Développement de tests diagnostiques par détection d'ADN extracellulaire

Rita Tanos

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie Santé

École doctorale Sciences Chimiques et Biologiques pour la Santé – CBS2

Institut de Recherche en Cancérologie de Montpellier – INSERM U1194

Développement de tests diagnostiques par détection d'ADN extracellulaire

Présentée par Rita TANOS
Le 25 Novembre 2019

Sous la direction de Alain THIERRY
et Mona DIAB ASSAF

Devant le jury composé de

Mme. Muriel MATHONNET, PU PH, CHU, Limoges	**Présidente**
Mr. Sébastien SALAS, PU PH, CHU Timone Adulte, Marseille	**Rapporteur**
Mme. Lynette FERNANDEZ-CUESTA, DR, Centre international de Recherche sur le Cancer, Lyon	**Rapporteur**
Mme. Florence LE CALVEZ-KELM, CR, Centre international de Recherche sur le Cancer, Lyon	**Examinatrice**
Mr. Didier TOUSCH, Maitre de conférences, Université de Montpellier, Montpellier	**Examinateur**
Mr. Stanislas DU MANOIR, CR, Institut de Recherche en Cancérologie de Montpellier, Montpellier	**Examinateur**
Mr. Alain THIERRY, DR, Institut de Recherche en Cancérologie de Montpellier, Montpellier	**Directeur de thèse**
Mme. Mona DIAB-ASSAF, Pr, Université Libanaise, Liban	**Co-directrice de thèse**

UNIVERSITÉ DE MONTPELLIER

Continue d'avancer, ne t'arrête pas. Car avancer, c'est aller vers la perfection. Marche, sans craindre les épines ou les pierres tranchantes dont est parsemé le sentier de la vie.

KHALIL GIBRAN

RESUME

Après sa découverte en 1948, l'ADN circulant (ADNcir) a été étudié dans divers domaines. Il est devenu un biomarqueur émergent, en particulier en oncologie, un domaine dans lequel plusieurs travaux ont récemment cherché à étudier son intérêt dans le dépistage et la détection précoce du cancer. La première partie de ma thèse a été consacrée à l'étude des caractéristiques quantitatives et structurelles de l'ADNcir, en prenant en compte son origine (ADNcir nucléaire et mitochondrial) et sa structure (fragmentation et profil de taille), pour le dépistage et la détection précoce du cancer. Deux paramètres, le Ref A 67 (concentration totale d'ADNcir nucléaire) et le MNR (Rapport entre la concentration de l'ADNcir mitochondrial et nucléaire), ont été quantifiés par q-PCR dans un modèle murin puis validés dans les milieux des cellules en culture pour évaluer leur potentiel à discriminer un état sain d'un état cancéreux. Ces deux paramètres ont été quantifiés chez l'Homme, en prenant en compte d'autres paramètres quantitatifs et structurels de l'ADNcir, après réajustement en fonction de l'âge, dans le plasma de 289 individus sains, 99 individus à risque de cancer colorectal (CCR) et 983 patients atteints de CCR (n = 791), de cancer du sein (n = 169) et d'autres cancers (hépatocellulaire, pancréatique, ovarien et lymphome) (n = 23). Par une approche d'apprentissage automatique, nous avons combiné ces différents paramètres dans un modèle de prédiction en utilisant des arbres de décision pour la classification des patients sains et cancéreux. Nous avons obtenu des résultats très encourageants, en particulier pour les cancers de stades précoces. Cette méthode semble prometteuse pour une détection précoce et non invasive du cancer. L'ajout d'autres biomarqueurs, comme le profil de taille ou le profil de méthylation de l'ADNcir, pourrait encore en augmenter le potentiel.

La deuxième partie de ma thèse a été consacrée à l'étude de la relation entre la quantité d'ADN extracellulaire d'origine nucléaire et mitochondriale dans le milieu de culture d'embryons, et la qualité de ces embryons lors d'une fécondation *in vitro* (FIV). En effet, il a été montré qu'un embryon libère de l'ADN extracellulaire dans le milieu de culture lors d'une FIV, et que cet ADN pourrait être un biomarqueur prédictif de la qualité de l'embryon et servir comme test génétique préimplantatoire (PGT) non invasif. Nous avons détecté le gène *SRY* dans le milieu de culture afin de déterminer le sexe de l'embryon, ce qui constitue une information importante dans les cas des maladies génétiques liées au sexe. Nous avons également entrepris de détecter la présence de la mutation Delta F508 du gène *CFTR* responsable de la mucoviscidose par analyse de l'ADN extracellulaire issu d'embryons à risque, afin d'évaluer son potentiel en tant que PGT non invasif.

ABSTRACT

After its discovery in 1948, circulating DNA (cirDNA) was studied in various fields. It has become an emerging biomarker, particularly in oncology, and several studies have recently sought to investigate its interest in cancer screening and early detection. The first part of my thesis was devoted to the study of the quantitative and structural characteristics of cirDNA, taking into account its origin (nuclear and mitochondrial cirDNA) and its structure (fragmentation and size profile), for the screening and early detection of cancer. Two cirDNA parameters, the Ref A 67 (total nuclear cirDNA concentration) and the MNR (Mitochondrial to Nuclear Ratio), were quantified by q-PCR in a mouse model and further validated in cell culture media to assess their potential to discriminate between a healthy and a cancerous state. These two variables were evaluated by taking into account other quantitative and structural parameters of cirDNA, after age adjustment, in the plasma of 289 healthy individuals, 99 individuals at risk of colorectal cancer (CRC) and 983 patients with CRC (n = 791), breast cancer (n = 169) and other cancers (hepatocellular, pancreatic, ovarian and lymphoma) (n = 23). Through a machine learning approach, we combined these different parameters into a prediction model using decision trees for the classification of healthy and cancer patients. We have obtained very encouraging results, especially for early-stage cancers. This method seems promising for early and non-invasive cancer detection. The addition of other biomarkers such as the size profile of the cirDNA or the detection of methylation markers could further increase its potential.

The second part of my thesis was devoted to the study of the relationship between the quantity of extracellular DNA of nuclear and mitochondrial origin in the embryo culture medium, and the quality of these embryos during *in vitro* fertilization (IVF). It has been shown that an embryo releases extracellular DNA into the culture medium during IVF, and that this DNA could be a predictive biomarker of embryo quality and thus be used as a non-invasive preimplantation genetic test (PGT). We detected, as well, the *SRY* gene in the culture medium to determine the sex of the embryo, which is an important information in the case of gender-related genetic disorders. We also tried to detect the presence of the Delta F508 mutation of the *CFTR* gene responsible for cystic fibrosis, by analyzing extracellular DNA from high-risk embryos to assess its potential as a non-invasive PGT.

REMERCIEMENTS

Je tiens tout d'abord à remercier les membres de mon jury de thèse.

A **Madame le Professeur Muriel MATHONNET,** je tiens à vous remercier de m'avoir fait l'honneur d'accepter de présider ce jury malgré votre emploi du temps chargé.

A **Monsieur le Professeur Sébastien SALAS** et à **Madame le Docteur Lynette FERNANDEZ-CUESTA**, je vous remercie pour l'intérêt que vous avez porté à mon travail en acceptant d'être rapporteurs. Je suis honorée que vous fassiez partie de mon jury de thèse.

A **Madame le Docteur Florence LE CALVEZ-KELM**, vous me faites l'honneur de juger mon travail de thèse en étant examinatrice. Je vous adresse mes sincères remerciements.

A **Monsieur le Docteur Didier TOUSCH**, vous m'avez accompagné depuis ma première année de thèse en étant membre de mon comité de suivi individuel. Je vous remercie pour les conseils et les recommandations que vous m'avez donnés, et également d'avoir participé à mon jury de thèse en étant examinateur.

A **Monsieur le Docteur Stanislas DU MANOIR**, je vous remercie pour l'intérêt que vous avez toujours porté à mon travail, et que vous montrez encore en acceptant d'être examinateur dans mon jury de thèse.

