite instability (MSI) in a significant fraction of IBD CRCs (5–10%) (6,7). At the time, we identified 17 subjects with MSI neoplasia out of a multicentric cohort of 205 patients, which constitutes, to our knowledge, the largest series of MSI IBD-associated CRCs and dysplastic lesions (7).

MSI, caused by a mismatch repair (MMR) deficiency, is an oncogenic process that leads to the accumulation of frequent somatic mutational events in cancer-related genes containing coding DNA repeats (for review, see Duval *et al.* (8)). MSI tumors usually occur in patients suffering from hereditary non-polyposis CRC (HNPCC)—recently renamed Lynch syndrome—with *MLH1, MSH2, MSH6* or *PMS2* bearing germline MMR defect (9,10). MSI is also, and mostly, observed in approximately 10–15% of sporadic colorectal, gastric and endometrial cancers (11–13), following epigenetic, bi-allelic silencing of *MLH1* expression by *de novo* methylation of its promoter (14). Our recent findings argued for the existence of alternative and yet unknown mechanisms underlying MMR deficiency in MSI tumors developed by IBD patients (6,7).

MicroRNAs (miRs) are small non-coding RNAs of 18–25 nucleotides in length that play a crucial role in the post-transcriptional regulation of genes and cellular homeostasis (15). Modifications of their expression and function have been found in various human tumors, including CRCs (16–18). In a recent study, the oncogenic miR-155 was proposed to play a key role in the modulation of the MMR system by downregulating the core MMR proteins (i.e. MSH2, MSH6 and MLH1) (19). In the same study, the authors reported that a number of MSI CRCs with unknown cause of MMR inactivation displayed miR-155 overexpression, providing support for miR-155 modulation of MMR as an alternative mechanism of cancer pathogenesis (19). Some additional studies reported the overexpression miR-155 in primary CRCs (20–22), and other studies have shown that MSI CRCs display higher levels of miR-155 than microsatellite stable (MSS) CRCs (MSS display microsatellite stability and chromosomal instability) (23–25). Besides miR-155, miR-21 is another miR reported as being clearly overexpressed in CRCs (26). It is thought to be a key player in colorectal carcinogenesis, targeting tumor suppressor genes such as *PDCD4, BCL2* and *TGFBR2* (27–29). It also targets *MSH2* and could contribute to induce resistance to 5-fluorouracil in CRCs (30).

Many arguments point to miR-155 as a link between both inflammation and cancer, notably by regulating many aspects of the innate and acquired immunity through antibody production, cytokine release and downregulation of different transcription factors (31). The increased expression of both miR-155 and miR-21 has been reported in many inflammatory diseases including UC and CD (32–34). Till now, no study has reported a role of miR-21 and miR-155 in the emergence of IBD CRCs, and more specifically the ones with MSI.

Here, we hypothesize that miR-155 and/or miR-21 overexpression might constitute an early event favoring the emergence of MMR-deficient clones in the colonic mucosa of IBD patients and thereby, the emergence of colon tumors displaying MSI. We also explored the possibility of miR-155 and/or miR-21 overexpression not being confined to neoplastic tissues but extending to non-neoplastic mucosa, creating an oncogenic "field defect" defined by the presence, within normal tissue, of genetic or epigenetic events that precede and predispose to the development of cancers.

Downloaded from http://carcin.oxfordjournals.org/ at INSERM on February 9, 2014

© The Author 2013. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Materials and methods

Patients and specimens

All CRC samples were collected from the Department of Pathology of the Saint-Antoine Hospital (AP-HP, Paris, France). Ethics approval was obtained from the human research ethics committee of the hospital. The following clinico-pathological data were collected for each patient: sex, age at surgery, anatomical distribution of the tumors [right colon (proximal to and including the splenic flexure), left colon (sigmoid and descending colon) and rectum] and tumor stage (Table I).

We analyzed 20 IBD-associated CRCs (10 MSI CRCs and 10 MSS CRCs) who were previously used and characterized for their MMR status (7). All were matched with non-neoplastic distant mucosa (non-neoplastic distant mucosa was located at least 10 cm away from the cancer). A total of 32 non-IBD-associated sporadic ($n = 19$) and inherited ($n = 13$) MSI and 30 non-IBD-associated MSS CRCs were also included in the study. Most were matched with distant normal mucosa ($n = 28$ for MSI and $n = 22$ for MSS CRCs, respectively). They were tested for MSI using a single fluorescent multiplex system comprising five quasi-monomorphic mononucleotide repeats (BAT-25, BAT-26, NR-21, NR-24 and NR-27) (35). MMR protein expression (MLH1, MSH2, MSH6 and PMS2) in CRCs displaying MSI was determined by immunohistochemistry (36). MSI CRCs were considered as likely sporadic or HNPCC according to the presence or absence of *MLH1* promoter methylation, respectively. Tumors displaying loss of MSH2, MSH6 or PMS2 expression were considered as likely HNPCC cases.

The control groups included specimens of non-inflammatory colon mucosa from 20 patients who underwent surgery for colonic diverticulitis (non-IBD controls) and colonic samples ($n = 24$) from 18 patients with a history of IBD but absence of neoplastic colonic lesions (IBD controls without CRCs) (37).

RNA preparation

Total RNA from formalin-fixed, paraffin-embedded (FFPE) tissues (IBDs, IBD-associated tumors and diverticulitis) was extracted using the RecoverAll™ Total Nucleic Acid Isolation Kit (Ambion) according to the manufacturer's protocol. RNA from frozen tissues (non-IBD-associated CRCs and matched normal mucosae) was prepared with TRIzol® reagent (Invitrogen). Briefly, 50–100 mg of frozen tissues was soaked in TRIzol®, then grinded in a TissueLyser with stainless steel beads (Qiagen). Lipids and proteins are eliminated with a 24:1 mixture of chloroform/isoamyl alcohol, and RNA was recovered by isopropyl alcohol precipitation. Nucleic acids quantification was performed using a ND-1000 NanoDrop spectrophotometer (NanoDrop Technologies).

Selected areas of neoplastic and non-neoplastic FFPE tissues were macrodissected under guidance from hematoxylin and eosin-stained sections. Frozen samples were included in the study and RNA extracted only if it was confirmed by hematoxylin and eosin staining performed on a frozen section whether or not they corresponded mostly to tumor or healthy tissues.

Reverse-transcription and real-time quantitative PCR

Quantitation of miR-155 and miR-21 expression was carried out using the TaqMan® MicroRNA Assay technology (Life Technologies). Complementary DNAs were generated from 10 ng of total RNA using miR-specific stem-loop reverse transcription primers (Life Technologies) together with the TaqMan® MicroRNA Reverse Transcription Kit according to the manufacturer's instructions. Real-time PCR amplifications were performed in triplicate on an ABI 7900 Sequence Detection System using Universal PCR Master Mix without AmpErase® UNG (Life Technologies) and TaqMan MicroRNA made-to-order probes. The thermal cycling conditions comprised an initial denaturation step at 95°C for 10 min followed by 45 cycles at 95°C for 15 s and 60°C for 1 min.

Ct, the cycle at which the amount of amplified target crosses a fixed threshold, was determined. miR expression was first normalized to that of a control transcript, the RNU48 (an abundant small nucleolar RNA), with $\Delta Ct = Ct_{miR} - Ct_{RNU48}$. An arbitrarily designated calibrator sample was used, according to which relative miR expressions were expressed as $2^{-\Delta\Delta Ct}$, where $-\Delta\Delta Ct = -(\Delta Ct_{sample} - \Delta Ct_{calibrator})$.

Many studies have pointed out consistent expression levels between FFPE and frozen tissues (38–41). Samples from both FFPE and frozen tissues used to study miR-155 and miR-21 expression were also profiled for miR-423, whose expression was found to be unchanged between healthy mucosas and CRC samples (NCBI GEO series accession no.GSE10259) and accordingly is used here as an additional control. This additional control is needed to prove that any difference in expression levels does not depend on the RNA extraction technique or the type of sample being used.

Statistical analysis

Kruskal–Wallis one-way analysis of variance followed by Dunn's multiple comparison test is performed to analyze the results presented as mean ± SD. Fisher's exact test is used to compare frequencies. A 95% confidence interval was adopted and differences were considered significant when $P < 0.05$.

Table I. Clinico-pathological features of patients

Variable	Number of patients	Sex F/M	Tumor location right colon/left colon/rectum/other[a]	Tumor stage I and II	Tumor stage III and IV	Mean age at the time of surgery (range)
Inherited MSI CRCs						
Neo	13	4/9	6/7/0/0	8	5	58.6 years (24–81)
D[b]	11					
Sporadic CRC						
MSI						
Neo	19	10/9	19/0/0/0	12	7	74.5 years (46–90)
D	17					
MSS						
Neo	30	9/21	10/20/0	13	17	68.6 years (47–92)
D	22					
Healthy controls (diverticulitis)	20	14/6	—	—	—	66.1 years (33–87)
IBD-associated CRC						
MSI[c]						
Neo	10	4/6	4/1/4/1	7	3	53.5 years (19–66)
D	10		—			
MSS[d]						
Neo	10	6/4	2/0/6/2	5	5	49.6 years (34–65)
D	10		—			
IBD controls (non-neoplastic IBD)	18 (24 samples)	9/9	—	—	—	37 years (19–78)

[a]Anastomotic relapse or perineal abcess.
[b]D: non-neoplastic distant mucosa.
[c]5 with UC and 5 with CD.
[d]6 with UC and 4 with CD.

Downloaded from http://carcin.oxfordjournals.org/ at INSERM on February 9, 2014

Results

MiR-155 and miR-21 are overexpressed in the neoplastic and non-neoplastic distant mucosa of IBD patients with CRC and in the colonic mucosa of IBD subjects without CRC

We studied the expression of miR-155 and miR-21 in IBD patients with MSI or MSS CRCs (IBD MSI or IBD MSS CRCs) and compared them to IBD subjects without CRC (IBD controls without CRCs) and non-IBD patients with MSI or MSS CRCs (non-IBD MSI or non-IBD MSS CRCs) and diverticulitis (non-IBD controls). The expression of both miRs was also studied in matched non-neoplastic distant mucosa of IBD and non-IBD patients with CRCs.

Clinico-pathological features of all patients are detailed in Table I and the loss of MMR proteins in MSI CRCs is summarized in Table II.

Relative expressions of miR-155 and miR-21 in neoplastic and matched normal distant mucosa of IBD patients were first compared with expressions in diverticulitis (non-IBD controls). Expression values in diverticulitis are normalized to 1, and all $2^{-\Delta\Delta Ct}$ values are calculated accordingly. MiR-155 and miR-21 appeared significantly overexpressed in CRCs complicating IBD ($P < 0.001$ and $P < 0.0001$, respectively; Figure 1A and B). A significant overexpression of these two miRs was also observed in non-neoplastic distant mucosa of IBD patients with CRC ($P < 0.0001$ versus diverticulitis in both cases) and was also clearly apparent in IBD subjects without CRC ($P < 0.0001$ versus diverticulitis in both cases).

In an identical manner, we compared miR expressions in both neoplastic and non-neoplastic distant mucosa samples of non-IBD patients to expressions in diverticulitis. In this context, miR-155 and miR-21 behaved differently. No overexpression of miR-155 was observed in hereditary and sporadic CRCs in comparison with diverticulitis ($2^{-\Delta\Delta Ct}$ values < 1), whereas overexpression of miR-21 was observed mostly in neoplastic tissues, and to a lesser extent in their associated non-neoplastic mucosa ($P < 0.0001$ and $P < 0.05$,

Table II. MMR proteins lost in MSI CRCs[a]

Variable	Number of patients	Sex, F/M	MLH1	MSH2	MSH6	PMS2
Inherited MSI CRCs	13	4/9	5	8	0	0
Sporadic MSI CRCs	19	10/9	19	0	0	0
MSI IBD CRCs	10	4/6	6	3	1	0

[a]MMR protein expression was determined by immunohistochemistry (7,37).

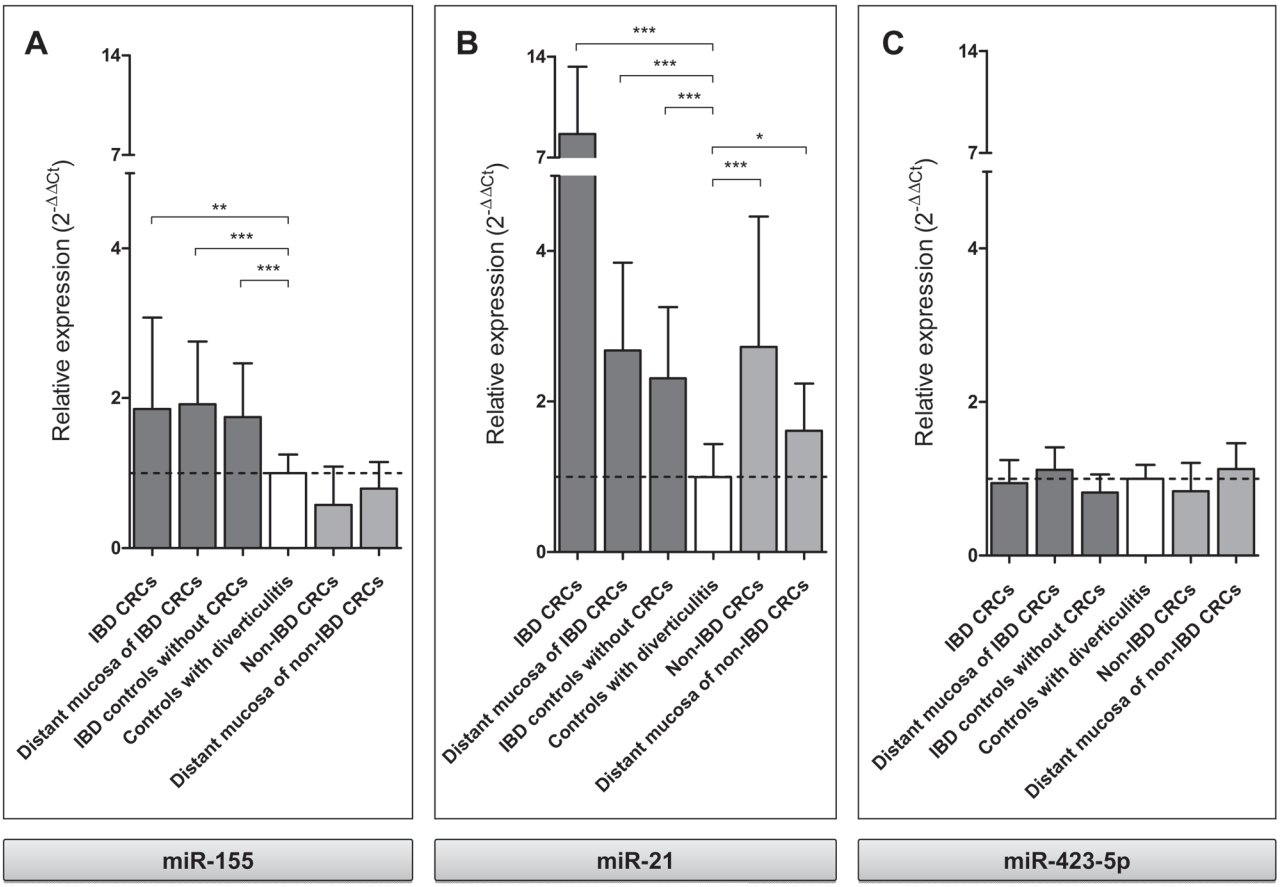


Fig. 1. Relative expression of miR-155 and miR-21 in different sets of healthy, neoplastic or inflamed colorectal mucosa. Relative expression of miR-155 (**A**) and miR-21 (**B**) is measured in two different settings associated (dark gray) or not associated (light gray) to IBD; both compared with healthy mucosa (controls with diverticulitis, $n = 20$). IBD-associated samples consist of IBD-associated CRCs ($n = 20$), their matched distant normal appearing mucosa ($n = 20$) and colonic mucosa from IBD patients without CRCs ($n = 24$). Non-IBD samples consist of hereditary (HNPCC) and sporadic CRCs ($n = 62$) and their matched normal appearing mucosa ($n = 50$). Statistical analysis is performed only when the samples overexpress the miR compared with diverticulitis, and statistics are shown only when P values are significant. Overexpression of miR-155 is noted in all IBD-associated tissues relative to diverticulitis, whereas miR-21 appeared globally overexpressed, regardless of the inflammatory or neoplastic context. The same samples are also tested for the control miR-423-5p (**C**). It displayed the same expression levels in all samples—healthy or pathological, FFPE or frozen—supporting the specificity of changes in miR-155 and miR-21 expressions. Kruskal–Wallis test ($P < 0.0001$ for miR-155 and miR-21), followed by Dunn's multiple comparison. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.

Downloaded from http://carcin.oxfordjournals.org/ at INSERM on February 9, 2014

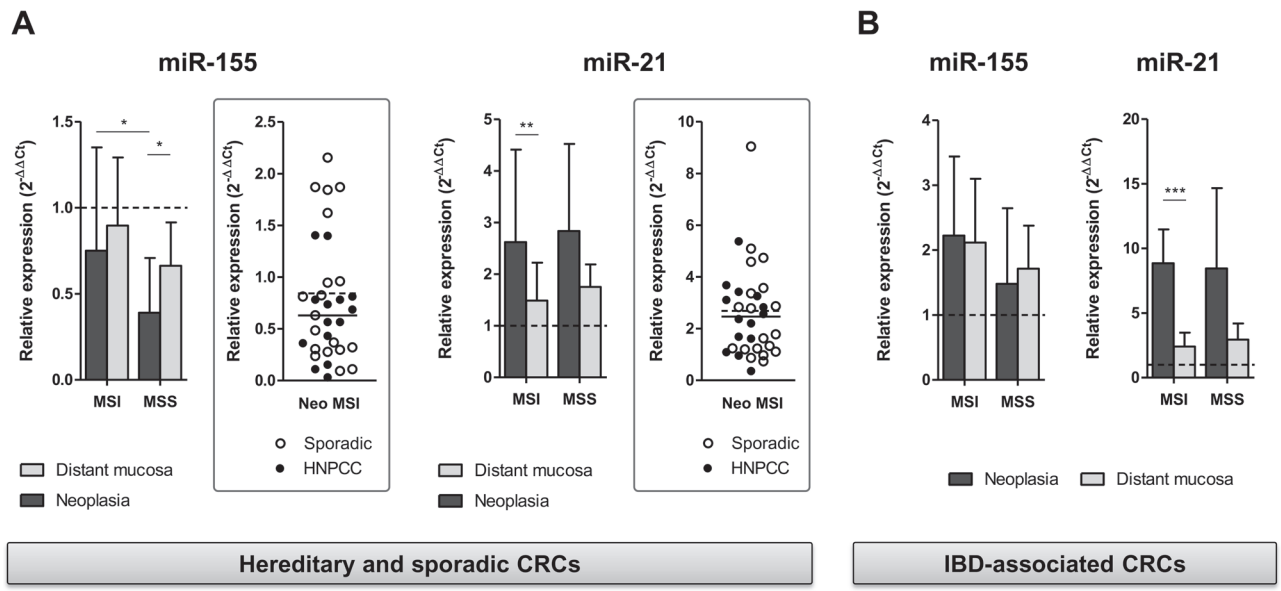


Fig. 2. Relative expression of miR-155 and miR-21 in MSI versus MSS CRCs associated or not associated to IBD. As in Figure 1, relative expression values are shown compared with diverticulitis (relative expression = 1, dotted line), and statistics are shown only when *P* values are significant. Non-IBD CRCs (**A**) and IBD CRCs (**B**) are each subdivided into MSI and MSS CRCs. The same applies to their matched normal appearing mucosa. On the side, MSI CRCs (called Neo MSI) are additionally separated into HNPCC (filled circles, the straight bar represents the average) and sporadic (empty circles, the dotted bar represents the average) CRCs. Note that no difference in relative expression of both miR was observed between hereditary and sporadic tumors. Although miR-21 displayed the same overexpression patterns in MSI and MSS CRCs, regardless of their inflammatory context, miR-155 was inversely expressed between CRCs associated or not associated to inflammation. The underexpression of miR-155 was clearly observed in MSS versus MSI CRCs for hereditary and sporadic CRCs ($P < 0.05$).

respectively; Figure 1A and B). MiR-423-5p, used as a control for an unbiased investigation, displayed expectedly unchanged expressions in all samples (Figure 1C).

Comparative analysis of miR-155 and miR-21 expression in MSI and MSS CRCs and their association to IBD

Non-IBD and IBD CRCs were each subdivided into MSI and MSS tumors, and analyzed for differences in miR expression (Figure 2). Expression in non-neoplastic distant mucosa matched to each CRC sample is presented as well. MSI CRCs were initially separated into sporadic and inherited (HNPCC) cases. However, expression of miR-155 and miR-21 did differ neither in the mean nor in the distribution in sporadic and HNPCC cases (Figure 2A, framed graphs). Accordingly, no further separation was made between these two subtypes of MSI CRCs.

Both non-IBD MSI and MSS neoplastic samples and their matched non-neoplastic distant mucosa (Figure 2A) displayed lower levels of miR-155 expression than healthy controls (diverticulitis; dotted line). In an IBD context, all samples—neoplastic and non-neoplastic, MSI and MSS—presented higher but similar levels of expression of miR-155 as compared with the non-IBD controls (Figure 2B). MiR-21, in both IBD and non-IBD settings, did not show any difference between MSI and MSS tumors, but it displayed significant overexpression once neoplastic samples were compared with their matched non-neoplastic colonic tissues, especially in the inflammatory context (Figure 2A and B).

To broaden our understanding of the differences between MSI and MSS CRCs, in an IBD environment or not, and in a sample-dependent manner, ratios of tumor versus normal [fold changes (FC)] were calculated. As a result, tumor samples are presented as displaying overexpression (positive FC log values) or underexpression (negative FC log values) of the miR in neoplastic tissues (Figure 3). Figure 3C and F shows the fraction of tumors (associated or not associated to IBD) that present an over- (dark gray) or under- (light gray) FC expression of miR-155 and miR-21. Overall, miR-21 did not show any difference between MSI and MSS CRCs regardless of their inflammatory context (Figure 3D–F). miR-155 in non-IBD-associated CRCs did not show any MSI-specific expression pattern or higher expression levels, being even underexpressed in most MSI and MSS tumor samples when compared with matching normal mucosa samples (Figure 3A). The only difference between MSS and MSI CRCs was in a context of IBD, where miR-155 displayed near-significant overexpression in MSI CRCs versus MSS CRCs (Figure 3B and C) ($P = 0.057$).

Discussion

This is the first study that investigates the expression of miR-155 and miR-21 in IBD CRCs. Our results are in accordance with previous data reporting overexpression of miR-155 and miR-21 in the inflamed colonic mucosa of patients with UC (33) and miR-21 in the inflamed colonic mucosa of patients with colonic CD (34). Here, we report results showing relative overexpression of both miRs in neoplastic tissues of IBD patients with CRCs. Importantly, in these patients, overexpression of miR-155 and miR-21 also extended to the distant normal appearing, non-neoplastic mucosa.

Valeri *et al.* (19) recently suggested that overexpression of miR-155—targeting core MMR proteins—might constitute an early step underlying the development of MSI colorectal tumors displaying MMR deficiency, but lacking any identifiable genetic or epigenetic defects. We, on the other hand, have reported colon tumors displaying MSI in IBD patients (7). Intriguingly, the mechanisms behind MMR deficiency in IBD patients are different from those involved in sporadic MSI CRCs—being unassociated to epigenetic methylation of *MLH1* promoter—and, although not the case, appear to be more related to Lynch syndrome with a loss of expression of MLH1, MSH2, MSH6 or PMS2. We thus propose that these tumors may have arisen through downregulation of the core MMR proteins by miR-155 deregulation. This seems to be specifically associated to MSI IBD neoplasias as suggested by the differences observed between MSI and MSS CRCs (Figure 3B and C). Although differences could not reach statistical significance ($P = 0.0573$), they were, however, particularly marked in an IBD context unlike non-IBD CRCs ($P = 0.32$). MiR-155 overexpression due to inflammation can also be proposed as a mechanism that may significantly contribute to the intriguing emergence of MSI colon tumors in IBD patients at a

Downloaded from http://carcin.oxfordjournals.org/ at INSERM on February 9, 2014

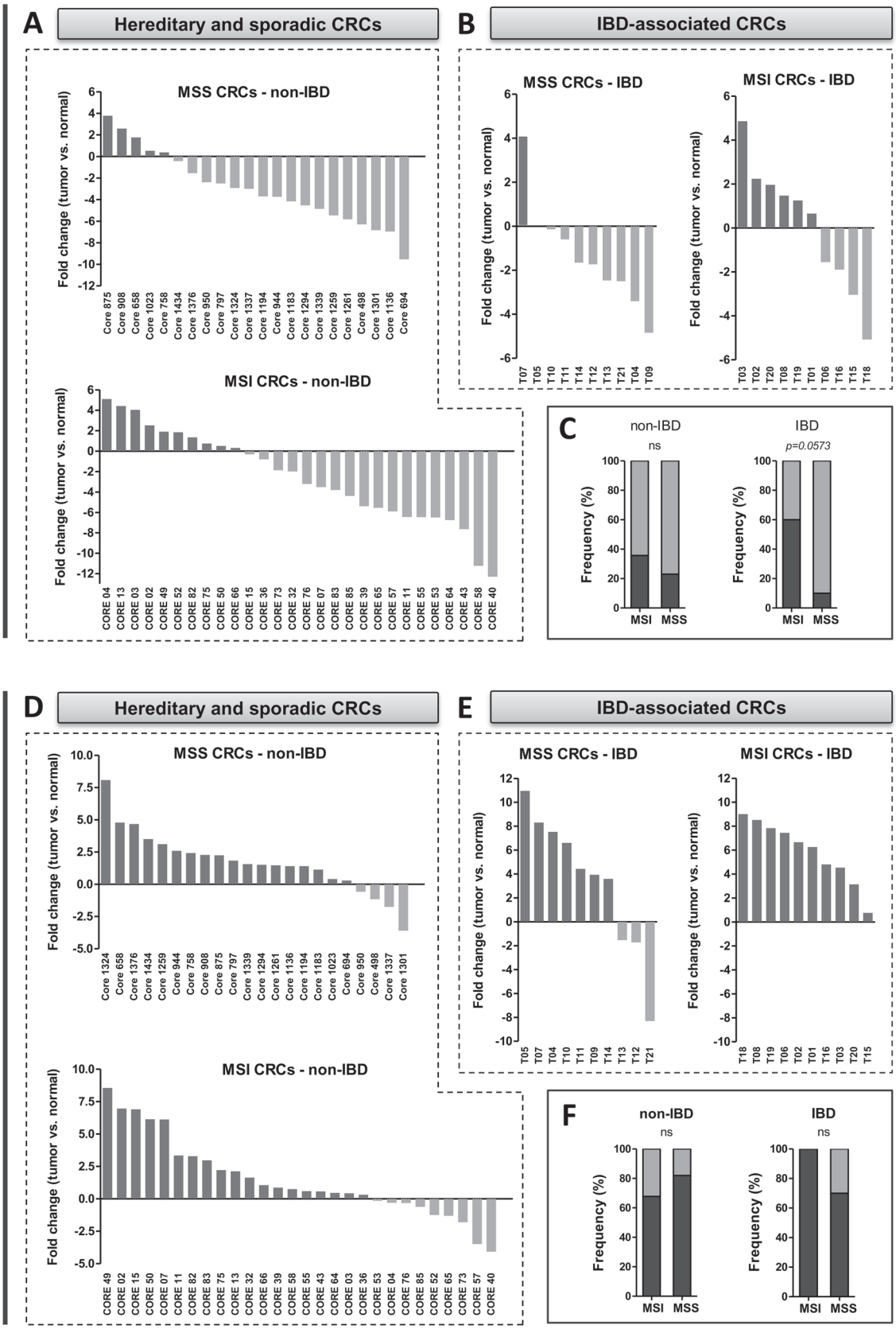


Fig. 3. Tumor versus normal ratios of miR-155 and miR-21 expressions in neoplastic MSI and MSI samples associated or not associated to IBDs. Ratios of tumor versus normal (FC) are calculated in order to assess the overexpression of miR-155 (**A** and **B**) and miR-21 (**D** and **E**) within the same patient. All ratios are log transformed and multiplied by 10. MiR-155 and miR-21 are considered overexpressed in a patient if the values of their respective FC are positive and, underexpressed, if the values of the FC are negative. The frequency of overexpression and underexpression is calculated for each category (**C** and **F**). Similar results of miR-21 expression are obtained for MSI and MSS tumors associated or not associated to IBD. The extent of overexpression is, however, greater in MSI than in MSS CRC in an inflammatory context. Interestingly, together with an overexpression in MSI IBD CRCs, miR-155 appeared clearly underexpressed in MSS IBD CRCs.

Downloaded from http://carcin.oxfordjournals.org/ at INSERM on February 9, 2014

young age despite the absence of any genetic predisposition (i.e. germline MMR defects as in Lynch patients). This model is reinforced by the recent findings that both miR-155 overexpression and pro-inflammatory environments could lead to the emergence of a mutator phenotype in tumor cells (31).

In contrast, our data are less conclusive concerning a specific role of miR-21 in the MMR pathway to cancer when associated to IBD. There was indeed no evidence for a specific pattern of expression of this miR in both non-neoplastic-matched distant mucosa and tumoral tissues of IBD patients with MSI or MSS colon tumors. However, miR-21 overexpression was more prominent in IBD-related CRCs than non-IBD CRCs. In light of the many oncogenic potentials of miR-21 (42,43), a possibly late but synergistic role of this miR during the progression of colorectal tumors complicating IBDs can be proposed.

An important observation was that overexpression of miR-155 occurred not only in neoplastic mucosa and matched non-neoplastic distant mucosa of IBD CRC patients when compared with non-IBD controls and non-IBD colon tumors but also in the inflamed mucosa of IBD patients without cancer. These results are compatible with the concept of epigenetic field defect in the IBD context regardless of the presence of tumors. It can thus be proposed that overexpression of miR-155, along with other putative inflammation-related epigenetic processes, might predispose to the establishment of proliferative MMR-deficient clones in the colonic mucosa of IBD patients that are yet to undergo malignant transformation due to MSI. This field defect has been previously reported by our group for other cancer-related genes (e.g. O^6-methylguanine DNA methyltransferase) (37). As mentioned, overexpression of miR-21—shown here to be mainly associated with both MSI and MSS neoplasia and particularly accentuated in an inflammatory setting—could also contribute along with miR-155 to the development of IBD-related MSI CRCs but in a non-specific manner. In contrast, our data indicate that non-IBD patients with CRC display lower expression of miR-155 in both non-neoplastic and neoplastic mucosa when compared with IBD and healthy controls. Even if it appeared more expressed in MSI than MSS tumors, a difference previously noted by some studies (19,23,24), an important role for miR-155 in MSI non-IBD settings is unlikely (including some Lynch tumors in which the second hit in MMR genes was not identified).

In conclusion, our findings argue in favor of the existence of some epigenetic mechanisms directly associated to inflammation implicating miRs (and particularly miR-155), which might be the reason behind the emergence of an MSI phenotype in IBD colorectal carcinogenesis, whereas these epigenetic mechanisms seem to be an epiphenomenon in sporadic and hereditary MSI colorectal carcinogenesis. We recently demonstrated that heterogeneous MMR defects with a low incidence of *MLH1* methylation also characterizes MSI post-transplant lymphomas (44,45). MiR-155 overexpression, observed in many types of malignancies of B cell or myeloid origin (46), could therefore be also implicated in the genesis of these types of lymphomas and this will be investigated in future studies.

CRC is a heterogeneous disease consisting of multiple tumor subtypes characterized by different molecular and morphological features and with varying responses to chemotherapy. This study highlights—as a proof of concept and for the first time—that overexpression of miR-155 may promote MSI-driven transformation of the colonic mucosa in a context of miR-155 field defect associated with IBDs. These preliminary data still need to be confirmed on larger cohorts of patients. The detection and monitoring of field defects in IBD may have implications for cancer prevention and treatment.

Funding

Ligue Nationale Contre le Cancer fellowship (to N.E.M.); Ligue Nationale Contre le Cancer (to A.D.).

Conflict of Interest Statement: None declared.

References

1. Ekbom,A. *et al.* (1990) Ulcerative colitis and colorectal cancer. A population-based study. *N. Engl. J. Med.*, **323**, 1228–1233.
2. Gillen,C.D. *et al.* (1994) Crohn's disease and colorectal cancer. *Gut*, **35**, 651–655.
3. Gillen,C.D. *et al.* (1994) Ulcerative colitis and Crohn's disease: a comparison of the colorectal cancer risk in extensive colitis. *Gut*, **35**, 1590–1592.
4. Rutter,M. *et al.* (2004) Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis. *Gastroenterology*, **126**, 451–459.
5. Ullman,T.A. *et al.* (2011) Intestinal inflammation and cancer. *Gastroenterology*, **140**, 1807–1816.
6. Schulmann,K. *et al.* (2005) Molecular phenotype of inflammatory bowel disease-associated neoplasms with microsatellite instability. *Gastroenterology*, **129**, 74–85.
7. Svrcek,M. *et al.* (2007) Specific clinical and biological features characterize inflammatory bowel disease associated colorectal cancers showing microsatellite instability. *J. Clin. Oncol.*, **25**, 4231–4238.
8. Duval,A. *et al.* (2002) Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer Res.*, **62**, 2447–2454.
9. Hampel,H. *et al.* (2005) Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N. Engl. J. Med.*, **352**, 1851–1860.
10. Truninger,K. *et al.* (2005) Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. *Gastroenterology*, **128**, 1160–1171.
11. Aaltonen,L.A. *et al.* (1993) Clues to the pathogenesis of familial colorectal cancer. *Science*, **260**, 812–816.
12. Ionov,Y. *et al.* (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*, **363**, 558–561.
13. Thibodeau,S.N. *et al.* (1993) Microsatellite instability in cancer of the proximal colon. *Science*, **260**, 816–819.
14. Veigl,M.L. *et al.* (1998) Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 8698–8702.
15. Croce,C.M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.*, **10**, 704–714.
16. Esquela-Kerscher,A. *et al.* (2006) Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer*, **6**, 259–269.
17. Lujambio,A. *et al.* (2012) The microcosmos of cancer. *Nature*, **482**, 347–355.
18. Wu,W.K. *et al.* (2011) MicroRNA in colorectal cancer: from benchtop to bedside. *Carcinogenesis*, **32**, 247–253.
19. Valeri,N. *et al.* (2010) Modulation of mismatch repair and genomic stability by miR-155. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 6982–6987.
20. Bakirtzi,K. *et al.* (2011) Neurotensin signaling activates microRNAs-21 and -155 and Akt, promotes tumor growth in mice, and is increased in human colon tumors. *Gastroenterology*, **141**, 1749–61.e1.
21. Ng,E.K. *et al.* (2009) Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. *Gut*, **58**, 1375–1381.
22. Volinia,S. *et al.* (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 2257–2261.
23. Earle,J.S. *et al.* (2010) Association of microRNA expression with microsatellite instability status in colorectal adenocarcinoma. *J. Mol. Diagn.*, **12**, 433–440.
24. Lanza,G. *et al.* (2007) mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer. *Mol. Cancer*, **6**, 54.
25. Sarver,A.L. *et al.* (2009) Human colon cancer profiles show differential microRNA expression depending on mismatch repair status and are characteristic of undifferentiated proliferative states. *BMC Cancer*, **9**, 401.
26. Luo,X. *et al.* (2011) MicroRNA signatures: novel biomarker for colorectal cancer? *Cancer Epidemiol. Biomarkers Prev.*, **20**, 1272–1286.
27. Asangani,I.A. *et al.* (2008) MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene*, **27**, 2128–2136.
28. Kim,Y.J. *et al.* (2009) MiR-21 regulates adipogenic differentiation through the modulation of TGF-beta signaling in mesenchymal stem cells derived from human adipose tissue. *Stem Cells*, **27**, 3093–3102.
29. Wickramasinghe,N.S. *et al.* (2009) Estradiol downregulates miR-21 expression and increases miR-21 target gene expression in MCF-7 breast cancer cells. *Nucleic Acids Res.*, **37**, 2584–2595.
30. Valeri,N. *et al.* (2010) MicroRNA-21 induces resistance to 5-fluorouracil by down-regulating human DNA MutS homolog 2 (hMSH2). *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 21098–21103.

Downloaded from http://carcin.oxfordjournals.org/ at INSERM on February 9, 2014

31. Tili,E. *et al.* (2011) Mutator activity induced by microRNA-155 (miR-155) links inflammation and cancer. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 4908–4913.
32. Fasseu,M. *et al.* (2010) Identification of restricted subsets of mature microRNA abnormally expressed in inactive colonic mucosa of patients with inflammatory bowel disease. *PLoS ONE*, **5**, e13160.
33. Takagi,T. *et al.* (2010) Increased expression of microRNA in the inflamed colonic mucosa of patients with active ulcerative colitis. *J. Gastroenterol. Hepatol.*, **25** (suppl. 1), S129–S133.
34. Wu,F. *et al.* (2010) Identification of microRNAs associated with ileal and colonic Crohn's disease. *Inflamm. Bowel Dis.*, **16**, 1729–1738.
35. Buhard,O. *et al.* (2004) Quasimonomorphic mononucleotide repeats for high-level microsatellite instability analysis. *Dis. Markers*, **20**, 251–257.
36. Zaanan,A. *et al.* (2011) Defective mismatch repair status as a prognostic biomarker of disease-free survival in stage III colon cancer patients treated with adjuvant FOLFOX chemotherapy. *Clin. Cancer Res.*, **17**, 7470–7478.
37. Svrcek,M. *et al.* (2010) Methylation tolerance due to an O6-methylguanine DNA methyltransferase (MGMT) field defect in the colonic mucosa: an initiating step in the development of mismatch repair-deficient colorectal cancers. *Gut*, **59**, 1516–1526.
38. Li,J. *et al.* (2007) Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. *BMC Biotechnol.*, **7**, 36.
39. Liu,A. *et al.* (2011) MicroRNA isolation from formalin-fixed, paraffin-embedded tissues. *Methods Mol. Biol.*, **724**, 259–267.
40. Xi,Y. *et al.* (2007) Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *RNA*, **13**, 1668–1674.
41. Zhang,X. *et al.* (2008) An array-based analysis of microRNA expression comparing matched frozen and formalin-fixed paraffin-embedded human tissue samples. *J. Mol. Diagn.*, **10**, 513–519.
42. Krichevsky,A.M. *et al.* (2009) miR-21: a small multi-faceted RNA. *J. Cell. Mol. Med.*, **13**, 39–53.
43. Liu,M. *et al.* (2011) miR-21 targets the tumor suppressor RhoB and regulates proliferation, invasion and apoptosis in colorectal cancer cells. *FEBS Lett.*, **585**, 2998–3005.
44. Borie,C. *et al.* (2009) The mechanisms underlying MMR deficiency in immunodeficiency-related non-Hodgkin lymphomas are different from those in other sporadic microsatellite instable neoplasms. *Int. J. Cancer*, **125**, 2360–2366.
45. Duval,A. *et al.* (2004) The mutator pathway is a feature of immunodeficiency-related lymphomas. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 5002–5007.
46. Tili,E. *et al.* (2009) miR-155: on the crosstalk between inflammation and cancer. *Int. Rev. Immunol.*, **28**, 264–284.

Received July 30, 2012; revised December 5, 2012; accepted December 10, 2012

Downloaded from http://carcin.oxfordjournals.org/ at INSERM on February 9, 2014

Article 2

MiRNA Genes Constitute New Targets for Microsatellite Instability in Colorectal Cancer

El-Murr N, Abidi Z, Wanherdrick K, Svrcek M, Gaub M-P, Fléjou J-F, Hamelin R, Duval A, Lesuffleur T (2012) PLoS ONE 7(2): e31862.

OPEN ACCESS Freely available online

PLoS one

MiRNA Genes Constitute New Targets for Microsatellite Instability in Colorectal Cancer

Nizar El-Murr[1,2], Zoulira Abidi[1,2], Kristell Wanherdrick[1,2], Magali Svrcek[1,2,3], Marie-Pierre Gaub[4], Jean-François Fléjou[1,2,3,5], Richard Hamelin[1,2], Alex Duval[1,2], Thécla Lesuffleur[1,2]*

1 INSERM, UMRS 938 – Centre de Recherche Saint-Antoine, Equipe ''Instabilité des Microsatellites et Cancers'', Paris, France, **2** UPMC – Université Pierre et Marie Curie, Paris, France, **3** AP-HP, Hôpital Saint-Antoine, Service d'Anatomie et Cytologie Pathologiques, Paris, France, **4** INSERM, U682, Développement et Physiopathologie de l'Intestin et du Pancréas, Strasbourg, France, **5** AP-HP, hôpital Saint-Antoine, Tumorothèque CancerEst, Paris, France

Abstract

Mismatch repair-deficient colorectal cancers (CRC) display widespread instability at DNA microsatellite sequences (MSI). Although MSI has been reported to commonly occur at coding repeats, leading to alterations in the function of a number of genes encoding cancer-related proteins, nothing is known about the putative impact of this process on non-coding microRNAs. In miRbase V15, we identified very few human microRNA genes with mono- or di-nucleotide repeats ($n = 27$). A mutational analysis of these sequences in a large series of MSI CRC cell lines and primary tumors underscored instability in 15 of the 24 microRNA genes successfully studied at variable frequencies ranging from 2.5% to 100%. Following a maximum likelihood statistical method, microRNA genes were separated into two groups that differed significantly in their mutation frequencies and in their tendency to represent mutations that may or may not be under selective pressures during MSI tumoral progression. The first group included 21 genes that displayed no or few mutations in CRC. The second group contained three genes, i.e., *hsa-mir-1273c*, *hsa-mir-1303* and *hsa-mir-567*, with frequent (≥80%) and sometimes bi-allelic mutations in MSI tumors. For the only one expressed in colonic tissues, *hsa-mir-1303*, no direct link was found between the presence or not of mono- or bi-allelic alterations and the levels of mature miR expression in MSI cell lines, as determined by sequencing and quantitative PCR respectively. Overall, our results provide evidence that DNA repeats contained in human miRNA genes are relatively rare and preserved from mutations due to MSI in MMR-deficient cancer cells. Functional studies are now required to conclude whether mutated miRNAs, and especially the miR-1303, might have a role in MSI tumorigenesis.

Citation: El-Murr N, Abidi Z, Wanherdrick K, Svrcek M, Gaub M-P, et al. (2012) *MiRNA* Genes Constitute New Targets for Microsatellite Instability in Colorectal Cancer. PLoS ONE 7(2): e31862. doi:10.1371/journal.pone.0031862

Editor: Hoguen Kim, Yonsei University College of Medicine, Republic of Korea

Received November 18, 2011; **Accepted** January 13, 2012; **Published** February 14, 2012

Copyright: © 2012 EL-Murr et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: N El-Murr is a recipient of a LNCC (Ligue Nationale Contre le Cancer) fellowship. This work was supported by Institut National de la Santé et de la Recherche Médicale (INSERM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: thecla.lesuffleur@inserm.fr

Introduction

During the last decade, microRNA (miRNA) genes have been extensively identified in mammals, plants and viruses. They encode short (~22 nucleotides) single-stranded mature RNA molecules (miRs) that regulate gene expression mostly by base pairing with the 3′ UTR of target mRNA [1,2]. In humans, it is estimated that more than one thousand miRs control the expression of about 60% of protein-coding genes. Numerous functional studies have reported the participation of miRs in various cellular processes, and subsequently, their deregulation in a number of human diseases including cancer [2,3]. MiRNA genes can be located within introns and less frequently within exons, or intergenic regions, their expression being regulated by independent or host gene promoters [4]. Excepting miRNA genes arising entirely from the spliced out introns of host's mRNA (named mirtrons [5]), each microRNA gene produces three types of molecules that will undergo successive cleavages by two RNAses, called Drosha and Dicer, to yield the fully functional miR: a large primary transcript (pri-miR, ~500 to 3000 bases), a hairpin-like intermediate precursor (pre-miR, ~60 to 80 bases), and a transient miRNA duplex from which two different mature miRs are usually produced at different (major miR/minor miR*) or equivalent (miR-5p/miR-3p) amounts [5,6]. Each maturation step relies heavily on crucial structural features that dictate a correct and reliable biogenesis of the mature miRNA. Both size and sequence variations in various regions of the miRNA hairpin (basal segment, stem, miRNA duplex and loop) can cause dysregulation of miR biogenesis and are believed to have tumorigenic consequences [7,8,9,10,11,12].

Several genetic and epigenetic mechanisms that lead to the alteration of miR expression and function have been described in human tumors, including colorectal cancer (CRC) [13,14,15]. A methylation of *miRNA-34b/c* CpG islands, for instance, has been frequently observed in CRC cell lines and primary tumors [16], and an increased expression of miR-17-92 cluster has been noted in conjunction with a chromosomal instability (CIN) at the chromosome band 13q31 containing this *miRNA* locus [17]. Importantly, CIN also called MSS (microsatellite stability) characterizes the main subset of CRC (representing 80–85% of all CRCs). Microsatellite instability (MSI), on the other hand, a particularity resulting from a DNA mismatch repair (MMR)

deficiency, has been reported in the remaining 15–20% of CRCs [18]. In MSI CRC, which usually does not display CIN, tumor progression is thought to result notably from the accumulation of secondary mutational events (deletion/insertion) affecting microsatellites, repeated sequences of short DNA motifs (1–6 bp), contained in cancer-related genes [19,20]. These mutations affect target genes involved in various biological pathways such as the regulation of cell cycle and/or cell proliferation (*TGFBR2, IGF2R, TCF4, AXIN2, PTEN, RIZ*…), the regulation of apoptosis (*BAX, CASP5, BCL10, APAF1, FAS*…), or the DNA damage signalling and repair pathways (*RAD50, BLM, MSH3, MSH6, MBD4, MLH3, CHK1, ATR*…). In the majority of target genes, frame-shift mutations were observed in exonic repeats, most often mononucleotide tracts, leading to the production of truncated proteins. More rarely, somatic mutations in intronic mononucleotide repeats of MSI target genes (*MRE11* and *HSP110*) were shown to lead to aberrant splicing and to the generation of altered proteins [21,22].

In this study, we investigated for the first time whether miRNA genes, regardless of their genomic location, might constitute new targets of MSI in CRC. All human miRNA genes containing mononucleotide (MNR) or dinucleotide repeats (DNR) ($\geq$7 repeats) in their hairpin sequences were screened for mutations (nucleotides additions/deletions) using a large series of MMR-deficient CRC cell lines and primary tumors, as well as lymphoblastoid cell lines (LBLs) and MSS CRC controls enabling the assessment of the polymorphic status of these sequences.

Results

Screening for microsatellite repeats in miRNA hairpins and determination of their polymorphism in MMR-proficient cell lines

Amongst the 940 human miRNA sequences listed in miRbase V15, 24 contain MNR with at least seven repeat units (2.5%) (Table 1). DNR ($\geq$7 repeats) are more rarely found in miRNA sequences (3/940, 0.3%) (Table 1). These miRNA genes are distributed on several chromosomes. The majority is found in protein-coding genes (22/27, 81.5%), mostly within intronic sequences (21/22, 95.5%). Repeated sequences vary in size and can reach up to 18 repeats. More than two-thirds of MNR (17/24; 71%) are small (7 to 8 bp) and A/T rich (16/24; 67%). They span many regions of the hairpin precursor (Figure 1) considered important for miR maturation and/or function. These regions consist of the basal segments, the stem, the miRNA duplex and the terminal loop [7,8,9,10]. Two miRNAs (*hsa-mir-511-1* and *hsa-mir-511-2*) are duplicates of the same gene and display a unique sequence indistinguishable in our analysis. Except *hsa-mir-1234* and *hsa-mir-3166*, all miRNA genes were successfully amplified using a set of fluorescent primers bordering the hairpin sequence. MiRNA genes were first analyzed in healthy individuals (LBLs, $n = 40$) for an evaluation of the inherent polymorphism (Table 2). Out of 21 MNR analyzed in normal DNA samples, the majority (17 genes) was shown to be monomorphic (Table 2). Length polymorphism (1 bp shift) in *hsa-mir-1303* was observed with the smaller allele having the highest allelic frequency (Table S1, Figure S1). Single nucleotide polymorphisms (SNP) *rs33982250* and *rs34889453*, both corresponding to an Adenine deletion, are reported for *hsa-mir-1303* and are located outside the MNR [23]. Rare alleles with up to 3-bp shifts were also identified in *hsa-mir-511*, *hsa-mir-543* and *hsa-mir-1302-7* genes (Table S1, Figure S1). Similar results were obtained in a series of 25 primary MSS colorectal tumors and 13 MSS CRC cell lines (Table S1).

Table 1. Intergenic and intragenic microRNA genes containing MNR or DNR.

miRNA gene	Chromosomic location	Genomic location	Repeat size
hsa-mir-1302-7	8q24.3	IG	A7
hsa-mir-511-1	10p12.33	I	T7
hsa-mir-511-2	10p12.33	I	T7
hsa-mir-543	14q32.31	I	T7
hsa-mir-548f-3	5q22.1	I	T7
hsa-mir-548f-5	Xp21.1	I	T7
hsa-mir-548u	6p11.2	I	T7
hsa-mir-3166	11q14.2	IG	T7
hsa-mir-328	16q22.1	I	G7
hsa-mir-1225	16p13.3	I*	G7
hsa-mir-4271	3p21.31	E	G7
hsa-mir-92b	1q22	I	C7
hsa-mir-152	17q21.32	I	C7
hsa-mir-296	20q13.32	I	C7
hsa-mir-1249	22q13.31	I	C7
hsa-mir-320c-1	18q11.2	I	A8
hsa-mir-525	19q13.42	IG	A8
hsa-mir-320b-2	1q42.11	I	A9
hsa-mir-644	20q11.22	I	T9
hsa-mir-4329	Xq23	IG	T9
hsa-mir-1273c	6q25.2	I	T11
hsa-mir-567	3q13.2	I	A13
hsa-mir-1303	5q33.2	IG	T13
hsa-mir-1234	8q24.3	I*	G18
hsa-mir-1277	Xq24	I	(AU)7
hsa-mir-620	2q24.21	I	(UA)11
hsa-mir-558	12p22.3	I	(GU)18

IG, intergenic; I, intronic; I*, mirtron (miRNAs from small intronic sequences); E, exonic.

doi:10.1371/journal.pone.0031862.t001

MiRNA genes with DNR appeared to be highly polymorphic (Table 2). Two of 3 genes displayed several alleles with important length variations: 6 and 15 alleles found respectively for *hsa-mir-620* and *hsa-mir-558* in healthy individuals (Table S1, Figure S1). Here, length polymorphisms have been reported to be localized within the microsatellite repeats [23].

Mutation analysis of miRNA genes with MNR and DNR

The instability of all miRNA repeats was investigated in a series of 41 primary MSI CRCs and 14 MSI CRC cell lines (Table 2). Primary MSS CRCs ($n = 25$) and MSS CRC cell lines ($n = 13$) are used as controls. With very few exceptions, the MSS controls did not show size alterations in any of the miRNA repeats (3/557 and 3/304 mutational events in primary tumors and cell lines, respectively), while 135/913 and 50/326 mutational events were observed in MSI primary tumors ($p = 2.2\times10^{-16}$) and cell lines ($p = 2.28\times10^{-10}$), respectively. Analysis of DNR showed that 2 miRNAs with 7 (*hsa-mir-1277*) and 11 repeats (*hsa-mir-620*) were unaltered in MSI samples. *Hsa-mir-558* with 18 repeats appeared mutated in one third of MSI cell lines (4/13; 30.8%) and MSI CRCs (12/38; 31.6%) (Table 2).

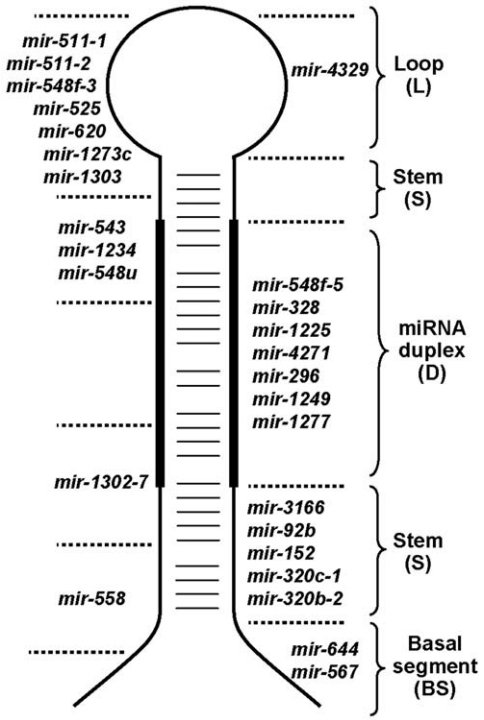


Figure 1. Representative scheme of miRNA hairpins with repeats spaning different locations. The basal segment (BS, single-stranded RNA), stem (S, double-stranded RNA) and terminal loop (L) are designated. The duplex (D, containing one or two potential miRs) is considered as a different entity and therefore distinguished from the stem region. Regions of the hairpin covered by MNRs or DNRs are noted for each miRNAs. To the left of the scheme are miRNA genes whose sequence repeats overlap two regions.
doi:10.1371/journal.pone.0031862.g001

Statistical studies were performed on miRNAs with MNR because of their largest number ($n = 21$). Based on the data reported in Table 2, we show that the majority of miRNA genes are rarely (3/21 with mutation rate $\leq$15%) or not altered (14/21, 66%) in MSI CRC cell lines, while very few genes (4/21; 19%) are found to be significantly mutated. Similar results were obtained in MSI CRCs (Table 2). However, alterations were less frequently encountered in MSI tumors compared to cell lines, an observation in accordance with what is reported for genes with coding MNRs [24]. Additionally, we noted a significant correlation between the size of miRNA MNRs and the mutation frequency in MSI tumors ($p<0.001$, $r = 0.76$) (Figure 2). This correlation was already observed for many microsatellite sequences independently of their genomic localizations (intergenic, coding/exonic, coding/5′ 3′ untranslated and intronic MNRs) (Figure S2).

miRNA genes with MNR segregate into different groups

Using a maximum likelihood statistical method as previously described (refer to Materials and Methods and [24]), we identified two distinct groups of miRNA genes containing MNR that differed in their mutability in MSI primary tumors (Figure 3). Briefly, the test considers that all genes belong to one group of frequency and opposes to this configuration the alternative hypothesis of two mutually exclusive groups of frequencies. The likelihood ratio calculates the chances of each hypothesis and gives the most "likely" to occur. The first, largest group comprises miRNA genes found to be not or not much mutated in MSI tumors (18/21; 86%). The second group contained 3 miRNA genes (*hsa-mir-1273c*, *hsa-mir-567* and *hsa-mir-1303*) frequently mutated in MSI CRCs ($\geq$75%). Similar groups were defined in MSI CRC cell lines (Figure S3).

Characterization of alterations and their impact on mature miRNA expression

Our analysis allowed the identification of 3 MSI-targeted miRNAs: *hsa-mir-1273c, hsa-mir-567 and hsa-mir-1303*. For these genes, the great majority of MSI cell lines presented up to 3 bp deletions (Figure 4, Table S2), and rarely a nucleotide addition (Table S2). A bi-allelic mutation was also noted in 36%, 57% and 83% of altered MSI CRC cell lines for mir-1273c, mir-567 or mir-1303, respectively (Figure 4, Table S2). Similar alterations were found in MSI primary tumors with the exception of *hsa-mir-1273c*, which generally displayed smaller levels of alterations in tumors (only 1 bp-deletion) (Figure 4, Table S2). Concerning the polymorphic *hsa-mir-1303* gene, whose major allele displays an A-deletion (Table S1), the hairpin precursor was sequenced in each MSI CRC cell line to determine the real extent of the deletion. Like LBL and MSS cell lines (Table S1), MSI cell lines appeared to be homozygous (50%; delA/delA) or heterozygous (50%, delA/A) for the "delA" allele. The relevance of the "delA" SNP in *hsa-mir-1303* is shown in Figure 5A that illustrates changes in the terminal loop of the mir-1303 hairpin structures. The dimension of the loop containing the microsatellite decreases when the alterations get bigger and whenever the hairpin structure contains the A-addition SNP (Figure 5A).

Based on these results, we hypothesized that nucleotide deletions encountered in miRNA precursors might affect the biogenesis of mature miRs, modifying their levels of expression and/or their sequence as reported for some other miRNA genes [7,8,9,10]. Using specific quantitative RT-PCR technologies, we determined the relative expression of each mature miR in wild-type (WT) and mutated CRC cell lines in comparison to expression in healthy colonic mucosae. Expression of miR-567 and miR-1273c was not detectable in colonic mucosae and colorectal cell lines ($C_T > 36$ cycles) (data not shown); whereas miR-1303 appeared to be fairly expressed in both normal and tumor colonic cells (Figure 5B). Taking into account the level of instability, miR-1303 expression was first analysed in MSI CRC cell lines. With normal colonic mucosae serving as controls, no difference in miR-1303 expression was noted between unaltered *hsa-mir-1303* MSI cell lines and the heterozygously mutated MSI cell lines (Figure 5B). An increase in the expression of miR-1303 expression was nevertheless observed in some of the MSI cell lines mutated on both alleles (Figure 5B). Furthermore, cell lines supposed to have identical structures of miRNA hairpins (Figure 5C), did not show comparable expression levels. It seems, therefore, as long as the size of the loop does not correlate to the levels of expression, that MSI alterations have no repercussion on mature miR expression. A significant increase of miR-1303 observed in MSS CRC cell lines in comparison to normal colonic mucosa supported this result (Figure 5B).

Discussion

To date, the only established indirect impact of MSI process on miRNA biogenesis is the targeting and therefore disruption of protein coding genes involved in miRNA processing and transport [25,26,27,28]. Our study is the first reporting somatic mutations in miRNA genes due to MSI in MMR-deficient CRCs. By screening the quasi totality of MNR and DNR ($\geq$7 repeat units) contained in miRbase-V15-annotated miRNA hairpin sequences, *hsa-mir-1273c, hsa-mir-1303 and hsa-mir-567* were demonstrated to be

Table 2. Polymorphism and somatic mutation frequency of microsatellite repeats in miRNA genes.

			CRC cell lines			Colorectal primary tumors		
miRNA gene	**LBL (%)**	**Polymorphism**	**MSS (%)**	**MSI (%)**	**P value**	**MSS (%)**	**MSI (%)**	**P value**
MNR								
hsa-mir-1302-7	1/33 (3%)[a]	yes	0/13 (0%)	0/14 (0%)[b]	NS	0/24 (0%)	1/36 (2.8%)[b]	NS
hsa-mir-511	3/40 (7.5%)[a]	yes	0/13 (0%)	0/14 (0%)[b]	NS	0/23 (0%)	0/36 (0%)[b]	NS
hsa-mir-543	6/35 (17.1%)[a]	yes	0/13 (0%)	1/14 (7%)[b]	NS	0/25 (0%)	2/39 (5%)[b]	NS
hsa-mir-548f-3	0/36 (0%)	no	0/13 (0%)	0/14 (0%)	NS	0/23 (0%)	0/36 (0%)	NS
hsa-mir-548f-5	0/38 (0%)	no	0/13 (0%)	0/14 (0%)	NS	0/24 (0%)	1/40 (2.5%)	NS
hsa-mir-548u	0/32 (0%)	no	0/13 (0%)	0/12 (0%)	NS	0/23 (0%)	0/39 (0%)	NS
hsa-mir-328	0/34 (0%)	no	0/12 (0%)	0/14 (0%)	NS	0/22 (0%)	0/40 (0%)	NS
hsa-mir-1225	0/37 (0%)	no	0/13 (0%)	0/13 (0%)	NS	1/22 (4,6%)	0/33 (0%)	NS
hsa-mir-4271	0/33 (0%)	no	0/10 (0%)	0/13 (0%)	NS	0/23 (0%)	1/38 (2.6%)	NS
hsa-mir-92b	0/38 (0%)	no	0/13 (0%)	0/14 (0%)	NS	0/24 (0%)	0/40 (0%)	NS
hsa-mir-152	0/40 (0%)	no	0/13 (0%)	0/14 (0%)	NS	0/20 (0%)	0/37 (0%)	NS
hsa-mir-296	0/36 (0%)	no	0/13 (0%)	0/14 (0%)	NS	0/24 (0%)	0/38 (0%)	NS
hsa-mir-1249	0/23 (0%)	no	1/13 (7.7%)	0/13 (0%)	NS	0/24 (0%)	2/39 (5%)	NS
hsa-mir-320c-1	0/38 (0%)	no	0/13 (0%)	0/14 (0%)	NS	0/23 (0%)	1/38 (2.6%)	NS
hsa-mir-525	0/38 (0%)	no	0/13 (0%)	2/14 (14.3%)	NS	0/21 (0%)	0/33 (0%)	NS
hsa-mir-320b-2	0/37 (0%)	no	0/13 (0%)	1/14 (7.1%)	NS	0/24 (0%)	3/40 (7.5%)	NS
hsa-mir-644	0/40 (0%)	no	0/12 (0%)	6/14 (42.9%)	0.03	0/23 (0%)	8/39 (20.5%)	0.053
hsa-mir-4329	0/40 (0%)	no	2/12 (16.7%)	0/13 (0%)	NS	0/22 (0%)	6/38 (15.8%)	NS
hsa-mir-1273c	0/32 (0%)	no	0/13 (0%)	9/11 (82%)	<0.0001	0/25 (0%)	33/39 (84.6%)	<0.0001
hsa-mir-567	0/39 (0%)	no	0/12 (0%)	14/14 (100%)	<0.0001	2/25 (8%)	31/39 (79.5%)	<0.0001
hsa-mir-1303	20/39 (51%)[a]	yes	0/13 (0%)	12/14 (85.7%)[b]	<0.0001	0/23 (0%)	33/40 (82.5%)[b]	<0.0001
DNR								
hsa-mir-1277	0/40 (0%)	no	0/13 (0%)	0/14 (0%)	NS	0/25 (0%)	1/41 (2.4%)	NS
hsa-mir-620	32/40 (80%)[a]	yes	0/13 (0%)	1/14 (7%)[b]	NS	0/21 (0%)	1/38 (2.6%)[b]	NS
hsa-mir-558	35/40 (87.5%)[a]	yes	0/12 (0%)	4/13 (30.8%)[b]	0.095	0/24 (0%)	12/38 (31.6%)[b]	0.006

NS, not significant;
[a]the polymorphism rate is the percentage of normal samples showing length variations when compared to the major peak (see Table S1, Figure S1);
[b]mutation rates were estimated by taking into account sizes that diverge from the normal polymorphism (refer to Figure S1).
doi:10.1371/journal.pone.0031862.t002

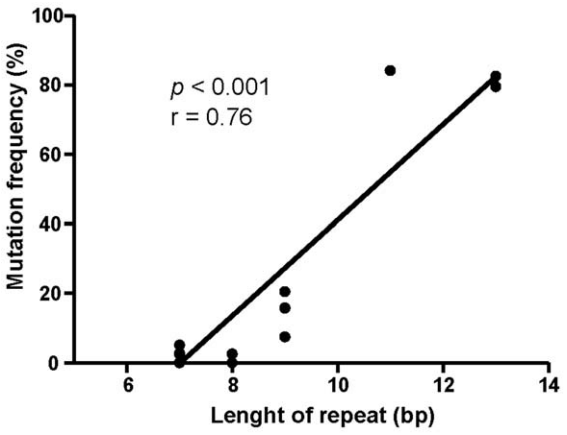


Figure 2. Correlation between the lengths of mononucleotide repeats in miRNAs and their mutation rates in MSI CRCs. Note the highly significant correlation observed.
doi:10.1371/journal.pone.0031862.g002

mutated at high frequency in both CRC cell lines and primary tumors. Since high mutation frequency is the first criterion currently taken into consideration to identify target genes that play a role in the MSI-driven pathway to cancer [19,20,29,30], these miRNA genes may thus constitute real target for MSI. In contrast, all other miRNA alterations we identified are likely to be the result of the background of genetic instability characterizing these tumors. They were found to be affected at low frequency and may play only a minor role in colon tumorigenesis, if any. Besides, some miRNA genes containing microsatellite were found to be never mutated in CRC cell lines and primary tumours. They might be considered as 'survivor' miRNA genes whose mutations are not selected for during tumor progression since they could be highly deleterious for cancer cells [19].

MSI is also influenced by sequence criteria, *e.g.* the length of the repeat. Expectedly, we observed an overall positive correlation between the length of miRNA repeats and their mutation rates in MSI CRC, corroborating the observation made for coding and non-coding MNR. As required for protein coding genes [29,30], functional criteria are necessary to assert that MSI-targeted miRNA genes have a role in MSI colon tumorigenesis. MiRNAs are deeply involved in the regulation of gene expression, being

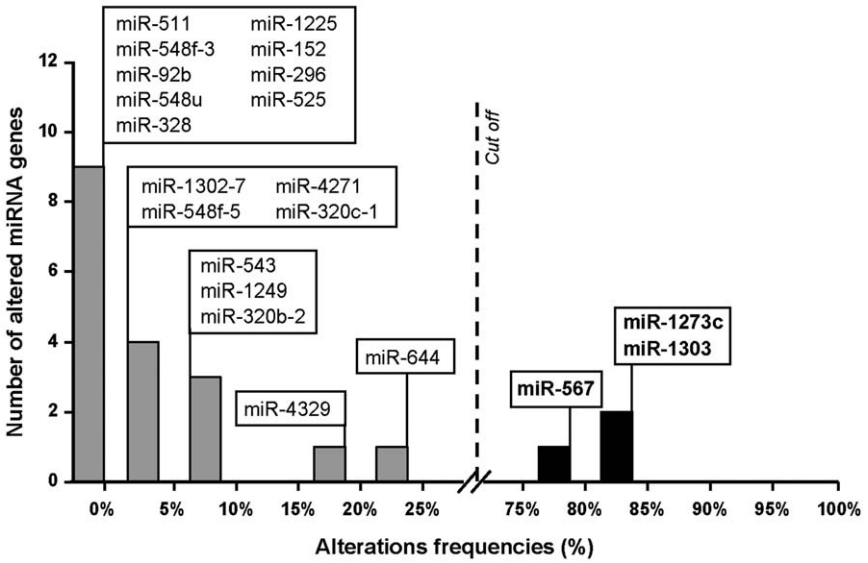


Figure 3. Classification of miRNAs with MNR according to their mutation frequencies in MSI CRCs. Two distinct groups of miRNAs with MNR are established based on their mutation frequencies in MSI primary tumors. The cut-off value is calculated by the ratio of likelihood statistical method and is marked by a dashed vertical line. Note that *hsa-mir-644* is included in the group of miRNAs rarely or not mutated in MSI CRCs ($n = 18$, frequency of mutation $<25\%$) whereas *hsa-mir-1273c*, *hsa-mir-567* and *hsa-mir-1303* constitute the group of miRNAs frequently altered ($n = 3$, frequency of mutation $>75\%$).
doi:10.1371/journal.pone.0031862.g003

therefore associated with various biological processes. No biological role has yet been attributed to *hsa-mir-1303* that is, amongst these, the only miR that was significantly expressed in the colonic mucosa. Several hundreds of mRNA can be assigned *in silico* to miR-1303 depending on the software used, and further analyses are necessary to define those whose *in vivo* expression really depends on *hsa-mir-1303*. Furthermore, the priority is to prove that MNR alterations due to MSI may impact mir-1303 function. Several teams have already highlighted the major role of the terminal loop within the primary miRNA hairpin in miRNA biogenesis and function [8,9]. In addition, a single base alteration (i.e. SNP) within the miRNA gene itself (pri-, pre- and mature miRNA sequences) is sometimes sufficient to alter miRNA expression and/or function in cancers, blocking the processing of pri-miRNA to pre-miRNA [23,31] or, conversely, increasing mature miR expression [32]. We failed here to observe any evident correlation between expression level of mature miR-1303 and mutation in the DNA repeat contained in its terminal loop in MSI CRC cells, regardless of the SNP adjacent to the loop. Further studies using appropriate plasmid constructions will be developed to know whether MNR mutations affecting this miRNA gene could modify the processing of the mature miR-1303 generating different miRs [33] or the function of the pri- or pre-miRNA molecules that are recently reported by Trujillo *et al.* [34] to be biologically active.

In conclusion, our findings have two main implications for the role of miRNA in MSI-driven carcinogenesis. They first provide evidence that MSI CRCs display only a few somatic mutations affecting a small number of miRNA genes containing DNA repeats with yet unclear consequence on the processing and function of these molecules. Functional studies are now required to enforce the idea that the miR-1303 might have a role in MSI tumorigenesis. Secondly, they show that DNA repeats contained in miRNA genes are relatively rare and usually preserved from mutations due to MSI in MMR-deficient cancer cells. This study focuses on the nucleotide repeats located in the hairpin sequences that contain at least the pre-miRNA, and could be enlarged to MNR located in other regions of miRNA genes important for the transcription and the processing of large primary miRNA transcripts (pri-miRNAs).

Materials and Methods

DNA samples

Normal DNA was obtained from 40 healthy individuals (lymphoblastoid cell lines) provided by CEPH (Centre d'Etude du Polymorphisme Humain, Paris, France). Primary tumor tissues (41 MSI and 25 MSS tumors) from patients undergoing surgery for CRC at either the Saint-Antoine hospital (Paris, France) or the hospital of Hautepierre (Strasbourg, France) were collected at the Biological Resources Centers of each institution. Written informed consent was obtained from all patients. Ethics approval was obtained from the Human Research Ethics Committee (Paris, France) and from the "Comité pour la Protection des Personnes de Strasbourg" (CPPEST IV, Strasbourg, France). The MSI status was determined by fluorescent multiplex PCR, as previously described [35], and allowed the evaluation of the percentage of epithelial carcinoma tissue in MSI samples. Only tumors with at least 40% of tumor material have been included in our study. DNA was extracted from 14 MSI and 13 MSS CRC cell lines using QIAamp DNA mini Kit (Qiagen, Courtaboeuf, France) according to manufacturer's instructions.

Identification of mono- and dinucleotide repeats

A systematic screening for mono and dinucleotide repeats was performed in miRbase (version 15, April 2010 release) (http://www.mirbase.org/) [6]. This database provides sequences of miRNA hairpin precursors and mature miRNAs in various species. We selected human miRNAs having at least 7 repeat units. This minimum number was chosen because microsatellites were rarely found to be unstable below 7 repeats (refer to SelTarbase, http://www.seltarbase.org/).

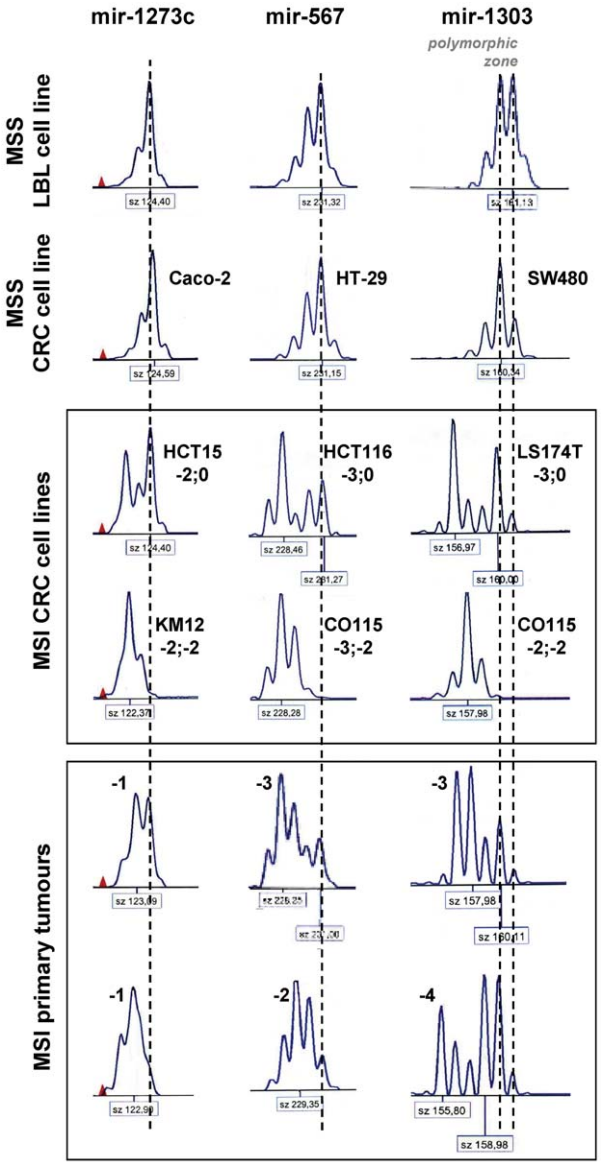


Figure 4. MNR instabilities in *hsa-mir-1273c* (T11), *hsa-mir-567* (A13) and *hsa-mir-1303* (T13). Allelic profiles for several MSI CRC cell lines and primary tumors are shown. Normal profiles are defined in LBL and MSS cell lines and primary tumors. For monomorphic genes, a dashed vertical line indicates the unique allele. The polymorphic zone for *hsa-mir-1303* is defined between two dashed vertical lines going along the 2 alleles (see Figure S1). Sizes (bp) are indicated in a box below each profile. Various allelic deletions ranging from 1 to 4 bp were observed in MSI CRC cell lines and primary tumors and are indicated in bold. The observed deletions were sometimes bi-allelic in MSI CRC cell lines. In MSI primary tumors, the allelic profiles were also highly suggestive of bi-allelic mutations. Due to the inherent polymorphism that can modify the length of the sequence, the hairpin sequence of *hsa-mir-1303* was determined for a correct and reliable evaluation of the alterations in MSI CRC cell lines (see Table S2).
doi:10.1371/journal.pone.0031862.g004

Mutation analysis

Specific primers flanking the hairpin sequences were designed using the Amplifix software for each miRNA candidate. PCR amplification was performed in a final volume of 25 µl with 5 ng of DNA, high or low $MgCl_2$ concentration, with or without Q solution, and Taq DNA polymerase (Qiagen). Primer sequences are listed in Table S3. The thermal cycling conditions comprised an initial denaturation step at 94°C for 3 min, followed by 35 cycles at 94°C for 45 sec; 60°C for 60 sec and 72°C for 60 sec. Finnzyme Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Illkirch, France) was also used in some cases according to manufacturer's protocol. Adequate dilutions of the fluorescent PCR products were mixed with formamide and GeneScan^{TM} 400HD ROX^{TM} Size Standard (Life Technologies, Courtaboeuf, France), heat-denatured and run on a short capillary containing GS Performance Optimized Polymer 7 on the ABI 3100 Genetic Analyzer. Data were visualized and annotated in the GeneMapper 3.7 software (Life Technologies).

Sequencing of hsa-mir-1303 hairpin precursor

A sequence containing the hsa-mir-1303 hairpin precursor was amplified by the Finnzyme Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific) using the following primers: F-GTGAACTAAACGCTGCCTCTGCTA and R-TGCAGGA-ACCGTACTAAGCACT (Tm = 66°C). PCR products were then purified on 96-well Multiscreen-PCR filtration plats (Millipore, Molsheim, France). Sequencing reaction was carried out using the Big Dye Terminator Kit V3.1 (Life Technologies) according to the manufacturer's protocol and using separately the forward or reverse primers. Sequences products were then purified on 96-well Multiscreen-DV plates (Millipore) with Sephadex G-50 Fine and analysed on an ABI 3100 Genetic Analyzer (Life Technologies)

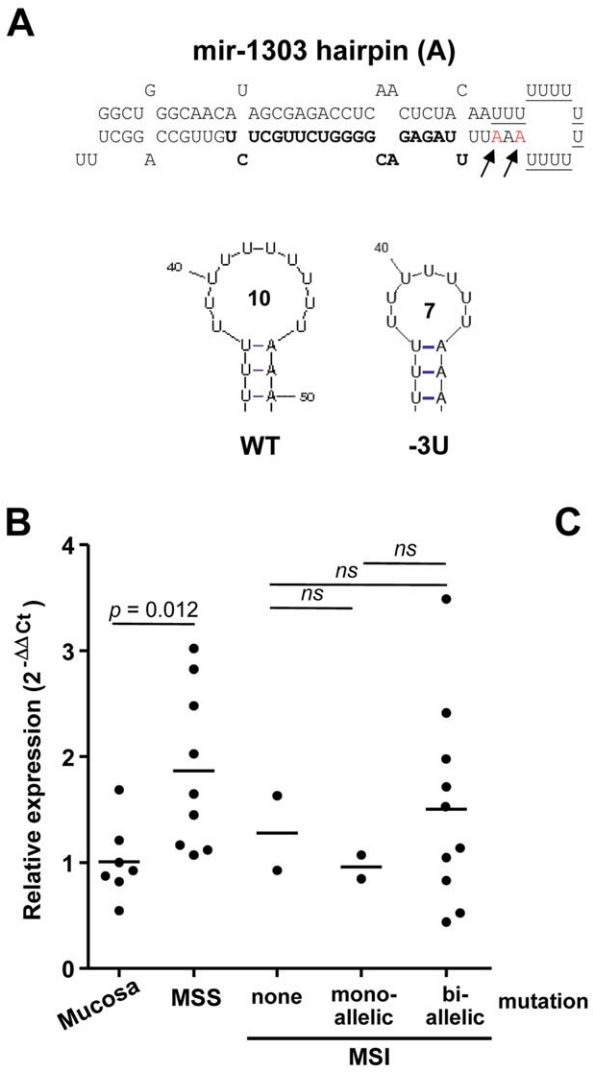
Reverse-Transcription and Real-Time quantitative PCR

Total RNA was prepared from exponentially grown cells (9 MSS and 14 MSI CRC cells) and normal colonic mucosa from patients with CRC ($n = 7$) using TRIzol reagent (Life Technologies) then quantified using a NanoDrop spectrophotometer. cDNAs were generated from 10 ng or 100 ng of total RNA using miRNA-specific stem loop RT primers (Life Technologies) for miR-1303 or miR-1273c, respectively. RT-qPCR assays were performed in triplicate on an ABI 7900 Sequence Detection System using the TaqMan MicroRNA assay according to the manufacturer's instructions (Life Technologies). For the detection of miR-567, the miScript PCR system (miScript RT and miScript SYBR Green PCR kits) was used according to the manufacturer's instructions (Qiagen). With both technologies, CT, the cycle number at which the amount of amplified target reaches a fixed threshold, was determined. CT above 36 were considered as false positives. Mature miRNA expression was normalized to that of RNU48 (Life Technologies) or RNU6B (Qiagen) ($\Delta CT = CT_{miR} - CT_{RNU}$). Comparative quantification was performed using a calibrator sample. Relative miRNA expression was expressed as $2^{-\Delta\Delta CT}$ ($2^{-(\Delta CT\ sample - \Delta CT calibrator)}$). The thermal cycling conditions comprised 45 cycles at 95°C for 15 s and 60°C for 1 min (Life Technologies) or 45 cycles at 94°C for 15 s, 55°C for 30 s and 70°C for 30 s, preceded by an initial activation step at 95°C for 15 min (Qiagen).

Statistical analyses

The differences between variables were assessed with the *Chi-2* or Fisher's exact test, when required. Student t-test was used to evaluate differences in miR expression levels.

As previously described by Duval *et al.*, a ratio of likelihood was calculated to assign each miRNA to a frequency group [24]. With N_i, the number of tumors tested at locus i and n_i, the number of tumors unstable at this locus, the *H1* alternative hypothesis assumes the presence of two types of loci (ie. miRNA genes) that differ in their mutation frequencies (a "stable" group with a low mutation rate, p_1; and an "unstable" group with a high mutation rate, p_2. α and 1-α are the proportions of sites with the p_1 and p_2



MSI cell line	Size of the loop Allele 1	Allele 2	miR-1303 expression
HCT-8	11	11	1,63
TC7			0,93
LS411	10	8	3,49
RKO			0,44
Lim2405	9	9	0,52
KM12			2,42
LoVo	9	8	1,72
HCT116			1,05

Figure 5. Secondary structures of WT and mutated *hsa-mir-1303* and expression levels of miR-1303 in CRC cell lines. A: Alterations in repeat sequences of *hsa-mir-1303* (A) and its variant (delA) did not seem to affect overall the secondary structure of the hairpin but the dimension of the loop (annoted inside) is slightly reduced as determined by mfold software (http://mfold.rna.albany.edu/). Mature miR (bold letters) and MNR (underlined letters) are shown in both hairpin sequences. The arrows indicate the potential positions of an Adenine deletion that leads to an enlargement of the loop. B: Comparison of the relative expressions of mature miR-1303 in MSS (unaltered MNR) and MSI CRC cell lines with none, mono- or bi-allelic mutations of *hsa-mir-1303*. MiR expression was normalized to the expression of RNU48. Means are shown for each group (black horizontal line). A significant increase in the expression of miR-1303 was observed between MSS cell lines and normal colonic mucosae ($p = 0.012$). C: Absence of correlation between the size of mir-1303 loop and the levels of mature miR-1303 expression in MSI cell lines with no (HCT-8, TC7) or bi-allelic mutations (LS411, RKO, LIM2405, KM12, LoVo, HCT116) in MNR of *hsa-mir-1303*. Note cell lines that produce hairpin precursors with the same size of the loop do express mature miR-1303 at various levels.
doi:10.1371/journal.pone.0031862.g005

instability respectively). Conversely, the *H0* null hypothesis assumes that all miRNA genes can be grouped in one frequency class, p_0, where no significant differences exist between each miRNA locus. The ratio of likelihood is given by: $2\ln\frac{L(H1)}{L(H0)}$

$$\text{where } L(H0)=\prod_i \binom{N_i}{n_i}[p_0^{n_i}(1-p_0)^{N_i-n_i}]$$

$$\text{and } L(H1)=\prod_i \binom{N_i}{n_i}[\alpha p_1^{n_i}(1-p_1)^{N_i-n_i}+(1-\alpha)p_2^{n_i}(1-p_2)^{N_i-n_i}]$$

This ratio follows a Chi-square distribution with two degrees of freedom.