A mes directeurs de thèse,

A **Monsieur Alain THIERRY**, je vous remercie pour la confiance que vous m'avez accordée pour réaliser ce travail. Grâce à vous j'ai eu l'opportunité de travailler avec des personnes agréables et très compétentes et de m'intéresser à un sujet très passionnant. Je vous remercie énormément pour votre encadrement et vos conseils tout en me laissant être en même temps autonome. J'ai beaucoup appris pendant mes quatre ans de thèse sous votre direction, sur le plan scientifique et personnel. Je vous suis reconnaissante pour le temps que vous m'avez accordé et l'attention de tout instant sur mes travaux, et je vous souhaite le mieux pour vos futurs travaux ainsi que votre vie personnelle.

A ma co-directrice de thèse le **Docteur Mona DIAB ASSAF**, je vous remercie de m'avoir accompagnée depuis mon Master en Cancérologie à l'Université Libanaise, et de m'avoir donné l'opportunité de venir à Montpellier pour mon stage, ce qui m'a permis par la suite de mener cette thèse. Je vous remercie pour votre aide ainsi que votre soutien, votre attention et vos encouragements.

Il me sera très difficile de remercier tout le monde car c'est grâce à l'aide de nombreuses personnes que j'ai pu mener cette thèse à son terme.

Je vais commencer par remercier les membres de mon équipe, vous savez toutes et tous à quel point je vous apprécie. **Marc**, je vous remercie de m'avoir accueillie dans cette magnifique équipe. **Corinne**, j'ai sincèrement apprécié tes sourires et ton aide au quotidien, et je te remercie surtout pour ton aide pour la correction de ce manuscrit. **Evelyne**, je t'ai toujours considérée comme une mère qui s'inquiète pour nous et qui demande de nos nouvelles. **Caroline**, merci pour ton aide et ta disponibilité chaque fois que je venais vers toi avec des questions (qui me semblaient très compliquées) concernant les biostats. **Antoine**, merci pour la bonne humeur que tu as apportée aux réunions du mardi après-midi. **Philippe** et **Thibault**, merci pour votre présence et pour la bonne ambiance.

Au groupe ADN circulant, à mes collègues de boulot, vous allez me manquer. C'est un peu comme quitter une famille que de vous quitter. **Laurence**, tu n'es pas une collègue, tu es plutôt une amie et une sœur. Je te remercie pour ton aide dans les manips ainsi que dans la vie de tous les jours. **Jean-Daniel**, j'ai beaucoup apprécié ton implication, ton souci du détail ainsi que nos discussions scientifiques et non scientifiques, et je te remercie d'avoir pris le temps pour la relecture de mon manuscrit. **Ekaterina** et **Andrei** спасибо за все то хорошее время, что мы провели вместе. **Amaëlle**, merci pour ta présence, ton soutien et tes prières, et merci de m'avoir appris comment être zen et patiente.

Finalement à ceux qui étaient là depuis le début et qui sont devenus de vrais amis. **Romain**, merci de m'avoir considérée comme une petite sœur et d'avoir toujours pris soin de moi. **Cynthia**, sans toi ça aurait été difficile de gérer le stress de cette thèse, même si des fois tu étais la cause du stress (c'est la CATA :P). **Brice**, avoir travaillé avec un collègue, devenu ami, comme toi, a été un vrai bonheur et une belle chance. Merci pour les fous rires, les blagues et les beaux souvenirs. La bise au chat :D.

Je tiens à remercier également toutes les personnes qui ont fait partie de l'équipe au passé : **Geoffroy, Audrey, Simon, Nico, Benoit R, Julien, Alexandre, Baptiste, Safia, Barbara, Hajar et Han**. Merci pour tous les beaux moments qu'on a passés ensemble.

Je remercie ensuite l'ensemble de l'institut, en commençant par le directeur **Monsieur Claude Sardet**, ainsi que tous les chefs d'équipes, les chercheurs et les ingénieurs. J'ai considéré l'IRCM comme ma deuxième maison grâce à la bonne ambiance et la bonne entente entre toutes les personnes. Je tiens particulièrement à remercier les membres des équipes AD et EJ, les équipes du couloir, que j'ai croisés tous les jours pendant ces quatre ans, et qui m'ont toujours accueillie avec un grand sourire.

Aux potes du couloir, **Yannick** et **Julie**, je vous remercie énormément pour votre amitié. Je n'oublierai jamais les moments qu'on a passés ensemble, la raquette dans le bureau et les mots-fléchés, les nombreux accidents (oui Yannick je parle de toi :P), les parties de pétanque, les pizzas et les bières tard le soir au labo… Aux résidents du bureau des étudiants que j'ai considéré comme mon deuxième bureau :D ; **Miriam, Nabiya, Yezza, Yasser, Leslie, Vera, Lou** et **Katrin**, merci de m'avoir supportée toutes les fois que je suis venue squatter chez vous et que je me comportais comme si j'étais chez moi☺.

A tous les doctorants, post-docs et ingénieurs de l'institut : **Valomanda, Guillaume, Sara B, Gabriel, Maeva, Emilia, Mélanie, Adrien, Céléstine, Tristan, Vincent, Hadjer, Alice, Valentin, Antoine, Madi, Jaime, Imène, Mehdi, Ambre, Ander, Bilguun, Carlo, Jordi, Simon, Ghita, Habib,** et **Timothée**. Merci pour la bonne ambiance, la bonne entente, la bonne humeur et les beaux moments et souvenirs que je garderai pour toujours.

Je remercie également les « anciens IRCM », **Rana, Racha, Emile, Mona, Mohammad, Joelle A, Laetitia, Benoit B, Nour, Augusto, Emilie, Hanane, Leila, Sarah C et Romain L**, pour m'avoir accueillie quand je suis arrivée dans l'institut et de m'avoir aidée à intégrer. Je remercie surtout **Amanda**, une amie et une sœur qui a supporté toutes mes folies et mes histoires bizarres et ma vie pleine « d'action » comme elle l'a décrit :P.

Je tiens ensuite à exprimer toute ma gratitude envers le personnel administratif de l'IRCM (**Daniela, Sandrine, Nadia, Nadège, Déborah et Carmen**) pour leur aide, leur disponibilité et leur gentillesse.

A mes amis, ou plutôt ma famille à Montpellier,

I will never forget all the good times and adventures that we had together, and I know that our friendship will last even if we end up in different corners of the globe. **Jovon**, Madame casual herself, since the first day I met you and you called me "lej2a", I knew that we will get along very well :P. The other **Jovon** (ma enti), my Marvel twin, even though we're on opposite teams :P, but I "love you 3000" :*. **Willi**, Mr Chilling, you always have your own way for making us laugh, and your love and passion for science are inspiring. **Laza**, I hesitated between calling you "Mr rezz 3a djej" or "Mr influencer", but in both cases thank you for all the laughs, the "soirées" and the "service terminé". **Joss**, you were the last one to join our group but we felt like you were always one of us. **Toto**, **"**arabte", tu étais là depuis mon arrivée à Montpellier pour mon stage, et tu es encore là pour la dernière ligne droite. Merci pour ton amitié et ton soutien <3. **Ziad**, "jare l3aziz wal monkez fi l2awkat l7arija", ça va me manquer d'entendre d'un coup les voix de Wadih Lsheikh et Hssein Ldeek venant de l'autre bout du bâtiment :P. **Ayman** "Ewaaa" et **Mohammad** "Mokhtar", mon **"**sandwich" préféré :D, vous avez une place particulière dans mon coeur, merci pour les moments où on a rigolé ensemble, ainsi que tous les temps où vous m'avez fait chier :P. **Zahra** et **Diala**, vous m'avez accompagnée tous les jours, au labo et hors-labo, pendant les moments les plus durs et les plus beaux. Vous étiez toujours là quand j'en avais besoin, et vous m'avez toujours soutenue et supportée. Sans vous ça aurait été très difficile d'arriver là. Je vous aime et merci pour tout !

Je tiens aussi à remercier les amis que j'ai rencontrés à Montpellier et qui ont rendu le temps que j'ai passé ici beaucoup plus agréable. Merci à **Kamar, Jamal, Jacinthe, Baraa, Amanie, Hassan, Rana, Sana, Diala K** et **Lara**.