For all tests, a 95% confidence interval was applied and $P<0.05$ was considered as significant.

Supporting Information

Figure S1 Allelic profiles of polymorphic miRNA genes in LBLs. For *hsa-mir-1303* (T_{13}), *hsa-mir-620* ($(TA)_{11}$) and *hsa-mir-558* ($(GT)_{18}$) genes, the polymorphic zone is determined between the smallest and the largest alleles (located between the two dashed vertical lines) observed in a large series of 40 lymphoblastoid cell lines from healthy individuals. The length of the predominant alleles (bp) is indicated in a box below each profile.
(TIF)

Figure S2 Comparison of mutation frequencies of miRNA MNRs to those of exonic, untranslated, intronic or intergenic MNRs. MNRs with sizes between 7 and 13 bp and different genomic locations were included in this comparison. These MNRs are taken from SelTarbase (http://www.seltarbase.org/, October 2010 release), an open database of human mononucleotidic microsatellite mutations in MSI cancers.
(TIF)

Figure S3 Classification of miRNAs with MNR based on to their mutation frequencies in MSI CRC cell lines. Two distinct groups of miRNAs with MNR are determined according to the frequency of mutation in MSI cell lines. The cut-off value is calculated by the ratio of likelihood method and is signalled by a dashed vertical line. Note that *hsa-mir-644* is incorporated in the group of miRNAs frequently altered, that also includes *hsa-mir-1273c*, *hsa-mir-567* and *hsa-mir-1303* ($n = 4$, frequency of mutation $>45\%$). All the other miRNAs constitute the group of miRNAs rarely altered or not altered at all ($n = 17$, frequency of mutation $<15\%$).
(TIF)

Table S1 Allelic distribution of polymorphic miRNA genes in LBLs and MSS colorectal tumors and cell lines.
(DOC)

Table S2 Size alterations of miRNA loci in MSI CRC cell lines.
(DOC)

Table S3 Primers sequences for mutation analysis of miRNAs.
(DOC)

Acknowledgments

We thank Lina El-Murr for critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: AD TL. Performed the experiments: NEM ZA. Analyzed the data: NEM KW RH TL. Contributed reagents/materials/analysis tools: MS MPG JFF. Wrote the paper: NEM AD TL.

References

1. Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet 9: 102–114.
2. Melo SA, Esteller M (2011) Dysregulation of microRNAs in cancer: Playing with fire. FEBS Lett in press.
3. Esquela-Kerscher A, Slack FJ (2006) Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer 6: 259–269.
4. Olena AF, Patton JG (2010) Genomic organization of microRNAs. J Cell Physiol 222: 540–545.
5. Winter J, Jung S, Keller S, Gregory RI, Diederichs S (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol 11: 228–234.
6. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36: D154–158.
7. Zeng Y, Cullen BR (2005) Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. J Biol Chem 280: 27595–27603.
8. Zeng Y, Yi R, Cullen BR (2005) Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. Embo J 24: 138–148.
9. Zhang X, Zeng Y (2010) The terminal loop region controls microRNA processing by Drosha and Dicer. Nucleic Acids Res 38: 7689–7697.
10. Han J, Lee Y, Yeom KH, Nam JW, Heo I, et al. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 125: 887–901.
11. Wu M, Jolicoeur N, Li Z, Zhang L, Fortin Y, et al. (2008) Genetic variations of microRNAs in human cancer and their effects on the expression of miRNAs. Carcinogenesis 29: 1710–1716.
12. Ryan BM, Robles AI, Harris CC (2010) Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer 10: 389–402.
13. Calin GA, Croce CM (2006) MicroRNAs and chromosomal abnormalities in cancer cells. Oncogene 25: 6202–6210.
14. Breving K, Esquela-Kerscher A (2009) The complexities of microRNA regulation: mirandering around the rules. Int J Biochem Cell Biol 42: 1316–1329.
15. Krol J, Loedige I, Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet 11: 597–610.
16. Toyota M, Suzuki H, Sasaki Y, Maruyama R, Imai K, et al. (2008) Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer. Cancer Res 68: 4123–4132.
17. Diosdado B, van de Wiel MA, Terhaar Sive Droste JS, Mongera S, Postma C, et al. (2009) MiR-17–92 cluster is associated with 13q gain and c-myc expression during colorectal adenoma to adenocarcinoma progression. Br J Cancer 101: 707–714.
18. Boland CR, Goel A (2010) Microsatellite instability in colorectal cancer. Gastroenterology 138: 2073–2087.
19. Duval A, Hamelin R (2002) Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. Cancer Res 62: 2447–2454.
20. Woerner SM, Benner A, Sutter C, Schiller M, Yuan YP, et al. (2003) Pathogenesis of DNA repair-deficient cancers: a statistical meta-analysis of putative Real Common Target genes. Oncogene 22: 2226–2235.
21. Giannini G, Ristori E, Cerignoli F, Rinaldi C, Zani M, et al. (2002) Human MRE11 is inactivated in mismatch repair-deficient cancers. EMBO Rep 3: 248–254.
22. Dorard C, de Thonel A, Collura A, Marisa L, Svrcek M, et al. (2011) Expression of a mutant HSP110 sensitizes colorectal cancer cells to chemotherapy and improves disease prognosis. Nat Med 17: 1283–1289.
23. Duan S, Mi S, Zhang W, Dolan ME (2009) Comprehensive analysis of the impact of SNPs and CNVs on human microRNAs and their regulatory genes. RNA Biol 6: 412–425.
24. Duval A, Rolland S, Compoint A, Tubacher E, Iacopetta B, et al. (2001) Evolution of instability at coding and non-coding repeat sequences in human MSI-H colorectal cancers. Hum Mol Genet 10: 513–518.
25. Melo SA, Ropero S, Moutinho C, Aaltonen LA, Yamamoto H, et al. (2009) A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. Nat Genet 41: 365–370.
26. Melo SA, Moutinho C, Ropero S, Calin GA, Rossi S, et al. (2010) A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells. Cancer Cell 18: 303–315.
27. Garre P, Perez-Segura P, Diaz-Rubio E, Caldes T, de la Hoya M (2010) Reassessing the TARBP2 mutation rate in hereditary nonpolyposis colorectal cancer. Nat Genet 42: 817–818.
28. Kim MS, Oh JE, Kim YR, Park SW, Kang MR, et al. (2010) Somatic mutations and losses of expression of microRNA regulation-related genes AGO2 and TNRC6A in gastric and colorectal cancers. J Pathol 221: 139–146.
29. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, et al. (1998) A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 58: 5248–5257.
30. Perucho M (1999) Correspondence re: C.R. Boland et al., A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 58: 5248–5257, 1998. Cancer Res 59: 249–256.
31. Duan R, Pak C, Jin P (2007) Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. Hum Mol Genet 16: 1124–1131.
32. Hu Z, Chen J, Tian T, Zhou X, Gu H, et al. (2008) Genetic variants of miRNA sequences and non-small cell lung cancer survival. J Clin Invest 118: 2600–2608.
33. Jazdzewski K, Liyanarachchi S, Swierniak M, Pachucki J, Ringel MD, et al. (2009) Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. Proc Natl Acad Sci U S A 106: 1502–1505.
34. Trujillo RD, Yue SB, Tang Y, O'Gorman WE, Chen CZ (2010) The potential functions of primary microRNAs in target recognition and repression. Embo J 29: 3272–3285.
35. Buhard O, Cattaneo F, Wong YF, Yim SF, Friedman E, et al. (2006) Multipopulation analysis of polymorphisms in five mononucleotide repeats used to determine the microsatellite instability status of human tumors. J Clin Oncol 24: 241–251.

Table S1. Allelic distribution of polymorphic miRNA genes in LBLs and MSS colorectal tumors and cell lines

miRNA gene	Alleles (bp)	LBL (%)	Cell lines (%)	Primary tumors (%)
hsa-mir-1302-7	a (285 bp)	32/33 (97%)	13/13 (100%)	24/24 (100%)
	b (282 bp)	0/33 (0%)	0/13 (0%)	0/24 (0%)
	a/b	1/33 (3%)	0/13 (0%)	0/24 (0%)
hsa-mir-511	a (195 bp)	37/40 (93%)	13/13 (100%)	19/23 (83%)
	b (194 bp)	3/40 (7%)	0/13 (0%)	4/23 (17%)
	a/b	0/40 (0%)	0/13 (0%)	0/23 (0%)
hsa-mir-543	a (218 bp)	29/35 (83%)	13/13 (100%)	25/25 (100%)
	b (219 bp)	0/35 (0%)	0/13 (0%)	0/25 (0%)
	a/b	6/35 (17%)	0/13 (0%)	0/25 (0%)
hsa-mir-1303*	a (160 bp)	19/39 (49%)	6/13 (46%)	10/23 (43%)
	b (161 bp)	4/39 (10%)	2/13 (15%)	3/23 (13%)
	a/b	16/39 (41%)	5/13 (39%)	10/23 (43%)
hsa-mir-620*	a (356 bp)	8/40 (20%)	3/13 (23%)	2/21 (10%)
	b to f	32/40 (80%)	10/13 (77%)	19/21(90%)
hsa-mir-558*	a (337 bp)	5/40 (12%)	4/12 (33%)	2/24 (8%)
	b to j	35/40 (88%)	8/12 (67%)	22/24 (92%)

The letter “a” is assigned to the major, most abundant alleles; letters “b” to “j” are assigned to the minor, more or less rare shifted forms of the amplicon; *, miRNA genes reported as being polymorphic in Patrocles miRNA database (http://www.patrocles.org) and by Duan et al. [23]. On the basis of the percentage obtained for *hsa-mir-1303*, the major allele “a” displays an A-deletion (delA).

Table S2. Size alterations of miRNA loci in MSI CRC cell lines

Cell lines	mir-1273c	mir-567	mir-1303
Co115	-1 ; -1	-3 ; -2	-3 ; -2
HCT116	-1 ; 0	-3 ; 0	-3 ; -2
HCT15	0 ; -2	0 ; +1	-1 ; 0
HCT8	0 ; 0	0 ; +1	0 ; 0
KM12	-2 ; -2	-1 ; 0	-3 ; -1
LIM1215	0 ; 0	-3 ; -1	-3 ; -1
LIM2405	-1 ; +1	-2 ; 0	-2 ; -2
LoVo	-1 ; 0	-2 ; -1	-3 ; -2
LS174T	-1 ; 0	-2 ; 0	-3 ; -1
LS411	-1 ; -3	-2 ; -2	-2 ; -1
RKO	-1 ; 0	-2 ; -1	-2 ; -1
SW48	-1 ; 0	-2 ; -1	-3 ; 0
TC71	-1 ; 0	-2 ; -1	-2 ; -1
TC7	0 ; 0	-3 ; -2	0 ; 0

"-" or "+" represent nucleotide deletions or additions, respectively.
Following numbers characterize the extent of the deletion or addition. Sequencing of *hsa-mir-1303* hairpin sequence was performed, and accordingly many cell lines having the "A" allele displayed a more extended deletion contrary to what was initially thought by genotyping (refer to Figure 4, LS174T and Co115).
Note that the alterations observed are bi-allelic in more than half of MSI CRC cell lines for mir-567 and mir-1303.

Table S3. Primers sequences for mutation analysis of miRNAs

miRNA gene	Forward and Reverse Primers (5'-3')	Amplicon size	Amplicon location (GRCh37.p2)
hsa-mir-525	CTGGAGATGGTCGTTTTAGG	242 bp	19: 26468859-26469100
	ACGTTGCTTTACCCAAACCG		
hsa-mir-567	TAGGTACTAGGTGGAAGGAG	230 bp	3: 18326730-18326959
	CCACTGAAAGGCAATGGAAG		
hsa-mir-1302-7	CTAGGTCCTGTCCATGCTC	284 bp	8: 56141340-56141057
	CAGGATTCCACAGATGGTGA		
hsa-mir-320b-2	GCTGGGTTGAGAGGGCAAAAGGAA	237 bp	1: 17962547-17962311
	CCAGACCTTGGAGCCATGCTTGA		
hsa-mir-320c-1	ATGAGGCCTTCTCTTCCCAGTTCT	220 bp	18: 752585-752797
	AAGGAAATTCCCTTAGTCTGTGTTCATTCA		
hsa-mir-328	GTCGAAGTCCTCCTGTGTAG	182 bp	16: 20850632-20850451
	GGGCTGTATGCACTTTCTC		
hsa-mir-1225	GATGGGCCTCAGCAAGGTCAA	153 bp	16: 2080310-2080158
	AGCGGCTCCATCCCTTCAA		
hsa-mir-1249	AGCTGCCAGGACTTTGGTG	196 bp	22: 24987574-24987379
	TCTTGAAGGACGAGTAGGAG		
hsa-mir-152	GGGCATGCTTCTGGAGTCTA	174 bp	17: 11388867-11388695
	AGTTCTGTCATGCACTGACTG		
hsa-mir-296	GGGAAGATCCTGAGTGAAATTG	195 bp	20: 27588901-27588707
	GTATCGACTGTCCTGTCTCCA		
hsa-mir-92b	CCGCCAATATTGCACTCG	218 bp	1: 6653664-6653880
	GAGGTGCTGGATGGAGTTAAG		
hsa-mir-1303	AGATCAGGCTGGGCAACATAGCGA	162 bp	5: 15228257-15228418
	TCACGAGGTAGCTTGGGGAAGCA		
hsa-mir-511-1	GACACCCATCGTGTCTTTTGCTCT	194 bp	10: 273903-274096
	TTCAGCTGATGGACTTCCTTCGAG		
hsa-mir-511-2	GACACCCATCGTGTCTTTTGCTCT	194 bp	10: 17827112-17827305
	TTCAGCTGATGGACTTCCTTCGAG		
hsa-mir-543	CACGGCCATATCTTTGTCAC	217 bp	14 : 82498204-82498420
	CGTCTTCAAGGCAGAATAGG		
hsa-mir-548f-3	AATGCCCAGAGATTGGGAGTGCAGA	169 bp	5: 18163560-18163393

	ACGGCAATTGTATTAGGTTGGTGCAAAAG		
hsa-mir-548f-5	TGGTTTAGCGATTAGAGAAATGA	285 bp	X: 30541628-30541344
	TTTGATGTTGCTATTAGGTTGGT		
hsa-mir-644	CCCCTCATTCTGTTAATGTG	229 bp	20: 3250087-3250315
	TAGGGAACCATCTTTGCTC		
hsa-mir-4271	GCAGATCCTTTGTGACCTCTTG	292 bp	3: 49251494-49251785
	CTGCGATGCTCGTGAGTCATTA		
hsa-mir-4329	GACAAGAGCTGGAAGGGGAGCCTA	248 bp	X: 35320446-35320199
	TCTAAGCCTTGGCCTCAATCGGTA		
hsa-mir-548u	TGTCAGGCACCAGTTCATTTGGTC	157 bp	6: 57194851-57195007
	AGGTTGGCGCAAAAGTAATTGCAG		
hsa-mir-1273c	GCTTCAGTGAGCTATGTTCACG	126 bp	6: 59343920-59344045
	AGAGGTGAGATCATGCACTCCA		
hsa-mir-558	TCCTGTTTCAGGTGCCATTTCC	320 bp	2: 11578890-11579209
	GTTTTAAGCCAGCCCGTTTGTG		
hsa-mir-620	AATGAGGAGTGCTACTGCAAGTGG	356 bp	12: 7162739-7163094
	GCAAAGCCTCCTTTCTTAGG		
hsa-mir-1277	TCCAGCAGAAATGCAGTATCC	315 bp	X: 1788039-1788353
	GATTTAAGCAAGGGGCTGAA		

Figure S1

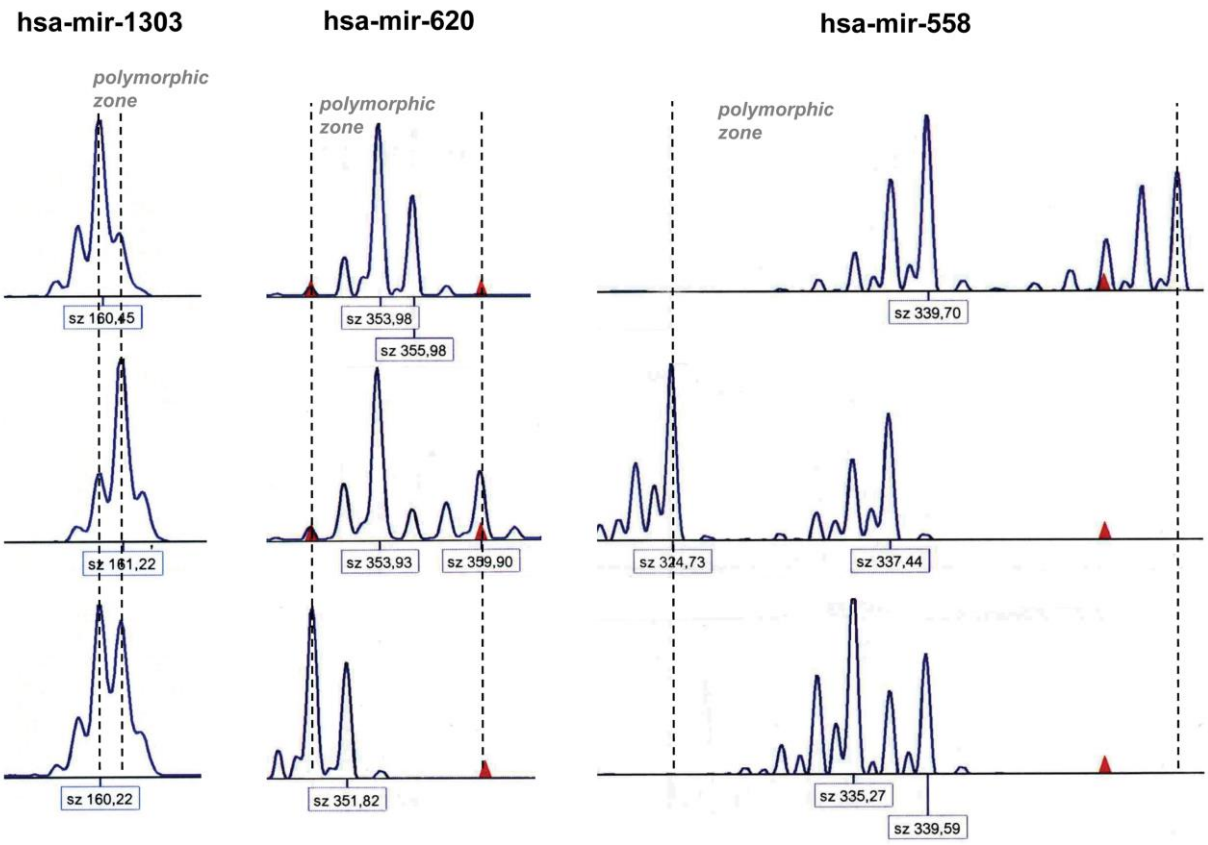

hsa-mir-1303
hsa-mir-620
hsa-mir-558
polymorphic zone
polymorphic zone
polymorphic zone
sz 160,45
sz 161,22
sz 160,22
sz 353,98
sz 355,98
sz 353,93
sz 359,90
sz 351,82
sz 339,70
sz 324,73
sz 337,44
sz 335,27
sz 339,59

Figure S2

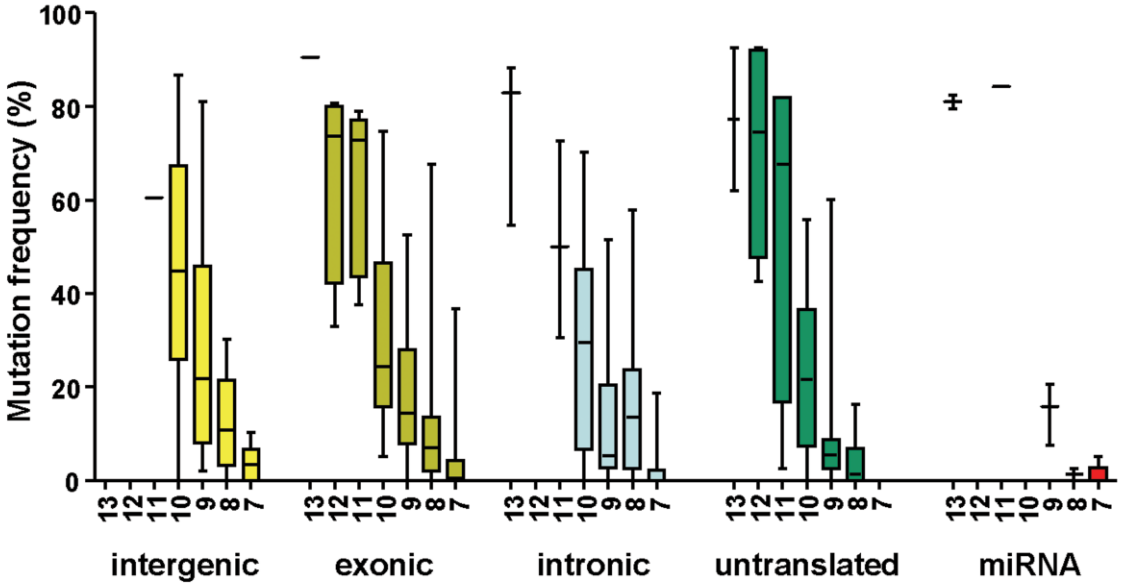
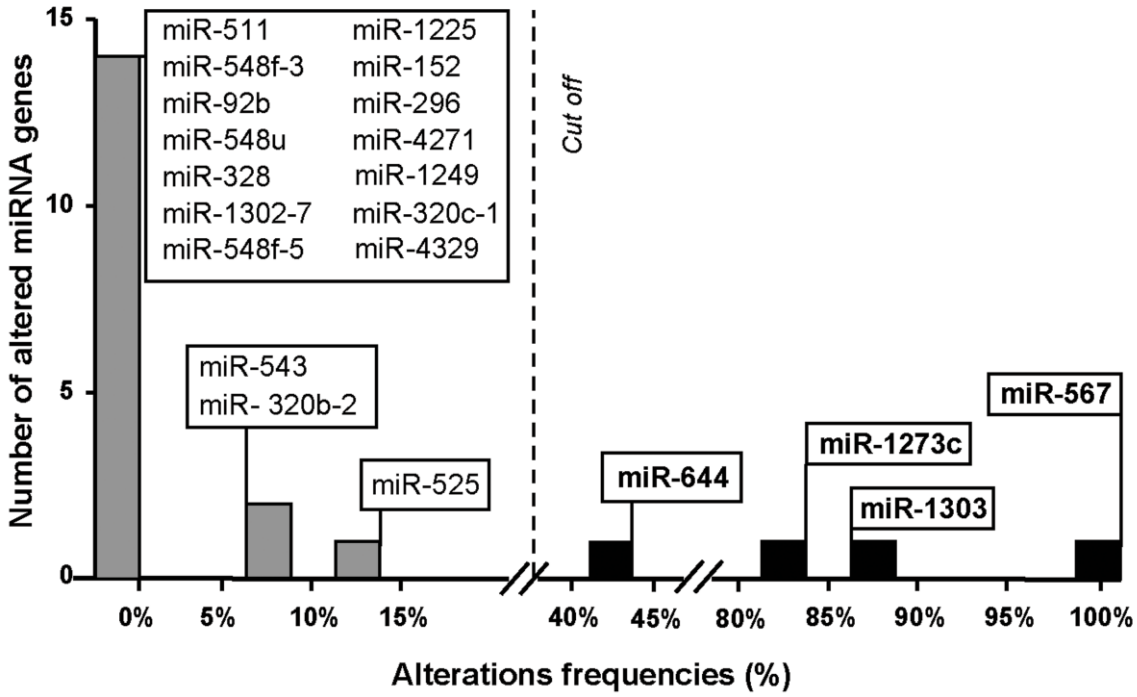

Mutation frequency (%)
100
80
60
40
20
0
13 12 11 10 9 8 7
intergenic
exonic
intronic
untranslated
miRNA

Figure S3


Number of altered miRNA genes
15
10
5
0
miR-511
miR-548f-3
miR-92b
miR-548u
miR-328
miR-1302-7
miR-548f-5
miR-1225
miR-152
miR-296
miR-4271
miR-1249
miR-320c-1
miR-4329
Cut off
miR-543
miR- 320b-2
miR-525
miR-644
miR-1273c
miR-1303
miR-567
0%
5%
10%
15%
40%
45%
80%
85%
90%
95%
100%
Alterations frequencies (%)

Article 3

A somatic mutation in miRNA hairpin precursor affects the biogenesis and expression of miR-3613-5p in MMR-deficient colorectal tumors.

El-Murr N, Marisa M, Buhard O, Wanherdrick K, Gaub M-P, Fléjou J-F, Svrcek M, Lesuffleur T, Duval D.

Manuscrit en préparation

Article 4

Low mutation frequencies characterize miRNA-associated protein coding genes in microsatellite unstable colorectal tumors

El-Murr N, Buhard O, Wanherdrick K, Gaub M-P, Lesuffleur T, Fléjou J-F, Svrcek M, Duval D.

Manuscrit en préparation

DISCUSSION

Le but de ce travail de thèse a été d'étudier le rôle des microARNs dans la carcinogenèse colique de type MSI. Un rôle qui peut intervenir soit à l'étape d'initiation et de mise en place des CCRs MSI (perte de fonction du système MMR), soit à l'étape de progression tumorale (mutations somatiques du fait de MSI dans des gènes cibles de l'instabilité microsatellitaire).

Surexpression des miRs 155 et 21, défaut de champ et initiation des cancers MSI

Nos résultats montrent une surexpression des miRs 155 et 21 dans la muqueuse normale et dysplasique de patients MICI ayant développé un cancer. De manière intéressante, nous montrons également la présence d'un défaut de champs caractérisé par une surexpression de ces deux miRs dans la muqueuse inflammatoire de patients MICI sans cancer. Ce défaut de champs pourrait donc constituer une lésion pré-néoplasique prédisposant à l'émergence d'un cancer. Seule la surexpression du miR-155 semblerait être associée au phénotype MSI, alors que la surexpression du miR-21 contribuerait à la tumorigenèse colorectale de type MSI de manière MMR-indépendante.

Surexpression du miR-155, un évènement pré-tumoral ? L'une des limites de notre étude, en dehors du faible nombre d'échantillons MICI-CCRs de type MSI analysés (qui constituent néanmoins l'une des plus grandes séries publiées), est qu'elle se restreint à rapporter une différence d'expression du miR-155 entre les tumeurs MSI et MSS, sans investiguer le rôle de cette surexpression dans l'initiation du processus tumorigénique MSI. A cet effet, une étude *in vivo* sur modèle murin transgénique de surexpression du miR-155 est en train d'être réalisée dans le laboratoire d'accueil. Elle permettra d'asseoir le fait qu'une dérégulation du miR-155 dans un contexte inflammatoire pourrait constituer une réelle étape pré-tumorale dans l'oncogenèse MSI. Une étude de l'association de cette surexpression avec les cancers MSI MICI-CCRs devrait également être réalisée sur une cohorte plus large de patients MICI.

Défaut de champ, microARNs et carcinogenèse colorectale. Les miRs 155 et 21 ne sont pas les seuls à être impliqué dans des défauts de champs épigénétiques. De manière générale et indépendamment d'un contexte inflammatoire et d'instabilité microsatellitaire, il a été

montré que dans la muqueuse colique saine de nombreux patients avec un CCR, plusieurs miARNs étaient dérégulées. Deng *et al* ont montré que la méthylation des *miR-124a* et *miR-34b/c*, très fréquente dans les cancers colorectaux, survenait tôt dans la carcinogenèse colorectale et qu'elle affectait la muqueuse saine de patients ayant développé un cancer **(Deng 2011)**. Les miRs 34b et 34c, membres de la famille miR-34, sont par ailleurs très connus pour leur rôle de suppresseurs de tumeurs **(Corney 2007; Ji 2008)**. Le miR-124a jouerait lui aussi un rôle suppresseur de tumeur : dans les glioblastomes par exemple, sa sous-expression corrèle avec une augmentation des capacités de migration et d'invasion des tumeurs **(Fowler 2011)**. Le miR-342 est aussi intéressant. Il est inclus dans une séquence intronique d'EVL *(Ena/Vasp-like)* et c'est la méthylation du promoteur du gène hôte qui explique sa sous-expression. La méthylation de ce gène s'étend à la muqueuse normale des patients atteints de CCRs, et serait liée à une résistance des cellules à l'apoptose **(Grady 2008)**. En dehors du contexte colorectal, il a été également suggéré que des effets champs de microARNs joueraient un rôle dans le développement d'autres cancers du tube digestif. Le miR-34b/c par exemple a aussi été trouvé méthylé et donc sous-exprimé dans les cancers gastriques et leurs muqueuses adjacentes non cancéreuses **(Suzuki 2010)**. Le modèle qui semble donc se dégager se baserait sur l'existence de larges réseaux de régulation par les microARNs dont les miRs 155 et 21 feraient partie et dont la dérégulation dans la muqueuse normale prédisposerait à l'émergence d'un cancer. Des réseaux différents dans des contextes différents seraient donc potentiellement capables d'expliquer les caractéristiques moléculaires et clinico-pathologiques des tumeurs colorectales en générale et MSI en particulier.

Importance clinique des défauts de champs. Plusieurs implications cliniques découlent du fait qu'un champ, qui macroscopiquement a une apparence normale, soit capable d'évoluer en cancer. Après résection chirurgicale de la tumeur primaire, il n'est qu'une question de temps pour que le champ qui entoure cette tumeur et qui n'a pas été prélevé, donne naissance à un autre cancer (Figure 23). Cette nouvelle tumeur est qualifiée de « récidive locale » dans le cas où elle se développe à partir de cellules tumorales qui n'ont pas été élimées, ou à partir du même champ duquel est issue la première tumeur (1er cas de figure). Elle est qualifiée de « seconde tumeur primaire » quand elle se développe indépendamment

de la première (c'est-à-dire dans un champ différent) (2ème cas de figure). D'un point de vue clinique, cette différence est importante car qui dit étiologies différentes, dit conséquences cliniques différentes (réponses aux traitements, agressivité, survie des patients, etc.) **(Braakhuis 2003)**. Il existe donc un besoin urgent de trouver des techniques fiables et pratiques qui permettent de suivre l'évolution des défauts de champs et d'essayer de mesurer le risque de développer un cancer chez un patient. Suite à notre étude, plusieurs microARNs peuvent être utilisé dans cette optique ; particulièrement les miRs 155 et 21, dont la quantification est relativement assez facile sur des biopsies ou les échantillons tumoraux et dont la surexpression est assez claire et significative comparée à des échantillons normaux ($p<0,001$).

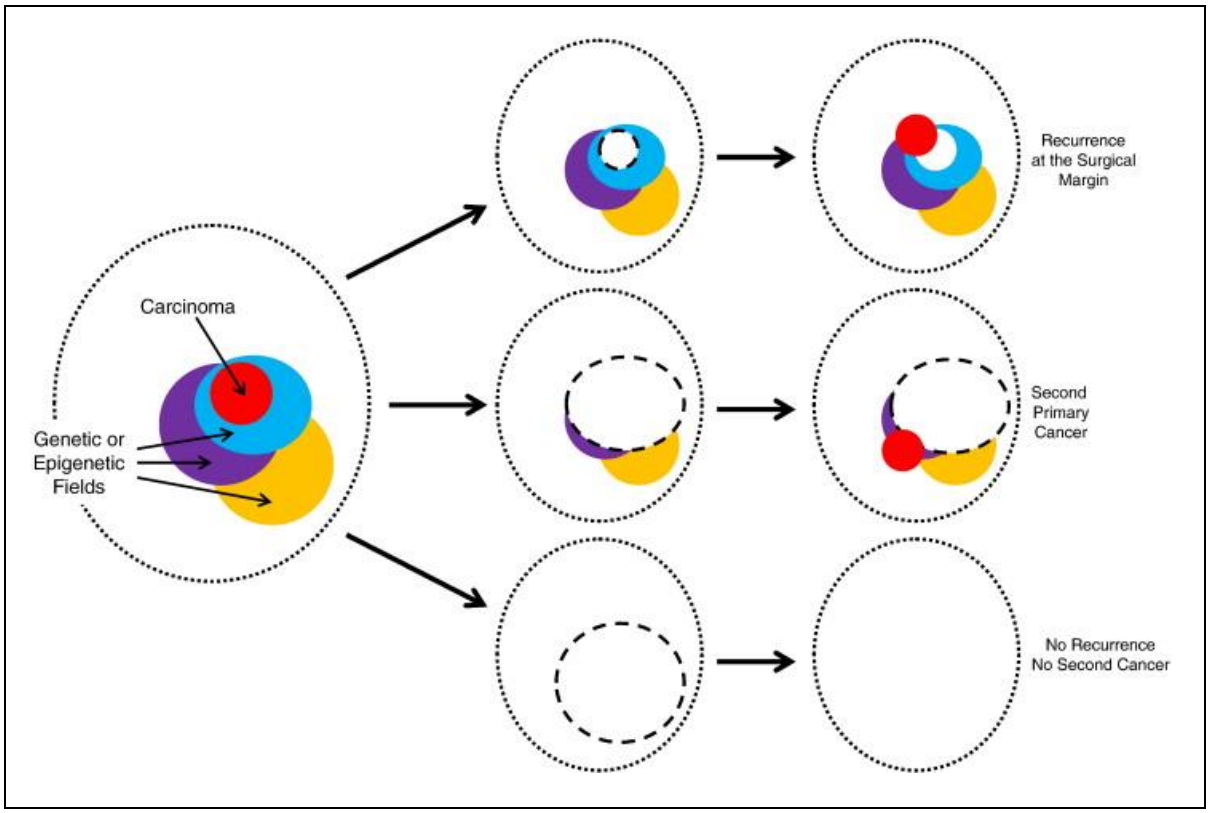


Figure 23. Conséquences cliniques de la persistance d'un défaut de champs après résection de la tumeur primaire. Comme nous venons de le voir, un défaut de champ (plusieurs champs sont représentés en bleu, violet et orange) consiste en une lésion pré-néoplasique qui, en accumulant des mutations supplémentaires, évolue en cancer (en rouge). Le chirurgien peut choisir de réséquer uniquement la tumeur ou inclure une part de la muqueuse normale adjacente (marge de sécurité) ; cependant si tout le champ altéré n'est pas retiré, un cancer réapparaitra. Il est très difficile de délimiter un champ puisque macroscopiquement, il se présente comme de la muqueuse saine. Dans certains cas de cancers colorectaux héréditaires (polypose adénomateuse familiale, par exemple) un colectomie prophylactique (exérèse du colon) est conseillée en pratique. D'après Rivenbark *et al* **(Rivenbark 2012)**.

Mutations somatiques liées à MSI et microARNs

La compréhension des mécanismes moléculaires de la tumorigenèse passe par l'identification de mutations somatiques qui différencient les cellules cancéreuses des cellules normales. Ces mutations somatiques peuvent influer sur le développement d'un cancer et sa réponse aux traitements. A l'origine de ce travail, l'hypothèse a été que les microARNs, à l'instar des gènes codant pour les protéines, peuvent également être ciblés par l'instabilité microsatellitaire avec des conséquences potentielles sur la tumorigenèse colique de type MSI. Globalement, ce que l'on tire à la fois de nos résultats et de la littérature peut se résumer comme suit :

1. Les études réalisées par Saunders *et al.* et par Quach *et al.* montrent que les séquences de microARNs sont relativement peu polymorphes du fait d'une forte pression sélective **(Saunders 2007; Quach 2009)**. A cette observation s'ajoute le fait que très peu d'études ont réussi à mettre en évidence des mutations somatiques ou germinales dans les gènes de microARNs et cela malgré les avancées majeures dans les techniques de nouvelle génération et de séquençage profond des microARNs.

2. Dans un contexte MSI, nos résultats laissent suggérer que les répétitions microsatellitaires sont plutôt faiblement représentées dans les séquences *hairpin* des microARNs. En accord avec le modèle précédemment cité, ces répétitions, constituant des séquences difficiles à maintenir à un état monomorphe, ont de ce fait probablement été contre sélectionnées au cours de l'évolution de par l'impact néfaste que leur instabilité risque d'avoir sur la physiologie cellulaire.

3. De manière intéressante, la majorité des séquences *hairpin* contenant des séquences répétées identifiées dans notre étude ont été retrouvées mutées à des fréquences basses dans les lignées cellulaires et tumeurs primaires MSI. Ces faibles taux de mutations, particulièrement dans un contexte hypermutateur, laissent suggérer que les microARNs constitueraient généralement des gènes « *survivors* » (comme certains gènes codants dont les mutations ne sont pas compatibles avec une survie cellulaire prolongée **(Duval 2002)**). De même que précédemment, cette notion sous-entend que toute

mutation ou polymorphisme seraient probablement délétères à la survie cellulaire et seraient donc fortement contre-sélectionnés, particulièrement dans le contexte d'un développement tumoral.

4. A la marge, certaines séquences répétées de gènes de microARNs sont mutées à des fréquences très élevées et donc, de premier abord, pourraient être considérées comme des vraies cibles MSI. L'absence d'effet sur l'expression du miR mature laisse néanmoins présager qu'il s'agirait plutôt d'évènements neutres ou *« bystanders »* sans effet évident sur l'oncogenèse **(Duval 2002)**.
5. Finalement, la seule exception à ce modèle pourrait être *hsa-miR-3613*. Il est très fréquemment muté dans les lignées cellulaires et les tumeurs primaires MSI, et la mutation a un effet direct sur la biogenèse du miR-3613-5p. Nos résultats préliminaires laissent suggérer que la mutation est susceptible de perturber le *fine-tuning* de l'expression de plusieurs transcrits ciblés par le miR-3613-5p mais ses conséquences fonctionnelles restent encore à déterminer.

Une question de séquence ou de structure ? Dans un premier criblage, 3 gènes de miARNs sont ressortis comme pouvant constituer des vraies cibles MSI par argument de fréquence (*hsa-miR-1303*, *hsa-miR-1273c* et *hsa-miR-567*). Seul le miR-1303 a été retrouvé exprimé dans les lignées cellulaires et les tumeurs MSI, mais son expression ne semblait pas être affectée par l'instabilité. Ce résultat laisse sous-entendre un effet assez négligeable de l'instabilité sur la biogenèse et la fonction des miRs qui contiennent des séquences répétées. De manière assez étrange, et comme il a été décrit dans l'introduction de cette thèse, des changements de séquences (des polymorphismes, de l'*editing*, …) qui affectent la séquence ou la structure du miR peuvent avoir des conséquences moléculaires et fonctionnelles assez importantes (§4.1 ; §5.4.2). Des délétions de nucléotides dans la boucle terminale ont été rapportées comme pouvant inhiber le *processing* des pri-miRs en pre-miRs (§2.3.2). L'explication que l'on peut trouver serait donc que les séquences répétées des microARNs se trouvent au sein de régions assez peu structurées de l'épingle à cheveux (région basale simple brin pour le miR-567 et boucle terminale pour les miRs 1303 et 1273c). Se basant sur le modèle proposé par Auyeung *et al.* (§2.3.2), la taille de la séquence répétée ne ferait pas partie du module « séquence du précurseur » du fait de la redondance des nucléotides. Par

ailleurs, le peu de changements structuraux survenant suite à l'instabilité microsatellitaire, laisse sous-entendre que le module « structure » ne serait pas affecté (Figure 24).

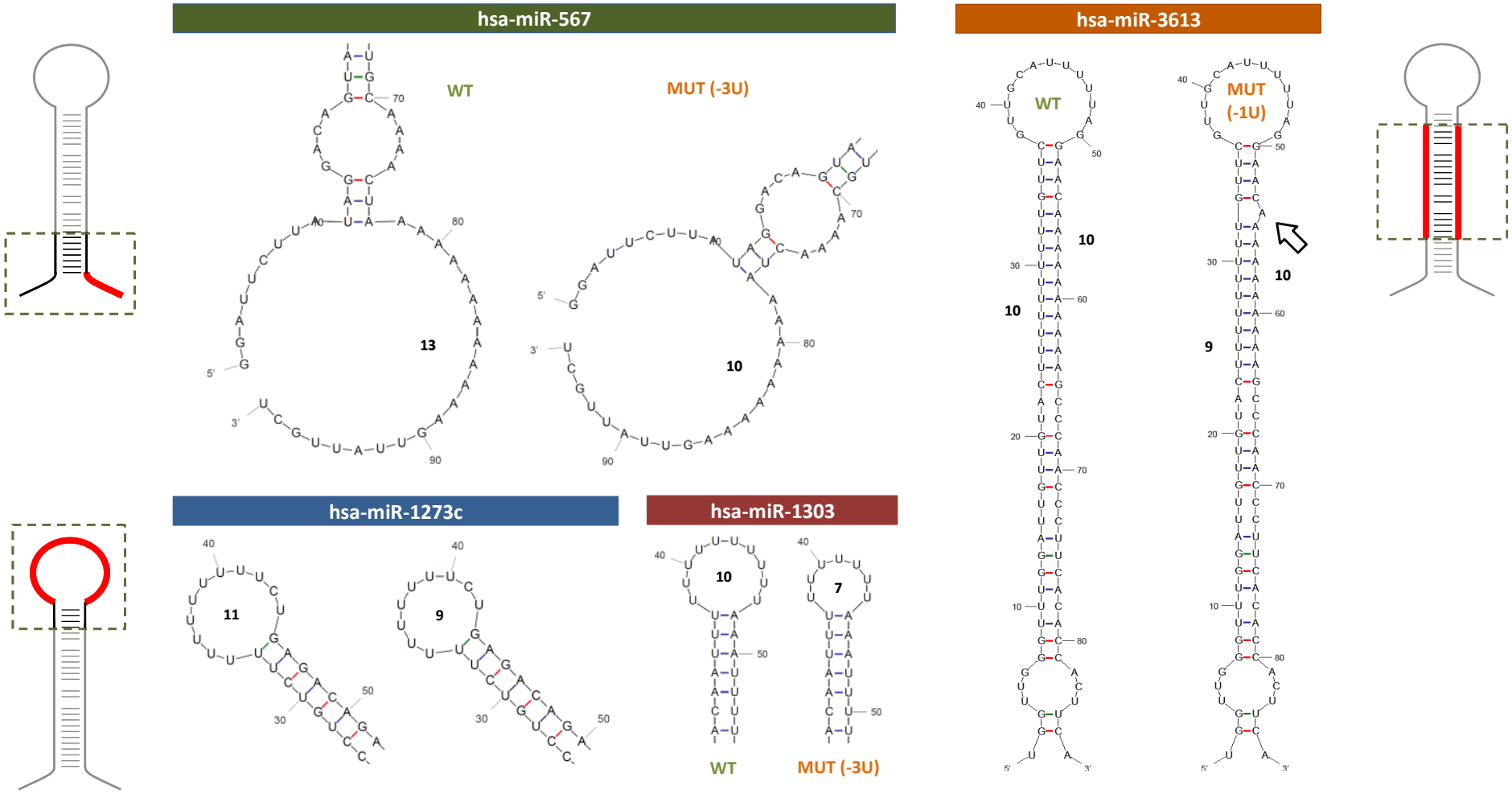


Figure 24. Modifications structurales survenant au sein de l'épingle à cheveux suite à MSI. L'emplacement de la séquence répétée au sein du précurseur est représenté en rouge sur les deux schémas à gauche. La structure secondaire a été modélisée en utilisant le logiciel Mfold **(Zuker 2003)**. Seules les structures ayant la plus faible énergie sont représentées.

Mutations somatiques, IsomiRs, physiologie et physiopathologie : le débat reste encore ouvert. Nous avons pu mettre en évidence une mutation somatique liée à l'instabilité microsatellitaire dans la séquence de deux miRs matures produits à partir d'un même précurseur : hsa-miR-3613-5p et hsa-miR-3613-3p (qui s'est avéré ne pas être exprimé dans les lignées cellulaires et tissus coliques) (Figure 24). Ce gène est fréquemment muté (environ 66%), et les mutations conduisent à une diminution des niveaux d'expression du miR-3613-5p et à la production de variants d'instabilité ou isomiRs qui diffèrent de la forme sauvage par l'ajout de un ou deux nucléotides à l'extrémité 3' du miR. Ces isomiRs sont uniquement produits dans un contexte MSI et sont spécifiques des tumeurs mutées sur *hsa-miR-3613*. Le rôle de ces isomiRs reste à établir, cependant nos résultats laissent suggérer qu'ils auraient une signature transcriptomique qui les rapproche de la forme sauvage pour certaines cibles (régulation concordante), et les différencie de la forme sauvage pour d'autres (régulation discordante). Ces résultats, qui pour le moment ne montrent pas une association définitive entre l'instabilité, les microARNs et la tumorigenèse colique de type MSI, mettent surtout l'accent sur la présence dans un contexte physiopathologique particulier de certaines isoformes de miARN qu'on ne retrouve pas ailleurs. Ils ouvrent également la voie à de futures études sur le rôle des isomiRs, leur physiologie et leurs implications dans divers processus pathologiques.

De manière plus générale, le séquençage profond de petits ARNs a révélé l'hétérogénéité de taille et de séquence qui caractérise la grande majorité des microARNs (§5). Il devient donc de plus en plus clair que de cette diversité devrait découler une certaine complexité fonctionnelle. Chez la drosophile par exemple, il a été montré que le niveau d'expression des isomiRs changeait en fonction des tissus et au cours du développement, ce qui laisse suggérer qu'une variabilité dans l'expression des isomiRs serait d'une certaine façon régulée et donc biologiquement significative **(Fernandez-Valverde 2010)**. Chez l'homme, dans les cellules endothéliales sujettes à l'hypoxie, l'abondance de certaines isoformes varie en fonction de l'état cellulaire : certains isomiRs devenant même plus abondant que le miR d'origine (ex. miR-30b-5p et miR-455-3p) **(Voellenkle 2012).** Dans un modèle simulant la progression des leucémies, Kuchenbauer *et al* montrent qu'au fur et à mesure que la maladie progresse, on assiste à la fois à des changements d'expression de plusieurs

microARNs et à celle des isomiRs qui leur sont associés **(Kuchenbauer 2008)**. Plus récemment, deux études ont montré des changements dans les profils d'expression de plusieurs isomiRs entre des tissus contrôles et des tissus tumoraux (mélanome et carcinome de la tyroïde) **(Kozubek 2013; Swierniak 2013)**. A cela s'ajoute une étude réalisée par Cloonan *et al.* qui montre que les isomiRs de manière globale, sont des molécules biologiquement actives qui coopèrent avec les miRs d'origine pour mieux contrôler l'expression de nombreux transcrits **(Cloonan 2011)**. Cette dernière étude est assez intéressante puisqu'elle met l'accent sur le rôle biologique des isomiRs de manière assez générale. Les isomiRs sont considérés comme des régulateurs fins de l'expression génique *(fine-tuning)* dont la fonction varierait en fonction des besoins de chaque type cellulaire. Pour le miR-3613-5p et ses isoformes, aucune conséquence fonctionnelle n'a encore été établie. Toutefois, ceci n'empêche pas qu'ils puissent coopérer entre eux et avec d'autres isomiRs spécifique aux cancers MSI. Ainsi, une conséquence fonctionnelle ou un rôle physiologique qui sont difficilement observables à l'échelle d'un isomiR, le seront probablement plus à l'échelle d'un réseau coopératif formé par plusieurs miRs matures, leurs isomiRs et les divers ARNm qu'ils ciblent.

Remarque sur le taux d'expression des isomiRs. Comme le montrent certaines publications, les isomiRs peuvent parfois être fortement exprimés **(Morin 2008; Chiang 2010)**. Par contre, ces isomiRs sont loin de constituer la majorité, et pour un certain nombre, dont les isomiRs 3613-5p, les niveaux d'expression dépassent à peine le bruit de fond expérimental. Alon *et al* se sont intéressés particulièrement à cette question dans un cadre différent. En étudiant l'*editing* des miARNs dans le cerveau, ils remarquent que beaucoup des évènements qu'ils répertorient sont à peine détectables. Ils constatent en même temps que chez la souris, les mêmes microARNs sont édités au même endroit et au même niveau, laissant suggérer que, même faiblement exprimées, ces modifications seraient biologiquement signifiantes (l'*editing* pouvant changer le répertoire de cibles d'un microARN) **(Alon 2012)**. En se basant sur cette étude et sur l'hypothèse de la coopération qui peut exister entre différents isomiRs, la faible expression d'un isomiR ne serait donc pas un argument permettant d'éliminer la possibilité qu'il puisse avoir une importance fonctionnelle.

Qu'en est-il des protéines de biogenèse des miRs ? Melo *et al.* ont rapporté la présence de séquences répétées dans les gènes de deux protéines impliquées dans la biogenèse des microARNs : *TARBP2* et *XPO-5* **(Melo 2009; Melo 2010)**. Selon ces auteurs, ces deux gènes sont assez fréquemment mutées à la fois dans les tumeurs coliques et gastriques de type MSI avec des effets sur la prolifération cellulaire, la tumorigenèse et la signature miARN de ces tumeurs. *XPO-5* est muté entre 22 et 26% des cas dans les tumeurs coliques MSI et dans 28% des cas dans les tumeurs gastriques MSI, alors que *TARBP2* est muté entre 25 et 43% dans les tumeurs coliques MSI et dans 14% les tumeurs gastriques MSI. Cependant, ces résultats sont loin de faire l'unanimité. Dans notre analyse effectuée sur 67 tumeurs primaires MSI et reprenant ces mêmes gènes, les fréquences mutationnelles de *TARBP2* et *XPO-5* étaient aux alentours de 5% pour les deux gènes. Nos résultats rejoignent ceux de Garre *et al.* qui, reprenant l'étude pour *TARBP2*, n'ont pas réussi à avoir le même taux de mutation (un seul CCR MSI muté sur 51 analysés, soit seulement 2%) **(Garre 2010)**. Kim el al., étudiant l'instabilité de plusieurs gènes impliqués dans le *processing* des miRs, rapportent eux aussi des fréquences assez faibles pour *XPO-5* (2% dans les tumeurs coliques) et *TARBP2* (4% dans les tumeurs gastriques) **(Kim 2010)**. Ces résultats remettent donc en question le fait que les mutations de *XPO-5* et *TARBP2* puissent avoir un rôle physiopathologique important dans la tumorigenèse de type MSI. Quand on regroupe les gènes cibles de l'instabilité en y incluant *XPO-5* et *TARBP2* (tenant compte des fréquences obtenues dans notre étude), ces deux gènes se retrouvent classés dans un groupe de gènes très rarement mutés, qui constituent le bruit de fond d'instabilité. D'un point de vue moléculaire, ces gènes ne constitueraient donc pas des « vraies cibles » MSI, étant donné que leurs mutations ne semblent pas être sélectionnées au cours de la progression tumorale MSI. Répondant à la lettre de Garre *et al.*, Melo *et al.* mettent l'accent sur la différence entre les techniques utilisées pour analyser les fréquences mutationnelles. Garre *et al.*, Kim *et al.*, et nous même utilisons une technique basée sur l'analyse de la taille de fragments de PCR par électrophorèse capillaire. Melo *et al.* jugent cette technique comme étant très peu sensible pour des délétions de 1 nucléotide, du fait de la contamination du matériel tumoral par du tissu sain et de l'hétérogénéité clonale au sein des tumeurs. Ils proposent le clonage et le séquençage pour chaque ADN tumoral de 12 produits PCR afin d'identifier les mutations **(Garre 2010)**. Cependant, cette technique a également ces limites. Les tumeurs MSI se caractérisent tout d'abord et principalement par leur phénotype hypermutateur. C'est

d'ailleurs cette propriété qui rend assez difficile l'identification des gènes cibles. Ainsi, par clonage des produits de PCR et séquençage des clones, toute répétition microsatellite ressort mutée dans une petite proportion de cellules tumorales, qu'elle soit codante ou non codante, et qu'elle ait un rôle ou pas dans la tumorigenèse MSI. Nous avons par ailleurs identifié plusieurs autres gènes impliqués dans la biogenèse des miRs, qui contiennent des microsatellites dans leur exons et qui pour certains peuvent être instables dans les cancers colorectaux de type MSI. Les fréquences restent toutefois assez faibles (entre 0% et 25,4%). Deux de ces gènes, *AGO2* et *TNRC6A*, ont été précédemment étudiés par Kim *et al.* qui rapportent des taux de mutations dans les cancers MSI du côlon et de l'estomac ne dépassant pas les 10% (environ les mêmes fréquences obtenues que dans notre étude) **(Kim 2010)**.

PERSPECTIVES

Surexpression du miR-155 et cancers colorectaux de type MSI

Un projet est en cours de réalisation dans le laboratoire d'accueil qui vise à développer une approche fonctionnelle visant à investiguer le rôle d'une surexpression du miR-155 dans l'initiation du processus tumorigénique MSI. A cet effet, le laboratoire d'accueil a récemment acquis une souris transgénique *miR-155* (construite par l'équipe de Juan Iovanna, Centre de Recherche en Cancérologie de Marseille). Cette souris (fond mixte C57B16/CD1/FVB) surexprime le miR-155 de manière inductible (transgène *BIC* sous promoteur inductible P-Tight), est viable et ne développe pas de tumeurs. L'utilisation de cette souris aura lieu dans deux contextes :

(1) Dans un contexte inflammatoire. Chez la souris, il a été mis en évidence qu'une exposition continue au DSS (Dextran Sodium Sulfate, produit chimique utilisé pour induire des colites chez les rongeurs) conduisait à des dysplasies et éventuellement à l'apparition de cancers **(Cooper 2000; Itzkowitz 2004)**. Notre but est donc d'investiguer le rôle d'une surexpression du miR-155 dans un contexte inflammatoire induit par le DSS. Plus particulièrement, il s'agit d'étudier l'apparition des cancers colorectaux et leur phénotype (MSI ou MSS) chez des souris traitées au DSS et surexprimant le miR-155, comparées à des souris traitées de la même façon mais ne surexprimant par le miR-155. Par ailleurs, l'équipe d'accueil a mis en évidence une déficience de la O^6-methylhuanine-DNA methyltransferase (MGMT) comme évènement pré-tumoral chez des patients atteints de CCRs MSI **(Svrcek 2010)**. Un défaut de champs MGMT a également été observé dans la muqueuse non transformée de patients atteints de MICIs. Il serait donc d'intérêt d'étudier l'effet synergique des défauts de champs miR-155 et MGMT dans un contexte proche de celui des MICI. Le laboratoire d'accueil possède à cet effet une lignée murine *Mgmtko* (fond pur FVB, construite par l'équipe japonaise du Dr M. Sekiguchi) qui pourrait être croisée avec les souris transgéniques *miR-155*. La souris *Mgmtko* est viable et ne développe pas de tumeurs.

(2) Dans un contexte de prédisposition génétique aux CCRs MSI (Syndrome de Lynch). Le Syndrome de Lynch est un syndrome de prédisposition aux cancers de type MSI, où les patients héritent d'un allèle muté et d'un allèle sauvage pour l'une des protéines du

système MMR. Selon le modèle de Knudson, un deuxième évènement est nécessaire pour déclencher le processus de transformation cellulaire. Ce deuxième *« hit »* peut être une mutation ponctuelle, une perte d'hétérozygotie, ou une hyperméthylation de l'un des promoteurs des gènes du système MMR. Notre hypothèse est qu'une surexpression du miR-155 pourrait constituer ce deuxième évènement. Les souris transgéniques miR-155 seront alors croisées avec des souris *MSH2wt*, *MSH2+/-* et *MSH2ko* disponibles au laboratoire et l'apparition de cancers MSI (lymphomes MSI) sera suivie pour chaque cohorte. Nous nous attendons à ce que les souris *MSH2+/-* surexprimant le miR-155 développent (1) un lymphome MSI et (2) à un âge plus jeune que celles ne surexprimant pas le miR-155. Par ailleurs, il est connu d'un point de vue clinique, que pour plusieurs patients atteints du Syndrome de Lynch ayant développé un CCR de type MSI, aucune mutation constitutionnelle n'a été retrouvée sur le gène correspondant à la protéine MMR perdue. Dans la continuité de ce projet, notre hypothèse est que le miR-155 pourrait dans une certaine mesure expliquer l'apparition de ces cancers. Les niveaux d'expression du miR-155 seront alors quantifiés par qRT-PCR sur une collection unique de patients opérés à l'hôpital Saint-Antoine et disponible au service d'anatomie pathologique de l'hôpital (M. Svrcek).

Mutations somatiques dans les gènes de miARNs, isomiRs et CCRs de type MSI

Comme nous l'avons vu, l'effet direct de l'instabilité microsatellitaire sur la biogenèse des miRs semble assez réduit, le miR-3613-5p étant jusqu'à présent la seule exception. Les résultats obtenus sur ce miR montrent cependant dans un sous-type particulier de CCRs – à savoir les cancers MSI – l'existence d'isomiRs qui ne sont pas retrouvé ailleurs. La question se pose de savoir si les isomiRs 3613-5p constituent un cas isolé, ou si au contraire, l'expression d'isomiRs dérégulés entre les cancers de type MSI et ceux de type MSS est un phénomène redondant. De manière intéressante, nos résultats suggèrent qu'un isomiR qui diffère par son extrémité 3', a les mêmes ARNm cibles que la séquence de référence mais possède également un petit groupe de cibles qui lui sont spécifiques. L'importance clinique et moléculaire de cette observation reste à déterminer, mais si les CCRs de type MSI se distinguent par un profil d'isomiRs, l'effet cumulé de ces isomiRs risque d'être important et

pourrait avoir des conséquences sur la tumorigenèse colique de type MSI et ses caractéristiques clinico-pathologiques. Malgré une séquence *« seed »* identique, la différence qui existe entre un isomiR et la séquence sauvage peut être due à plusieurs éléments. Certaines études montrent en effet que l'extrémité 3' et les séquences en dehors du *« seed »* jouent un rôle important dans la stabilité des miRs matures, leur *turnover*, voire même leur localisation cellulaire **(Hwang 2007; Bail 2010; Zhang 2011)**. Pour répondre à la problématique des isomiRs, le laboratoire d'accueil possède, en collaboration avec la Ligue Nationale Contre le Cancer (LNCC), une collection de tumeurs primaires MSI, MSS et de muqueuses saines qui ont été étudiées par séquençage profond de petits ARNs et dont l'analyse des données est en cours. Les techniques de nouvelle génération et le séquençage profond en particulier versent un nouveau jour sur l'étude des microARNs. Ils permettront à terme de répondre à ces questions.

Les microARNs : biomarqueurs des cancers colorectaux MSI ?

Le séquençage profond ne se limite pas à la détection des variations de séquences ou des mutations, il constitue en même temps un outil de quantification de l'ensemble des miARNs présents dans un échantillon. Les données obtenues permettent alors de construire des profils d'expression de miARNs capables de caractériser moléculairement différentes entités clinico-pathologiques. C'est en se basant sur ces données qu'il a été proposé d'utiliser les miARN comme bio-marqueurs des cancers colorectaux. Les différentes applications qui en découlent (diagnostique, pronostique et thérapeutique) permettraient éventuellement une meilleure prise en charge des patients **(Manne 2010; Vilar 2011; Hrasovec 2012)**. De récentes études, dont les principales ont été réalisées par, ou en conjonction avec le *« Cancer Genome Atlas Network »*, montrent cependant qu'une meilleure classification pourrait être obtenue en combinant différentes techniques de nouvelles génération (*whole-exome sequencing, DNA copy number analysis,* méthylation des promoteurs, étude des expressions des ARNm et des miARN, …) **(Cancer Genome Atlas Network 2012; Donehower 2013)**. C'est l'intégration de l'ensemble de ces données qui sera réalisée dans le laboratoire d'accueil en collaboration avec la LNCC dans le cadre du programme CIT (Carte d'Identité des

tumeurs). Ceci permettra de dégager une signature méta-analytique des CCRs MSI et MSS avec des applications potentielles en clinique.

Protéines régissant la biogenèse des miRs et MSI

Notre étude réalisée sur les protéines impliquées dans la biogenèse des miRs semble indiquer que l'existence d'un profil d'expression caractérisant les cancers MSI ne dépendrait pas d'une inactivation de ces protéines par le biais de MSI. *XPO-5* et *TARBP2* qui étaient rapportés comme étant mutées à des fréquences assez élevées sont ressortis comme étant très faiblement mutés dans notre criblage **(Melo 2009; Melo 2010)**. Se pose alors la question de savoir ce qui permet d'expliquer le profil miARN si particulier des CCRs de type MSI par rapport au CCRs MSS. Notre hypothèse est que cette signature est d'origine épigénétique, associée au phénotype CIMP *(CpG Island Methylator Phenotype).* Plus particulièrement, il s'agit de tirer profit des données de méthylome (programme CIT, LNCC) afin d'expliquer les différences d'expression des microARNs. De plus, la participation de processus épigénétiques conduisant à déréguler de manière coordonnée l'expression de gènes dans certaines régions du génome indépendamment de leur nombre de copie d'ADN a été particulièrement rapporté dans les cancers colorectaux de type MSI (mécanisme CIGD pour *« Coordinated copy-number Independent Gene Deregulation »*) **(Frigola 2006; Hitchins 2007)**. Notre hypothèse de travail est que les microARNs contenus dans de telles régions pourraient eux aussi voir leur expression dérégulée de manière coordonnée dans les CCR MSI.

CONCLUSION

Les objectifs de ce travail de thèse ont été d'étudier le rôle des microARNs dans la carcinogenèse colorectale de type MSI. Ces cancers se caractérisent par une déficience du système MMR qui conduit à l'accumulation de mutations dans les séquences d'ADN répétées du génome. Ce rôle a été investigué durant les deux étapes clés de la tumorigenèse MSI : (1) l'initiation ou la mise en place de la déficience MMR et (2) la progression tumorale et l'accumulation de mutations dans les gènes cibles de MSI. Nos résultats suggèrent dans un premier temps qu'un défaut de champs de certains miRs, notamment une surexpression du miR-155 qui cible les protéines du système MMR, pourraient constituer des évènements pré-tumoraux favorisant l'apparition d'un phénotype MSI, notamment dans le contexte de maladies inflammatoires chroniques intestinales. Dans un deuxième temps, la possibilité que les gènes de microARNs soient eux-mêmes ciblés par l'instabilité microsatellitaire a été étudiée. Notre criblage nous a permis d'identifier une trentaine de gènes contenant des séquences microsatellites. Parmi ces gènes, seule l'instabilité au niveau de *hsa-miR-3613* avait des conséquences sur l'expression du miR mature 3613-5p et sa biogenèse. Nos résultats nous ont permis d'identifier des isomiRs 3613-5p, spécifiquement dans un contexte MSI où *hsa-miR-3613* est muté. Ces isomiRs varient à leur extrémité 3' et auraient un répertoire de cibles qui pour certaines sont communes à la forme sauvage et pour d'autres spécifiques à chaque variant. Cette dualité rend difficile la mise en évidence et l'étude des conséquences fonctionnelles d'une telle mutation et des variations de séquences qu'elle engendre. Un rôle des isomiRs 3613-5p dans la tumorigenèse colique de type MSI ne serait donc probablement observé qu'en synergie avec d'autres isomiRs putatifs. Dans sa totalité, ce criblage montre qu'une dérégulation de l'expression des microARNs dans les cancers colorectaux instables sur les microsatellites ne peut pas être expliquée par des mutations somatiques dues à MSI dans les gènes et les séquences matures des microARNs. Dans cette problématique, un modèle récemment établi, propose que l'instabilité au niveau des gènes codants pour les protéines impliquées dans la biogenèse des miRs soit à l'origine d'une telle dérégulation. Notre criblage réalisé sur une dizaine de gènes, incluant *XPO-5*, *TARBP2*, *AGO-2* et *TNRC6A* montre cependant que ces gènes sont rarement mutés dans les lignées cellulaires et tumeurs primaires MSI et que ces mutations ne constitueraient donc pas des pistes fonctionnelles à retenir. L'origine de la dérégulation de l'expression des miRs dans ce type de cancers reste donc dans une grande mesure inexpliquée. La continuité de ce travail, par l'utilisation des techniques de nouvelles génération et l'étude des profils de méthylation,

caractérisant ces cancers, fera certainement la lumière sur cette question, avec des implications à la fois moléculaires (compréhension des mécanismes d'initiation et de progression tumorale et intégration des différents processus génétiques et épigénétiques) et cliniques (diagnostic, pronostic et thérapeutique).

ANNEXE

Article 5

Behavior of a hammerhead ribozyme in aqueous solution at medium to high temperatures

El-Murr N, Maurel MC, Rihova M, Vergne J, Hervé G, Kato M, Kawamura K. (2012)
Naturwissenschaften 99 (9): 731-738.

Travail réalisé au cours du M2 Biologie Moléculaire et Cellulaire – Spécialité Génétique

Osaka Prefecture University, Osaka, Japon
Université Pierre et Marie Curie, Paris, France

Naturwissenschaften (2012) 99:731–738
DOI 10.1007/s00114-012-0954-9

ORIGINAL PAPER

Behavior of a hammerhead ribozyme in aqueous solution at medium to high temperatures

Nizar El-Murr · Marie-Christine Maurel · Martina Rihova · Jacques Vergne · Guy Hervé · Mikio Kato · Kunio Kawamura

Received: 13 June 2012 / Revised: 17 July 2012 / Accepted: 19 July 2012 / Published online: 23 August 2012
© Springer-Verlag 2012

Abstract The "RNA world" hypothesis proposes that—early in the evolution of life—RNA molecules played important roles both in information storage and in enzymatic functions. However, this hypothesis seems to be inconsistent with the concept that life may have emerged under hydrothermal conditions since RNA molecules are considered to be labile under such extreme conditions. Presently, the possibility that the last common ancestor of the present organisms was a hyperthermophilic organism which is important to support the hypothesis of the hydrothermal origin of life has been subject of strong discussions. Consequently, it is of importance to study the behavior of RNA molecules under hydrothermal conditions from the viewpoints of stability, catalytic functions, and storage of genetic information of RNA molecules and determination of the upper limit of temperature where life could have emerged. In the present work, self-cleavage of a natural hammerhead ribozyme was examined at temperatures 10–200 °C. Self-cleavage was investigated in the presence of Mg^{2+}, which facilitates and accelerates this reaction. Self-cleavage of the hammerhead ribozyme was clearly observed at temperatures up to 60 °C, but at higher temperatures self-cleavage occurs together with hydrolysis and with increasing temperature hydrolysis becomes dominant. The influence of the amount of Mg^{2+} on the reaction rate was also investigated. In addition, we discovered that the reaction proceeds in the presence of high concentrations of monovalent cations (Na^+ or K^+), although very slowly. Furthermore, at high temperatures (above 60 °C), monovalent cations protect the ribozyme against degradation.

Keywords Origin of life · RNA world · Hammerhead ribozyme · Hydrothermal environments · Avocado viroid · Hydrothermal reaction · Prebiotic earth · Hydrothermal flow reactor system

Communicated by: Sven Thatje

Electronic supplementary material The online version of this article (doi:10.1007/s00114-012-0954-9) contains supplementary material, which is available to authorized users.

N. El-Murr · M.-C. Maurel · M. Rihova · J. Vergne
ER12, ANBioPhy, Fonctions et Interactions des Acides Nucléiques, UPMC Univ Paris 6,
75005 Paris, France

G. Hervé
ER3, Laboratoire BIOSIPE, UPMC Univ Paris 6,
75005 Paris, France

M. Kato
Department of Biological Science, Graduate School of Science,
Osaka Prefecture University,
Gakuen-cho 1-1, Naka-ku, Sakai,
Osaka 599-8531, Japan

N. El-Murr · K. Kawamura (✉)
Department of Applied Chemistry,
Graduate School of Engineering, Osaka Prefecture University,
Gakuen-cho 1-1, Naka-ku, Sakai,
Osaka 599-8531, Japan
e-mail: kawamura@shudo-u.ac.jp

Present Address:
K. Kawamura
Department of Human Environmental Studies,
Hiroshima Shudo University,
1-1-1 Ozuka-higashi, Asaminami-ku,
Hiroshima 731-3195, Japan

Introduction

The RNA world hypothesis proposes that a pre-cellular world based on RNA, consisting of a self-replicating system endowed with catalytic properties, preceded our modern DNA–RNA–Protein World (Cech 1986; Gilbert 1986; Orgel 2004). However, one of the problems of this hypothesis is

the fact that the RNA lability seems not to be consistent with the hypothesis of the hydrothermal origin of life, which is supported by numerous studies on thermophilic organisms (Corliss et al. 1981; Nisbet 1986). On the basis of evolutionary evidence, it has been proposed that the last universal common ancestor (LUCA) of the present organisms could be close to hyperthermophiles (Nisbet 1986; Pace 1991). The formation of genetic code would have taken place at high tempeatures for its late stage (Di Giulio 2000, 2005). However, this conclusion has been strongly disputed so that the nature of LUCA is not yet unequivocal (Forterre 1996; Galtier et al. 1999; Boussau et al. 2008). The experimental simulation of primitive earth conditions also supports the idea that hydrothermal vent environments could have played important roles in the chemical evolution of peptide-like molecules. So, coupling these two hypotheses is still problematic owing to the RNA thermolability (White 1984; Miller and LAzcano 1995; Lazcano and Miller 1996; Kawamura 2004), the difficulty of prebiotic formation of RNA at high temperatures (Kawamura and Umehara 2001; Kawamura and Maeda 2007, 2008), and to the reduction of biologically important weak interactions under these conditions (Kawamura 2005; Kawamura and Nagayoshi 2007). Experimental verification and accumulation of knowledge regarding the thermal stability, interactions, and prebiotic functions of RNA molecules under such extreme conditions are thus important to evaluate whether the hydrothermal origin-of-life hypothesis and the RNA world hypothesis are compatible. The structural, informational, and catalytic diversity added to the extreme sequence adaptability of RNA molecules reveal their fundamental importance and emphasize their primary role in the origins of life. The concept of a pre-cellular RNA world is built on this interesting diversity.

Our previous studies using the hydrothermal flow reactor system (HFRS) (Kawamura 1999, 2000) suggested that short and simple RNA molecules survive a few hours at 80 °C, but survive merely a few seconds at 200 °C in the presence of NaCl and $MgCl_2$ (Kawamura 2003). This does not straightforwardly support the revoking of the hydrothermal origin of life hypothesis (Kawamura 2004), but draws attention to the fact that RNA molecules are not as temperature sensitive as was expected, at least in the medium temperature range. However, the RNA molecules studied were obviously functionally simple, possessing non-biologically important and complex secondary structures with exception of the hairpin structure. At the same time, however, these data indicate that the formation of secondary structures enhances the stability of the RNA oligomer at 65–80 °C and increases the activation enthalpy change at 150–200 °C even for relatively simple RNA molecules being 4- to 23-mers in length, whether or not they possess hairpin structures (Kawamura 2003). Consequently, in order to verify the compatibility of the RNA world hypothesis and that of the hydrothermal origin of life, the thermal stability and cleavage behavior of a functional RNA molecule should be evaluated under simulated primitive earth conditions. For this purpose, we used a catalytic RNA, the HHR:ASBVd (−), a 79-nucleotide-long hammerhead ribozyme deduced from the minus strand of the Avocado Sunblotch Viroid (ASBVd; Fig. 1) that catalyzes the self-cleavage of RNA molecules in which it is embedded (Cochrane and Strobel 2008; Delan-Forino et al. 2011). This ribozyme is obviously long as compared to our previous model RNA oligomers and forms a three-dimensional structure. The fact that ribozymes are usually found in viroids or virus-like pathogens suggests a very early event of evolution (Diener 2001; Daròs et al. 2006; Góra-Sochacka 2004) and explains the reason why they were chosen as models for the present investigation. In the present study, the stability of the ribozyme was studied at medium and high temperatures using conventional and HFRS techniques, providing some insight into the possible hydrothermal behavior of prebiotic RNA molecules in the frame of the RNA world hypothesis.

Materials and methods

Preparation of HHR:ASBVd(−)

Studies on the self-cleavage and degradation were carried out under hydrothermal conditions, using the HHR:ASBVd(−) prepared by in vitro transcription using a TranscriptAid™ T7 High Yield Transcription Kit (Fermentas™). In vitro transcription was carried out through the double-helical DNA with T7 RNA polymerase. The transcription mixture contained the TranscriptAid™ Enzyme Mix, rNTP and of synthetic DNA produced by PCR. After incubation at 37 °C for 7–8 h, the reaction was stopped by addition of DNase I. Proteins were removed with an equal volume of the mixture phenol/chloroform/isoamyl alcohol (25:24:1), and then twice with an equal volume of chloroform/isoamyl alcohol (24:1). After ethanol precipitation, the RNA was purified by electrophoresis on a

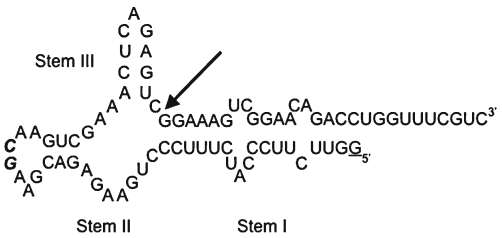


Fig. 1 Secondary structure representation of the hammerhead ribozyme HHR:ASBVd(−) designed from the minus strand of ASBVd. The *arrow* points to the scissile bond where cleavage occurs. A phosphodiester bond (*italic bold*) is artificially introduced to connect both strands, and a 5′ terminal G (*underlined*) is added for efficient transcription

10 % denaturing polyacrylamide gel [5 % bisacrylamide and 95 % acrylamide, Tris–borate–EDTA buffer pH 8.3 and 7 M urea; 10 % denaturing polyacrylamide gel electrophoresis (PAGE)].

The band corresponding to the non-cleaved ribozyme was localized by UV shadowing, cut out, crushed, and the RNA eluted from the gel in a 0.3 M sodium acetate solution for 15 min. The gel was then totally eliminated by double filtration on 0.45 and 0.20 μm microfilters. Finally, the RNA was ethanol precipitated, washed with 70 % ethanol, vacuum dried, dissolved in water, and quantified by measuring its absorbance at 260 nm.

Cleavage of HHR:ASBVd(−)

The cleavage of HHR:ASBVd(−) was followed on a 10 % denaturing PAGE, and its structural integrity studied using the HFRS. The catalytic activity was analyzed under different physicochemical conditions: Mg^{2+} concentration (0–50 mM), temperature (10–80 °C), and pH (7–9). Prior to the reaction the RNA was denatured in water for 1 min at 95 °C and slowly cooled to 25 °C (3 °C/min). Buffer was then added, and the cleavage reaction was initiated by addition of $MgCl_2$ at the chosen temperature. All cleavage reactions were stopped after 2 h by addition of equal volume of stop buffer consisting of 50 mM EDTA, 80 % formamide, and 0.5 % bromophenol blue, and by cooling on ice. Substrate and digestion products were separated on 10 % denaturing PAGE. The gel was then stained with ethidium bromide and visualized under UV light. The band density of the PAGE was analyzed using the ImageJ software. Cleavage yield is presented as the fraction of the substrate converted to product as shown by Eq. 1,

$$\text{Cleavage yield} = 3)\text{density } band\ 2 + \text{density } band \quad / \quad \sum(\text{density } bands 1, 2, 3) \tag{1}$$

where *band 1* corresponds to non-cleaved ribozyme (79 nt) and *band 2* and *3* to the cleavage products (51 and 28 nt, respectively). In addition, to clarify the effect of monovalent cations, self-cleavage of the hammerhead ribozyme was attempted at 60–70 °C in the presence of 50 mM $MgCl_2$ with addition of 4 M NaCl or 4 M KCl.

Time-dependent activity of the hammerhead ribozyme was monitored in the presence of 50 mM Tris, pH 7.5, and 50 mM $MgCl_2$ or 4 M NaCl or 4 M KCl. Aliquots were removed at different times and quenched with an equal volume of stop buffer. To compare the reaction rates, we calculated first-order rate constants k_{obs} from data fitted to the single-exponential Eq. 2,

$$A_t = A_0 e^{-k_{obs}t} \tag{2}$$

where A_t and A_0 are the fraction of the hammerhead ribozyme at time t and initial, respectively.

Hydrothermal flow reactor system and hydrothermal degradation of HHR:ASBVd(−)

The stability of HHR:ASBVd(−) was investigated using the hydrothermal flow reactor system (HFRS) (Kawamura 2000). It allows monitoring the hydrothermal degradation of biological macromolecules at temperatures up to 400 °C within 0.002–200 s at pressures up to 30 MPa. The device consists of a high-pressure HPLC pump (model PU-980, JASCO Corporation, Japan), a high-temperature reactor, a temperature controller, a pressure regulator, a sample injector, and a 1,000-mL double-distilled water reservoir. The sample, once injected, travels through a hydrothermal reactor in a stainless steel reaction tubing. After leaving the heated reactor, the sample is cooled and finally collected. The pressure in the reactor was held at 10 MPa to keep the water in the reactor. The residence time of the sample inside the reactor was controlled by changing the flow rate of the pump.

For each reaction, 10 μg of HHR:ASBVd(−) in 20 μL was used. The standard "no salt" reactions were carried out in 150 mM KCl and 50 mM Tris; reactions in NaCl and KCl were carried out in 50 mM Tris and 2 or 4 M NaCl or KCl. As in all previous experiments, the RNA was denatured in water for 1 min at 95 °C and slowly cooled to 25 °C. The temperature of the reactor was fixed at 80 °C, and a flow rate of 0.2 mL/min that allows 35 s of residence time. The sample was then analyzed by analytical high-performance liquid chromatography (HPLC) LC10 A (Shimadzu, Japan) with a DNA-NPR anion-exchange column from TOSOH Co., Japan using a gradient of 0.3–1.2 M NaCl at pH 9 with 0.02 M Tris buffer to determine the extent of RNA degradation and the compensatory effect of salt.

Thermal denaturation of HHR:ASBVd(−) followed by circular dichroism

To evaluate the structure of the ribozyme at medium temperatures below 100 °C, thermal melting of the ribozyme was monitored by circular dichroism, using a Jasco Model J-820 spectropolarimeter with a Jasco PTC-423L Peltier thermoelectric heater/cooler. Prior to structural analysis, HHR:ASBVd(−) was heated to 90 °C for 2 min and slowly cooled to adopt its three-dimensional structure. The concentration of RNA solutions was fixed at 360 ng μL^{-1}, dissolved in 50 mM Tris buffer containing 150 mM KCl as control and in a buffer containing 4 M NaCl. A cell length of 1 mm was used to limit absorbance. Wavelength was varied between 210 and 310 nm. The temperature dependence of the ribozyme structure was monitored by CD at a maximum of 272 nm with temperature increase rate of 1 °C/min.

Results

According to our previous studies, the hydrolytic degradation of polynucleotides is enhanced 100-fold by 0.1 M $MgCl_2$ (Kawamura et al. 1999). Mg^{2+} is necessary for the enzymatic cleavage of RNA molecules, and self-cleavage is accelerated with increasing the Mg^{2+} concentration. In addition, it appeared interesting to test substitutes of Mg^{2+} (Curtis and Bartel 2001), and monovalent cations were tested. High salt concentrations near hydrothermal springs, mineral deposits encapsulating and concentrating small abiogenically formed organic molecules can somehow exhibit a diminished water effect (Russell et al. 1988; Russell and Hall 1997; Dundas 1998).

The importance of Mg^{2+} is shown in Figs. 2 and 3 and in the Electronic supplementary materials (ESM) where the kinetics of self-cleavage is stimulated in the presence of $MgCl_2$ (Fig. 2), and the formation of self-cleaved products as a function of time is shown in the presence of different cations (Fig. 3). Self-cleavage proceeds in the presence of high concentrations of NaCl or KCl even without $MgCl_2$, although with less efficient rate constants in 4 M NaCl (k_{obs} 0.0020±0.0001 h^{-1}) and in 4 M KCl (k_{obs} 0.0063±0.0001 h^{-1}) compared to (k_{obs} 0.86± 0.03 h^{-1}) in 50 mM Mg^{2+}.