A mes amis au Liban,

Eddy, I know you will always have my back no matter what. Merci d'avoir toujours été à mes côtés quand les temps étaient durs et de m'avoir fait rire quand je ne voulais même pas sourire. **Céline** and **Lamis**, my childhood friends, even though we do not see each other often, and even when a long time passes without us talking, I know you will always be here for me and I will always be there for you too. **Sabine**, **Zaza** and **Saleh**, thank you for being such

amazing friends since the day we first met at the university. Even though we did not talk every day, our friendship stayed strong, and always will, despite the geographical distance. My favorite couple, **Jhonny** and **Chacha**, I love you both so much. You were always there for me, taking care of me and asking how I was doing. Thank you for everything!

A ma grande famille,

A mes oncles et leurs femmes, à mes cousins et cousines, à mes tantes, et à mes grands-mères (Allah ytawele bi 3omrkon), même de loin vous étiez toujours proches de moi, et je sais que vous m'avez toujours gardée dans vos pensées et vos prières. A mes grands-pères, vous resterez toujours dans mon cœur. Je vous aime énormément.

Aux deux personnes les plus proches de moi, my backbones and my bestfriends. Je vous ai mises avec la famille parce que vous n'êtes pas mes amies, vous êtes mes sœurs. **Pascale** et **Rim**, ma vie n'aurait pas été la même sans vous. Merci de m'avoir supportée pendant mes moments fous, et il y en avait plein ! Merci de ne pas m'avoir laissé faire des choses stupides…seule ☺ Je ne me souviens même plus qui parmi nous était la mauvaise influence.

A ma petite famille, sans laquelle je ne serai pas là.

A mon frère, je n'avais jamais imaginé que le frère contre lequel j'étais si en colère serait le même frère que je me sentirai bénie d'avoir. Même si tu veux que je me marie et que je te laisse tranquille, je sais que tu ne pourras jamais vivre sans moi ☺ I love you so much little brother ! You're the best brother a girl can ask for !

A mon père et ma mère, vous êtes mes idoles. Je vous dois tout, et si j'en suis là aujourd'hui, c'est grâce à vous, à votre présence et vos sacrifices, à votre tendresse et votre amour. Je voulais vous dire merci du plus profond de mon cœur. Merci pour tous les encouragements, pour tous les applaudissements, pour votre soutien et votre confiance qui m'ont permis de me lancer dans ma vie adulte. Vous avez toujours été là et je sais que vous le serez toujours. J'espère pouvoir vous rendre le quart de ce que vous m'avez donné. Je vous aime de tout mon cœur !

TABLE DES MATIERES

LISTE DES ABBREVIATIONS

°C : degré celsius

µl : microlitre

µM : micromolaire

ADN : acide désoxyribonucléique

ADNcf : ADN extracellulaire ou cell-free DNA

ADNcir : ADN extracellulaire circulant / cirDNA : circulating DNA

ADNmt : ADN d'origine mitochondriale

ADNnu : ADN génomique d'origine nucléaire

ARN : acide ribonucléique

ARNm : ARN méssager

ARNr : ARN ribosomal

ARNt : ARN de transfert

ART : techniques de procréation médicalement assistée ou Assisted Reproductive Technology

ATP : Adénosine triphosphate

AUC : area under curve

CCR : cancer colorectal / CRC : colorectal cancer

CFTR : pour Cystic fibrosis transmembrane conductance regulator

COI à COIII : cytochrome C oxydase sous-unités 1 à 3

DAMP : damage-associated molecular pattern

DNase : déoxyribonucléase

EGFR : Epidermal growth factor receptor

FACS : Fluorescence-activated cell sorting

FIV : fécondation *in vitro* / IVF: *in vitro* fertilization

FSH : hormone folliculostimulante

H_2O_2 : péroxyde d'hydrogène

HLA : human leukocyte antigen

hMG : gonadotrophine ménopausique humaine

HPLC : chromatographie en phase liquide à haute performance

ICSI : injection intracytoplasmique de spermatozoïde ou intracytoplasmic sperm injection

IL-26 : interleukine 26

J1, J3, J5 et J6 : jours 1, 3, 5 et 6 du développement embryonnaire

KRAS : Kirsten RAt Sarcoma virus

MTHR : méthylènetétrahydrofolate réductase

ml : millilitre

n : nombre

NADH : 1,4-Dihydronicotinamide adenine dinucleotide

Nb copies/ml plasma ou copy nb/ml : nombres de copies par millilitre de plasma

NET : neutrophil extracellular DNA traps

NGS : next-generation sequencing

nM : nanomolaire

Pb : paires de base

PCR : polymerase chain reaction

Pg : picogramme

PGT : test génétique préimplantatoire ou preimplantation genetic testing

PGT-A : Preimplantation Genetic Testing for aneuploidy ou test génétique préimplantatoire pour les aneuploïdies

PGT-M : Preimplantation Genetic Testing for monogenic diseases ou test génétique préimplantatoire pour les maladies monogéniques

PGT-SR : Preimplantation Genetic Testing for structural rearrangements

pmol/µl : picomol par microlitre

Q-PCR : quantitative polymerase chain reaction

Ref A 145 : la concentration des fragments d'ADNcir nucléaire ayant une taille ≥ 145 paires de bases

Ref A 320 : la concentration des fragments d'ADNcir nucléaire ayant une taille ≥ 320 paires de bases

Ref A 67 : la concentration totale d'ADN extracellulaire ou circulant d'origine nucléaire

Ref M 67 : la concentration totale d'ADN extracellulaire ou circulant d'origine mitochondrial

MNR : Mitochondrial to Nuclear Ratio ou le rapport entre la concentration totale d'ADNcir d'origine mitochondriale et nucléaire

RhD : Rh blood group D antigen

ROC : receiver operating characteristic

ROS : reactive oxygen species ou espèces réactives de l'oxygène

Se : sensibilité

sHLA-G : soluble human leukocyte antigen-G

SNP : single-nucleotide polymorphism

SNV : variantes mononucléotidiques

Sp : specificité

SRY : Sex-determining Region of Y chromosome

SSP-S : Whole-genome sequencing of single-stranded DNA library preparation

TLR-9 : Toll-like receptor 9

TP53 : tumor protein p53

TPNI : tests prénatal non invasif

α-thalassemia-^{SEA} : α-thalassémie au Sud-Est de l'Asie et au sud de la chine (SEA pour Southeast Asian deletion)

LISTE DES FIGURES

LISTE DES TABLEAUX

I. Introduction générale : L'ADN circulant

A. Découverte, historique, et généralités

L'ADN extracellulaire (ADNcf pour cell-free DNA) provient de L'ADN nucléaire (ADNnu) et/ou mitochondrial (ADNmt) libéré hors de la cellule. *In vivo*, ces molécules peuvent se retrouver dans les fluides non circulants (lait, urine…), et *in vitro* dans le surnageant de culture de cellules, d'organoïdes ou d'embryons **(Figure 1)**. Ils sont également présents dans les fluides biologiques circulants, comme le sang et la lymphe, et constituent ainsi l'ADN extracellulaire circulant (ADNcir).

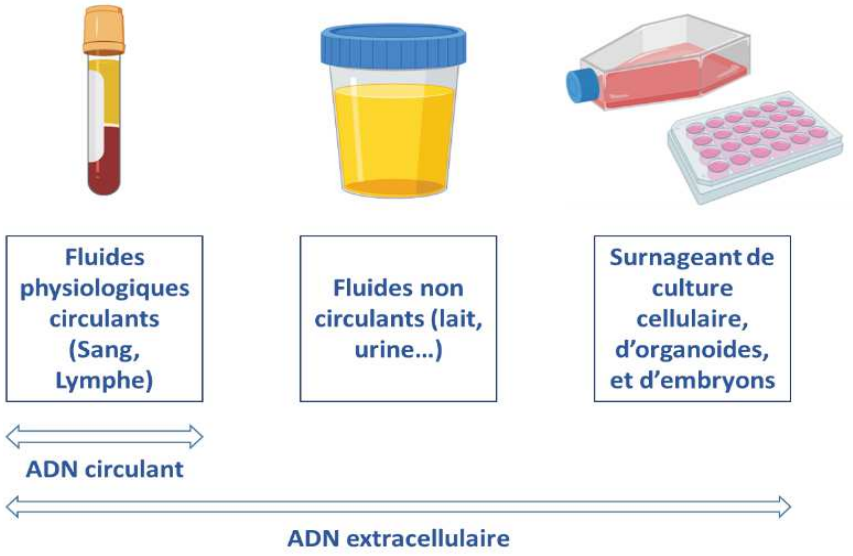


Figure 1: *L'ADN extracellulaire*

Il y a plus de 70 ans, *Mandel* et *Metais* furent les premiers à démontrer la présence d'acides nucléiques dans le sang d'individus sains et de patients atteints de cancer *(1)*, mais ce n'est que trente ou quarante ans plus tard que certains groupes se sont vraiment intéressés à ce composé biologique.

En 1963, *Stroun* et *Anker* ont montré qu'une hétérogreffe peut provoquer des modifications susceptibles de se transmettre à la descendance chez des plantes génétiquement stables *(2, 3)*. Ils ont ainsi émis l'hypothèse que ces modifications sont dues à un transfert d'ADN de la

plante « mentor » à la plante « pupille », et que cette information est transmise à la descendance. Les auteurs ont corrélé cette observation aux « gemmules » de Darwin. En effet, Darwin avait suggéré que toutes les unités du corps, en plus d'avoir le pouvoir de croître par division cellulaire, rejettent des « gemmules » minuscules qui sont dispersées dans le système *(4)*. Cette hypothèse était basée sur l'observation selon laquelle, dans le cas d'une hybridation par greffe, les tissus des plantes contiennent une matière formatrice, connue pour être l'ADN aujourd'hui, capable de se combiner avec celle d'un autre individu, et de reproduire chaque unité de l'organisme entier d'origine.

En 1966, Tan *et al.* ont montré que l'ADN circulant était présent dans le sang de patients atteints de lupus érythémateux disséminé *(5)*, entraînant la formation d'anticorps anti-ADN. Par la suite, l'intérêt pour l'ADN circulant a augmenté dans le domaine du diagnostic prénatal, en particulier lorsque Lo *et al.* ont mis en évidence, en 1997, de l'ADN d'origine fœtale dans la circulation sanguine de femmes enceintes *(6)*.

B. Domaines d'intérêt de l'analyse de l'ADN circulant

La découverte des ADN circulants en 1948 *(1)* a conduit à des recherches approfondies quant à son utilisation dans divers domaines. Il est en particulier devenu un biomarqueur émergent en oncologie, et a fait l'objet de nombreuses recherches translationnelles et cliniques. L'analyse de l'ADN circulant est actuellement utilisée en pratique clinique dans le diagnostic prénatal et constitue un biomarqueur prometteur dans d'autres domaines, tels que les maladies auto-immunes, le trauma, le sepsis, l'infarctus du myocarde et autres **(Figure 2)**.

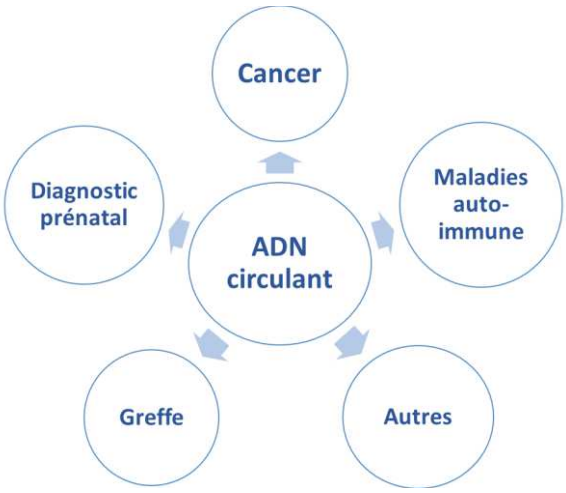


Figure 2: *Les domaines d'intérêt de l'ADN circulant*

1. L'ADN circulant en oncologie

En 1977, Leon *et al. (7)* ont démontré que la concentration d'ADN circulant chez les patients atteints de cancer est supérieure à celle des individus sains. Stroun *et al. (8)* ont ensuite identifié et caractérisé l'ADN circulant d'origine tumorale dans le sang des patients cancéreux. En 1989, Stroun et Anker *(9)* ont montré que chez ces patients, l'ADN circulant est composé en partie d'ADN circulant d'origine tumorale portant les altérations génétiques et épigénétiques propres aux cellules tumorales, caractérisant ainsi, à partir de cet ADN, les altérations moléculaires propres à la tumeur.

L'intérêt de la communauté scientifique en oncologie pour les ADN circulants a donc grandi à mesure que cet analyte permettait de fournir des informations relatives au diagnostic, au

pronostic et au théranostic, faisant naître la notion de « biopsie liquide » pour la gestion du cancer et la prise en charge des patients cancéreux *(10, 11)*. **Cette partie sera développée dans la partie III du manuscrit « ADN circulant et dépistage du cancer ».**

L'analyse de la biopsie liquide est une analyse non invasive de la tumeur, disponible en une simple prise de sang, et pouvant être examinée régulièrement au cours du traitement, ce qui permet de contourner les difficultés rencontrées aujourd'hui en pratique clinique, concernant les limites de l'analyse du tissu tumoral *(12)* ou encore le suivi des résistances acquises en réponse aux thérapies ciblées *(13, 14)*. De plus, elle fournit une idée globale de la tumeur en prenant en compte son hétérogénéité. Plusieurs techniques sont en cours de développement afin de détecter et caractériser ces ADNcir chez les patients cancéreux, comme le séquençage, la PCR digitale et d'autres approches *(15, 16)*.

2. Le diagnostic prénatal

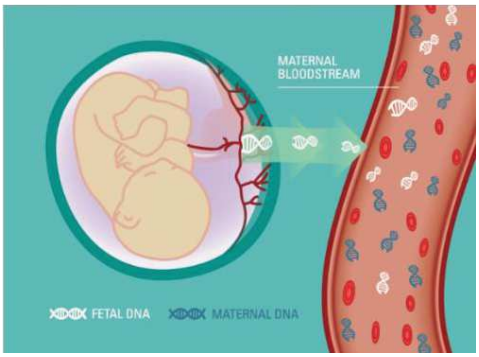


Figure 3: *La présence d'ADN d'origine fœtale dans le circulation des femmes enceintes (17)*

La présence d'ADN d'origine fœtale dans le plasma des femmes enceintes a été montrée par Lo *et al. (6)* **(Figure 3)**. Cette découverte a permis par la suite de développer des méthodes non invasives pour le diagnostic prénatal *(18)*, de génotyper le fœtus chez les femmes enceintes *(19, 20)*, et d'éviter ainsi les risques associés à certaines pratiques comme les amniocentèses par exemple *(21)*. Cette découverte a favorisé la recherche dans ce domaine et a contribué à la réalisation de certains tests cliniques comme la détection d'aneuploïdies chromosomales chez le fœtus *(22, 23)*, l'évaluation du sexe du fœtus *(24)*, et le génotypage du rhésus D fœtal *(25)*. **Cette partie sera développée dans la partie IV du manuscrit.**

3. Autres domaines

Outre le cancer et le diagnostic prénatal *(26)*, l'analyse des ADN circulants a été évaluée dans d'autres domaines comme celui des maladies auto-immunes. Plusieurs groupes ont en effet démontré, chez les patients atteints de lupus érythémateux systémique, une maladie auto-immune chronique touchant de nombreux organes et caractérisée par la formation d'auto-anticorps contre l'ADN double brin et d'autres composants nucléaires par exemple, que le taux d'ADN circulant dans le plasma était supérieur à celui des individus sains *(27–29)*. En outre, les variations de sa concentration reflétaient les changements dans la progression de la maladie, particulièrement lors d'une détérioration de l'état de santé des malades *(30)*.

L'analyse de l'ADN circulant s'est également révélée utile dans les cas du rejet ou des lésions des greffes d'organes solides grâce à la détection de l'ADN provenant du donneur dans le sang du receveur *(31)*. L'ADN du donneur a été quantifié en détectant des différences de SNP (single-nucleotide polymorphism) *(32)* ou d'autres signatures génétiques *(33)* dans le plasma des receveurs de transplantation cardiaque par exemple. Cet ADN augmente considérablement avant les événements de rejet. Des résultats similaires ont été observés pour la greffe du foie où la détermination de l'ADN circulant relatif au donneur dans le plasma reflétait la santé de l'organe *(34)* et a permis une discrimination plus précoce et sensible d'un rejet aigu *(35)*. Dans le cas de la transplantation rénale, des études ont également montré que cet ADN peut être utilisé pour évaluer les lésions et le rejet des allogreffes *(36, 37)*.