The pH of the solutions in the present study was generally adjusted at pH 7.5 to be near physiological conditions. Additionally, the influence of pH was investigated in the range 7–9 since the hammerhead ribozyme does not show activity below pH 7 (Kaddour et al. 2011). The pH is adjusted by taking into account the pK_a variation as a function of temperature, where the ΔpK_a/°C of HEPES is =−0.014. The overall disappearance of ribozyme and the percentage of cleavage were measured (Fig. 4). The amount of remaining ribozyme in the solution was decreasing with increasing pH, and the contribution of cleavage to the disappearance also decreased with pH.

In parallel, quantitative evolution of ribozyme catalysis was studied at different temperatures (Fig. 5). Increased cleavage activity was observed up to 60 °C. Above 60 °C,

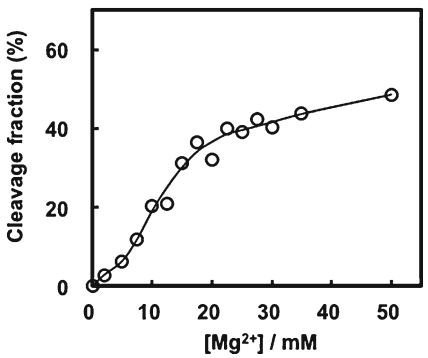


Fig. 2 Kinetics of in vitro self-cleavage of HHR:ASBVd(−) depending on Mg^{2+} concentration. HHR:ASBVd(−) was incubated for 2 h at increasing concentrations of MgCl2 (0–50 mM). The reaction was carried out in 150 mM KCl and 50 mM HEPES (pH 7.5) at 45 °C

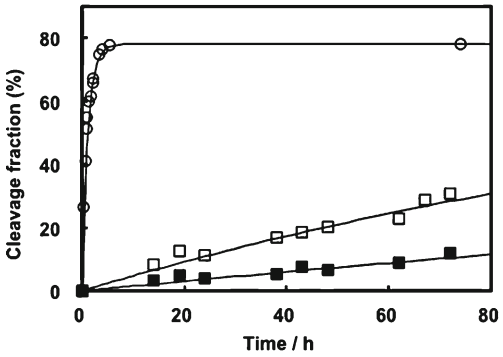


Fig. 3 Kinetics of in vitro self-cleavage of HHR:ASBVd(−) in the presence of monovalent (Na^+ and K^+) and divalent (Mg^{2+}) cations. HHR:ASBVd(−) was incubated for 72 h at 45 °C, in 50 mM Tris, 25 mM EDTA, pH 7.5, 4 M NaCl (*open squares*) or 4 M KCl (*closed squares*) for the monovalent salt experiments, and in 50 mM Tris pH 7.5, 150 mM KCl and 50 mM $MgCl_2$ (*open circles*) for the divalent salt experiment. Samples were taken at the indicated times and quenched. Data were fitted by Eq. 2, and first-order rate constants k_{obs} were calculated: 50 mM Mg^{2+} (k_{obs} 0.86±0.03 h^{-1}); 4 M K^+ (k_{obs} 0.0063±0.0001 h^{-1}); and 4 M Na+(k_{obs} 0.0020±0.0001 h^{-1})

hardly any traces of both uncleaved ribozyme and cleavage products were detected since nonspecific RNA degradation becomes predominant. These two reasons behind the decrease of the intact HHR:ASBVd(−) can be illustrated by the following Eqs. 3-1 and 3-2,

$$\begin{aligned}\text{HHR}:\text{ASBVd}(-) &\rightarrow \text{Product 1}(5'-\text{end 51 nt})\\ &\quad + \text{Product 2}(3'-\text{end 28 nt})\\ &\quad (3-1)(\text{self}-\text{cleavage})\end{aligned}$$

$$\begin{aligned}\text{HHR}:\text{ASBVd}(-)&\\ &\rightarrow \text{mono, di, tri, }\cdots\text{ short oligonucleotides}\\ &\quad (3-2)(\text{nonspecific degradation})\end{aligned}$$

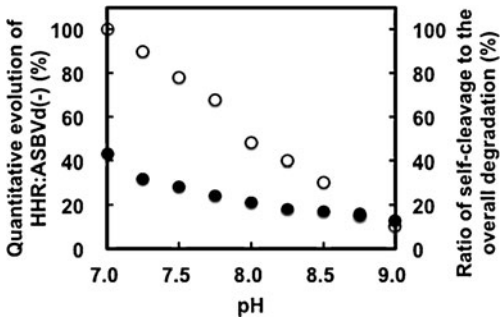


Fig. 4 pH effect on the in vitro self-cleavage of HHR:ASBCd(−). HHR:ASBVd(−) (0.5 μg) (*CfRNA*=0.033 μg μL^{-1}) was cleaved for 2 h at increasing pH (7–9). The reaction was carried out in 150 mM KCl, 50 mM Tris, and 50 mM Mg^{2+} at 45 °C. *Open circles*: a plot showing the quantitative evolution of the remaining ribozyme. The intensity of the band at pH 7.0 was arbitrarily assigned as 100 %. *Closed circles*: the ratio of self-cleavage to the overall disappearance of the ribozyme

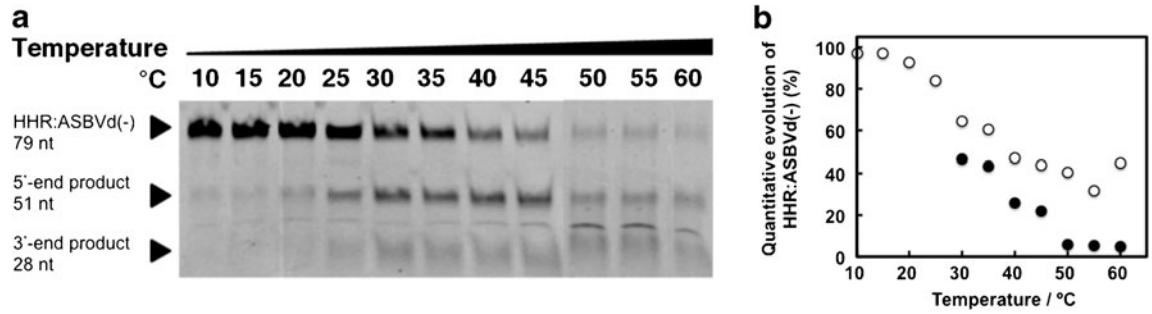


Fig. 5 Temperature effect on the in vitro self-cleavage of HHR:ASBVd(−). HHR:ASBVd(−) (0.5 μg) was cleaved for 2 h at increasing temperatures (10–80 °C). The reaction was carried out in 150 mM KCl, 50 mM HEPES (pH fixed at 7.5 for each desired temperature), and 50 mM Mg^{2+}. The data at 65–80 °C are not shown since all the ribozyme was destroyed. **a** 10 % denaturing PAGE. **b** Plot showing the quantitative evolution of the HHR:ASBVd(−) based on: (1) the self-cleavage activity alone (*open circles*) and (2) on both self-cleavage and nonspecific degradation (*closed circles*). The intensity of the band at 10 °C being was arbitrarily assigned as 100 %

the upper part corresponding to specific cleavage at the scissile bond (Fig. 1) and the lower part to nonspecific degradation due to many external factors, such as high temperature. The ratio of the self-cleavage activity alone and the overall degradation of the ribozyme including self-cleavage and nonspecific degradation of the ribozyme at 10–60 °C is shown in Fig. 5b. This indicates that self-cleavage not only intensifies with increasing temperature but also the ratio of the nonspecific degradation to self-cleavage rises with temperature.

Therefore, we performed experiment at higher temperatures over 60 °C focusing on ribozyme stability and the influence of monocations. HHR:ASBVd(−) was incubated for 2 h with or without 4 M monovalent salts (NaCl or KCl) at three different temperatures (60, 65, and 70 °C). In Fig. 6, we can observe the amount of remaining ribozyme after self-cleavage and nonspecific degradation. These results indicate that degradation of the HHR:ASBVd(−) ribozyme can be reduced by high concentrations of monocations.

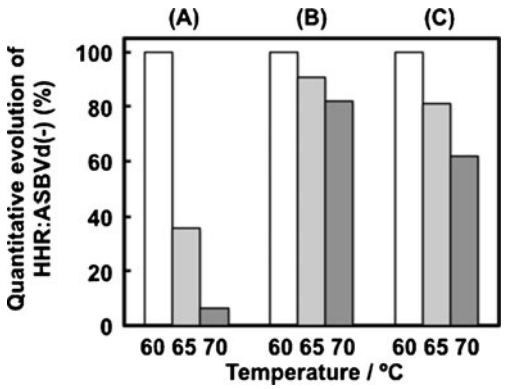


Fig. 6 In vitro activity and structural integrity of HHR:ASBVd(−) at high temperature. HHR:ASBVd(−) (0.5 μg) was cleaved for 2 h at increasing temperatures (60, 65, and 70 °C). **a** The "no salt" reactions were carried out in 150 mM KCl, 50 mM Tris pH 7.5, and 50 mM Mg^{2+}; **b** the "4 M NaCl" and **c** "4 M KCl" reactions were performed in 50 mM Tris pH 7.5, 50 mM Mg^{2+}, and 4 M NaCl or KCl. *Bars* involve both self-cleavage and nonspecific degradation. The intensity of the band at 60 °C was arbitrarily assigned as 100 %

Consequently, we have attempted to see an enhancement of the stability of the ribozyme in the presence of high concentration of monovalent cations using HFRS to monitor the self-cleavage. Ribozyme samples were incubated for 38 s at 80 °C in the HFRS and were analyzed by HPLC. The results (Fig. 7) suggest that in the absence of salt, degradation products, in the form of monomers, dimers, and trimers immediately appear after heating (a of Fig. 7a, b). Initial non-degraded and partially degraded ribozymes are still detectable despite being very sparse. Incubation of HHR:ASBVd(−) with 2 or 4 M salt (NaCl/KCl) readily changes the degradation profile: non-degraded and partially degraded ribozymes populations are preserved, and degradation products are fairly reduced (with both NaCl and KCl).

The potential protective effect of monovalent cations on HHR:ASBVd(−), at much higher temperatures in the presence of 4 M NaCl, was investigated using HFRS (Fig. 7d). Control sample with no hydrothermal incubation exhibits a large peak appearing at 37.9 min—the intact ribozyme—surrounded by two other small peaks (Fig. 7c). Simple comparison of the results suggests that the exposure of the ribozyme to 150–200 °C caused its rapid and total destruction.

Experiments using circular dichroism to study the structure as a function of temperature showed that the magnitude of CD decreases with increasing temperature almost proportionally (Fig. 8). The CD spectra in the absence and in the presence of NaCl are shown in the ESM. The CD spectra for both the experiments with and without NaCl behave similarly with increasing temperature. The temperature dependence of CD signal at 272 nm decreased with increasing temperature for both cases (Fig. 8), and no clear melting points were observed for either of them.

Discussion

The present study shows that catalysis involving the HHR:ASBVd(−) hammerhead ribozyme could occur in high salt

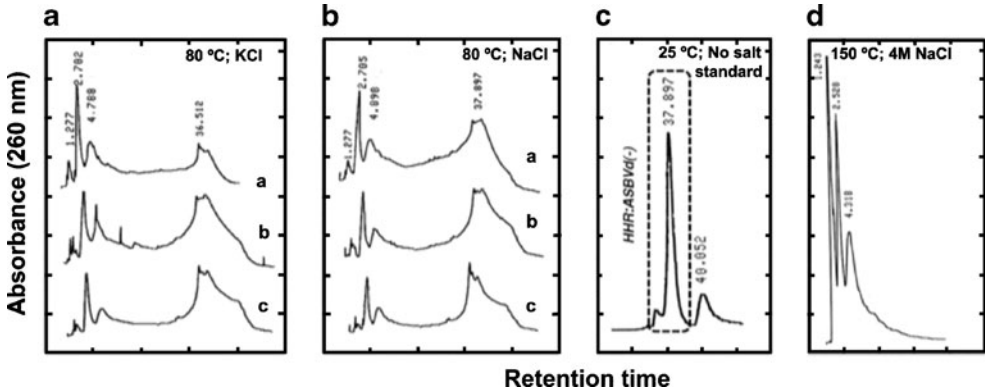


Fig. 7 Analytical HPLC trace displaying the hydrothermal degradation of HHR:ASBVd(−). **a** Effect of KCl. **b** Effect of NaCl. 5 μg of HHR:ASBVd(−) (RNA=1 μg μL^{-1}) was incubated for 38 s at 80 °C in the hydrothermal flow reactor system (HFRS). The RNA was then analyzed by HPLC using a DNA-NPR anion exchange column. The standard "no salt" reactions (**a**) were carried out in 50 mM Tris pH 7.5; the "2 M salt" reactions (**b**) in 50 mM Tris pH 7.5 and 2 M NaCl or 2 M KCl, and the "4 M salt" reactions (**c**) in 50 mM Tris pH 7.5 and 4 M NaCl or 4 M KCl. **c** HPLC control trace of HHR:ASBVd(−) with no hydrothermal incubation. **d** Unique HPLC trace obtained at 150 °C with 4 M NaCl. Chromatograms are not to scale

environments and is supported by increasing temperatures up to 60 °C. Preservation of the ribozyme in the presence of high monovalent cation concentrations (Fig. 6) suggests that the hammerhead ribozyme might be stabilized by high salt concentrations against heat degradation, although the salt effect was not clearly observed at temperatures exceeding 70 °C.

Our data suggest that the overall requirement for Mg^{2+} primarily reflects structural contributions to catalysis of HHR:ASBVd(−) hammerhead cleavage reaction. In natural environments where Mg^{2+} is scarce, catalytic RNAs should be able to function with monovalent cations (Russell et al. 1988; Russell and Hall 1997; Dundas 1998). Yet, because catalytic RNAs are extremely thermolabile, the development of an RNA world in thermophilic environments is quite difficult unless specific mechanisms are involved to stabilize the RNA under these conditions. Interestingly, monovalent salts (Na^+ or K^+) can be suitable for this role as well as was already shown with DNA (Marguet and Forterre 1994) and tRNA (Tehei et al. 2002). Thus, the influence of monovalent cations demonstrated in the present study supports the assumption that they are plausible factors to preserve some biological structures at high temperatures.

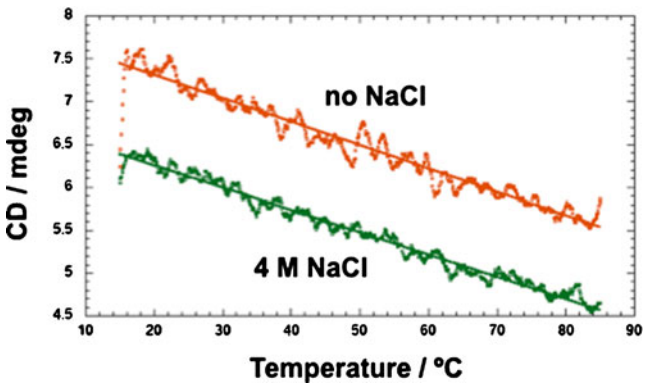


Fig. 8 Thermal melting study by CD (circular dichroism). Effect of temperature (15 to 85 °C) on the CD signal. The signal at 272 nm was measured as a function of temperature for sample without additive (*top curve*) and with 4 M NaCl (*bottom curve*)

Furthermore, our results suggest that monovalent salts prevent degradation essentially by enhancing tertiary contacts and rigidifying structures. This would lead to reduced nonspecific 2′-OH nucleophile attacks on the ribose–phosphate backbone of the RNA (Peracchi 1999). Although this might inactivate specific cleavage of the ribozyme, no such pronounced effect was observed (Fig. 6 and ESM). The CD measurements reveal that the spectral changes for ribozyme samples with and without NaCl are very similar, an observation which would indicate that the conformation of the ribozyme is both the same in these two cases (ESM). This suggests that the influence of NaCl with increasing temperatures would not be due to the conformational change but might be credited to the inhibition of nonspecific self-cleavage at the target phosphodiester bond of the ribozyme in the presence of high concentration of monocations. In addition, the temperature dependence of CD signal did not show clear melting point for the ribozyme, which is unusual for this type of molecule (Jonson and Söll 1971). This observation suggests the following characteristics of the ribozyme. First, a transition of three-dimensional structure does not occur at a particular melting temperature, and the ribozyme is gradually denatured at 10–80 °C. Second, the three-dimensional structure is poorly dependent on the presence of NaCl as the same three-dimensional structure is preserved in the presence or absence of NaCl during the self-cleavage. Presumably, the structure is held by Na^+ ions instead of Mg^{2+} ions. This is consistent with the temperature dependence of the self-cleavage kinetics as shown in

Fig. 5b, which express similar attributes to that of the CD signal shown in Fig. 8. Since the decrease in amount of HHR:ASBVd(−) can have two origins, (1) catalysis and (2) nonspecific degradation by temperature, it was interesting to inspect and compare the effect of temperature on both (Fig. 5b). The ratio of nonspecific degradation to self-cleavage increased with temperature.

The effect of pH on cleavage is shown in ESM; it indicates that the overall rate of ribozyme disappearance increases with increasing pH. This is consistent with the fact that the cleavage rate of phosphodiester bond for simple RNA molecules (Järvinen et al. 1991) increases with increasing pH in the pH range studied; consistent with the proposal that the cleavage is promoted by acid/base catalysis. Furthermore, it is known that the clevage rate constant of some hammerhead ribozymes (Zhuo et al. 1996; Feig et al. 1998; Murray et al. 2002) is proportional to concentration of OH^- at the pH range. This fact suggests that the remaining ribozyme at pH 9 would be 1/100 of that at pH 7 if the clevage follows the simple alkaline hydrolysis. However, the density for product 1 and product 2 on gel electrophoresis seems not much changed between pH 7 and 8. This might be due to minor changes in 3D structure induced by increasing alkaline environment, which may cause deprotonation of a number of bases and phosphate groups resulting in slow gradual changes of the whole 3D structure. The hydrolytic cleavage percentage decreases with increasing pH which also supports our assumption. The pH dependence study is very unique especially in the case of functional RNA, and not many measurements were carried out for the ribozyme (Kaddour et al. 2011).

The reaction products obtained at high temperatures, identified by HFRS, reflect the nonspecific degradation of the ribozyme. Interestingly, HHR:ASBVd(−) incubated with or without salt showed no resistance to temperature degradation above 100 °C. Identical HPLC chromatograms were obtained with RNA incubated at 150 and 200 °C without salt and with 4 M NaCl (Fig. 7d). This result indicates that nonspecific degradation proceeds much more rapidly beyond the time window of the HFRS. It appears that small RNAs and especially ribozymes show a pronounced plasticity and adaptability to various and extreme conditions (Ztouti et al. 2009; Chval et al. 2011).

Conclusively, the ribozyme survives well at temperatures below 60 °C, but it is degraded immediately at high temperatures around 150–200 °C. Thus, the ribozyme is obviously labile at temperatures near submarine hydrothermal vent systems.

At first sight, one can imagine that the temperature susceptibility of RNA molecules might rule out the hypothesis of the hydrothermal origin of life: an RNA world cannot thrive at very high temperatures. Less extreme environments are however known and, for the first time, the stability of a self-cleaving RNA molecule is established in hydrothermal salt conditions. It was found that self-cleavage of the hammerhead ribozyme proceeds at temperatures up to 70 °C with 4 M NaCl or KCl. This strongly confirms the fact that salt is essential to maintain the structural and functional integrity of RNA in hydrothermal conditions, an integrity that is mandatory for the establishment of an RNA world.

Acknowledgments Nizar El-Murr was the recipient of the Jasso Scholarship for short-term studies in Japan. This research was supported by 21200004 on Innovative Areas from MEXT in Japan, a JSPS Fellowship for Marie-Christine Maurel research in Japan (S10186), and a 20540476 Grant-in-Aid for Scientific Research from JSPS in Japan. We are grateful for EPOV program of CNRS and CNES grants for supporting this work. We also thank Dr. Anne-Lise Haenni for English improvements.

References

Boussau B, Blanquart S, Necsulea A, Lartillot N, Gouy M (2008) Parallel adaptations to high temperatures in the Archaean eon. Nature 456:942–945

Cech TR (1986) A model for the RNA-catalyzed replication of RNA. Proc Natl Acad Sci USA 83:4360–4363

Chval Z, Chvalova D, Leclerc F (2011) Modeling the RNA 2′OH activation: possible roles of metal ion and nucleobase as catalysts in self-cleaving ribozymes. J Phys Chem B 115:10943–10956

Cochrane JC, Strobel SA (2008) Catalytic strategies of self-cleaving ribozymes. Acc Chem Res 41:1027–1035

Corliss JB, Baross JA, Hoffman SE (1981) An hypothesis concerning the relationship between submarine hot springs and the origin of life on Earth. Oceanol Acta (Suppl) 4:59–69

Curtis EA, Bartel DP (2001) The hammerhead cleavage reaction in monovalent cations. RNA 7:546–552

Daròs JA, Elena SF, Flores R (2006) Viroids: an Ariadne's thread into the RNA labyrinth. EMBO Rep 7:593–598

Delan-Forino C, Maurel MC, Torchet C (2011) Replication of Avocado Sunblotch Viroid in the yeast *Saccharomyces cerevisiae*. J Virology 85:3229–3238

Di Giulio M (2000) The late stage of genetic code structuring took place at a high temperature. Gene 261:189–195

Di Giulio M (2005) The ocean abysses witnessedthe origine of the genetic code. Gene 346:7–12

Diener TO (2001) The viroid: biological oddity or evolutionary fossil? Adv Virus Res 57:137–184

Dundas I (1998) Was the environment for primordial life hypersaline? Extremophiles 2:375–377

Feig AL, Ammons GE, Uhlenbeck OC (1998) Cryoenzymology of the hammerhead ribozyme. RNA 4:1251–1258

Forterre P (1996) A hot topic: the origin of hyperthermophiles. Cell 85:789–792

Galtier N, Tourasse N, Gouy M (1999) A nonhyperthermophilic common ancestor to extant life forms. Science 283:220–221

Gilbert W (1986) Origin of life: the RNA world. Nature 319:618–618

Góra-Sochacka A (2004) Viroids: unusual small pathogenic RNAs. Acta Biochim Pol 51:587–607

Järvinen P, Oivane M, Lönnberg H (1991) Interconversion and phosphoester hydrolysis of 2′,5′- and 3′,5′-dinucleotide monophosphate: kinetics and mechanisms. J Org Chem 56:5396–5401

Jonson L, Söll D (1971) Temperature dependence of the aminoacylation of tRNA by *Bacillus stearothermophilus*. Bioplymers 10:2209–2221

Kaddour H, Vergne J, Hervé G, Maurel M-C (2011) High-pressure analysis of a hammerhead ribozyme from Chrysanthemum chlorotic mottle viroid reveals two different populations of self-cleaving molecule. FEBS J 278:3739–3747

Kawamura K (1999) Monitoring of hydrothermal reactions in 3 ms using fused-silica capillary tubing. Chem Lett 28:125–126

Kawamura K (2000) Monitoring hydrothermal reactions on the millisecond time scale using a micro-tube flow reactor and kinetics of ATP hydrolysis for the RNA world hypothesis. Bull Chem Soc Jpn 73:1805–1811

Kawamura K (2003) Kinetic analysis of cleavage of ribose phosphodiester bond within guanine and cytosine rich oligonucleotides and dinucleotides at 65–200°C and its implications on the chemical evolution of RNA. Bull Chem Soc Jpn 76:153–162

Kawamura K (2004) Behavior of RNA under hydrothermal conditions and the origins of life. Int J Astrobiol 3:301–309

Kawamura K (2005) A new probe for the indirect measurement of the conformation and interaction of biopolymers at extremely high temperatures using a capillary flow hydrothermal reactor system for UV–visible spectrophotometry. Anal Chim Acta 543:236–241

Kawamura K, Maeda J (2007) Kinetic analysis of oligo(C) formation from the 5′-monophosphorimidazolide of cytidine with Pb(II) ion catalyst at 10–75°C. Orig Life Evol Biosph 37:153–165

Kawamura K, Maeda J (2008) Kinetics and activation parameter analysis for the prebiotic oligocytidylate formation on Na+-montmorillonite at 0–100°C. J Phys Chem A 112:8015–8023

Kawamura K, Nagayoshi H (2007) Behavior of DNA under hydrothermal conditions with MgCl2 additive using an in situ UV–visible spectrophotometer. Thermochim Acta 466:63–68

Kawamura K, Umehara M (2001) Kinetic analysis of the temperature dependence of the template-directed formation of oligoguanylate from the 5′-phosphorimidazolide of guanosine on a poly(C) template with Zn2+. Bull Chem Soc Jpn 74:927–935

Kawamura K, Kameyama N, Matumoto O (1999) Kinetics of hydrolysis of ribonucleotide polymers in aqueous solution ate elevated temperatures: implications of chemical evolution of RNA and primitive ribonuclease. Viva Origino 27:107–118

Lazcano A, Miller SL (1996) The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. Cell 85:793–798

Marguet E, Forterre P (1994) DNA stability at temperatures typical for hyperthermophiles. Nucleic Acids Res 22:1681–1686

Miller SL, Lazcano A (1995) The origin of life—did it occur at high temperatures? J Mol Evol 41:689–692

Murray JB, Dunham CM, Scott WG (2002) A pH-dependent conformational change, rather than the chemical step, appears to be rate-limiting in the hammerhead ribozyme cleavage reaction. J Mol Biol 315:121–130

Nisbet EG (1986) Origin of life: RNA and hot-water springs. Nature 322:206–206

Orgel LE (2004) Prebiotic chemistry and the origin of the RNA world. Crit Rev Biochem Mol Biol 39:99–123

Pace NR (1991) Origin of life—facing up to the physical setting. Cell 65:531–533

Peracchi A (1999) Origins of the temperature dependence of hammerhead ribozyme catalysis. Nucl Acids Res 27:2875–2882

Russell MJ, Hall AJ (1997) The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. J Geol Soc 154:377–402

Russell MJ, Hall AJ, Cairns-Smith AG, Braterman PS (1988) Submarine hot springs and the origin of life. Nature 336:117–117

Tehei M, Franzetti B, Maurel MC, Vergne J, Hountondji C, Zaccai G (2002) The search for traces of life: the protective effect of salt on biological macromolecules. Extremophiles 6:427–430

White RH (1984) Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250°C. Nature 310:430–432

Zhuo DM, Usman N, Wincott FE, Matulic-Adamic J, Orita M, Zhang LH, Komiyama M, Kumar PKR, Taira K (1996) Evidence for the rate-limiting departure of the 5′-oygen in nonenzymatic and hammerhead ribozyme-catalyzed reactions. J Am Chem Soc 118:5862–5866

Ztouti M, Kaddour H, Miralles F, Simian C, Vergne J, Hervé G, Maurel M-C (2009) Adenine, a hairpin ribozyme cofactor—high-pressure and competition studies. FEBS J 276:2574–2588

Article 6

NRAS Mutation Is the Sole Recurrent Somatic Mutation in Large Congenital Melanocytic Nevi

Charbel C, Fontaine RH, Malouf GG, Picard A, Kadlub N, El-Murr N, How-Kit A, Su X, Coulomb-L'hermine A, Tost J, Mourah S, Aractingi S, Guégan S. (2013) *J Invest Dermatol.* [Epub ahead of print].

NRAS Mutation Is the Sole Recurrent Somatic Mutation in Large Congenital Melanocytic Nevi

Christelle Charbel[1,2], Romain H. Fontaine[1,2,13], Gabriel G. Malouf[2,3,13], Arnaud Picard[2,4], Natacha Kadlub[2,4], Nizar El-Murr[1,2], Alexandre How-Kit[5], Xiaoping Su[6], Aurore Coulomb-L'Hermine[2,7], Jorg Tost[5,8], Samia Mourah[9], Selim Aractingi[1,10,11,14] and Sarah Guégan[1,2,12,14]

Congenital melanocytic nevus (CMN) is a particular melanocytic *in utero* proliferation characterized by an increased risk of melanoma transformation during infancy or adulthood. *NRAS* and *BRAF* mutations have consistently been reported in CMN samples, but until recently results have been contradictory. We therefore studied a series of large and giant CMNs and compared them with small and medium CMNs using Sanger sequencing, pyrosequencing, high-resolution melting analysis, and mutation enrichment by an enhanced version of *ice*-COLD-PCR. Large–giant CMNs displayed *NRAS* mutations in 94.7% of cases (18/19). At that point, the role of additional mutations in CMN pathogenesis had to be investigated. We therefore performed exome sequencing on five specimens of large–giant nevi. The results showed that *NRAS* mutation was the sole recurrent somatic event found in such melanocytic proliferations. The genetic profile of small–medium CMNs was significantly different, with 70% of cases bearing *NRAS* mutations and 30% showing *BRAF* mutations. These findings strongly suggest that *NRAS* mutations are sufficient to drive melanocytic benign proliferations *in utero*.

Journal of Investigative Dermatology advance online publication, 14 November 2013; doi:10.1038/jid.2013.429

INTRODUCTION

Congenital melanocytic nevus (CMN) is a benign clonal melanocytic proliferation that develops *in utero*. CMN is classified by its projected adult size (PAS), may be associated with neurocutaneous melanosis (DeDavid *et al.*, 1996), and moreover exhibits a risk of melanoma transformation. Indeed, prospective studies have reported a rate of 4–5% of melanomas in patients with CMNs >20 cm PAS after 5 years (Ruiz-Maldonado *et al.*, 1992; Marghoob *et al.*, 1996). The risk of melanomas appears proportional to the lesion size, reaching up to a 10–15% risk for lesions >40 cm in diameter (Krengel *et al.*, 2006). Neurological abnormalities, numerous satellite nevi, and a truncal location are also reported in patients with CMNs >40 cm PAS (Hale *et al.*, 2005; Bett, 2006). On the other hand, familial cases of CMNs have only rarely been reported (Frieden and Williams, 1994; de Wijn *et al.*, 2010). CMN does not follow a Mendelian pattern and is rather considered to be the consequence of a causative somatic mutation. CMNs have consistently been reported to harbor *NRAS* and *BRAF* mutations (Papp *et al.*, 1999, 2005; Ichii-Nakato *et al.*, 2006; Bauer *et al.*, 2007; Dessars *et al.*, 2009; Qi *et al.*, 2011; Wu *et al.*, 2011). Several studies have described *NRAS* mutations in CMNs (Bauer *et al.*, 2007; Dessars *et al.*, 2009; Wu *et al.*, 2011; Kinsler *et al.*, 2013), whereas others detected the presence of *BRAF* mutations (Papp *et al.*, 1999, 2005; Wu *et al.*, 2007; Qi *et al.*, 2011). In view of these discrepancies, we aimed to study *NRAS* and *BRAF* mutations in our series of CMNs using different powerful sequencing techniques. Moreover, in melanomagenesis, *NRAS* mutations are known to lead to transformation by cooperating with other genetic events. Similarly, additional mutations may intervene in CMN pathogenesis. In order to answer the question of other pathways involved in CMN, we performed unbiased next-generation whole-exome sequencing using large CMNs and matched control blood specimens.

[1]*Saint Antoine Research Center, U938, Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France;* [2]*Université Pierre et Marie Curie-Paris VI, Paris, France;* [3]*Department of Medical Oncology, Groupe Hospitalier Pitié-Salpêtrière, Assistance Publique–Hôpitaux de Paris (AP-HP), Paris, France;* [4]*Department of Maxillofacial and Plastic Surgery, Hôpital Necker, APHP, Paris, France;* [5]*Laboratory for Functional Genomics, Fondation Jean Dausset–CEPH, Paris, France;* [6]*Department of Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA;* [7]*Department of Pathology, Hôpital Trousseau, APHP, Paris, France;* [8]*Laboratory for Epigenetics and Environment, Centre National de Génotypage, CEA–Institut de Génomique, Evry, France;* [9]*Laboratory of Pharmacology–Genetics, Hôpital Saint-Louis, APHP, Paris, France;* [10]*Université René Descartes-Paris V, Paris, France;* [11]*Department of Dermatology, Hôpital Cochin, APHP, Paris, France and* [12]*Department of Dermatology, Hôpital Tenon, APHP, Paris, France*

[13]*These authors contributed equally to this work and should be considered as second co-authors.*

[14]*These authors contributed equally to this work and should be considered as senior co-authors.*

Correspondence: Sarah Guegan, Université Pierre et Marie Curie, INSERM UMR_S 938, Centre de Recherche Saint-Antoine, 27 Rue de Chaligny, 75012 Paris, France. E-mail: sarah.guegan@tnn.aphp.fr

Abbreviations: CMN, congenital melanocytic nevus; HRM, high-resolution melting; PAS, projected adult size; SNV, single-nucleotide variation; WT, wild type

Received 5 June 2013; revised 25 August 2013; accepted 25 September 2013; accepted article preview online 15 October 2013

© 2013 The Society for Investigative Dermatology

RESULTS

NRAS mutations in large and giant CMNs analyzed by four different sequencing techniques

A total of 19 patients diagnosed with large or giant CMNs were included in the study. The clinical characteristics are reported in Table 1. The median age of the patients was 8 months (range, 2–60 months). The male/female sex ratio was 0.73 (8 males/11 females). All body localizations were represented: 7 of 19 nevi (36.9%) involved the head, 8 of 19 nevi (42.1%) involved the arm and/or trunk, and 4 of 19 (21%) the legs. Neurological symptoms were found in none of the 19 patients; ultrasonographic evaluations and magnetic resonance images were normal in all patients who underwent such examinations. No history of familial CMN was recorded in any patient.

DNA samples were processed using three standard techniques. *NRAS* and *BRAF* mutations were screened using Sanger sequencing. Mutations in *NRAS* exon 3 were also assessed using pyrosequencing and high-resolution melting (HRM) analysis in order to further improve the sensitivity of mutation detection. Using these three techniques, only one large CMN harbored a *BRAF* mutation. We showed that 16 large or giant CMNs bore an *NRAS* mutation, with 8 c.181C>A, p.Q61K mutations and 8 c.182A>G, p.Q61R mutations. Two nevi (patients 7 and 12) displayed neither *BRAF* nor *NRAS* mutations (Table 2).

A minor *NRAS* clone could still remain undetectable at this stage. To definitely solve this question, we harvested additional specimens (from 4 other areas for patient 12) as well as performed enhanced version of *ice*-COLD-PCR (E-*ice*-COLD-PCR; for patients 7 and 12). This technique relies on a chemically modified, nonextendable oligonucleotide complementary to the wild-type (WT) allele sequences that blocks amplification of WT alleles and thereby ''enrich'' mutated alleles. These mutated alleles were subsequently analyzed by pyrosequencing. Interestingly, a clonal *NRAS* Q61R mutation was found by E-*ice*-COLD-PCR in both of these large CMNs (patients 7 and 12). Therefore, this high-sensitivity sequencing technique detected *NRAS* mutations in two large CMNs (10.5%) previously classified as *BRAF* and *NRAS* WT using standard techniques. Overall, an *NRAS* mutation was detected in 94.7% of large or giant CMNs (18/19); 8 out of 19 nevi (42.1%) bore the c.181C>A, p.Q61K mutation whereas 10 out of 19 nevi (52.6%) displayed the c.182A>G, p.Q61R mutation. Finally, one large CMN (patient 9) harbored a *BRAF* mutation without any *NRAS* mutation. Of note, the first sample of patient 12 nevus did not display a *BRAF* or *NRAS* mutation even using E-*ice*-COLD-PCR. However, when four additional nevus areas (specimens) were selected and screened, two out of four revealed the presence of a c.182A>G, p.Q61R *NRAS* mutated clone.

Table 1. Clinical characteristics of the 19 large and giant CMNs

Patient	Age	Sex	Tissue site	PAS (cm)	Ruiz-Maldonado classification[1]	NCM/neurological symptoms	Familial CMN	Histological features
1	3.5 Years	Female	Head	>20–30	Giant	No	No	Compound
2	5 Years	Female	Arm	>20–30	Giant	No	No	Compound
3	2 Months	Female	Leg	>10–20	Large	No	No	Compound
4	19 Months	Male	Arm	>10–20	Large	No	No	Compound
5	21 Months	Female	Leg	>10–20	Large	No	No	Compound
6	5 Years	Male	Head	>10–20	Large	No	ND	Compound
7	3 Months	Male	Trunk	>10–20	Large	ND	ND	Compound
8	7 Months	Male	Trunk	>40–60	Giant	No	ND	Compound
9	25 Months	Male	Leg	>10–20	Large	No	No	Compound
10	4 Months	Male	Head	>10–20	Large	No	No	Compound
11	4 Months	Female	Trunk	>20–30	Giant	No	ND	Compound
12	5 Months	Female	Head	>10–20	Large	No	No	Compound
13	9 Months	Female	Head	>20–30	Giant	No	No	Dermal
14	13 Months	Male	Head	>10–20	Large	No	No	Compound
15	3 Months	Male	Head	>20–30	Giant	ND	ND	Compound
16	8 Months	Female	Arm	>10–20	Large	No	ND	Compound
17	6 Months	Female	Arm	>20–30	Giant	No	No	Compound
18	4 Months	Female	Trunk	>10–20	Large	No	No	Dermal
19	9 Months	Female	Leg	>10–20	Large	No	No	Compound

Abbreviations: CMN, congenital melanocytic nevus; NCM, neurocutaneous melanosis; ND, nondetermined; PAS, projected adult size.
[1]Classification by PAS as defined in Ruiz-Maldonado (2004).

Table 2. Genetic characteristics of the 19 large and giant CMNs

Patient	Sanger	NRAS exon 3		
		Pyrosequencing	HRM[1]	E-*ice*-COLD-PCR
1	WT	WT	NRAS3 mutated	NRAS3 Q61K
2	NRAS3 Q61R	NRAS3 mutated	NRAS3 mutated	
3	NRAS3 Q61K	NRAS3 mutated	NRAS3 mutated	
4	NRAS3 Q61R	NRAS3 mutated	NRAS3 mutated	
5	NRAS3 Q61K	NRAS3 mutated	NRAS3 mutated	
6	NRAS3 Q61K	NRAS3 mutated	NRAS3 mutated	
7	ND	WT	WT	NRAS3 Q61R
8	NRAS3 Q61R	NRAS3 mutated	NRAS3 mutated	
9	BRAF V600E	WT	WT	
10	NRAS3 Q61R	NRAS3 mutated	NRAS3 mutated	
11	NRAS3 Q61K	NRAS3 mutated	NRAS3 mutated	
12	WT	WT	WT	NRAS3 Q61R
13	NRAS3 Q61R	NRAS3 mutated	NRAS3 mutated	
14	NRAS3 Q61R	NRAS3 mutated	NRAS3 mutated	
15	NRAS3 Q61K	NRAS3 mutated	NRAS3 mutated	
16	NRAS3 Q61K	NRAS3 mutated	NRAS3 mutated	
17	NRAS3 Q61R	NRAS3 mutated	NRAS3 mutated	
18	NRAS3 Q61R	NRAS3 mutated	NRAS3 mutated	
19	NRAS3 Q61K	NRAS3 mutated	NRAS3 mutated	

Abbreviations: CMN, congenital melanocytic nevus; E-*ice*-COLD-PCR, enhanced version of *ice*-COLD-PCR; HRM, high-resolution melting analysis; ND, nondetermined; WT, wild type.

[1]High-resolution melting analysis. Sanger sequencing was used to screen *NRAS* exon 2 and 3 and *BRAF* exon 15 mutations. Pyrosequencing and high-resolution melting analysis were used to screen *NRAS* exon 2 and 3 mutations. Enhanced-*ice*-COLD-PCR was used to screen *NRAS* exon 3 mutation.

Table 3. Results of whole-exome sequencing of five CMN samples and matched blood DNA

Gene	Chr		cDNA shift	Peptide shift	SIFT prediction	Patient 6	Patient 12	Patient 14	Patient 15	Patient 16
NRAS	1	SNP	c.181C>A	p.Q61K	Damaging	1	0	0	1	1
NRAS	1	SNP	c.182A>G	p.Q61R	Damaging	0	0	1	0	0

Abbreviations: cDNA, complementary DNA; CMN, congenital melanocytic nevus; SNP, single-nucleotide polymorphism.