En outre, il a été démontré que l'ADN circulant fournit une grande précision pronostique chez les patients atteints de sepsis grave *(38, 39)*, infection générale de l'organisme résultant de la présence d'agents pathogènes dans le sang ou les tissus.

Enfin, chez des patients atteints d'infarctus du myocarde, les taux d'ADN circulant nucléaire *(40)* et mitochondrial *(41)* étaient supérieurs à ceux des contrôles. Des données disponibles suggèrent aussi que les taux d'ADNcir pourraient fournir des informations pronostiques pour l'évaluation de la gravité et de l'issue d'un accident vasculaire cérébral *(42, 43)*.

A. Origines, structures et fonctions des ADN circulants

1. Origines et structures des ADN circulants

L'ADN circulant est présent dans la circulation des individus sains comme dans la circulation des femmes enceintes et des patients atteints de différentes pathologies (cancer et autres) *(44, 45)*. Il provient majoritairement des cellules tumorales, des cellules saines, des cellules du système immunitaire, des cellules fœtales, et même des cellules du microenvironnement tumoral.

Bien que l'origine de l'ADN circulant soit étudiée depuis plusieurs années, le mécanisme de libération n'est pas encore élucidé. Une hypothèse est celle de plusieurs sources et mécanismes qui conduiraient à la présence de cet ADN sous différentes formes et structures **(Figure 4)** *(44–47)*:

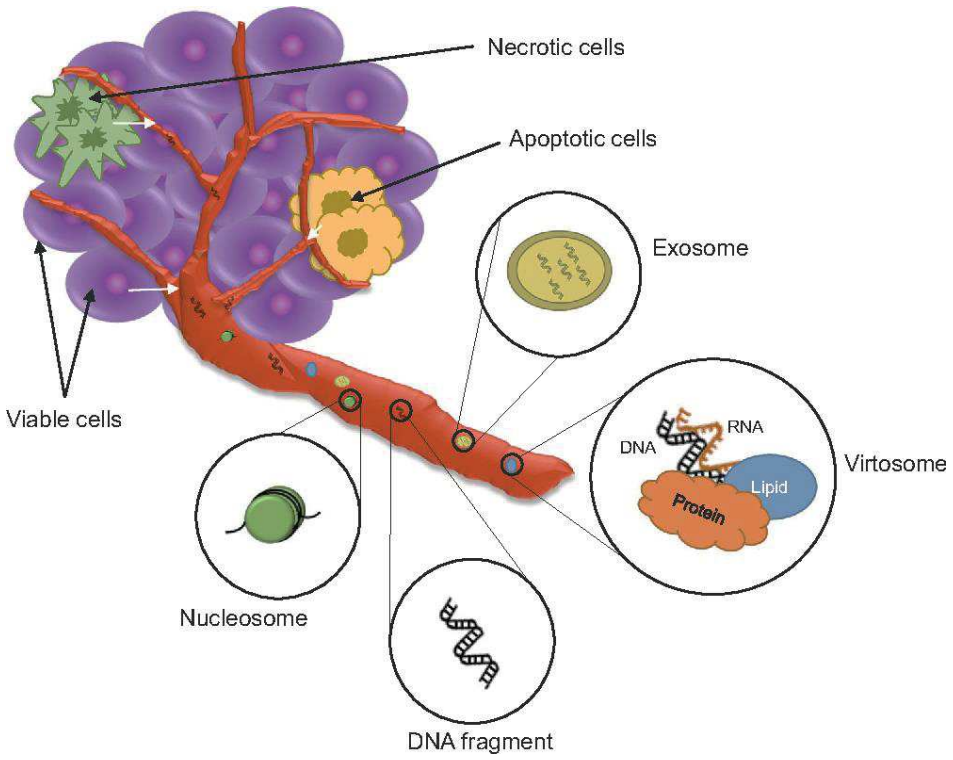


Figure 4: *Différentes origines et structures des ADN circulants (48)*

a) *La mort cellulaire*

- **Apoptose**

De nombreuses études suggèrent que la mort cellulaire par apoptose est l'une des principales sources d'ADN circulant. Ce processus de mort cellulaire programmée se caractérise par des changements morphologiques de la cellule et des mécanismes biochimiques ATP-dépendants *(49)*. L'apoptose provoque le clivage systématique de l'ADN génomique en multiples de 160 à 180 paires de bases (pb) *(50)*, et par la suite, la libération de mono- (~166pb) et poly-nucléosomes (~332pb, 498pb) dans la circulation *(46, 51, 52)*. L'ADNcir dans le plasma présente fréquemment une disposition en échelle (ladder pattern) en électrophorèse qui rappelle souvent celle des cellules apoptotiques. D'ailleurs, en 1990, Rumore *et al*. *(53)* ont montré que les ADN circulants présents dans le plasma de patients atteints de Lupus érythémateux disséminé présentent une taille comprise entre 160 et 180 pb.

Le nucléosome est donc l'une des principales structures sous laquelle se présentent les ADN circulants. Un nucléosome est composé d'un octamère d'histones (2H2A, 2H2B, 2H3 et 2H4) associé à de l'ADN double brin enroulé autour de ce complexe protéique contenant l'histone H1. Chaque nucléosome est relié à l'autre par de l'ADN double brin : le linker DNA **(Figure 5)**. L'ADN enroulé autour de l'octamère d'histones mesure 147 pb, et le linker DNA qui assure la liaison entre chaque nucléosome, mesure de 20 à 90 pb.

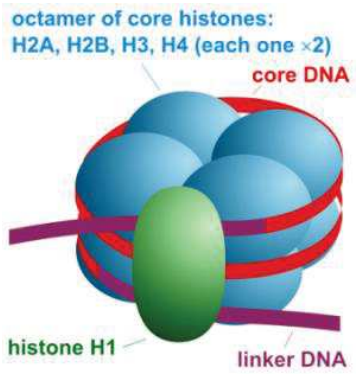


Figure 5: *Représentation schématique d'un nucléosome (source Wikipedia)*

Les corps apoptotiques constituent une autre structure des ADNcir. Ils sont formés lors de la phase tardive de l'apoptose, mesurent entre 1 et 5 µm et contiennent de l'ADN dégradé lors de l'apoptose *(54)*.

- **Nécrose**

La nécrose est une autre source possible d'ADNcir. Elle survient plus rapidement que l'apoptose, mais l'élimination des cellules nécrotiques est plus lente. Elle entraine une digestion non spécifique de l'ADN, produisant de larges fragments qui sont généralement supérieurs à 10 000pb *(46, 47, 51)*.

L'ADN mitochondrial (ADNmt) peut être libéré par les deux mécanismes de mort cellulaire (apoptose et nécrose). En outre, les mitochondries peuvent être éliminées par autophagie, appelée mitophagie *(55)*, soit pour réguler leur nombre, soit pour éliminer spécifiquement celles qui sont endommagées, ce mécanisme contribuant à la libération d'ADN mitochondrial dans la circulation *(56)*.

- **Phagocytose**

La phagocytose joue un rôle important dans la libération des ADN extracellulaires par tous les mécanismes de dégradation et de mort cellulaire *(45)*. Des études *in vivo (57)* et *in vitro (58)* ont montré l'importance des macrophages pour la génération d'ADNcf par des cellules mortes et mourantes. Les cellules nécrotiques, les vésicules autophagiques, les corps apoptotiques, etc… subissent une phagocytose, entraînant la libération de fragments d'ADN génomique et mitochondrial dans la circulation *(45)*. Parfois les macrophages ne réussissent pas à les digérer, ce qui peut causer leur mort et la libération de leur propre ADN ainsi que celui des cellules phagocytées dans la circulation.

- NETose

La NETose est un processus rapide de désintégration nucléaire et de mort cellulaire. Elle est limitée aux neutrophiles et survient lorsque le système immunitaire est confronté à des attaques infectieuses ou à des états physiologiques particuliers *(45)*. Elle induit la formation de NET (neutrophil extracellular DNA traps) *(59)*, des structures en forme de filet composées de chromatine remodelée d'origine nucléaire et mitochondrial et contenant des granules antimicrobiennes des neutrophiles *(44)*. Ces structures ont pour fonction principale de capter et d'éliminer les pathogènes, et il a été montré qu'elles ont des effets biologiques et

physiopathologiques particuliers. Une production inappropriée ou une exposition prolongée aux NETs pourrait mener à la production d'auto-anticorps, et leur persistance, due à une faible dégradation à cause de la présence d'une mutation au niveau du gène de la *DNase1* ou la *DNase1-like 3* par exemple, est associée au Lupus érythémateux disséminé *(60, 61)*. De nouvelles preuves suggèrent que les NETs sont aussi impliqués dans la propagation du cancer en piégeant les cellules tumorales circulantes, facilitant ainsi le processus de métastase *(62)*. Récemment, il a également été mis en évidence qu'après un exercice physique intense, les neutrophiles sécrétaient beaucoup de NETs, en plus de la sécrétion d'ADNcir par les myocytes, et de fortes concentrations d'ADN circulant sont observées *(63)*.

b) La sécrétion active

Le mécanisme de la sécrétion active de l'ADN circulant a été démontré par Stroun *et al. (52, 64, 65)*. Les auteurs ont en effet mis en évidence la synthèse et la sécrétion spontanée par des cellules saines telles que les lymphocytes et les cellules d'oreillette de grenouille d'un complexe ADN-ARN-glycoprotéines qu'ils ont appelé virtosome *(66)*. Le virtosome est un complexe nucléolipoprotéique présent dans le cytoplasme et contenant de l'ADN, de l'ARN ainsi que l'ADN et l'ARN polymérases. La fraction d'ADN du virtosome est nouvellement synthétisée dans le noyau et partage des caractéristiques communes avec l'ADN métabolique qui correspond à des copies de gènes supplémentaires formés pour la production rapide d'ARN messagers et qui seront détruits ultérieurement *(67)*. Il a donc été proposé que l'ADN métabolique servirait de précurseur de l'ADN présent dans les virtosomes *(68)*.

Des études *in vitro* ont également montré que les cellules normales *(69)* en culture peuvent libérer activement de l'ADN dans le milieu de culture. La concentration de cet ADN atteint un certain plateau quel que soit le temps d'incubation, même après le lavage des cellules et leur ré-incubation dans un milieu neuf, ce qui suggère qu'un mécanisme de régulation pourrait être impliqué pour contrôler la libération d'ADN extracellulaire.

Une autre forme de libération active de l'ADN circulant est celle de l'ADN contenu dans des vésicules extracellulaires, comme les exosomes et les microvesicules. Les exosomes sont libérés par exocytose, ont une taille comprise entre 30 et 100 nm et portent des fragments

d'ARN et d'ADN *(70)*. Une étude a montré que l'ADN exosomal représente l'ensemble du génome et reflète le statut mutationnel des cellules tumorales parentales *(71)*. Guescini *et al.* ont également rapporté que les exosomes dérivés d'astrocytes et de cellules de glioblastome peuvent transporter l'ADN mitochondrial *(72)*. Les microvésicules ont une taille de 50 à 1000 nm, et sont libérés dans l'espace extracellulaire par le bourgeonnement et la fission de la membrane plasmique *(73)*. Ceux provenant des cellules tumorales et contenant des substances oncogéniques sont appelées oncosomes *(74)*.

La découverte de la sécrétion active de l'ADNcir a ainsi remis en cause l'hypothèse selon laquelle l'apoptose serait l'origine majeure et la plus pertinente de l'ADN dans la circulation.

c) La dégradation des cellules sanguines

Lui *et al.* ont émis l'hypothèse que la mort des cellules hématopoïétiques serait principalement responsable de la présence de l'ADNcir chez les individus sains *(75)*. Pour tester cette hypothèse, les auteurs ont utilisé un modèle de greffe de moelle osseuse de sexe opposé. Sun *et al.*, quant à eux, ont effectué un séquençage au bisulfite du génome sur des échantillons de plasma provenant de femmes enceintes, de patients atteints d'un carcinome hépatocellulaire et de sujets ayant subi une greffe de moelle osseuse ou de foie, et ont comparé le profil de méthylation avec plusieurs types de tissus *(76)*. Dans la majorité des échantillons, les globules blancs étaient les contributeurs majeurs de l'ADNcir dans le plasma. D'autre part, Snyder *et al.* ont montré que le profil de fragmentation de l'ADNcir, selon le positionnement des nucléosomes, peut être utilisé pour identifier son origine *(77)*. Ils ont observé que, chez les sujets sains, l'espacement des nucléosomes de l'ADNcir était le plus fortement corrélé avec les caractéristiques épigénétiques des cellules lymphoïdes et myéloïdes.

d) *La dégradation d'agents pathogènes*

La dégradation d'agents pathogènes, comme les bactéries et les virus, peut être à l'origine de faibles niveaux d'ADN dans la circulation. Les ADN du virus d'Epstein Barr et du virus du papillome humain par exemple ont été détectés dans des cas de carcinome nasopharyngé *(78, 79)* et de col de l'utérus *(80)* respectivement, et sont devenus des marqueurs potentiels pour le dépistage et le suivi des patients atteints de ces cancers.

e) *ADN lié à la surface des cellules*

De l'ADN peut se lier à des récepteurs membranaires, comme le récepteur TLR-9 (Toll-like receptor 9) présent à la surface des leucocytes *(81)* et des érythrocytes *(82)*. Cet ADN lié à la surface peut être soit nu, soit associé à des structures vésiculaires, soit lié à des macromolécules *(44)*, et peut être, par la suite, internalisé dans la cellule. Il a été montré, *in vitro*, que des fragments d'ADN se trouvent à la surface des cellules HeLa (lignée cellulaire du carcinome cervical humain) et des cellules A431 (lignée cellulaire du carcinome squameux) *(83)*. Une hypothèse a été émise suggérant que l'ADNcir activement libéré par les cellules se fixe d'abord aux membranes cellulaires en se liant aux protéines de liaison, et lorsque la capacité des cellules à se lier à l'ADN devient saturée, il se détache des cellules et pénètre dans la circulation sanguine *(84)*.

2. L'ADN mitochondrial

L'ADN circulant est constitué non seulement d'ADN génomique d'origine nucléaire (ADNnu) mais aussi d'ADN mitochondrial (ADNmt), d'où l'importance d'étudier ce dernier quand il s'agit d'une étude de l'ADN circulant global.

a) Propriétés de l'ADN mitochondrial

Les mitochondries jouent un rôle essentiel dans le métabolisme énergétique de la cellule. Ces organelles fonctionnent de manière semi-autonome: elles possèdent leur propre génome dont elles sont capables d'assurer la réplication et la transcription *(85, 86)*.

L'ADN mitochondrial (ADNmt) est un ADN circulaire bicaténaire (brin lourd et brin léger), de 16 569 pb. Il est présent à raison de plusieurs centaines de copies par cellule : en effet, il existe 2 à 10 copies d'ADNmt dans chaque mitochondrie, et des centaines de mitochondries par cellule. Cet ADN comprend 37 gènes qui codent pour 13 polypeptides de la chaine respiratoire mitochondriale, ainsi que pour 22 ARN de transfert (ARNt) et 2 ARN ribosomaux (ARNr) nécessaires à la synthèse de ces polypeptides dans la mitochondrie **(Figure 6)**.

Ces gènes comprennent :

- le complexe I de la NADH déshydrogénase constitué de 7 sous-unités ND1 à ND6 et ND4L

- le complexe III du cytochrome b

- le complexe IV du cytochrome c oxydase (complexe IV de la chaîne respiratoire) constitué de 3 sous-unités COI à COIII

- le complexe V des ATP synthases 6 à 8.