***NRAS* mutation as the sole recurrent somatic mutation confirmed by whole-exome sequencing of large and giant CMNs**

We asked therefore whether additional mutations beside *NRAS* were needed to drive CMN pathogenesis. To answer this question, we selected DNA from five patients with large or giant CMNs with matched nontumor peripheral blood DNA available. Whole-exome sequencing was performed on these five sample pairs in order to detect somatic mutations in the nevi (Table 3). Median coverage of the exome sequencing was 54× (range, 51–60). Analysis revealed a total of nine somatic nonsynonymous mutations involving eight genes. As expected, 4 of the 5 tumors harbored the previously confirmed *NRAS* mutation (patients 6, 14, 15, and 16). Visual inspection of the reads did not identify any other mutations in the vicinity of the *NRAS* Q61 mutation. We also evaluated the percentage of mutated cells within a nevus harboring the *NRAS* mutation. It ranged from 10 to 40% in these 4 patients analyzed using whole-exome sequencing: 25% in patient 6 (13 mutated/51 nonmutated tags), 10% in patient 14 (8 mutated/86 nonmutated tags), 40% in patient 15 (22 mutated/55 nonmutated tags), and 19% in patient 16 (14 mutated/74 nonmutated tags). Interestingly, patient 12 nevus, which revealed a NRAS mutated subclone only after an extensive screening using E-*ice*-COLD-PCR sequencing of several areas, displayed 1 mutated tag of *NRAS* out of 79 tags in the nevus versus no identified mutated tag in matched blood DNA (number of total tags detected 64). Therefore, both E-*ice*-COLD-PCR and exome sequencing techniques led to corresponding results. Besides *NRAS*, and using our criteria (3 mutated tags in nevus, and a total of 10 tags covered in nevus and normal at minimum), exome sequencing revealed the presence of seven other non-*NRAS*, non-*BRAF*, nonsynonymous mutations

Table 4. Clinical and genetic characteristics of the 20 small and medium CMNs

Sample	Age	Sex	Tissue site	PAS (cm)	Histological features	Sequencing[1]	E-*ice*-COLD-PCR
20	35 Months	Female	Arm	1.5–10	Compound	WT	NRAS3 Q61K
21	4 Months	Female	Head	1.5–10	Compound	NRAS3 Q61R	
22	10 Months	Male	Arm	1.5–10	Compound	NRAS3 Q61R	
23	16 Years	Female	Leg	1.5–10	Dermal	NRAS3 Q61K	
24	11 Months	Male	Head	1.5–10	Compound	WT	NRAS Q61R
25	9 Months	Female	Trunk	1.5–10	Compound	BRAF V600E	NRAS3 WT
26	16 Months	Female	Head	1.5–10	Compound	WT	NRAS3 Q61R
27	44 Months	Male	Trunk	1.5–10	Compound	ND	NRAS3 Q61K
28	12 Years	Male	Head	1.5–10	Compound	BRAF V600E	NRAS3 WT
29	38 Months	Female	Arm	1.5–10	Compound	WT	NRAS3 Q61K
30	8.5 Years	Female	Trunk	1.5–10	Compound	BRAF V600E	
31	6 Years	Female	Arm	1.5–10	Compound	NRAS3 Q61K	NRAS3 Q61K
32	16 Months	Male	Leg	1.5–10	Compound	BRAF V600E	
33	5 Years	Female	Leg	1.5–10	Compound	WT	NRAS3 Q61R
34	10 Years	Male	Head	<1.5	Compound	BRAF V600E	
35	4 Months	Female	Arm	1.5–10	Compound	BRAF V600E	
36	22 Months	Female	Trunk	1.5–10	Compound	WT	NRAS3 Q61K
37	6 Years	Female	Leg	1.5–10	Compound	ND	NRAS3 Q61K
38	33 Months	Male	Leg	1.5–10	Compound	WT	NRAS3 Q61R
39	7 Months	Male	Trunk	1.5–10	Compound	NRAS3 Q61K	NRAS3 Q61K

Abbreviations: CMN, congenital melanocytic nevus; E-*ice*-COLD-PCR, enhanced version of *ice*-COLD-PCR; ND, nondetermined; PAS, projected adult size; WT, wild type.
Mutation enrichment by enhanced-*ice*-COLD-PCR was used to screen *NRAS* exon 3 mutations.
[1]Sanger sequencing was used to screen *NRAS* exon 2 and 3 and *BRAF* exon 15 mutations followed by pyrosequencing and high-resolution melting analysis in order to screen *NRAS* exon 2 and 3 mutations.

involving *BAGE5*, *C5orf42*, *EPS15*, *ETV2*, *KRT77*, *NKIRAS2*, and *TRPV4*. At this stage, we had to confirm whether these results corresponded to real mutations or to nonreproducible artifacts. Using two different sequencing techniques, i.e., Sanger sequencing (Supplementary Table S1 online) and a sensitive allele-specific PCR approach, we failed to confirm the presence of these mutations, suggesting that they corresponded to false positives of the technique. Therefore, *NRAS* mutation appears to be the sole recurrent somatic mutation identified in large and giant CMNs.

Mixed pattern of *NRAS* and *BRAF* mutations in small and medium CMNs

A second group of smaller CMNs was compared with the previous group of CMNs: 20 patients diagnosed with small or medium CMNs were also studied. The median age of the patients was 50 months, whereas the mean age was 35 months. The male/female sex ratio was 0.67 (8 males/12 females). All body localizations were represented: 5 of 20 nevi (25%) involved the head, 10 of 20 nevi (50%) involved the arm and/or trunk, and 5 of 20 (25%) the leg. Therefore, in terms of sex ratio and body localization of nevi, both groups were comparable.

Small and medium CMNs were analyzed for *NRAS* and *BRAF* mutations using the various techniques previously described: Sanger sequencing, pyrosequencing, HRM analysis, and mutation enrichment by E-*ice*-COLD-PCR (Table 4).

An *NRAS* mutation was detected in 14 out of 20 (70%) small–medium CMN: 8 nevi (40%) bore the c.181C>A, p.Q61K mutation and 6 nevi (30%) bore the c.182A>G, p.Q61R mutation. A total of 30% (6/20) of small–medium CMNs harbored a *BRAF* mutation. Therefore, the mutational profile of these nevi was different from the large CMN mutational profile, with a lower percentage of *NRAS* mutations and a higher percentage of *BRAF* mutations. Of note, only 5 *NRAS* mutated nevi (25%) were detected using the three conventional techniques. The high-sensitivity sequencing technology showed low-frequency mutated clones in small–medium lesions as it revealed *NRAS* mutations in 9 nevi (45%) previously classified as *NRAS* WT using standard technologies, whereas this was the case in only 2 large CMNs (10.5%).

DISCUSSION

NRAS is one of the three major isoforms of the RAS family of GTPases proteins. Through the downstream activation of four major signaling pathways mitogen-activated protein kinase,

phosphatidylinositide 3 kinase, phospholipase C/protein kinase C, and RalGDS, NRAS is involved in cell growth. All reported *NRAS* gene mutations occur in exon 2 or exon 3, with 65% of mutations occurring in exon 3 at codon 61. The most frequent amino acid substitutions are Q61K and Q61R (Takata and Saida, 2006). These substitutions result in a constitutively active NRAS protein, unable to cleave guanosine triphosphate.

Using an unbiased series of patients, we found the presence of *NRAS* mutations in 94.7% and 70% of large–giant and small–medium CMN cases, respectively. *NRAS* mutations have been detected at low percentage in malignant melanoma as well as in acquired melanocytic nevi (Poynter *et al.*, 2006). However, *NRAS* mutations have been closely associated with CMNs. Previous studies detected variable levels of *NRAS* mutation in CMNs. These vary between 22 and 80% in affected cutaneous and neurological tissues of patients (Carr and Mackie, 1994; Ichii-Nakato *et al.*, 2006; Dessars *et al.*, 2009; Wu *et al.*, 2011; Kinsler *et al.*, 2013). In view of these differences, our study confirms that CMNs display a high incidence of *NRAS* mutations, highlighting the role of this pathway in the disease.

Malignant tumors are usually featured by several mutations that accumulate during neoplastic proliferation (Hanahan and Weinberg, 2011). It was therefore mandatory to assess the eventual intervention of other mutated pathways. We therefore selected 5 patients out of our series, including patient 12 initially negative for *BRAF* and *NRAS* mutations when screened by four sequencing techniques. The nevus specimens were sequenced using next-generation whole-exome analysis and compared with the patient's peripheral blood concomitantly analyzed. This technique is currently the most powerful one to assess the presence of any somatic mutation in an acquired disease (Shyr and Liu, 2013). Surprisingly, the unique recurrent mutation detected was *NRAS*. Patient 12, who was found positive for an *NRAS* mutated subclone only after screening of 4 additional nevus areas using E-*ice*-COLD-PCR, also displayed an *NRAS* clone (1 mutated tag in 79) in whole-exome sequencing, consistent with the fact that a minor *NRAS* mutated clone existed in this case.

Our study focuses on exonic regions, and further studies are needed to assess that intronic mutations are not involved. Indeed, exome sequencing only allows the identification of somatic mutations present in coding areas. Our study does not provide information concerning intronic regions, nor was it conducted in a way to analyze the significance of germline mutations. Nevertheless, Kinsler *et al.* (2012) have reported a role for germline *MC1R* genotype in CMNs, dismissing *MC1R* as a somatic mosaic candidate. Exome sequencing cannot depict DNA copy number changes such as microdeletions or duplications. These request performing comparative genomic hybridization array. However, previous studies from Bastian (2003) and Bastian *et al.* (1999) have indicated the absence of such abnormalities in congenital nevi and acquired nevi, except in Spitz nevi. Altogether, our data combined with these other results indicate that no other somatic event than the *NRAS* mutation at codon 61 is responsible for the development of congenital nevi to the exclusion of possible intronic mutations. Of note, other examples showed that tumors including malignant ones may rely on a unique mutation. Indeed, in rhabdoid child tumors, exome sequencing analysis indicated that biallelic loss of a subunit of the SWI/SNF chromatin remodeling complex was solely responsible for this malignancy (Lee *et al.*, 2012). Moreover, the recent demonstration that identical somatic codon 61 *NRAS* mutations were found in different CMN lesions, melanomas as well as in neurological lesions from the same patients, also supports the causative role of the *NRAS* mutation in CMN genesis (Kinsler *et al.*, 2013). In the same study, Kinsler *et al.* (2013) report that the *NRAS* mutation found in affected patients segregated only in neural crest–derived cells or tumors such as choroid plexus papilloma, neurocristic hamartoma, meningioma, or diffuse leptomeningeal melanocytosis as the same *NRAS* mutation was found in these separate anatomical sites. Conversely, *NRAS* mutation was absent from other tissues such as the blood of patients. This supports the hypothesis of a single-mutated precursor cell giving rise to the nevus as well as rare neural crest–derived tumors.

We report *BRAF* mutations in 5.2% (1/19) and 30% of large–giant and small–medium CMNs, respectively. Previous studies found conflicting data indicating the presence of a *BRAF* mutation between 0% (Bauer *et al.*, 2007; Wu *et al.*, 2011) and 86% in nevi documented either at birth by parents testimony (Ichii-Nakato *et al.*, 2006), medical records (Wu *et al.*, 2007; Qi *et al.*, 2011), or selected by histological criteria (Pollock *et al.*, 2003). Such histological selection is controversial as it is known that acquired melanocytic nevi may mimic histologically a congenital pattern. Using our sensitive techniques with nevi assessed both by parents' testimony and medical records, we were able to find that the frequency of *BRAF* mutation does not exceed 30%, showing that it was less associated with CMN than *NRAS* mutation.

In our study, patients have been categorized using Ruiz-Maldonado classification (Ruiz-Maldonado, 2004). The risk of melanoma appears nearly absent in small and medium nevi that have been grouped here. In contrast, malignant transformation develops in up to 5% in series of nevi with PAS >20 cm (Ruiz-Maldonado *et al.*, 1992; Marghoob *et al.*, 1996). Importantly, in the literature, a precise estimation of such risk is not known for the large nevi (>10–20 cm) that we have therefore grouped with the giant nevi (>20 cm). Interestingly, the rate of *NRAS* mutation appears similar in these large (11/12, 92%) and giant (7/7, 100%) specimens.

CMNs may have different genetic signatures, with either *NRAS* or *BRAF* mutations, *BRAF* mutations having a lower incidence rate in CMN >20 cm PAS, and a higher incidence rate in small CMNs. On the other hand, $\sim$79% (376 of 479) of acquired nevi harbor *BRAF* mutations (Pollock *et al.*, 2003; Ichii-Nakato *et al.*, 2006; Ross *et al.*, 2011). Of note, these acquired nevi do not transform more frequently than melanocytes interspersed in normal epidermis outside a nevus. Therefore, the risk of malignant transformation does not parallel the expression of *BRAF* mutation. Interestingly, as compared with *BRAF* mutated melanomas or *NRAS*/*BRAF* WT melanomas, melanomas with *NRAS* mutations are more likely to be thicker tumors and to have a higher mitotic rate

(Devitt *et al.*, 2011; Ascierto *et al.*, 2013). One may therefore consider that *BRAF* mutations may not be as important as *NRAS* mutations in the risk of melanoma arising from benign nevi in general and CMNs in particular. Nevertheless, in order to fully answer that question, studies comparing melanoma incidence rate in *NRAS* mutated CMNs and in *BRAF* mutated CMNs would be needed.

In conclusion, our study strongly suggests that codon 61 *NRAS* mutations in melanocytes are sufficient to drive *in utero* proliferation. This result should encourage to examine alternative medical treatments to surgery.

MATERIALS AND METHODS

Study approval

The study was approved by the institutional independent ethics committee (Comité de Protection des Personnes Ile-de-France V) and complied with the Declaration of Helsinki Principles. The patients' guardians provided written informed consent before their participation.

Subjects

Clinical phenotyping was performed by AP, NK, and SG. Classification was done using the PAS of the largest lesion (Ruiz-Maldonado, 2004). A total of 19 patients with large or giant CMNs and 20 patients with small or medium CMNs were included prospectively in the study and nevus samples collected. Tumor tissues and matched blood from five patients diagnosed with large or giant CMNs were collected for the whole-exome sequencing analysis.

DNA extraction and genotyping

DNA was extracted using standard techniques: from fresh tissue using the DNeasy Blood and Tissue Kit (Qiagen, Courtaboeuf, France) and from paraffin-embedded tissue using QIAamp DNA FFPE Tissue Kit (Qiagen, France). The DNA was quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). Several genotyping methods were used to identify BRAF exon 15 and NRAS exons 2 and 3 mutations using specific primer sets. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) primers were used for control of DNA quality. Positive, negative, and no-DNA controls were included in each run.

Sanger sequencing

A sequence of ~500 nucleotides surrounding the mutation site was amplified using the Finnzyme Phusion High-Fidelity DNA polymerase (Thermo Scientific, Villebon-sur-Yvette, France). Depending on the CG content of the amplified sequence, DMSO was added to the reaction. PCR products were then purified on 96-well Multiscreen-PCR filtration plates (Millipore, Molsheim, France). The following NRAS and BRAF primers were used: BRAF_F, 5′-TGCTCTGATAGGAAAATGAGATCTAC-3′ and BRAF_R, 5′-GGCCAAAAATTTAATCAGTGG-3′; NRAS2_F, 5′-GAACCAAATGGAAGGTCACA-3′ and NRAS2_R, 5′-TGGGTAAAGATGATCCGACA-3′; NRAS3_F, 5′-CCAGATAGGCAGAAATGGGCT-3′ and NRAS3_R, 5′-GCTCTATCTTCCCTAGTGTGGT-3′. A listing of the other primers used, annealing temperatures, and expected length of PCR products is given in Supplementary Table S1 online. Sanger sequencing analyses were conducted according to the manufacturer's protocol using respectively the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Courtaboeuf, France), on an ABI3130 running ABI Prism DNA Sequence Analysis Software (Applied Biosystems).

Pyrosequencing

Pyrosequencing was performed according to standard procedures using PyroMark Gold Q96 (Qiagen, Hilden, Germany) reagents and a PyroMarkTM Q96 ID instrument. Pyrograms outputs were analyzed with PyroMarkTM Q96 ID Software (Qiagen, Germany) using the allele quantification mode.

High-resolution melting

HRM analyses were conducted according to the manufacturer's protocol using the LC480 HRM Scanning Master (Roche, Boulogne-Billancourt, France) on a LightCycler 480 (Roche).

Enrichment of NRAS mutations by E-ice-COLD-PCR

For codon 61 *NRAS* mutation amplification, an E-*ice*-COLD-PCR assay was designed following the standard protocol (How Kit *et al.*, 2013). PCR primers (NRAS_F 5′-GTTGGACATACTGGATAC-3′ and NRAS_R Biotin-5′-ATGACTTGCTATTATTGATG-3′) were designed with Beacon Designer 8 (Premier Biosoft, Palo Alto, CA) and purchased from Biotez (Berlin, Germany). Primers amplified a 113-bp PCR product flanking the codons of interest. The nonextendable blocker probe of 50-nucleotide length (NRAS_BK_LNA 5′-GATACAGCTGGA+C+A+AGAAGAGTACAGTGCCATGAGAGACCAATACATGAG-3′-Phosphate), purchased from TibMolBiol (Berlin, Germany), overlaps 5 nucleotides with the forward PCR primer and presents a triplet of LNA residues centered on codon 61 in order to improve the ΔTM of homoduplexes compared with heteroduplexes when the mismatches are located on the three nucleotides of interest. The critical temperature of 79 °C was determined as described in How Kit *et al.* (2013) using a NRAS c.182A>G mutated cell line (BLM melanoma cells) and WT DNA (Promega, Lyon, France). E-*ice*-COLD-PCR reaction was performed in a Mastercycler Pro S (Eppendorf, Le Pecq, France) and contained 5–25 ng of genomic DNA, 1× HotStar Taq DNA polymerase Buffer, 1.6 mM of additional $MgCl_2$, 200 μM of dNTPs, 200 nM of each primer, 2 U of HotStar Taq DNA polymerase (Qiagen, France), and 10–50 nM of Blocker probe in a final volume of 25 μl. An initial denaturation step was performed for 10 minutes at 95 °C, followed by 6 cycles of standard PCR (30-second denaturation at 95 °C, 20-second annealing at 55 °C, and 10-second elongation at 72 °C) and 44 cycles of E-*ice*-COLD-PCR (20-second denaturation at 95 °C, 30-second blocker annealing at 70 °C, 20-second at 79 °C (critical temperature), 20-second primer annealing at 55 °C, and 10-second elongation at 72 °C) to enrich the mutations. Mutation detection, identification, and quantification were performed by pyrosequencing. Next, 10 μl of amplified E-*ice*-COLD-PCR products were purified and rendered single-stranded on a PyroMark Q96 Vacuum Workstation (Qiagen, France) as previously described (Tost and Gut, 2007). The sequencing primer (NRAS_Gen_F, 5′-TGGACATACTGGATACAGC-3′) was annealed to the target sequence after incubation at 80 °C for 2 minutes.

Sensitivity of the applied sequencing techniques

Using given amounts of mutated DNA as standards, the mutation detection sensitivity for each of the applied sequencing techniques for the *BRAF* and *NRAS* mutations is as follows: *BRAF* screening by

Sanger, 15%; *NRAS* exon 2 and 3 screening by Sanger, 15%; *NRAS* exon 2 and 3 screening by HRM, 5%; and *NRAS* exon 3 screening by *ice*-COLD-PCR, 0.5–1%.

Whole-exome sequencing

Exome capture was performed using Agilent (Santa Clara, CA) SureSelect Human All Exon 50 Mb according to the manufacturer's instructions. Briefly, 3 μg of DNA from each sample was used to prepare the sequencing library through shearing of the DNA followed by ligation of sequencing adaptors. Whole-exome sequencing was performed, and paired-end sequencing (2×76 bp) was carried out using the Illumina (San Diego, CA) HiSeq 2000; the resulting data were analyzed with the Illumina pipeline to generate raw fastq files. The coverage of our samples varied between $51\times$ and $54\times$.

Somatic mutation detection from whole-exome sequencing

After raw paired-end reads from whole-exome resequencing were aligned/mapped to the human genome reference (hg19) and PCR duplicate reads were removed by Mosaik aligner, we then analyzed the resulting alignments using the Bayesian model-based software GigaBayes/FreeBayes (National Institutes of Health, Bethesda, MD) that enables the efficient analysis of billions of aligned short-read sequences (Marth *et al.*, 1999). The program evaluates each aligned base and base quality value at each position to indicate putative single-nucleotide variations (SNVs) and short insertions/deletions (indels), and their corresponding SNV probability value. Base quality values are converted to base probabilities corresponding to each of the four possible nucleotides. Using a Bayesian formulation, a SNV probability value (or indel probability value, as appropriate) is calculated as the likelihood that multiple different alleles are present between the reference genome sequence and the reads aligned at that position. If the probability value exceeds a prespecified threshold, the SNV or indel candidate is reported in the output. In this study, we used a SNV probability cutoff value (0.9) to define a high-confidence SNV or short indel candidate. We also filtered out all known SNVs/indels in the University of California, Santa Cruz (UCSC) dbSNP 135 and 1000 human genome project SNP database, and kept any mutations that are in the Catalogue of Somatic Mutations in Cancer (COSMIC) database curated by Wellcome Trust Sanger Institute. We then determined the somatic status of each SNV (or indel) by comparing the genotypes and likelihood between matched normal and tumor samples. The somatic status of a specific SNV/indel was reported if the matched normal had wild allele–based homozygous genotype and the tumor had heterozygous or mutant allele–based homozygous genotype with a certain cutoff of genotype likelihood/*P*-value (0.99). Finally, each somatic mutation or indel was annotated with its functional effect by SIFT (http://sift.jcvi.org, J Craig Venter Institute, San Diego, CA) to determine whether a mutation candidate was synonymous or nonsynonymous (benign or deleterious). Five nevus samples and five matched blood samples were processed and analyzed using whole-exome sequencing (Otogenetics, Norcross, GA). Library construction followed the procedure of previous publications (Stransky *et al.*, 2011; Berger *et al.*, 2011; Chapman *et al.*, 2011). Exome sequencing data were then validated by Sanger sequencing and when necessary allele-specific PCR. The allele-specific PCR approach was used to validate the 7 non-*NRAS*, non-*BRAF* mutations using WT and mutant primer couples for each gene. Primer annealing temperature was selected to allow WT allele amplification with WT complementary primers and not with the mutant complementary primer.

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

This work was funded by the Société Française de Dermatologie and the Société de Recherche Dermatologique. CC was supported by the Fondation René Touraine, France, and Lebanese Conseil National Recherche Scientifique (CNRS) and AREMPH. We are grateful to all the patients and families who participated in this research.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/jid

REFERENCES

Ascierto PA, Schadendorf D, Berking C *et al.* (2013) MEK162 for patients with advanced melanoma harbouring NRAS or Val600 BRAF mutations: a non-randomised, open-label phase 2 study. *Lancet Oncol* 14:249–56

Bastian BC (2003) Understanding the progression of melanocytic neoplasia using genomic analysis: from fields to cancer. *Oncogene* 22:3081–6

Bastian BC, Wesselmann U, Pinkel D *et al.* (1999) Molecular cytogenetic analysis of Spitz nevi shows clear differences to melanoma. *J Invest Dermatol* 113:1065–9

Bauer J, Curtin JA, Pinkel D *et al.* (2007) Congenital melanocytic nevi frequently harbor NRAS mutations but no BRAF mutations. *J Invest Dermatol* 127:179–82

Berger MF, Lawrence MS, Demichelis F *et al.* (2011) The genomic complexity of primary human prostate cancer. *Nature* 470:214–20

Bett BJ (2006) Large or multiple congenital melanocytic nevi: occurrence of neurocutaneous melanocytosis in 1008 persons. *J Am Acad Dermatol* 54:767–77

Carr J, Mackie RM (1994) Point mutations in the N-ras oncogene in malignant melanoma and congenital naevi. *Br J Dermatol* 131:72–7

Chapman MA, Lawrence MS, Keats JJ *et al.* (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature* 471:467–72

De Wijn RS, Zaal LH, Hennekam RC *et al.* (2010) Familial clustering of giant congenital melanocytic nevi. *J Plast Reconstr Aesthet Surg* 63: 906–13

DeDavid M, Orlow SJ, Provost N *et al.* (1996) Neurocutaneous melanosis: clinical features of large congenital melanocytic nevi in patients with manifest central nervous system melanosis. *J Am Acad Dermatol* 35: 529–38

Dessars B, De Raeve LE, Morandini R *et al.* (2009) Genotypic and gene expression studies in congenital melanocytic nevi: insight into initial steps of melanotumorigenesis. *J Invest Dermatol* 129:139–47

Devitt B, Liu W, Salemi R *et al.* (2011) Clinical outcome and pathological features associated, with NRAS mutation in cutaneous melanoma. *Pigment Cell Melanoma Res* 24:666–72

Frieden IJ, Williams ML (1994) Familial site-specific congenital melanocytic nevus: report of two families. *Arch Dermatol* 130:1075–6

Hale EK, Stein J, Ben-Porat L *et al.* (2005) Association of melanoma and neurocutaneous melanocytosis with large congenital melanocytic naevi–results from the NYU-LCMN registry. *Br J Dermatol* 152: 512–7

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144:646–74

How Kit A, Mazaleyrat N, Daunay A *et al.* (2013) Sensitive detection of KRAS mutations using E-ice-COLD-PCR mutation enrichment and direct sequence identification. *Hum Mutat* 34:1568–80

Ichii-Nakato N, Takata M, Takayanagi S *et al.* (2006) High frequency of BRAFV600E mutation in acquired nevi and small congenital nevi, but low

frequency of mutation in medium-sized congenital nevi. *J Invest Dermatol* 126:2111–8

Kinsler VA, Thomas AC, Ishida M *et al.* (2013) Multiple congenital melanocytic nevi and neurocutaneous melanosis are caused by postzygotic mutations in codon 61 of NRAS. *J Invest Dermatol* 133: 2229–36

Kinsler VA, Abu-Amero S, Budd P *et al.* (2012) Germline melanocortin-1-receptor genotype, is associated with severity of cutaneous phenotype in congenital melanocytic nevi: a role for, MC1R in human fetal development. *J Invest Dermatol* 132:2026–32

Krengel S, Hauschild A, Schafer T (2006) Melanoma risk in congenital melanocytic naevi: a systematic review. *Br J Dermatol* 155:1–8

Lee RS, Stewart C, Carter SL *et al.* (2012) A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. *J Clin Invest* 122:2983–8

Marghoob AA, Schoenbach SP, Kopf AW *et al.* (1996) Large congenital melanocytic nevi and the risk for the development of malignant melanoma. A prospective study. *Arch Dermatol* 132:170–5

Marth GT, Korf I, Yandell MD *et al.* (1999) A general approach to single-nucleotide polymorphism discovery. *Nat Genet* 23:452–6

Papp T, Pemsel H, Zimmermann R *et al.* (1999) Mutational analysis of the N-ras, p53, p16INK4a, CDK4, and MC1R genes in human congenital melanocytic naevi. *J Med Genet* 36:610–4

Papp T, Schipper H, Kumar K *et al.* (2005) Mutational analysis of the BRAF gene in human congenital and dysplastic melanocytic naevi. *Melanoma Res* 15:401–7

Pollock PM, Harper UL, Hansen KS *et al.* (2003) High frequency of BRAF mutations in nevi. *Nat Genet* 33:19–20

Poynter JN, Elder JT, Fullen DR *et al.* (2006) BRAF and NRAS mutations in melanoma and melanocytic nevi. *Melanoma Res* 16:267–73

Qi RQ, He L, Zheng S *et al.* (2011) BRAF exon 15 T1799A mutation is common in melanocytic nevi, but less prevalent in cutaneous malignant melanoma, in Chinese Han. *J Invest Dermatol* 131:1129–38

Ross AL, Sanchez MI, Grichnik JM (2011) Molecular neurogenesis. *Dermatol Res Pract* 2011:463184

Ruiz-Maldonado R (2004) Measuring congenital melanocytic nevi. *Pediatr. Dermatol.* 21:178–9

Shyr D, Liu Q (2013) Next generation sequencing in cancer research and clinical application. *Biol Proced Online* 15:4

Stransky N, Egloff AM, Tward AD *et al.* (2011) The mutational landscape of head and neck squamous cell carcinoma. *Science* 333:1157–60

Takata M, Saida T (2006) Genetic alterations in melanocytic tumors. *J Dermatol Sci* 43:1–10

Tost J, Gut IG (2007) DNA methylation analysis by pyrosequencing. *Nat Protoc* 2:2265–75

Wu D, Wang M, Wang X *et al.* (2011) Lack of BRAF(V600E) mutations in giant congenital melanocytic nevi in a Chinese population. *Am J Dermatopathol* 33:341–4

Wu J, Rosenbaum E, Begum S *et al.* (2007) Distribution of BRAF T1799A(V600E) mutations across various types of benign nevi: implications for melanocytic tumorigenesis. *Am J Dermatopathol* 29:534–7

Article 7

Cap polyposis and colitis cystica profunda: a rare association

Arana R, Fléjou JF, Parc Y, El-Murr N, Cosnes J, Svrcek M. (2014) *Histopathology* 64 (4): 604-607.

diagnosis in 16–27% of cases, resulting in one in five patients being given a histopathological grade different from the consensus (gold) standard. This problem may be alleviated by having multiple pathologists assess difficult cases, which can easily be done using WSI. The results of the present study show that alternative WSI-based assessment of laryngeal premalignancies will not compromise diagnostic accuracy.

Bart Sturm
Stijn J H M Fleskens[1]
Fredrik J Bot[2]
Marie-Louise van Velthuysen[3]
Ernst-Jan Speel[2]
Piet J Slootweg
Jeroen A W M van der Laak

Department of Pathology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, [1]*Department of Otorhinolaryngology and Head and Neck Surgery, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands,* [2]*Department of Pathology, Maastricht University Medical Center, Maastricht, The Netherlands, and* [3]*Department of Pathology, The Netherlands Cancer Institute/Antoni van Leeuwenhoek Ziekenhuis, Amsterdam, The Netherlands*

1. van der Post RS, van der Laak JAWM, Sturm B *et al.* The evaluation of colon biopsies using virtual microscopy is reliable. *Histopathology* 2013; **63**; 114–121.
2. Pantanowitz L, Sinard JH, Henricks WH *et al.* Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch. Pathol. Lab. Med.* 2013. (in press).
3. Fleskens SA, Bergshoeff VE, Voogd AC *et al.* Interobserver variability of laryngeal mucosal premalignant lesions: a histopathological evaluation. *Mod. Pathol.* 2011; **24**; 892–898.
4. Gale N, Pilch BZ, Sidransky D, Westra W, Califano J. Tumours of the hypopharynx, larynx and trachea (epithelial precursor lesions). In Barnes L, Eveson JW, Reichart P, Sidransky D eds. *World Health Organization classification of tumours. Pathology & genetics. Head and neck tumours*. Lyon: IARC Press, 2005.

Cap polyposis and colitis cystica profunda: a rare association

DOI: 10.1111/his.12292

© 2013 John Wiley & Sons Ltd.

Sir: Cap polyposis and colitis cystica profunda (CCP) are two rare, benign and distinct colorectal conditions whose aetiology remains unclear. A rare association between cap polyposis and CCP has already been described; in those cases, cystic glands were present in the submucosal layer of the cap polyps.[1,2] Interestingly, in the present case, lesions of CCP were not intermingled with cap polyps, but were distant from them.

A 49-year-old man with chronic colitis located in the left colon and rectum, diagnosed as ulcerative colitis 5 years previously, was referred to our institution for persistent mucous bloody stools refractory to immunosuppressive therapies (infliximab). Prior colonoscopy revealed the presence of ~50 sessile reddish polyps, mainly located in the rectosigmoid area. Biopsy specimens of the polyps showed superficial erosion with elongated hyperplastic glands and a mixed inflammatory infiltrate in the lamina propria. There was no atypia or dyplasia. Biopsy specimens of the intervening mucosa were not available. Laboratory values, including complete blood count, electrolytes, and serum protein, were within normal limits. Stool culture, ova and parasite examination and PCR for *Clostridium difficile* toxin B and cytomegalovirus were all negative. Despite the intake of infliximab, abdominal symptoms persisted, and the patient underwent a total proctocolectomy with anal preservation and end-ileostomy.

On gross examination, the proctocolectomy specimen contained numerous sessile reddish polyps measuring up to 30 mm in diameter (Figure 1A). These polyps were all on the apices of transverse mucosal folds (Figure 1B). A 20-mm cystic intramural lesion was identified in the rectum (Figure 1C,D). Further examination revealed no signs of diverticulosis. Histologically, the surface of the polyps was ulcerated and covered by a cap of fibrinopurulent exudate (Figure 2A). Polyps were composed of elongated crypts, without dysplasia. The lamina propria contained a mixed inflammatory cell infiltrate and smooth muscle bundles arranged perpendicularly to the muscularis mucosae (Figure 2B). The cystic structure corresponded to a localized form of colitis cystica profunda (CCP)-type lesion, being located in the submucosa (Figure 2C). Its wall was lined by partially denuded, benign mucinous epithelium, and contained some calcifications (Figure 2D). The mucosa between the polypoid lesions was normal, which excluded the diagnosis of ulcerative colitis made preoperatively. The loop ileostomy was closed 3 months after proctocolectomy.

Cap polyposis is a poorly recognized condition with distinct clinical, colonoscopic and pathological features. This entity is characterized by multiple inflammatory colonic polyps covered by a cap of granulation tissue; hence the name 'cap polyposis'.[3] Its

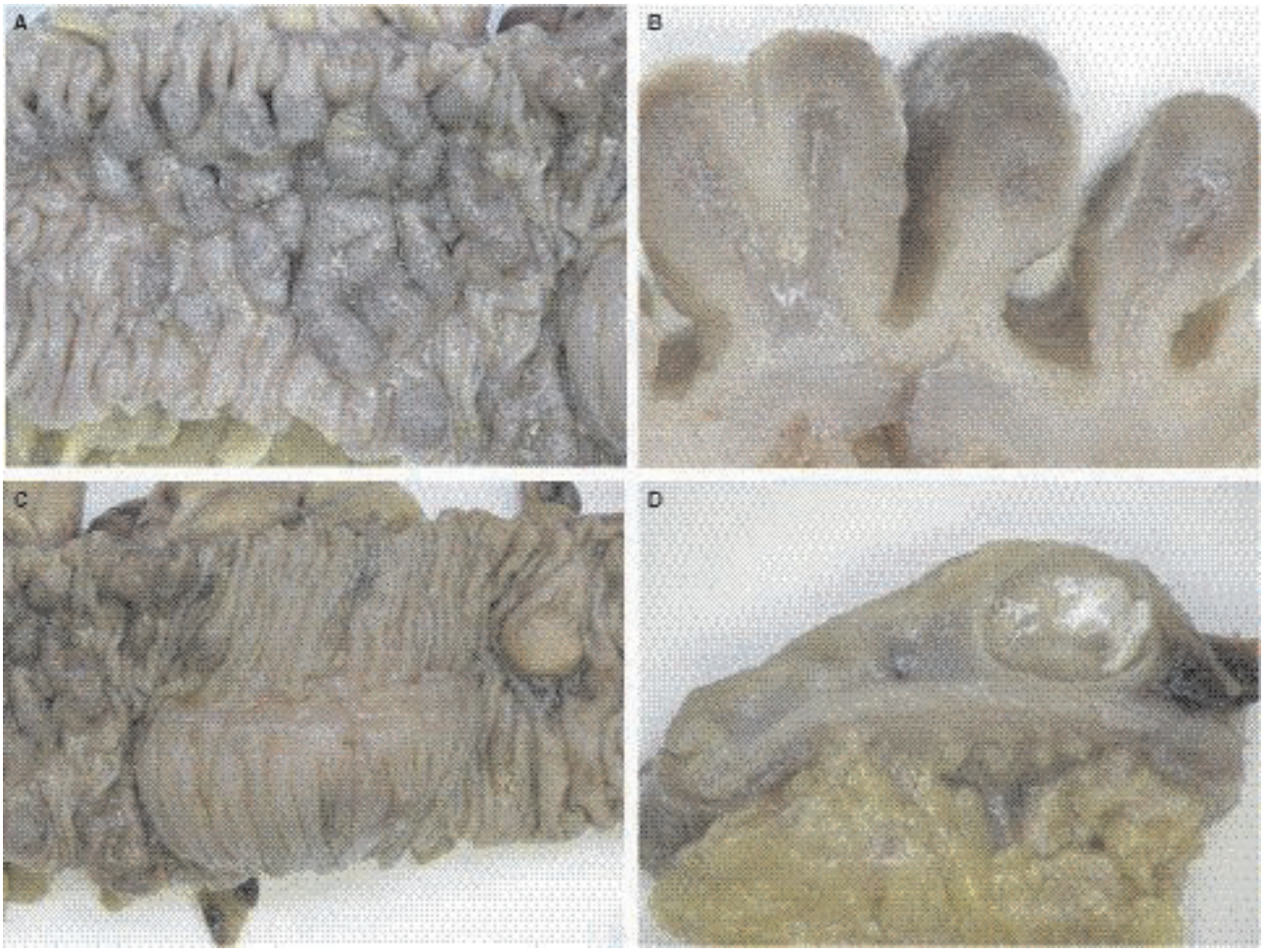


Figure 1. **A**, Proctocolectomy specimen showing rectosigmoid polyps. These polyps are covered by a cap of fibrinopurulent exudate. **B**, A cross-section of the colonic wall with three transverse mucosal folds. The polypoid mucosa is thicker than the non-polypoid mucosa. The crypts of the former are longer than those of the latter. **C**, Proctocolectomy specimen: the right half of the field shows a 20-mm polypoid and rubbery mass in the rectum, and the other half shows cap polyps in the rectosigmoid area. **D**, A cross-section of the rectal mass showing a submucosal cystic lesion containing mucoid material.

epidemiology has not been well investigated, and data have been obtained from occasional case series.[1,4] It has been described in both men and women, aged 12–76 years. The most frequent symptoms include mucous diarrhoea, rectal bleeding, and chronic straining at stool. Cap polyposis predominantly affects the rectosigmoid area, but has been reported to extend as far as the caecum. Histopathological examination of representative biopsies can confirm the diagnosis. The differential diagnosis mainly includes pseudomembranous colitis, idiopathic chronic inflammatory bowel disease, juvenile polyps, and any other disorders where mucosal prolapse is the underlying mechanism.

Clues supporting cap polyps are the typical gross appearance, the muscularization of the lamina propria, the absence of dysplasia, and the absence of ulcerative colitis in the background mucosa.

On the basis of the histological similarities between cap polyposis and 'mucosal prolapse syndrome', chronic mechanical stimulation of the mucosa by abnormal rectosigmoid motility has been proposed as an important aetiological factor for the pathogenesis of cap polyposis.[5,6] However, the improvement of some cases after treatment with anti-inflammatory agents has led some authors to suggest mucosal inflammation as another aetiological theory.

Cap polyposis is not known to be a premalignant condition. The optimal treatment has not been established, and management mainly relies on endoscopic polypectomy or proctocolectomy, depending on the

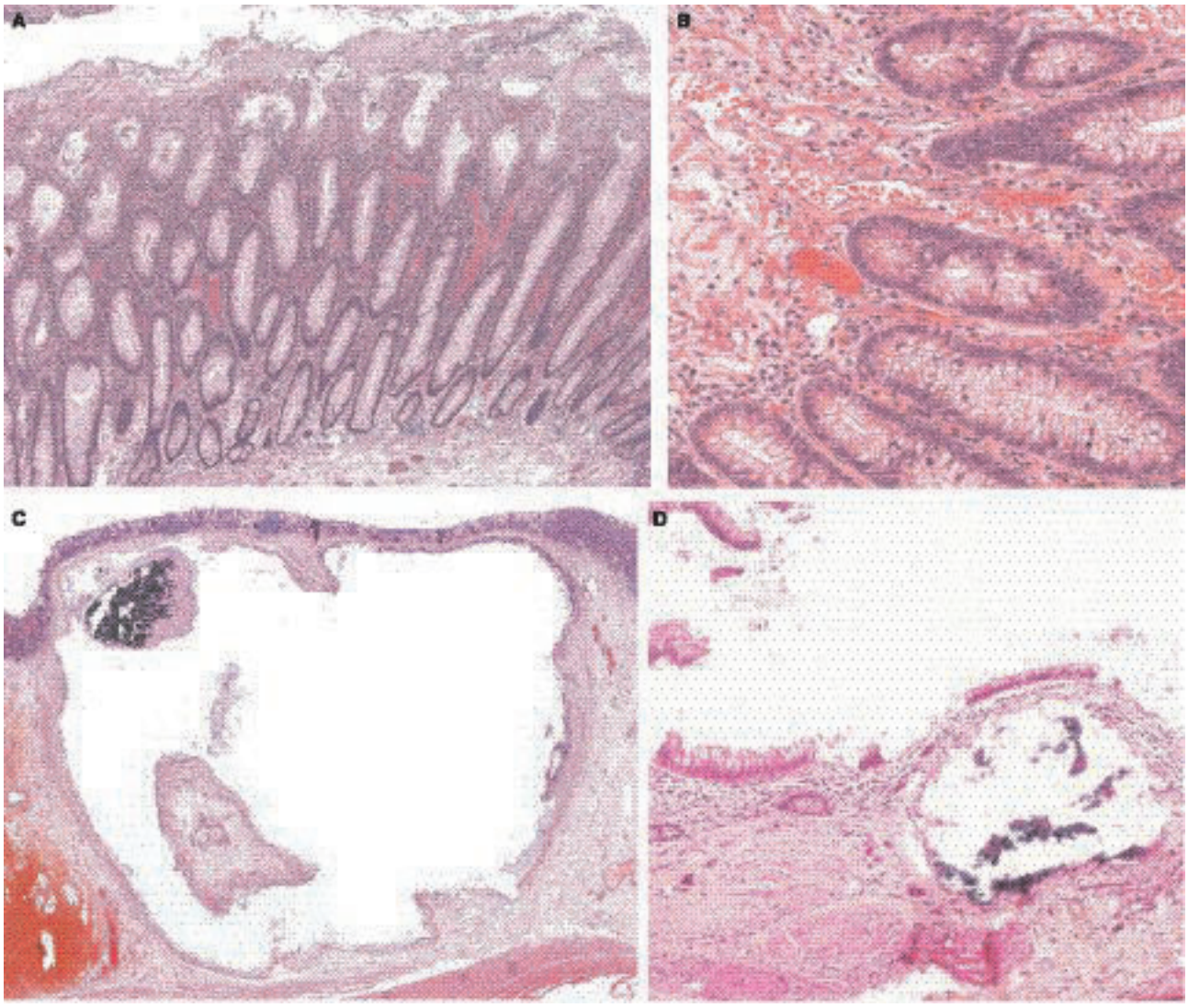


Figure 2. **A**, This low-power view shows the histological appearance of a colonic polyp with the cap of fibrinopurulent exudate. The polyp is composed of elongated crypts without dysplasia. **B**, This high-power view shows a mixed inflammatory cell infiltrate in the lamina propria, and smooth muscle bundles arranged perpendicularly to the muscularis mucosae. **C**, Very low-power view of the rectal cyst showing its relationship with the overlying mucosa. **D**, The wall of the rectal cyst is lined by partially denuded, benign mucinous epithelium, and contains some calcifications.

number of polyps (<10 versus >10). Symptomatic recurrence is, however, not uncommon.

Colitis cystica profunda is characterized by the presence of mucin-filled cysts in the submucosal layer of the colon.[7] Different theories exist for the aetiology of this disease. Because CCP is usually found in association with inflammatory bowel diseases, mucosal prolapse syndrome, or infectious colitis, its pathogenesis is thought to be related to mucosal ulceration resulting from physical trauma, ischaemia, or chronic mucosal inflammation. During the process of healing and re-epithelialization of ulcers, misplacement and entrapment of regenerating glands in the colon may occur.[8] Herniation of colonic mucosa into the submucosa because of congenital or acquired defects of the muscularis mucosae has also been postulated. Regardless of the mechanism, glands entrapped in the walls of the intestine may undergo dilatation and cystic change. They may show a loss of epithelium because of pressure atrophy. Acellular mucin pools are then left behind, and may undergo fibrosis, calcium deposition, or ossification.

Our case report is interesting, because the association of cap polyposis and a localized form of CCP supports the established mucosal prolapse theory as the aetiology for both conditions.[9]

Riad Arana[1,2]
Jean-François Fléjou[1,2]
Yann Parc[2,3]
Nizar El-Murr[2]
Jacques Cosnes[2,4]
Magali Svrcek[1,2]

[1]*AP-HP, Hôpitaux Universitaires Est Parisien, Hôpital Saint-Antoine, Service d'Anatomie et Cytologie Pathologiques,* [2]*Université Pierre et Marie Curie—Paris 6,* [3]*AP-HP, Hôpital Saint-Antoine, Service de Chirurgie Viscérale, and* [4]*AP-HP, Hôpital Saint-Antoine, Service de Gastroentérologie et Nutrition, Paris, France*

1. Papaconstantinou I, Karakatsanis A, Benia X, Polymeneas G, Kostopoulou E. Solitary rectal cap polyp: case report and review of the literature. *World J. Gastrointest. Surg.* 2012; **4**; 157–162.
2. Peny MO, Noel JC, Haot J *et al.* Cap polyposis: a rare syndrome. *Gastroenterol. Clin. Biol.* 1998; **22**; 349–352.
3. Williams GT, Bussey HJR, Morson BC. Inflammatory cap polyps of the large intestine. *Br. J. Surg.* 1985; **72**; S133.
4. Ng KH, Mathur P, Kumarasinghe MP, Eu KW, Seow-Choen F. Cap polyposis: further experience and review. *Dis. Colon Rectum* 2004; **47**; 1208–1215.
5. Campbell AP, Cobb CA, Chapman RW *et al.* Cap polyposis—an unusual cause of diarrhoea. *Gut* 1993; **34**; 562–564.
6. Singh B, Mortensen NJ, Warren BF. Histopathological mimicry in mucosal prolapse. *Histopathology* 2007; **50**; 97–102.
7. Kim WH, Choe GY, Kim YI, Kim JP. Localized form of colitis cystica profunda—a case of occurrence in the descending colon. *J. Korean Med. Sci.* 1992; **7**; 76–78.
8. De Petris G, Leung ST. Pseudoneoplasms of the gastrointestinal tract. *Arch. Pathol. Lab. Med.* 2010; **134**; 378–392.
9. Konishi T, Watanabe T, Takei Y *et al.* Cap polyposis: an inflammatory disorder or a spectrum of mucosal prolapse syndrome? *Gut* 2005; **54**; 1342–1343.

BIBLIOGRAPHIE

A

Aaltonen, L. A., P. Peltomaki, F. S. Leach, P. Sistonen, L. Pylkkanen, J. P. Mecklin, H. Jarvinen, S. M. Powell, J. Jen, S. R. Hamilton and et al. (1993). "Clues to the pathogenesis of familial colorectal cancer." *Science* **260**(5109): 812-816.

Alon, S., E. Mor, F. Vigneault, G. M. Church, F. Locatelli, F. Galeano, A. Gallo, N. Shomron and E. Eisenberg (2012). "Systematic identification of edited microRNAs in the human brain." *Genome Res* **22**(8): 1533-1540.

Ambros, V. (2008). "The evolution of our thinking about microRNAs." *Nat Med* **14**(10): 1036-1040.

Aravin, A. A., N. M. Naumova, A. V. Tulin, V. V. Vagin, Y. M. Rozovsky and V. A. Gvozdev (2001). "Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline." *Curr Biol* **11**(13): 1017-1027.

Asangani, I. A., S. A. Rasheed, D. A. Nikolova, J. H. Leupold, N. H. Colburn, S. Post and H. Allgayer (2008). "MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer." *Oncogene* **27**(15): 2128-2136.

Aust, D. E., J. P. Terdiman, R. F. Willenbucher, C. G. Chang, A. Molinaro-Clark, G. B. Baretton, U. Loehrs and F. M. Waldman (2002). "The APC/beta-catenin pathway in ulcerative colitis-related colorectal carcinomas: a mutational analysis." *Cancer* **94**(5): 1421-1427.

Auyeung, V. C., I. Ulitsky, S. E. McGeary and D. P. Bartel (2013). "Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing." *Cell* **152**(4): 844-858.

Azuma-Mukai, A., H. Oguri, T. Mituyama, Z. R. Qian, K. Asai, H. Siomi and M. C. Siomi (2008). "Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing." *Proc Natl Acad Sci U S A* **105**(23): 7964-7969.

B

Baek, D., J. Villen, C. Shin, F. D. Camargo, S. P. Gygi and D. P. Bartel (2008). "The impact of microRNAs on protein output." *Nature* **455**(7209): 64-71.

Bail, S., M. Swerdel, H. Liu, X. Jiao, L. A. Goff, R. P. Hart and M. Kiledjian (2010). "Differential regulation of microRNA stability." *RNA* **16**(5): 1032-1039.

Balaguer, F., L. Moreira, J. J. Lozano, A. Link, G. Ramirez, Y. Shen, M. Cuatrecasas, M. Arnold, S. J. Meltzer, S. Syngal, E. Stoffel, R. Jover, X. Llor, A. Castells, C. R. Boland, M. Gironella and

A. Goel (2011). "Colorectal cancers with microsatellite instability display unique miRNA profiles." *Clin Cancer Res* **17**(19): 6239-6249.

Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." *Cell* **136**(2): 215-233.

Baumgart, D. C. and S. R. Carding (2007). "Inflammatory bowel disease: cause and immunobiology." *Lancet* **369**(9573): 1627-1640.

Bernstein, E., A. A. Caudy, S. M. Hammond and G. J. Hannon (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." *Nature* **409**(6818): 363-366.

Bian, Z., L. Li, J. Cui, H. Zhang, Y. Liu, C. Y. Zhang and K. Zen (2011). "Role of miR-150-targeting c-Myb in colonic epithelial disruption during dextran sulphate sodium-induced murine experimental colitis and human ulcerative colitis." *J Pathol* **225**(4): 544-553.

Bohnsack, M. T., K. Czaplinski and D. Gorlich (2004). "Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs." *RNA* **10**(2): 185-191.

Boland, C. R. and A. Goel (2010). "Microsatellite instability in colorectal cancer." *Gastroenterology* **138**(6): 2073-2087 e2073.

Borchert, G. M., B. L. Gilmore, R. M. Spengler, Y. Xing, W. Lanier, D. Bhattacharya and B. L. Davidson (2009). "Adenosine deamination in human transcripts generates novel microRNA binding sites." *Hum Mol Genet* **18**(24): 4801-4807.

Borie, C., C. Colas, P. Dartigues, T. Lazure, P. Rince, O. Buhard, P. Folliot, A. Chalastanis, M. Muleris, R. Hamelin, D. Mercier, C. Oliveira, R. Seruca, A. Chadburn, V. Leblond, S. Barete, G. Gaidano, A. Martin, P. Gaulard, J. F. Flejou, M. Raphael and A. Duval (2009). "The mechanisms underlying MMR deficiency in immunodeficiency-related non-Hodgkin lymphomas are different from those in other sporadic microsatellite instable neoplasms." *Int J Cancer* **125**(10): 2360-2366.

Bouma, G. and W. Strober (2003). "The immunological and genetic basis of inflammatory bowel disease." *Nat Rev Immunol* **3**(7): 521-533.

Braakhuis, B. J., M. P. Tabor, J. A. Kummer, C. R. Leemans and R. H. Brakenhoff (2003). "A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications." *Cancer Res* **63**(8): 1727-1730.

Brennecke, J., A. Stark, R. B. Russell and S. M. Cohen (2005). "Principles of microRNA-target recognition." *PLoS Biol* **3**(3): e85.

Brest, P., P. Lapaquette, M. Souidi, K. Lebrigand, A. Cesaro, V. Vouret-Craviari, B. Mari, P. Barbry, J. F. Mosnier, X. Hebuterne, A. Harel-Bellan, B. Mograbi, A. Darfeuille-Michaud and P. Hofman (2011). "A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease." *Nat Genet* **43**(3): 242-245.

Brodersen, P. and O. Voinnet (2009). "Revisiting the principles of microRNA target recognition and mode of action." *Nat Rev Mol Cell Biol* **10**(2): 141-148.

Burroughs, A. M., Y. Ando, M. J. de Hoon, Y. Tomaru, T. Nishibu, R. Ukekawa, T. Funakoshi, T. Kurokawa, H. Suzuki, Y. Hayashizaki and C. O. Daub (2010). "A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness." *Genome Res* **20**(10): 1398-1410.

Burroughs, A. M., Y. Ando, M. J. de Hoon, Y. Tomaru, H. Suzuki, Y. Hayashizaki and C. O. Daub (2011). "Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin." *RNA Biol* **8**(1): 158-177.

C

Calin, G. A., M. Ferracin, A. Cimmino, G. Di Leva, M. Shimizu, S. E. Wojcik, M. V. Iorio, R. Visone, N. I. Sever, M. Fabbri, R. Iuliano, T. Palumbo, F. Pichiorri, C. Roldo, R. Garzon, C. Sevignani, L. Rassenti, H. Alder, S. Volinia, C. G. Liu, T. J. Kipps, M. Negrini and C. M. Croce (2005). "A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia." *N Engl J Med* **353**(17): 1793-1801.

Calin, G. A., C. Sevignani, C. D. Dumitru, T. Hyslop, E. Noch, S. Yendamuri, M. Shimizu, S. Rattan, F. Bullrich, M. Negrini and C. M. Croce (2004). "Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers." *Proc Natl Acad Sci U S A* **101**(9): 2999-3004.

Canavan, C., K. R. Abrams and J. Mayberry (2006). "Meta-analysis: colorectal and small bowel cancer risk in patients with Crohn's disease." *Aliment Pharmacol Ther* **23**(8): 1097-1104.

Cancer Genome Atlas Network (2012). "Comprehensive molecular characterization of human colon and rectal cancer." *Nature* **487**(7407): 330-337.

Capel, E., J. F. Flejou and R. Hamelin (2007). "Assessment of MLH1 promoter methylation in relation to gene expression requires specific analysis." *Oncogene* **26**(54): 7596-7600.

Cenik, E. S., R. Fukunaga, G. Lu, R. Dutcher, Y. Wang, T. M. Tanaka Hall and P. D. Zamore (2011). "Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease." *Mol Cell* **42**(2): 172-184.

Cenik, E. S. and P. D. Zamore (2011). "Argonaute proteins." *Curr Biol* **21**(12): R446-449.

Chai, H. and R. E. Brown (2009). "Field effect in cancer-an update." *Ann Clin Lab Sci* **39**(4): 331-337.

Chang, C. L., G. Marra, D. P. Chauhan, H. T. Ha, D. K. Chang, L. Ricciardiello, A. Randolph, J. M. Carethers and C. R. Boland (2002). "Oxidative stress inactivates the human DNA mismatch repair system." *Am J Physiol Cell Physiol* **283**(1): C148-154.

Chang, T. C., E. A. Wentzel, O. A. Kent, K. Ramachandran, M. Mullendore, K. H. Lee, G. Feldmann, M. Yamakuchi, M. Ferlito, C. J. Lowenstein, D. E. Arking, M. A. Beer, A. Maitra and J. T. Mendell (2007). "Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis." *Mol Cell* **26**(5): 745-752.

Cheloufi, S., C. O. Dos Santos, M. M. Chong and G. J. Hannon (2010). "A dicer-independent miRNA biogenesis pathway that requires Ago catalysis." *Nature* **465**(7298): 584-589.

Chen, C. Z., L. Li, H. F. Lodish and D. P. Bartel (2004). "MicroRNAs modulate hematopoietic lineage differentiation." *Science* **303**(5654): 83-86.

Chiang, H. R., L. W. Schoenfeld, J. G. Ruby, V. C. Auyeung, N. Spies, D. Baek, W. K. Johnston, C. Russ, S. Luo, J. E. Babiarz, R. Blelloch, G. P. Schroth, C. Nusbaum and D. P. Bartel (2010). "Mammalian microRNAs: experimental evaluation of novel and previously annotated genes." *Genes Dev* **24**(10): 992-1009.

Chiba, T., H. Marusawa and T. Ushijima (2012). "Inflammation-associated cancer development in digestive organs: mechanisms and roles for genetic and epigenetic modulation." *Gastroenterology* **143**(3): 550-563.

Chin, L. J., E. Ratner, S. Leng, R. Zhai, S. Nallur, I. Babar, R. U. Muller, E. Straka, L. Su, E. A. Burki, R. E. Crowell, R. Patel, T. Kulkarni, R. Homer, D. Zelterman, K. K. Kidd, Y. Zhu, D. C. Christiani, S. A. Belinsky, F. J. Slack and J. B. Weidhaas (2008). "A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk." *Cancer Res* **68**(20): 8535-8540.

Chiosea, S., E. Jelezcova, U. Chandran, J. Luo, G. Mantha, R. W. Sobol and S. Dacic (2007). "Overexpression of Dicer in precursor lesions of lung adenocarcinoma." *Cancer Res* **67**(5): 2345-2350.

Choudhury, Y., F. C. Tay, D. H. Lam, E. Sandanaraj, C. Tang, B. T. Ang and S. Wang (2012). "Attenuated adenosine-to-inosine editing of microRNA-376a* promotes invasiveness of glioblastoma cells." *J Clin Invest* **122**(11): 4059-4076.

Christensen, B. C., B. J. Moyer, M. Avissar, L. G. Ouellet, S. L. Plaza, M. D. McClean, C. J. Marsit and K. T. Kelsey (2009). "A let-7 microRNA-binding site polymorphism in the KRAS 3' UTR is associated with reduced survival in oral cancers." *Carcinogenesis* **30**(6): 1003-1007.

Cloonan, N., S. Wani, Q. Xu, J. Gu, K. Lea, S. Heater, C. Barbacioru, A. L. Steptoe, H. C. Martin, E. Nourbakhsh, K. Krishnan, B. Gardiner, X. Wang, K. Nones, J. A. Steen, N. A. Matigian, D. L. Wood, K. S. Kassahn, N. Waddell, J. Shepherd, C. Lee, J. Ichikawa, K. McKernan, K. Bramlett, S. Kuersten and S. M. Grimmond (2011). "MicroRNAs and their isomiRs function cooperatively to target common biological pathways." *Genome Biol* **12**(12): R126.

Cogoni, C. and G. Macino (1999). "Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase." *Nature* **399**(6732): 166-169.

Contreras, J. and D. S. Rao (2012). "MicroRNAs in inflammation and immune responses." *Leukemia* **26**(3): 404-413.

Cooper, H. S., S. Murthy, K. Kido, H. Yoshitake and A. Flanigan (2000). "Dysplasia and cancer in the dextran sulfate sodium mouse colitis model. Relevance to colitis-associated neoplasia in the human: a study of histopathology, B-catenin and p53 expression and the role of inflammation." *Carcinogenesis* **21**(4): 757-768.

Corney, D. C., A. Flesken-Nikitin, A. K. Godwin, W. Wang and A. Y. Nikitin (2007). "MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth." *Cancer Res* **67**(18): 8433-8438.

Coskun, M., J. T. Bjerrum, J. B. Seidelin and O. H. Nielsen (2012). "MicroRNAs in inflammatory bowel disease--pathogenesis, diagnostics and therapeutics." *World J Gastroenterol* **18**(34): 4629-4634.

Croce, C. M. (2009). "Causes and consequences of microRNA dysregulation in cancer." *Nat Rev Genet* **10**(10): 704-714.

Cuilliere-Dartigues, P., J. El-Bchiri, A. Krimi, O. Buhard, P. Fontanges, J. F. Flejou, R. Hamelin and A. Duval (2006). "TCF-4 isoforms absent in TCF-4 mutated MSI-H colorectal cancer cells colocalize with nuclear CtBP and repress TCF-4-mediated transcription." *Oncogene* **25**(32): 4441-4448.

Czech, B. and G. J. Hannon (2011). "Small RNA sorting: matchmaking for Argonautes." *Nat Rev Genet* **12**(1): 19-31.

Czech, B., R. Zhou, Y. Erlich, J. Brennecke, R. Binari, C. Villalta, A. Gordon, N. Perrimon and G. J. Hannon (2009). "Hierarchical rules for Argonaute loading in Drosophila." *Mol Cell* **36**(3): 445-456.

D

Dalal, S. R. and J. H. Kwon (2010). "The Role of MicroRNA in Inflammatory Bowel Disease." *Gastroenterol Hepatol (N Y)* **6**(11): 714-722.

Davis-Dusenbery, B. N. and A. Hata (2010). "MicroRNA in Cancer: The Involvement of Aberrant MicroRNA Biogenesis Regulatory Pathways." *Genes Cancer* **1**(11): 1100-1114.

Davis, B. N., A. C. Hilyard, G. Lagna and A. Hata (2008). "SMAD proteins control DROSHA-mediated microRNA maturation." *Nature* **454**(7200): 56-61.

de Hoon, M. J., R. J. Taft, T. Hashimoto, M. Kanamori-Katayama, H. Kawaji, M. Kawano, M. Kishima, T. Lassmann, G. J. Faulkner, J. S. Mattick, C. O. Daub, P. Carninci, J. Kawai, H. Suzuki and Y. Hayashizaki (2010). "Cross-mapping and the identification of editing sites in mature microRNAs in high-throughput sequencing libraries." *Genome Res* **20**(2): 257-264.

Deng, G., S. Kakar and Y. S. Kim (2011). "MicroRNA-124a and microRNA-34b/c are frequently methylated in all histological types of colorectal cancer and polyps, and in the adjacent normal mucosa." *Oncol Lett* **2**(1): 175-180.

Diederichs, S. and D. A. Haber (2006). "Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing." *Cancer Res* **66**(12): 6097-6104.

Doench, J. G. and P. A. Sharp (2004). "Specificity of microRNA target selection in translational repression." *Genes Dev* **18**(5): 504-511.

Dominissini, D., N. Amariglio and G. Rechavi (2012). "Micro-editing mistake translates into a devastating brain tumor." *J Clin Invest* **122**(11): 3842-3845.

Dominissini, D., S. Moshitch-Moshkovitz, N. Amariglio and G. Rechavi (2011). "Adenosine-to-inosine RNA editing meets cancer." *Carcinogenesis* **32**(11): 1569-1577.

Donehower, L. A., C. J. Creighton, N. Schultz, E. Shinbrot, K. Chang, P. H. Gunaratne, D. Muzny, C. Sander, S. R. Hamilton, R. A. Gibbs and D. Wheeler (2013). "MLH1-silenced and non-silenced subgroups of hypermutated colorectal carcinomas have distinct mutational landscapes." *J Pathol* **229**(1): 99-110.

Dorard, C., A. de Thonel, A. Collura, L. Marisa, M. Svrcek, A. Lagrange, G. Jego, K. Wanherdrick, A. L. Joly, O. Buhard, J. Gobbo, V. Penard-Lacronique, H. Zouali, E. Tubacher, S. Kirzin, J. Selves, G. Milano, M. C. Etienne-Grimaldi, L. Bengrine-Lefevre, C. Louvet, C. Tournigand, J. H. Lefevre, Y. Parc, E. Tiret, J. F. Flejou, M. P. Gaub, C. Garrido and A. Duval (2011). "Expression of a mutant HSP110 sensitizes colorectal cancer cells to chemotherapy and improves disease prognosis." *Nat Med* **17**(10): 1283-1289.

Duan, R., C. Pak and P. Jin (2007). "Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA." *Hum Mol Genet* **16**(9): 1124-1131.

Dueck, A., C. Ziegler, A. Eichner, E. Berezikov and G. Meister (2012). "microRNAs associated with the different human Argonaute proteins." *Nucleic Acids Res* **40**(19): 9850-9862.

Duttagupta, R., S. DiRienzo, R. Jiang, J. Bowers, J. Gollub, J. Kao, K. Kearney, D. Rudolph, N. B. Dawany, M. K. Showe, T. Stamato, R. C. Getts and K. W. Jones (2012). "Genome-wide maps of circulating miRNA biomarkers for ulcerative colitis." *PLoS One* **7**(2): e31241.

Duval, A. and R. Hamelin (2002). "Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability." *Cancer Res* **62**(9): 2447-2454.

Duval, A., B. Iacopetta, G. N. Ranzani, R. A. Lothe, G. Thomas and R. Hamelin (1999). "Variable mutation frequencies in coding repeats of TCF-4 and other target genes in colon, gastric and endometrial carcinoma showing microsatellite instability." *Oncogene* **18**(48): 6806-6809.

Duval, A., S. Rolland, E. Tubacher, H. Bui, G. Thomas and R. Hamelin (2000). "The human T-cell transcription factor-4 gene: structure, extensive characterization of alternative splicings, and mutational analysis in colorectal cancer cell lines." *Cancer Res* **60**(14): 3872-3879.

Dyson, J. K. and M. D. Rutter (2012). "Colorectal cancer in inflammatory bowel disease: what is the real magnitude of the risk?" *World J Gastroenterol* **18**(29): 3839-3848.

E

Eaden, J. A., K. R. Abrams and J. F. Mayberry (2001). "The risk of colorectal cancer in ulcerative colitis: a meta-analysis." *Gut* **48**(4): 526-535.

Ekbom, A., C. Helmick, M. Zack and H. O. Adami (1990). "Increased risk of large-bowel cancer in Crohn's disease with colonic involvement." *Lancet* **336**(8711): 357-359.

Ekbom, A., C. Helmick, M. Zack and H. O. Adami (1990). "Ulcerative colitis and colorectal cancer. A population-based study." *N Engl J Med* **323**(18): 1228-1233.

Ekdahl, Y., H. S. Farahani, M. Behm, J. Lagergren and M. Ohman (2012). "A-to-I editing of microRNAs in the mammalian brain increases during development." *Genome Res* **22**(8): 1477-1487.

Elkayam, E., C. D. Kuhn, A. Tocilj, A. D. Haase, E. M. Greene, G. J. Hannon and L. Joshua-Tor (2012). "The structure of human argonaute-2 in complex with miR-20a." *Cell* **150**(1): 100-110.

Ender, C. and G. Meister (2010). "Argonaute proteins at a glance." *J Cell Sci* **123**(Pt 11): 1819-1823.

Esquela-Kerscher, A. and F. J. Slack (2006). "Oncomirs - microRNAs with a role in cancer." *Nat Rev Cancer* **6**(4): 259-269.

F

Fabbri, M., A. Bottoni, M. Shimizu, R. Spizzo, M. S. Nicoloso, S. Rossi, E. Barbarotto, A. Cimmino, B. Adair, S. E. Wojcik, N. Valeri, F. Calore, D. Sampath, F. Fanini, I. Vannini, G. Musuraca, M. Dell'Aquila, H. Alder, R. V. Davuluri, L. Z. Rassenti, M. Negrini, T. Nakamura, D. Amadori, N. E. Kay, K. R. Rai, M. J. Keating, T. J. Kipps, G. A. Calin and C. M. Croce (2011). "Association of a microRNA/TP53 feedback circuitry with pathogenesis and outcome of B-cell chronic lymphocytic leukemia." *JAMA* **305**(1): 59-67.

Faber, C., D. Horst, F. Hlubek and T. Kirchner (2011). "Overexpression of Dicer predicts poor survival in colorectal cancer." *Eur J Cancer* **47**(9): 1414-1419.

Faber, C., T. Kirchner and F. Hlubek (2009). "The impact of microRNAs on colorectal cancer." *Virchows Arch* **454**(4): 359-367.

Faggad, A., A. Kasajima, W. Weichert, A. Stenzinger, N. E. Elwali, M. Dietel and C. Denkert (2012). "Down-regulation of the microRNA processing enzyme Dicer is a prognostic factor in human colorectal cancer." *Histopathology*.

Fallik, D., F. Borrini, V. Boige, J. Viguier, S. Jacob, C. Miquel, J. C. Sabourin, M. Ducreux and F. Praz (2003). "Microsatellite instability is a predictive factor of the tumor response to irinotecan in patients with advanced colorectal cancer." *Cancer Res* **63**(18): 5738-5744.

Fasseu, M., X. Treton, C. Guichard, E. Pedruzzi, D. Cazals-Hatem, C. Richard, T. Aparicio, F. Daniel, J. C. Soule, R. Moreau, Y. Bouhnik, M. Laburthe, A. Groyer and E. Ogier-Denis (2010). "Identification of restricted subsets of mature microRNA abnormally expressed in inactive colonic mucosa of patients with inflammatory bowel disease." *PLoS One* **5**(10).

Fearon, E. R. and B. Vogelstein (1990). "A genetic model for colorectal tumorigenesis." *Cell* **61**(5): 759-767.

Fernandez-Valverde, S. L., R. J. Taft and J. S. Mattick (2010). "Dynamic isomiR regulation in Drosophila development." *RNA* **16**(10): 1881-1888.

Filipowicz, W., S. N. Bhattacharyya and N. Sonenberg (2008). "Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?" *Nat Rev Genet* **9**(2): 102-114.

Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver and C. C. Mello (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." *Nature* **391**(6669): 806-811.

Fishel, R., M. K. Lescoe, M. R. Rao, N. G. Copeland, N. A. Jenkins, J. Garber, M. Kane and R. Kolodner (1993). "The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer." *Cell* **75**(5): 1027-1038.

Fleisher, A. S., M. Esteller, N. Harpaz, A. Leytin, A. Rashid, Y. Xu, J. Liang, O. C. Stine, J. Yin, T. T. Zou, J. M. Abraham, D. Kong, K. T. Wilson, S. P. James, J. G. Herman and S. J. Meltzer (2000). "Microsatellite instability in inflammatory bowel disease-associated neoplastic lesions is associated with hypermethylation and diminished expression of the DNA mismatch repair gene, hMLH1." *Cancer Res* **60**(17): 4864-4868.

Fodde, R., R. Smits and H. Clevers (2001). "APC, signal transduction and genetic instability in colorectal cancer." *Nat Rev Cancer* **1**(1): 55-67.

Forstemann, K., M. D. Horwich, L. Wee, Y. Tomari and P. D. Zamore (2007). "Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1." *Cell* **130**(2): 287-297.

Fowler, A., D. Thomson, K. Giles, S. Maleki, E. Mreich, H. Wheeler, P. Leedman, M. Biggs, R. Cook, N. Little, B. Robinson and K. McDonald (2011). "miR-124a is frequently down-regulated in glioblastoma and is involved in migration and invasion." *Eur J Cancer* **47**(6): 953-963.

Frank, F., N. Sonenberg and B. Nagar (2010). "Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2." *Nature* **465**(7299): 818-822.

Friedman, R. C., K. K. Farh, C. B. Burge and D. P. Bartel (2009). "Most mammalian mRNAs are conserved targets of microRNAs." *Genome Res* **19**(1): 92-105.

Frigola, J., J. Song, C. Stirzaker, R. A. Hinshelwood, M. A. Peinado and S. J. Clark (2006). "Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band." *Nat Genet* **38**(5): 540-549.

G

Garre, P., P. Perez-Segura, E. Diaz-Rubio, T. Caldes and M. de la Hoya (2010). "Reassessing the TARBP2 mutation rate in hereditary nonpolyposis colorectal cancer." *Nat Genet* **42**(10): 817-818; author reply 818.

Gasche, C., C. L. Chang, J. Rhees, A. Goel and C. R. Boland (2001). "Oxidative stress increases frameshift mutations in human colorectal cancer cells." *Cancer Res* **61**(20): 7444-7448.

Ghildiyal, M., H. Seitz, M. D. Horwich, C. Li, T. Du, S. Lee, J. Xu, E. L. Kittler, M. L. Zapp, Z. Weng and P. D. Zamore (2008). "Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells." *Science* **320**(5879): 1077-1081.

Ghildiyal, M., J. Xu, H. Seitz, Z. Weng and P. D. Zamore (2010). "Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway." *RNA* **16**(1): 43-56.

Giovannucci, E. and S. Ogino (2005). "DNA methylation, field effects, and colorectal cancer." *J Natl Cancer Inst* **97**(18): 1317-1319.

Goel, A. and C. R. Boland (2012). "Epigenetics of colorectal cancer." *Gastroenterology* **143**(6): 1442-1460 e1441.

Gommans, W. M. (2012). "A-to-I editing of microRNAs: regulating the regulators?" *Semin Cell Dev Biol* **23**(3): 251-257.

Grady, W. M. and J. M. Carethers (2008). "Genomic and epigenetic instability in colorectal cancer pathogenesis." *Gastroenterology* **135**(4): 1079-1099.

Grady, W. M., R. K. Parkin, P. S. Mitchell, J. H. Lee, Y. H. Kim, K. D. Tsuchiya, M. K. Washington, C. Paraskeva, J. K. Willson, A. M. Kaz, E. M. Kroh, A. Allen, B. R. Fritz, S. D. Markowitz and M. Tewari (2008). "Epigenetic silencing of the intronic microRNA hsa-miR-342 and its host gene EVL in colorectal cancer." *Oncogene* **27**(27): 3880-3888.

Gregory, R. I., T. P. Chendrimada, N. Cooch and R. Shiekhattar (2005). "Human RISC couples microRNA biogenesis and posttranscriptional gene silencing." *Cell* **123**(4): 631-640.

Gregory, R. I., K. P. Yan, G. Amuthan, T. Chendrimada, B. Doratotaj, N. Cooch and R. Shiekhattar (2004). "The Microprocessor complex mediates the genesis of microRNAs." *Nature* **432**(7014): 235-240.

Griffiths-Jones, S., H. K. Saini, S. van Dongen and A. J. Enright (2008). "miRBase: tools for microRNA genomics." *Nucleic Acids Res* **36**(Database issue): D154-158.

Grimson, A., K. K. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim and D. P. Bartel (2007). "MicroRNA targeting specificity in mammals: determinants beyond seed pairing." *Mol Cell* **27**(1): 91-105.

Grosshans, H. and I. Bussing (2010). "MicroRNA biogenesis takes another single hit from microsatellite instability." *Cancer Cell* **18**(4): 295-297.

Gryfe, R., H. Kim, E. T. Hsieh, M. D. Aronson, E. J. Holowaty, S. B. Bull, M. Redston and S. Gallinger (2000). "Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer." *N Engl J Med* **342**(2): 69-77.

Guo, H., N. T. Ingolia, J. S. Weissman and D. P. Bartel (2010). "Mammalian microRNAs predominantly act to decrease target mRNA levels." *Nature* **466**(7308): 835-840.

Guttler, T. and D. Gorlich (2011). "Ran-dependent nuclear export mediators: a structural perspective." *EMBO J* **30**(17): 3457-3474.

H

Hamelin, R., A. Chalastanis, C. Colas, J. El Bchiri, D. Mercier, A. S. Schreurs, V. Simon, M. Svrcek, A. Zaanan, C. Borie, O. Buhard, E. Capel, H. Zouali, F. Praz, M. Muleris, J. F. Flejou and A. Duval (2008). "[Clinical and molecular consequences of microsatellite instability in human cancers]." *Bull Cancer* **95**(1): 121-132.

Hampel, H., W. L. Frankel, E. Martin, M. Arnold, K. Khanduja, P. Kuebler, H. Nakagawa, K. Sotamaa, T. W. Prior, J. Westman, J. Panescu, D. Fix, J. Lockman, I. Comeras and A. de la Chapelle (2005). "Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer)." *N Engl J Med* **352**(18): 1851-1860.

Han, J., Y. Lee, K. H. Yeom, J. W. Nam, I. Heo, J. K. Rhee, S. Y. Sohn, Y. Cho, B. T. Zhang and V. N. Kim (2006). "Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex." *Cell* **125**(5): 887-901.

Hartner, J. C., C. R. Walkley, J. Lu and S. H. Orkin (2009). "ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling." *Nat Immunol* **10**(1): 109-115.

He, L. and G. J. Hannon (2004). "MicroRNAs: small RNAs with a big role in gene regulation." *Nat Rev Genet* **5**(7): 522-531.

He, L., J. M. Thomson, M. T. Hemann, E. Hernando-Monge, D. Mu, S. Goodson, S. Powers, C. Cordon-Cardo, S. W. Lowe, G. J. Hannon and S. M. Hammond (2005). "A microRNA polycistron as a potential human oncogene." *Nature* **435**(7043): 828-833.

Heo, I., M. Ha, J. Lim, M. J. Yoon, J. E. Park, S. C. Kwon, H. Chang and V. N. Kim (2012). "Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs." *Cell* **151**(3): 521-532.

Hitchins, M. P., V. A. Lin, A. Buckle, K. Cheong, N. Halani, S. Ku, C. T. Kwok, D. Packham, C. M. Suter, A. Meagher, C. Stirzaker, S. Clark, N. J. Hawkins and R. L. Ward (2007). "Epigenetic inactivation of a cluster of genes flanking MLH1 in microsatellite-unstable colorectal cancer." *Cancer Res* **67**(19): 9107-9116.

Hrasovec, S. and D. Glavac (2012). "MicroRNAs as Novel Biomarkers in Colorectal Cancer." *Front Genet* **3**: 180.

Hu, H. Y., Z. Yan, Y. Xu, H. Hu, C. Menzel, Y. H. Zhou, W. Chen and P. Khaitovich (2009). "Sequence features associated with microRNA strand selection in humans and flies." *BMC Genomics* **10**: 413.

Hu, Z., J. Chen, T. Tian, X. Zhou, H. Gu, L. Xu, Y. Zeng, R. Miao, G. Jin, H. Ma, Y. Chen and H. Shen (2008). "Genetic variants of miRNA sequences and non-small cell lung cancer survival." *J Clin Invest* **118**(7): 2600-2608.

Huntzinger, E. and E. Izaurralde (2011). "Gene silencing by microRNAs: contributions of translational repression and mRNA decay." *Nat Rev Genet* **12**(2): 99-110.

Hussain, S. P., P. Amstad, K. Raja, S. Ambs, M. Nagashima, W. P. Bennett, P. G. Shields, A. J. Ham, J. A. Swenberg, A. J. Marrogi and C. C. Harris (2000). "Increased p53 mutation load in noncancerous colon tissue from ulcerative colitis: a cancer-prone chronic inflammatory disease." *Cancer Res* **60**(13): 3333-3337.

Hutvagner, G., J. McLachlan, A. E. Pasquinelli, E. Balint, T. Tuschl and P. D. Zamore (2001). "A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA." *Science* **293**(5531): 834-838.

Hwang, H. W., E. A. Wentzel and J. T. Mendell (2007). "A hexanucleotide element directs microRNA nuclear import." *Science* **315**(5808): 97-100.

I

Iborra, M., F. Bernuzzi, P. Invernizzi and S. Danese (2012). "MicroRNAs in autoimmunity and inflammatory bowel disease: crucial regulators in immune response." *Autoimmun Rev* **11**(5): 305-314.

Iliopoulos, D., H. A. Hirsch and K. Struhl (2009). "An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation." *Cell* **139**(4): 693-706.

Ionov, Y., M. A. Peinado, S. Malkhosyan, D. Shibata and M. Perucho (1993). "Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis." *Nature* **363**(6429): 558-561.

Issa, J. P., N. Ahuja, M. Toyota, M. P. Bronner and T. A. Brentnall (2001). "Accelerated age-related CpG island methylation in ulcerative colitis." *Cancer Res* **61**(9): 3573-3577.

Itzkowitz, S. H. and X. Yio (2004). "Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation." *Am J Physiol Gastrointest Liver Physiol* **287**(1): G7-17.

J

Jazdzewski, K., E. L. Murray, K. Franssila, B. Jarzab, D. R. Schoenberg and A. de la Chapelle (2008). "Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma." *Proc Natl Acad Sci U S A* **105**(20): 7269-7274.

Ji, Q., X. Hao, Y. Meng, M. Zhang, J. Desano, D. Fan and L. Xu (2008). "Restoration of tumor suppressor miR-34 inhibits human p53-mutant gastric cancer tumorspheres." *BMC Cancer* **8**: 266.

Johnson, S. M., H. Grosshans, J. Shingara, M. Byrom, R. Jarvis, A. Cheng, E. Labourier, K. L. Reinert, D. Brown and F. J. Slack (2005). "RAS is regulated by the let-7 microRNA family." *Cell* **120**(5): 635-647.

K

Kasinski, A. L. and F. J. Slack (2011). "Epigenetics and genetics. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy." *Nat Rev Cancer* **11**(12): 849-864.

Katoh, T., Y. Sakaguchi, K. Miyauchi, T. Suzuki, S. Kashiwabara and T. Baba (2009). "Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2." *Genes Dev* **23**(4): 433-438.

Kawahara, Y., M. Megraw, E. Kreider, H. Iizasa, L. Valente, A. G. Hatzigeorgiou and K. Nishikura (2008). "Frequency and fate of microRNA editing in human brain." *Nucleic Acids Res* **36**(16): 5270-5280.

Kawahara, Y., B. Zinshteyn, T. P. Chendrimada, R. Shiekhattar and K. Nishikura (2007). "RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex." *EMBO Rep* **8**(8): 763-769.

Kawahara, Y., B. Zinshteyn, P. Sethupathy, H. Iizasa, A. G. Hatzigeorgiou and K. Nishikura (2007). "Redirection of silencing targets by adenosine-to-inosine editing of miRNAs." *Science* **315**(5815): 1137-1140.

Kawamata, T., H. Seitz and Y. Tomari (2009). "Structural determinants of miRNAs for RISC loading and slicer-independent unwinding." *Nat Struct Mol Biol* **16**(9): 953-960.

Khvorova, A., A. Reynolds and S. D. Jayasena (2003). "Functional siRNAs and miRNAs exhibit strand bias." *Cell* **115**(2): 209-216.