L'ADN mitochondrial est caractérisé par l'absence quasi-totale d'introns. La région D-Loop est la seule région non codante de cet ADN. Cette région est constituée de 1121 pb et correspond à un site essentiel pour la réplication et la transcription du génome mitochondrial. Elle contient le point d'origine de la synthèse du brin lourd par où commence la réplication de

l'ADNmt, ainsi que la majorité des promoteurs de la transcription de ce génome et plusieurs sites de fixation de certains facteurs de transcription.

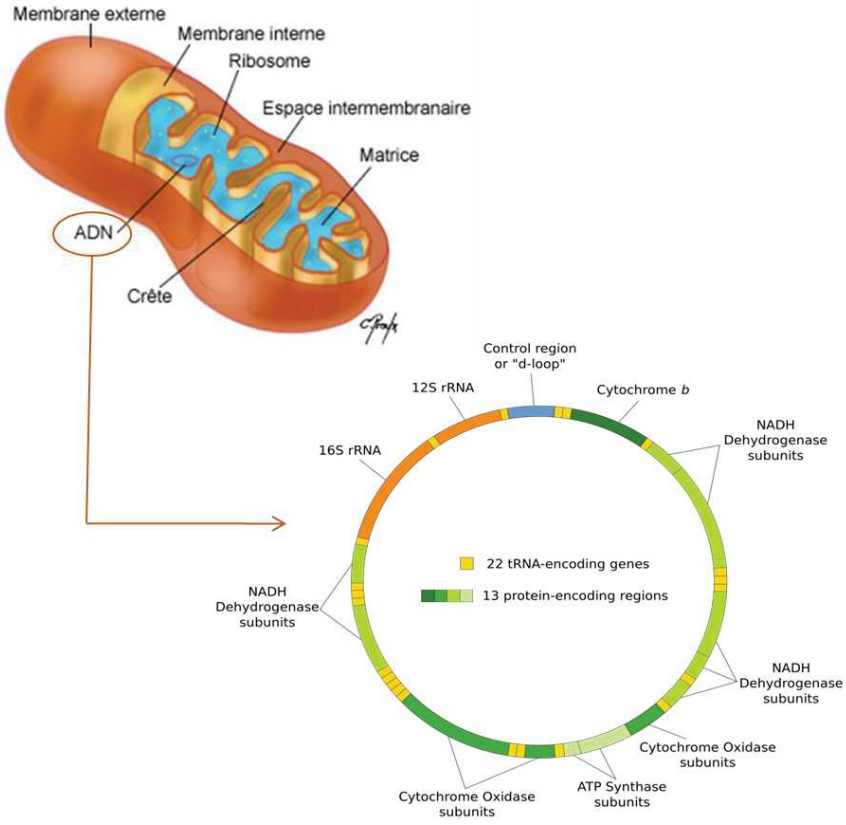


Figure 6: *La mitochondrie et les différents gènes codés par l'ADN mitochondrial*

Le taux de mutations de l'ADNmt est de dix à 200 fois supérieur à celui de l'ADNnu *(87)* ce qui est dû à plusieurs facteurs :

- Le taux important de ROS (reactive oxygen species ou espèces réactives de l'oxygène) générées dans la mitochondrie au cours de la phosphorylation oxydative, exposant d'avantage l'ADNmt à ces agents mutagènes.

- Les systèmes de réparation de l'ADNmt moins développés et moins efficaces que ceux de l'ADNnu.

- Les erreurs de réplication de l'ADN polymérase γ mitochondriale qui, bien que hautement fidèle (moins d'un nucléotide erroné inséré sur 250 000 nucléotides synthétisés), est responsable d'erreurs de réplication conduisant à des insertions ou des délétions et à un décalage du cadre de lecture.

b) ADN mitochondrial circulant

Outre les travaux menés sur l'ADNcir d'origine nucléaire, plusieurs études récentes ont démontré la présence d'ADN mitochondrial dans la circulation, la première étant celle démontrant la présence de la mutation au niveau du gène *tRNA$^{Leu(UUR)}$* de l'ADN mitochondrial, qui correspond à une substitution d'une adénine par une guanine à la position 3243, dans le plasma et le sérum de patients atteints de diabète du type 2 *(88)*.

Par la suite, des mutations somatiques de l'ADN mitochondrial circulant, ainsi que des altérations du nombre de copies, ont été détectées dans divers types de tumeurs *(89)* **(voir partie III A : Caractéristiques de l'ADN circulant tumoral/ ADNcir mitochondrial en oncologie)**. L'ADNcir mitochondrial a donc logiquement été étudié en tant que biomarqueur non invasif de diagnostic et de pronostic en oncologie. De plus, plusieurs études ont montré que le taux d'ADNcir mitochondrial diffère significativement entre patients atteints de cancer et individus sains *(90)*.

D'autres pathologies, comme en cas de traumatisme *(91)* ou de maladie mentale comme la dépression majeure *(92)*, ont été associées à une augmentation mesurable de l'ADNmt plasmatique. L'ADNcf mitochondrial semble également être associé à la neurodégénérescence, avec des diminutions significatives dans le liquide céphalo-rachidien de patients atteints de la maladie de Parkinson *(93)*, d'Alzheimer *(94)*, et de la sclérose en plaque *(95)*.

Enfin, en raison de sa similarité avec l'ADN bactérien et de sa capacité à stimuler les réponses immunitaires innées par l'intermédiaire du récepteur TLR-9, l'ADN mitochondrial circulant a suscité un intérêt en tant que déclencheur de l'inflammation *(96, 97)* **(voir partie I/C/3 : Aspects fonctionnels des ADN circulants)**.

3. Aspects fonctionnels des ADN circulants

Les ADN circulants peuvent jouer différents rôles dans l'organisme en tant que messagers intercellulaires, déclencheurs du système immunitaire ou facteur favorisant l'apparition de métastases dans le cas du cancer.

a) Messagers intercellulaires

L'ADNcir a la capacité d'agir comme messager intercellulaire. Mittra *et al. (98)* ont en effet montré que l'ADN isolé du plasma de patients sains ou atteints de différents types de cancers, puis utilisé pour traiter plusieurs lignées cellulaires murines, avait été incorporé dans ces cellules et s'était localisé dans le noyau. Cette incorporation avait causé des cassures double brin et une activation de l'apoptose des cellules traitées. Des résultats similaires ont été obtenus après l'injection d'ADN circulants d'origine humaine chez des souris Balb/c *(98)*. D'ailleurs, l'ADN extracellulaire peut être reconnu par diverses protéines de liaison à l'ADN localisées à la surface des cellules et peut être internalisé pour être soit dégradé en mononucléotides dans la cellule, soit transporté dans le noyau *(99)*.

Il a été suggéré que le virtosome contenant de l'ADNcir, peut entrer dans d'autres cellules et modifier leur biologie, en causant par exemple des changements immunologiques et même la transformation de cellules normales en cellules cancéreuses *(66, 100)*. Le rôle des vésicules extracellulaires dans la communication intercellulaire a aussi été démontré *(101)*. Ces vésicules peuvent être internalisées par les cellules réceptrices, et leur contenu dérivant de la cellule d'origine, dont l'ADN et l'ARN, peut y être transféré. Waldenström *et al.* ont par exemple montré que des vésicules extracellulaires dérivées de cardiomyocytes murins adultes peuvent transférer de l'ARNm, mais aussi de l'ADN, aux fibroblastes *(102)*.