Kim, M. S., J. E. Oh, Y. R. Kim, S. W. Park, M. R. Kang, S. S. Kim, C. H. Ahn, N. J. Yoo and S. H. Lee (2010). "Somatic mutations and losses of expression of microRNA regulation-related genes AGO2 and TNRC6A in gastric and colorectal cancers." *J Pathol* **221**(2): 139-146.

Kim, V. N., J. Han and M. C. Siomi (2009). "Biogenesis of small RNAs in animals." *Nat Rev Mol Cell Biol* **10**(2): 126-139.

Kim, Y. K., I. Heo and V. N. Kim (2010). "Modifications of small RNAs and their associated proteins." *Cell* **143**(5): 703-709.

Kim, Y. K. and V. N. Kim (2007). "Processing of intronic microRNAs." *EMBO J* **26**(3): 775-783.

Kohler, A. and E. Hurt (2007). "Exporting RNA from the nucleus to the cytoplasm." *Nat Rev Mol Cell Biol* **8**(10): 761-773.

Koizumi, K., S. Alonso, Y. Miyaki, S. Okada, H. Ogura, N. Shiiya, F. Konishi, T. Taya, M. Perucho and K. Suzuki (2012). "Array-based identification of common DNA methylation alterations in ulcerative colitis." *Int J Oncol* **40**(4): 983-994.

Kozubek, J., Z. Ma, E. Fleming, T. Duggan, R. Wu, D. G. Shin and S. S. Dadras (2013). "In-depth characterization of microRNA transcriptome in melanoma." *PLoS One* **8**(9): e72699.

Kuchenbauer, F., R. D. Morin, B. Argiropoulos, O. I. Petriv, M. Griffith, M. Heuser, E. Yung, J. Piper, A. Delaney, A. L. Prabhu, Y. Zhao, H. McDonald, T. Zeng, M. Hirst, C. L. Hansen, M. A. Marra and R. K. Humphries (2008). "In-depth characterization of the microRNA transcriptome in a leukemia progression model." *Genome Res* **18**(11): 1787-1797.

Kumar, M. S., J. Lu, K. L. Mercer, T. R. Golub and T. Jacks (2007). "Impaired microRNA processing enhances cellular transformation and tumorigenesis." *Nat Genet* **39**(5): 673-677.

L

Ladewig, E., K. Okamura, A. S. Flynt, J. O. Westholm and E. C. Lai (2012). "Discovery of hundreds of mirtrons in mouse and human small RNA data." *Genome Res* **22**(9): 1634-1645.

Lagos-Quintana, M., R. Rauhut, W. Lendeckel and T. Tuschl (2001). "Identification of novel genes coding for small expressed RNAs." *Science* **294**(5543): 853-858.

Lai, E. C. (2002). "Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation." *Nat Genet* **30**(4): 363-364.

Lambertz, I., D. Nittner, P. Mestdagh, G. Denecker, J. Vandesompele, M. A. Dyer and J. C. Marine (2010). "Monoallelic but not biallelic loss of Dicer1 promotes tumorigenesis in vivo." *Cell Death Differ* **17**(4): 633-641.

Landgraf, P., M. Rusu, R. Sheridan, A. Sewer, N. Iovino, A. Aravin, S. Pfeffer, A. Rice, A. O. Kamphorst, M. Landthaler, C. Lin, N. D. Socci, L. Hermida, V. Fulci, S. Chiaretti, R. Foa, J. Schliwka, U. Fuchs, A. Novosel, R. U. Muller, B. Schermer, U. Bissels, J. Inman, Q. Phan, M. Chien, D. B. Weir, R. Choksi, G. De Vita, D. Frezzetti, H. I. Trompeter, V. Hornung, G. Teng, G. Hartmann, M. Palkovits, R. Di Lauro, P. Wernet, G. Macino, C. E. Rogler, J. W. Nagle, J. Ju, F. N. Papavasiliou, T. Benzing, P. Lichter, W. Tam, M. J. Brownstein, A. Bosio, A. Borkhardt, J. J. Russo, C. Sander, M. Zavolan and T. Tuschl (2007). "A mammalian microRNA expression atlas based on small RNA library sequencing." *Cell* **129**(7): 1401-1414.

Lanza, G., M. Ferracin, R. Gafa, A. Veronese, R. Spizzo, F. Pichiorri, C. G. Liu, G. A. Calin, C. M. Croce and M. Negrini (2007). "mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer." *Mol Cancer* **6**: 54.

Lau, P. W., K. Z. Guiley, N. De, C. S. Potter, B. Carragher and I. J. MacRae (2012). "The molecular architecture of human Dicer." *Nat Struct Mol Biol* **19**(4): 436-440.

Lau, P. W., C. S. Potter, B. Carragher and I. J. MacRae (2009). "Structure of the human Dicer-TRBP complex by electron microscopy." *Structure* **17**(10): 1326-1332.

Leach, F. S., N. C. Nicolaides, N. Papadopoulos, B. Liu, J. Jen, R. Parsons, P. Peltomaki, P. Sistonen, L. A. Aaltonen, M. Nystrom-Lahti and et al. (1993). "Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer." *Cell* **75**(6): 1215-1225.

Lee, H. C., J. G. Kim, Y. S. Chae, S. K. Sohn, B. W. Kang, J. H. Moon, S. W. Jeon, M. H. Lee, K. H. Lim, J. Y. Park, G. S. Choi and S. H. Jun (2010). "Prognostic impact of microRNA-related gene polymorphisms on survival of patients with colorectal cancer." *J Cancer Res Clin Oncol* **136**(7): 1073-1078.

Lee, R. C., R. L. Feinbaum and V. Ambros (1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14." *Cell* **75**(5): 843-854.

Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim and V. N. Kim (2003). "The nuclear RNase III Drosha initiates microRNA processing." *Nature* **425**(6956): 415-419.

Lee, Y., K. Jeon, J. T. Lee, S. Kim and V. N. Kim (2002). "MicroRNA maturation: stepwise processing and subcellular localization." *EMBO J* **21**(17): 4663-4670.

Lee, Y., M. Kim, J. Han, K. H. Yeom, S. Lee, S. H. Baek and V. N. Kim (2004). "MicroRNA genes are transcribed by RNA polymerase II." *EMBO J* **23**(20): 4051-4060.

Lewis, B. P., C. B. Burge and D. P. Bartel (2005). "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." *Cell* **120**(1): 15-20.

Lewis, B. P., I. H. Shih, M. W. Jones-Rhoades, D. P. Bartel and C. B. Burge (2003). "Prediction of mammalian microRNA targets." *Cell* **115**(7): 787-798.

Ligtenberg, M. J., R. P. Kuiper, T. L. Chan, M. Goossens, K. M. Hebeda, M. Voorendt, T. Y. Lee, D. Bodmer, E. Hoenselaar, S. J. Hendriks-Cornelissen, W. Y. Tsui, C. K. Kong, H. G. Brunner, A. G. van Kessel, S. T. Yuen, J. H. van Krieken, S. Y. Leung and N. Hoogerbrugge (2009). "Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1." *Nat Genet* **41**(1): 112-117.

Lim, L. P., N. C. Lau, E. G. Weinstein, A. Abdelhakim, S. Yekta, M. W. Rhoades, C. B. Burge and D. P. Bartel (2003). "The microRNAs of Caenorhabditis elegans." *Genes Dev* **17**(8): 991-1008.

Lu, J., G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz and T. R. Golub (2005). "MicroRNA expression profiles classify human cancers." *Nature* **435**(7043): 834-838.

Lujambio, A. and S. W. Lowe (2012). "The microcosmos of cancer." *Nature* **482**(7385): 347-355.

Lund, E., S. Guttinger, A. Calado, J. E. Dahlberg and U. Kutay (2004). "Nuclear export of microRNA precursors." *Science* **303**(5654): 95-98.

M

Ma, E., K. Zhou, M. A. Kidwell and J. A. Doudna (2012). "Coordinated activities of human dicer domains in regulatory RNA processing." *J Mol Biol* **422**(4): 466-476.

MacRae, I. J. and J. A. Doudna (2007). "Ribonuclease revisited: structural insights into ribonuclease III family enzymes." *Curr Opin Struct Biol* **17**(1): 138-145.

Macrae, I. J., K. Zhou, F. Li, A. Repic, A. N. Brooks, W. Z. Cande, P. D. Adams and J. A. Doudna (2006). "Structural basis for double-stranded RNA processing by Dicer." *Science* **311**(5758): 195-198.

Manne, U., C. Shanmugam, L. Bovell, V. R. Katkoori and H. L. Bumpers (2010). "miRNAs as biomarkers for management of patients with colorectal cancer." *Biomark Med* **4**(5): 761-770.

Mao, G., S. Lee, J. Ortega, L. Gu and G. M. Li (2012). "Modulation of microRNA processing by mismatch repair protein MutLalpha." *Cell Res* **22**(6): 973-985.

Mao, G., X. Pan and L. Gu (2008). "Evidence that a mutation in the MLH1 3'-untranslated region confers a mutator phenotype and mismatch repair deficiency in patients with relapsed leukemia." *J Biol Chem* **283**(6): 3211-3216.

Markowitz, S., J. Wang, L. Myeroff, R. Parsons, L. Sun, J. Lutterbaugh, R. S. Fan, E. Zborowska, K. W. Kinzler, B. Vogelstein and et al. (1995). "Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability." *Science* **268**(5215): 1336-1338.

Matranga, C., Y. Tomari, C. Shin, D. P. Bartel and P. D. Zamore (2005). "Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes." *Cell* **123**(4): 607-620.

Melo, S. A., C. Moutinho, S. Ropero, G. A. Calin, S. Rossi, R. Spizzo, A. F. Fernandez, V. Davalos, A. Villanueva, G. Montoya, H. Yamamoto, S. Schwartz, Jr. and M. Esteller (2010). "A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells." *Cancer Cell* **18**(4): 303-315.

Melo, S. A., S. Ropero, C. Moutinho, L. A. Aaltonen, H. Yamamoto, G. A. Calin, S. Rossi, A. F. Fernandez, F. Carneiro, C. Oliveira, B. Ferreira, C. G. Liu, A. Villanueva, G. Capella, S. Schwartz, Jr., R. Shiekhattar and M. Esteller (2009). "A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function." *Nat Genet* **41**(3): 365-370.

Merritt, W. M., Y. G. Lin, L. Y. Han, A. A. Kamat, W. A. Spannuth, R. Schmandt, D. Urbauer, L. A. Pennacchio, J. F. Cheng, A. M. Nick, M. T. Deavers, A. Mourad-Zeidan, H. Wang, P. Mueller, M. E. Lenburg, J. W. Gray, S. Mok, M. J. Birrer, G. Lopez-Berestein, R. L. Coleman, M. Bar-Eli and A. K. Sood (2008). "Dicer, Drosha, and outcomes in patients with ovarian cancer." *N Engl J Med* **359**(25): 2641-2650.

Michael, M. Z., O. C. SM, N. G. van Holst Pellekaan, G. P. Young and R. J. James (2003). "Reduced accumulation of specific microRNAs in colorectal neoplasia." *Mol Cancer Res* **1**(12): 882-891.

Miyoshi, K., T. Miyoshi and H. Siomi (2010). "Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production." *Mol Genet Genomics* **284**(2): 95-103.

Mongroo, P. S. and A. K. Rustgi (2010). "The role of the miR-200 family in epithelial-mesenchymal transition." *Cancer Biol Ther* **10**(3): 219-222.

Morin, R. D., M. D. O'Connor, M. Griffith, F. Kuchenbauer, A. Delaney, A. L. Prabhu, Y. Zhao, H. McDonald, T. Zeng, M. Hirst, C. J. Eaves and M. A. Marra (2008). "Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells." *Genome Res* **18**(4): 610-621.

Morlando, M., M. Ballarino, N. Gromak, F. Pagano, I. Bozzoni and N. J. Proudfoot (2008). "Primary microRNA transcripts are processed co-transcriptionally." *Nat Struct Mol Biol* **15**(9): 902-909.

N

Nagel, R., C. le Sage, B. Diosdado, M. van der Waal, J. A. Oude Vrielink, A. Bolijn, G. A. Meijer and R. Agami (2008). "Regulation of the adenomatous polyposis coli gene by the miR-135 family in colorectal cancer." *Cancer Res* **68**(14): 5795-5802.

Nam, Y., C. Chen, R. I. Gregory, J. J. Chou and P. Sliz (2011). "Molecular basis for interaction of let-7 microRNAs with Lin28." *Cell* **147**(5): 1080-1091.

Neilsen, C. T., G. J. Goodall and C. P. Bracken (2012). "IsomiRs--the overlooked repertoire in the dynamic microRNAome." *Trends Genet* **28**(11): 544-549.

Nguyen, H. T., G. Dalmasso, Y. Yan, H. Laroui, S. Dahan, L. Mayer, S. V. Sitaraman and D. Merlin (2010). "MicroRNA-7 modulates CD98 expression during intestinal epithelial cell differentiation." *J Biol Chem* **285**(2): 1479-1489.

Nicholson, A. W. (2012). "Dissecting human dicer: some assembly required." *J Mol Biol* **422**(4): 464-465.

Nielsen, C. B., N. Shomron, R. Sandberg, E. Hornstein, J. Kitzman and C. B. Burge (2007). "Determinants of targeting by endogenous and exogenous microRNAs and siRNAs." *RNA* **13**(11): 1894-1910.

Nishikura, K. (2006). "Editor meets silencer: crosstalk between RNA editing and RNA interference." *Nat Rev Mol Cell Biol* **7**(12): 919-931.

Noffsinger, A., S. Kretschmer, J. Belli, F. Fogt and C. Fenoglio-Preiser (2000). "Microsatellite instability is uncommon in intestinal mucosa of patients with Crohn's disease." *Dig Dis Sci* **45**(2): 378-384.

Noland, C. L., E. Ma and J. A. Doudna (2011). "siRNA repositioning for guide strand selection by human Dicer complexes." *Mol Cell* **43**(1): 110-121.

O

O'Connell, M. A. and L. P. Keegan (2006). "Drosha versus ADAR: wrangling over pri-miRNA." *Nat Struct Mol Biol* **13**(1): 3-4.

O'Connell, R. M., K. D. Taganov, M. P. Boldin, G. Cheng and D. Baltimore (2007). "MicroRNA-155 is induced during the macrophage inflammatory response." *Proc Natl Acad Sci U S A* **104**(5): 1604-1609.

Okada, C., E. Yamashita, S. J. Lee, S. Shibata, J. Katahira, A. Nakagawa, Y. Yoneda and T. Tsukihara (2009). "A high-resolution structure of the pre-microRNA nuclear export machinery." *Science* **326**(5957): 1275-1279.

Okamura, K., J. W. Hagen, H. Duan, D. M. Tyler and E. C. Lai (2007). "The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila." *Cell* **130**(1): 89-100.

Olena, A. F. and J. G. Patton (2010). "Genomic organization of microRNAs." *J Cell Physiol* **222**(3): 540-545.

Olive, V., I. Jiang and L. He (2010). "mir-17-92, a cluster of miRNAs in the midst of the cancer network." *Int J Biochem Cell Biol* **42**(8): 1348-1354.

Olsen, P. H. and V. Ambros (1999). "The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation." *Dev Biol* **216**(2): 671-680.

Orom, U. A., F. C. Nielsen and A. H. Lund (2008). "MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation." *Mol Cell* **30**(4): 460-471.

P

Papachristou, D. J., A. Korpetinou, E. Giannopoulou, A. G. Antonacopoulou, H. Papadaki, P. Grivas, C. D. Scopa and H. P. Kalofonos (2011). "Expression of the ribonucleases Drosha, Dicer, and Ago2 in colorectal carcinomas." *Virchows Arch* **459**(4): 431-440.

Paraskevi, A., G. Theodoropoulos, I. Papaconstantinou, G. Mantzaris, N. Nikiteas and M. Gazouli (2012). "Circulating MicroRNA in inflammatory bowel disease." *J Crohns Colitis* **6**(9): 900-904.

Park, J. E., I. Heo, Y. Tian, D. K. Simanshu, H. Chang, D. Jee, D. J. Patel and V. N. Kim (2011). "Dicer recognizes the 5' end of RNA for efficient and accurate processing." *Nature* **475**(7355): 201-205.

Pasquinelli, A. E. (2012). "MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship." *Nat Rev Genet* **13**(4): 271-282.

Pasquinelli, A. E., B. J. Reinhart, F. Slack, M. Q. Martindale, M. I. Kuroda, B. Maller, D. C. Hayward, E. E. Ball, B. Degnan, P. Muller, J. Spring, A. Srinivasan, M. Fishman, J. Finnerty, J. Corbo, M. Levine, P. Leahy, E. Davidson and G. Ruvkun (2000). "Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA." *Nature* **408**(6808): 86-89.

Paz, N., E. Y. Levanon, N. Amariglio, A. B. Heimberger, Z. Ram, S. Constantini, Z. S. Barbash, K. Adamsky, M. Safran, A. Hirschberg, M. Krupsky, I. Ben-Dov, S. Cazacu, T. Mikkelsen, C. Brodie, E. Eisenberg and G. Rechavi (2007). "Altered adenosine-to-inosine RNA editing in human cancer." *Genome Res* **17**(11): 1586-1595.

Pekow, J. R., U. Dougherty, R. Mustafi, H. Zhu, M. Kocherginsky, D. T. Rubin, S. B. Hanauer, J. Hart, E. B. Chang, A. Fichera, L. J. Joseph and M. Bissonnette (2012). "miR-143 and miR-145 are downregulated in ulcerative colitis: putative regulators of inflammation and protooncogenes." *Inflamm Bowel Dis* **18**(1): 94-100.

Peltomaki, P., L. A. Aaltonen, P. Sistonen, L. Pylkkanen, J. P. Mecklin, H. Jarvinen, J. S. Green, J. R. Jass, J. L. Weber, F. S. Leach and et al. (1993). "Genetic mapping of a locus predisposing to human colorectal cancer." *Science* **260**(5109): 810-812.

Peng, Z., Y. Cheng, B. C. Tan, L. Kang, Z. Tian, Y. Zhu, W. Zhang, Y. Liang, X. Hu, X. Tan, J. Guo, Z. Dong, L. Bao and J. Wang (2012). "Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome." *Nat Biotechnol* **30**(3): 253-260.

Pillai, R. S., S. N. Bhattacharyya, C. G. Artus, T. Zoller, N. Cougot, E. Basyuk, E. Bertrand and W. Filipowicz (2005). "Inhibition of translational initiation by Let-7 MicroRNA in human cells." *Science* **309**(5740): 1573-1576.

Pino, M. S. and D. C. Chung (2010). "The chromosomal instability pathway in colon cancer." *Gastroenterology* **138**(6): 2059-2072.

Podolsky, D. K. (2002). "Inflammatory bowel disease." *N Engl J Med* **347**(6): 417-429.

Popat, S., R. Hubner and R. S. Houlston (2005). "Systematic review of microsatellite instability and colorectal cancer prognosis." *J Clin Oncol* **23**(3): 609-618.

Q

Qi, Y., A. M. Denli and G. J. Hannon (2005). "Biochemical specialization within Arabidopsis RNA silencing pathways." *Mol Cell* **19**(3): 421-428.

Quach, H., L. B. Barreiro, G. Laval, N. Zidane, E. Patin, K. K. Kidd, J. R. Kidd, C. Bouchier, M. Veuille, C. Antoniewski and L. Quintana-Murci (2009). "Signatures of purifying and local positive selection in human miRNAs." *Am J Hum Genet* **84**(3): 316-327.

R

Rand, T. A., S. Petersen, F. Du and X. Wang (2005). "Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation." *Cell* **123**(4): 621-629.

Reid, J. G., A. K. Nagaraja, F. C. Lynn, R. B. Drabek, D. M. Muzny, C. A. Shaw, M. K. Weiss, A. O. Naghavi, M. Khan, H. Zhu, J. Tennakoon, G. H. Gunaratne, D. B. Corry, J. Miller, M. T. McManus, M. S. German, R. A. Gibbs, M. M. Matzuk and P. H. Gunaratne (2008). "Mouse let-7 miRNA populations exhibit RNA editing that is constrained in the 5'-seed/ cleavage/anchor regions and stabilize predicted mmu-let-7a:mRNA duplexes." *Genome Res* **18**(10): 1571-1581.

Reinhart, B. J., F. J. Slack, M. Basson, A. E. Pasquinelli, J. C. Bettinger, A. E. Rougvie, H. R. Horvitz and G. Ruvkun (2000). "The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans." *Nature* **403**(6772): 901-906.

Ribic, C. M., D. J. Sargent, M. J. Moore, S. N. Thibodeau, A. J. French, R. M. Goldberg, S. R. Hamilton, P. Laurent-Puig, R. Gryfe, L. E. Shepherd, D. Tu, M. Redston and S. Gallinger (2003). "Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer." *N Engl J Med* **349**(3): 247-257.

Rivenbark, A. G. and W. B. Coleman (2012). "Field cancerization in mammary carcinogenesis - Implications for prevention and treatment of breast cancer." *Exp Mol Pathol* **93**(3): 391-398.

Rodriguez, A., S. Griffiths-Jones, J. L. Ashurst and A. Bradley (2004). "Identification of mammalian microRNA host genes and transcription units." *Genome Res* **14**(10A): 1902-1910.

Rodriguez, A., E. Vigorito, S. Clare, M. V. Warren, P. Couttet, D. R. Soond, S. van Dongen, R. J. Grocock, P. P. Das, E. A. Miska, D. Vetrie, K. Okkenhaug, A. J. Enright, G. Dougan, M. Turner and A. Bradley (2007). "Requirement of bic/microRNA-155 for normal immune function." *Science* **316**(5824): 608-611.

Ruby, J. G., C. Jan, C. Player, M. J. Axtell, W. Lee, C. Nusbaum, H. Ge and D. P. Bartel (2006). "Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans." *Cell* **127**(6): 1193-1207.

Rutter, M., B. Saunders, K. Wilkinson, S. Rumbles, G. Schofield, M. Kamm, C. Williams, A. Price, I. Talbot and A. Forbes (2004). "Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis." *Gastroenterology* **126**(2): 451-459.

Ryan, B. M., A. I. Robles and C. C. Harris (2010). "Genetic variation in microRNA networks: the implications for cancer research." *Nat Rev Cancer* **10**(6): 389-402.

S

Sarver, A. L., A. J. French, P. M. Borralho, V. Thayanithy, A. L. Oberg, K. A. Silverstein, B. W. Morlan, S. M. Riska, L. A. Boardman, J. M. Cunningham, S. Subramanian, L. Wang, T. C. Smyrk, C. M. Rodrigues, S. N. Thibodeau and C. J. Steer (2009). "Human colon cancer profiles show differential microRNA expression depending on mismatch repair status and are characteristic of undifferentiated proliferative states." *BMC Cancer* **9**: 401.

Saunders, M. A., H. Liang and W. H. Li (2007). "Human polymorphism at microRNAs and microRNA target sites." *Proc Natl Acad Sci U S A* **104**(9): 3300-3305.

Sawh, A. N. and T. F. Duchaine (2012). "Turning Dicer on its head." *Nat Struct Mol Biol* **19**(4): 365-366.

Schepeler, T., J. T. Reinert, M. S. Ostenfeld, L. L. Christensen, A. N. Silahtaroglu, L. Dyrskjot, C. Wiuf, F. J. Sorensen, M. Kruhoffer, S. Laurberg, S. Kauppinen, T. F. Orntoft and C. L. Andersen (2008). "Diagnostic and prognostic microRNAs in stage II colon cancer." *Cancer Res* **68**(15): 6416-6424.

Schetter, A. J. and C. C. Harris (2011). "Alterations of microRNAs contribute to colon carcinogenesis." *Semin Oncol* **38**(6): 734-742.

Schirle, N. T. and I. J. MacRae (2012). "The crystal structure of human Argonaute2." *Science* **336**(6084): 1037-1040.

Schulmann, K., Y. Mori, V. Croog, J. Yin, A. Olaru, A. Sterian, F. Sato, S. Wang, Y. Xu, E. Deacu, A. T. Berki, J. P. Hamilton, T. Kan, J. M. Abraham, W. Schmiegel, N. Harpaz and S. J. Meltzer (2005). "Molecular phenotype of inflammatory bowel disease-associated neoplasms with microsatellite instability." *Gastroenterology* **129**(1): 74-85.

Schwarz, D. S., G. Hutvagner, T. Du, Z. Xu, N. Aronin and P. D. Zamore (2003). "Asymmetry in the assembly of the RNAi enzyme complex." *Cell* **115**(2): 199-208.

Scott, D. D. and C. J. Norbury (2013). "RNA decay via 3' uridylation." *Biochim Biophys Acta*.

Seitz, H., M. Ghildiyal and P. D. Zamore (2008). "Argonaute loading improves the 5' precision of both MicroRNAs and their miRNA* strands in flies." *Curr Biol* **18**(2): 147-151.

Seitz, H., J. S. Tushir and P. D. Zamore (2011). "A 5'-uridine amplifies miRNA/miRNA* asymmetry in Drosophila by promoting RNA-induced silencing complex formation." *Silence* **2**: 4.

Shen, L., Y. Kondo, G. L. Rosner, L. Xiao, N. S. Hernandez, J. Vilaythong, P. S. Houlihan, R. S. Krouse, A. R. Prasad, J. G. Einspahr, J. Buckmeier, D. S. Alberts, S. R. Hamilton and J. P. Issa (2005). "MGMT promoter methylation and field defect in sporadic colorectal cancer." *J Natl Cancer Inst* **97**(18): 1330-1338.

Sinicrope, F. A. and D. J. Sargent (2012). "Molecular pathways: microsatellite instability in colorectal cancer: prognostic, predictive, and therapeutic implications." *Clin Cancer Res* **18**(6): 1506-1512.

Slaby, O., J. Bienertova-Vasku, M. Svoboda and R. Vyzula (2012). "Genetic polymorphisms and microRNAs: new direction in molecular epidemiology of solid cancer." *J Cell Mol Med* **16**(1): 8-21.

Slaughter, D. P., H. W. Southwick and W. Smejkal (1953). "Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin." *Cancer* **6**(5): 963-968.

Slezak-Prochazka, I., J. Kluiver, D. de Jong, G. Kortman, N. Halsema, S. Poppema, B. J. Kroesen and A. van den Berg (2013). "Cellular Localization and Processing of Primary Transcripts of Exonic MicroRNAs." *PLoS One* **8**(9): e76647.

Starega-Roslan, J., E. Koscianska, P. Kozlowski and W. J. Krzyzosiak (2011). "The role of the precursor structure in the biogenesis of microRNA." *Cell Mol Life Sci* **68**(17): 2859-2871.

Stewart, M. (2009). "Cell biology. Nuclear export of small RNAs." *Science* **326**(5957): 1195-1196.

Stratmann, J., C. J. Wang, S. Gnosa, A. Wallin, D. Hinselwood, X. F. Sun and H. Zhang (2011). "Dicer and miRNA in relation to clinicopathological variables in colorectal cancer patients." *BMC Cancer* **11**: 345.

Sun, G., J. Yan, K. Noltner, J. Feng, H. Li, D. A. Sarkis, S. S. Sommer and J. J. Rossi (2009). "SNPs in human miRNA genes affect biogenesis and function." *RNA* **15**(9): 1640-1651.

Suzuki, H., E. Yamamoto, M. Nojima, M. Kai, H. O. Yamano, K. Yoshikawa, T. Kimura, T. Kudo, E. Harada, T. Sugai, H. Takamaru, T. Niinuma, R. Maruyama, H. Yamamoto, T. Tokino, K. Imai, M. Toyota and Y. Shinomura (2010). "Methylation-associated silencing of microRNA-34b/c in gastric cancer and its involvement in an epigenetic field defect." *Carcinogenesis* **31**(12): 2066-2073.

Svrcek, M., O. Buhard, C. Colas, F. Coulet, S. Dumont, I. Massaoudi, A. Lamri, R. Hamelin, J. Cosnes, C. Oliveira, R. Seruca, M. P. Gaub, M. Legrain, A. Collura, O. Lascols, E. Tiret, J. F. Flejou and A. Duval (2010). "Methylation tolerance due to an O6-methylguanine DNA methyltransferase (MGMT) field defect in the colonic mucosa: an initiating step in the development of mismatch repair-deficient colorectal cancers." *Gut* **59**(11): 1516-1526.

Svrcek, M., J. El-Bchiri, A. Chalastanis, E. Capel, S. Dumont, O. Buhard, C. Oliveira, R. Seruca, C. Bossard, J. F. Mosnier, F. Berger, E. Leteurtre, A. Lavergne-Slove, M. P. Chenard, R. Hamelin, J. Cosnes, L. Beaugerie, E. Tiret, A. Duval and J. F. Flejou (2007). "Specific clinical and biological features characterize inflammatory bowel disease associated colorectal cancers showing microsatellite instability." *J Clin Oncol* **25**(27): 4231-4238.

Swierniak, M., A. Wojcicka, M. Czetwertynska, E. Stachlewska, M. Maciag, W. Wiechno, B. Gornicka, M. Bogdanska, L. Koperski, A. de la Chapelle and K. Jazdzewski (2013). "In-depth characterization of the microRNA transcriptome in normal thyroid and papillary thyroid carcinoma." *J Clin Endocrinol Metab* **98**(8): E1401-1409.

T

Takagi, T., Y. Naito, K. Mizushima, I. Hirata, N. Yagi, N. Tomatsuri, T. Ando, Y. Oyamada, Y. Isozaki, H. Hongo, K. Uchiyama, O. Handa, S. Kokura, H. Ichikawa and T. Yoshikawa (2010). "Increased expression of microRNA in the inflamed colonic mucosa of patients with active ulcerative colitis." *J Gastroenterol Hepatol* **25 Suppl 1**: S129-133.

Takamizawa, J., H. Konishi, K. Yanagisawa, S. Tomida, H. Osada, H. Endoh, T. Harano, Y. Yatabe, M. Nagino, Y. Nimura, T. Mitsudomi and T. Takahashi (2004). "Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival." *Cancer Res* **64**(11): 3753-3756.

Thibodeau, S. N., G. Bren and D. Schaid (1993). "Microsatellite instability in cancer of the proximal colon." *Science* **260**(5109): 816-819.

Thornton, J. E. and R. I. Gregory (2012). "How does Lin28 let-7 control development and disease?" *Trends Cell Biol* **22**(9): 474-482.

Tili, E., C. M. Croce and J. J. Michaille (2009). "miR-155: on the crosstalk between inflammation and cancer." *Int Rev Immunol* **28**(5): 264-284.

Tili, E., J. J. Michaille and C. M. Croce (2013). "MicroRNAs play a central role in molecular dysfunctions linking inflammation with cancer." *Immunol Rev* **253**(1): 167-184.

Tili, E., J. J. Michaille, D. Wernicke, H. Alder, S. Costinean, S. Volinia and C. M. Croce (2011). "Mutator activity induced by microRNA-155 (miR-155) links inflammation and cancer." *Proc Natl Acad Sci U S A* **108**(12): 4908-4913.

Tomari, Y., T. Du and P. D. Zamore (2007). "Sorting of Drosophila small silencing RNAs." *Cell* **130**(2): 299-308.

Tomari, Y., C. Matranga, B. Haley, N. Martinez and P. D. Zamore (2004). "A protein sensor for siRNA asymmetry." *Science* **306**(5700): 1377-1380.

Tomari, Y. and P. D. Zamore (2005). "MicroRNA biogenesis: drosha can't cut it without a partner." *Curr Biol* **15**(2): R61-64.

Tsutsumi, A., T. Kawamata, N. Izumi, H. Seitz and Y. Tomari (2011). "Recognition of the pre-miRNA structure by Drosophila Dicer-1." *Nat Struct Mol Biol* **18**(10): 1153-1158.

U

Ullman, T. A. and S. H. Itzkowitz (2011). "Intestinal inflammation and cancer." *Gastroenterology* **140**(6): 1807-1816.

V

Valeri, N., P. Gasparini, C. Braconi, A. Paone, F. Lovat, M. Fabbri, K. M. Sumani, H. Alder, D. Amadori, T. Patel, G. J. Nuovo, R. Fishel and C. M. Croce (2010). "MicroRNA-21 induces resistance to 5-fluorouracil by down-regulating human DNA MutS homolog 2 (hMSH2)." *Proc Natl Acad Sci U S A* **107**(49): 21098-21103.

Valeri, N., P. Gasparini, M. Fabbri, C. Braconi, A. Veronese, F. Lovat, B. Adair, I. Vannini, F. Fanini, A. Bottoni, S. Costinean, S. K. Sandhu, G. J. Nuovo, H. Alder, R. Gafa, F. Calore, M. Ferracin, G. Lanza, S. Volinia, M. Negrini, M. A. McIlhatton, D. Amadori, R. Fishel and C. M. Croce (2010). "Modulation of mismatch repair and genomic stability by miR-155." *Proc Natl Acad Sci U S A* **107**(15): 6982-6987.

Vasudevan, S., Y. Tong and J. A. Steitz (2007). "Switching from repression to activation: microRNAs can up-regulate translation." *Science* **318**(5858): 1931-1934.

Vilar, E. and S. B. Gruber (2010). "Microsatellite instability in colorectal cancer-the stable evidence." *Nat Rev Clin Oncol* **7**(3): 153-162.

Vilar, E., J. Tabernero and S. B. Gruber (2011). "Micromanaging the classification of colon cancer: the role of the microRNAome." *Clin Cancer Res* **17**(23): 7207-7209.

Voellenkle, C., J. Rooij, A. Guffanti, E. Brini, P. Fasanaro, E. Isaia, L. Croft, M. David, M. C. Capogrossi, A. Moles, A. Felsani and F. Martelli (2012). "Deep-sequencing of endothelial cells exposed to hypoxia reveals the complexity of known and novel microRNAs." *RNA* **18**(3): 472-484.

Vogelstein, B. and K. W. Kinzler (2002). *The genetic basis of human cancer*. New York, McGraw-Hill, Medical Pub. Division.

W

Wang, H. W., C. Noland, B. Siridechadilok, D. W. Taylor, E. Ma, K. Felderer, J. A. Doudna and E. Nogales (2009). "Structural insights into RNA processing by the human RISC-loading complex." *Nat Struct Mol Biol* **16**(11): 1148-1153.

Weinberg, R. A. (2007). *The biology of cancer*. New York, Garland Science.

Welker, N. C., T. S. Maity, X. Ye, P. J. Aruscavage, A. A. Krauchuk, Q. Liu and B. L. Bass (2011). "Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode." *Mol Cell* **41**(5): 589-599.

Wiesen, J. L. and T. B. Tomasi (2009). "Dicer is regulated by cellular stresses and interferons." *Mol Immunol* **46**(6): 1222-1228.

Wightman, B., T. R. Burglin, J. Gatto, P. Arasu and G. Ruvkun (1991). "Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development." *Genes Dev* **5**(10): 1813-1824.

Wightman, B., I. Ha and G. Ruvkun (1993). "Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans." *Cell* **75**(5): 855-862.

Willenbucher, R. F., D. E. Aust, C. G. Chang, S. J. Zelman, L. D. Ferrell, D. H. Moore, 2nd and F. M. Waldman (1999). "Genomic instability is an early event during the progression pathway of ulcerative-colitis-related neoplasia." *Am J Pathol* **154**(6): 1825-1830.

Winter, J., S. Jung, S. Keller, R. I. Gregory and S. Diederichs (2009). "Many roads to maturity: microRNA biogenesis pathways and their regulation." *Nat Cell Biol* **11**(3): 228-234.

Wu, F., N. J. Guo, H. Tian, M. Marohn, S. Gearhart, T. M. Bayless, S. R. Brant and J. H. Kwon (2011). "Peripheral blood microRNAs distinguish active ulcerative colitis and Crohn's disease." *Inflamm Bowel Dis* **17**(1): 241-250.

Wu, F., S. Zhang, T. Dassopoulos, M. L. Harris, T. M. Bayless, S. J. Meltzer, S. R. Brant and J. H. Kwon (2010). "Identification of microRNAs associated with ileal and colonic Crohn's disease." *Inflamm Bowel Dis* **16**(10): 1729-1738.

Wu, F., M. Zikusoka, A. Trindade, T. Dassopoulos, M. L. Harris, T. M. Bayless, S. R. Brant, S. Chakravarti and J. H. Kwon (2008). "MicroRNAs are differentially expressed in ulcerative colitis and alter expression of macrophage inflammatory peptide-2 alpha." *Gastroenterology* **135**(5): 1624-1635 e1624.

Wyman, S. K., E. C. Knouf, R. K. Parkin, B. R. Fritz, D. W. Lin, L. M. Dennis, M. A. Krouse, P. J. Webster and M. Tewari (2011). "Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity." *Genome Res* **21**(9): 1450-1461.

Y

Yamamoto, H., Y. Adachi, H. Taniguchi, H. Kunimoto, K. Nosho, H. Suzuki and Y. Shinomura (2012). "Interrelationship between microsatellite instability and microRNA in gastrointestinal cancer." *World J Gastroenterol* **18**(22): 2745-2755.

Yang, J. H., X. Luo, Y. Nie, Y. Su, Q. Zhao, K. Kabir, D. Zhang and R. Rabinovici (2003). "Widespread inosine-containing mRNA in lymphocytes regulated by ADAR1 in response to inflammation." *Immunology* **109**(1): 15-23.

Yang, J. S. and E. C. Lai (2011). "Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants." *Mol Cell* **43**(6): 892-903.

Yang, W., T. P. Chendrimada, Q. Wang, M. Higuchi, P. H. Seeburg, R. Shiekhattar and K. Nishikura (2006). "Modulation of microRNA processing and expression through RNA editing by ADAR deaminases." *Nat Struct Mol Biol* **13**(1): 13-21.

Yates, L. R. and P. J. Campbell (2012). "Evolution of the cancer genome." *Nat Rev Genet* **13**(11): 795-806.

Ye, Y., K. K. Wang, J. Gu, H. Yang, J. Lin, J. A. Ajani and X. Wu (2008). "Genetic variations in microRNA-related genes are novel susceptibility loci for esophageal cancer risk." *Cancer Prev Res (Phila)* **1**(6): 460-469.

Yekta, S., I. H. Shih and D. P. Bartel (2004). "MicroRNA-directed cleavage of HOXB8 mRNA." *Science* **304**(5670): 594-596.

Yi, R., Y. Qin, I. G. Macara and B. R. Cullen (2003). "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs." *Genes Dev* **17**(24): 3011-3016.

Yoda, M., T. Kawamata, Z. Paroo, X. Ye, S. Iwasaki, Q. Liu and Y. Tomari (2010). "ATP-dependent human RISC assembly pathways." *Nat Struct Mol Biol* **17**(1): 17-23.

Z

Zahm, A. M., M. Thayu, N. J. Hand, A. Horner, M. B. Leonard and J. R. Friedman (2011). "Circulating microRNA is a biomarker of pediatric Crohn disease." *J Pediatr Gastroenterol Nutr* **53**(1): 26-33.

Zeng, Y. and B. R. Cullen (2005). "Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences." *J Biol Chem* **280**(30): 27595-27603.

Zeng, Y., R. Yi and B. R. Cullen (2005). "Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha." *EMBO J* **24**(1): 138-148.

Zhang, B., X. Pan, G. P. Cobb and T. A. Anderson (2007). "microRNAs as oncogenes and tumor suppressors." *Dev Biol* **302**(1): 1-12.

Zhang, W., T. Winder, Y. Ning, A. Pohl, D. Yang, M. Kahn, G. Lurje, M. J. Labonte, P. M. Wilson, M. A. Gordon, S. Hu-Lieskovan, D. J. Mauro, C. Langer, E. K. Rowinsky and H. J. Lenz (2011). "A let-7 microRNA-binding site polymorphism in 3'-untranslated region of KRAS gene predicts response in wild-type KRAS patients with metastatic colorectal cancer treated with cetuximab monotherapy." *Ann Oncol* **22**(1): 104-109.

Zhang, Z., J. Zou, G. K. Wang, J. T. Zhang, S. Huang, Y. W. Qin and Q. Jing (2011). "Uracils at nucleotide position 9-11 are required for the rapid turnover of miR-29 family." *Nucleic Acids Res* **39**(10): 4387-4395.

Zisman, T. L. and D. T. Rubin (2008). "Colorectal cancer and dysplasia in inflammatory bowel disease." *World J Gastroenterol* **14**(17): 2662-2669.

Zorc, M., D. J. Skok, I. Godnic, G. A. Calin, S. Horvat, Z. Jiang, P. Dovc and T. Kunej (2012). "Catalog of microRNA seed polymorphisms in vertebrates." *PLoS One* **7**(1): e30737.

Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." *Nucleic Acids Res* **31**(13): 3406-3415.