Ermakov *et al*. ont, quant à eux, indiqué que l'ADN extracellulaire peut participer au « Bystander effect » induit par une exposition des lymphocytes et des cellules endothéliales aux rayons X *(103)*. Cet effet fait référence au transfert d'information de cellules exposées à des agents nocifs physiques ou chimiques vers des cellules adjacentes non exposées *(67)*. En effet, en cas de stress oxydatif, les cellules mourantes libèrent leur ADN « oxydé » endommagé

(104). Cet ADN extracellulaire serait un médiateur de signal de stress qui favoriserait le « Bystander effect ». On pourra supposer qu'un processus similaire se produit dans le corps humain suite à une radiothérapie, et que lorsque l'ADNcir oxydé est libéré par les cellules tumorales mourantes irradiées, il pénètre dans la circulation et est transporté vers les organes éloignés, exerçant ainsi un effet systémique *(104)*.

b) Immunité

L'ADN extracellulaire est capable d'activer le système immunitaire et d'induire une réponse inflammatoire. En raison de son origine bactérienne, l'ADN mitochondrial peut par exemple servir de DAMP (damage-associated molecular pattern) lorsqu'il est libéré des cellules *(105)*. Cet ADN peut en particulier se lier au récepteur TLR9 des leucocytes par l'intermédiaire des motifs CpG hypométhylés *(81)* et déclencher une réaction inflammatoire. Poli *et al.* ont par ailleurs montré que l'interleukine 26 (IL-26) qui s'exprime dans les cellules musculaires lisses, peut se lier à l'ADN extracellulaire génomique ou mitochondrial libéré par les cellules mourantes, ainsi qu'aux NETs. Le complexe déclenche la sécrétion des cytokines pro-inflammatoires par les monocytes par l'intermédiaire de la voie STING et de l'inflammasome *(106)* **(Figure 7)** pour attirer les cellules immunitaires.

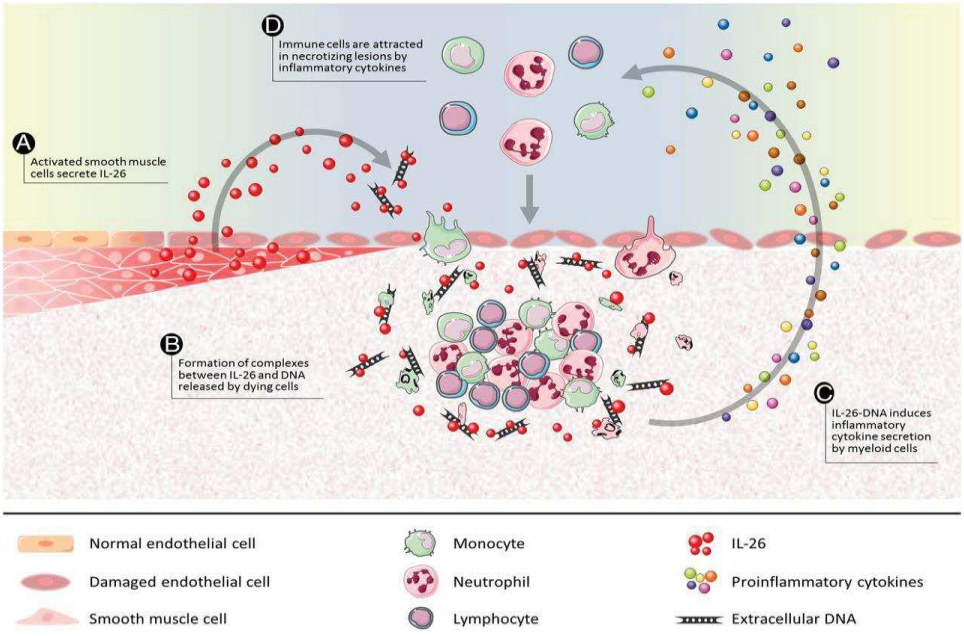


Figure 7: *Le rôle de l'IL-26 dans l'induction d'une réponse inflammatoire par liaison à l'ADN extracellulaire (106)*

c) Génométastase

Il a été suggéré que la propagation métastatique du cancer peut éventuellement être favorisée par de l'ADN d'origine tumorale circulant dans le sang *(107)*. Différentes études ont en effet montré que des fragments d'ADNcir peuvent pénétrer dans les cellules et modifier la biologie de ces cellules réceptrices*(108)*. En particulier, il a été montré que l'ADNcir présent dans le plasma de patients cancéreux peut permettre le transfert d'oncogènes à des cellules en culture *in vitro* et favoriser leur transformation *(109)*. De même, l'ADN extracellulaire présent dans le sérum de patients atteints de cancer du côlon ou dans le milieu de culture de cellules SW480 (cellules humaines de cancer colorectal), est capable d'induire la transformation de cellules murines NIH3T3 par leur simple incubation avec ce sérum ou ce surnageant *(110)*. García-Olmo *et al.* ont ainsi proposé la théorie de « génométastase » *(111)* selon laquelle les métastases pourraient survenir par transfection de cellules sensibles situées dans des organes cibles éloignés avec des oncogènes dominants circulants issus de la tumeur. Cette théorie est en accord avec le mécanisme de transfert horizontal d'ADN entre différentes cellules de l'organisme *(112, 113)*.

II. Objectifs de la thèse

Mon projet de thèse a pour but général de valider l'analyse combinée des ADN circulants d'origine nucléaire et mitochondriale à des fins diagnostiques. Deux modèles d'études ont été utilisés pour atteindre cet objectif : d'une part l'étude des ADN extracellulaires dans le domaine du cancer, et d'autre part dans le domaine de la procréation assistée.

Objectif n°1 :

Le cancer figure parmi les premières causes de décès dans le monde en raison de sa détection à des stades tardifs, sans apparition de symptômes *(114)*. Les mécanismes responsables de sa progression ne sont pas encore complètement élucidés, mais une détection précoce pourrait considérablement augmenter les chances de succès du traitement et donc, de survie des patients. Actuellement, les tests pour la détection précoce du cancer ne sont pas efficaces, du fait d'une faible sensibilité en partie en raison de l'hétérogénéité inter et intra-tumorale.

L'ADN circulant s'est avéré être un biomarqueur émergent pour la détection précoce du cancer ce qui a poussé plusieurs équipes à travailler sur la mise au point de tests de dépistage du cancer à partir d'une simple analyse sanguine *(115)*. Différentes approches ont été utilisées, comme, par exemple, l'association de la détection d'altérations génétiques au niveau de l'ADNcir avec des marqueurs protéiques *(116)*, la détection d'ADN viral dans le plasma *(79)*, ou encore l'analyse du méthylome de l'ADNcir *(117)*. Toutefois, les questions concernant leur sensibilité et leur spécificité, en particulier dans le cas des stades précoces du cancer, peuvent affecter leur fiabilité.

Notre équipe a pour ambition de développer un test de dépistage ou de détection précoce du cancer selon une approche différente. **Le premier objectif de ma thèse était donc d'étudier les caractéristiques quantitatives et structurelles de l'ADNcir pour le dépistage et la détection précoce du cancer, en prenant en compte l'origine (ADNcir nucléaire et mitochondrial) et la structure (fragmentation et profil de taille) de cet ADN.**

Les objectifs spécifiques sont détaillés dans la partie III « ADN circulant et dépistage du cancer » du manuscrit.

Objectif n°2 :

Les technologies de procréation assistée, notamment la fécondation *in vitro* (FIV), permettent une grossesse quand le couple rencontre des problèmes de fertilité. Elles consistent à faciliter la fécondation d'un ovule par un spermatozoïde dans un milieu favorable à leur survie *ex-utero*. L'embryon formé quelques jours après fécondation est alors implanté dans l'utérus de la mère. Actuellement, l'évaluation de la qualité de l'embryon et le choix de l'embryon à implanter sont basés sur sa morphologie *(118)*. Cette simple observation est probablement la cause du taux élevé d'échecs d'implantation, et l'utilisation d'autres biomarqueurs capables de refléter la qualité de l'embryon et son potentiel d'implantation serait pertinent. Or, il a été montré que l'embryon libère de l'ADN extracellulaire dans le milieu de culture lors d'une FIV *(119, 120)*, cet ADN pouvant être un biomarqueur prédictif pour la qualité de l'embryon. **Le premier aspect du deuxième objectif de ma thèse était donc d'étudier la relation entre la quantité d'ADN extracellulaire d'origine nucléaire et mitochondriale dans le milieu de culture des embryons, et la qualité et la viabilité de ces embryons lors d'une FIV.**

En outre, une FIV peut également être indiquée pour des couples à risque vis-à-vis de certaines maladies génétiques graves afin d'éviter leur transmission à la descendance. Dans ce cas, un test génétique préimplantatoire (PGT pour preimplantation genetic testing) est utilisé pour détecter, chez l'embryon, des altérations génétiques ou des anomalies chromosomiques connues pour être à l'origine de l'apparition de ces maladies *(121)*. Ce test est souvent invasif et s'effectue en réalisant une biopsie de cellule de l'embryon, une procédure assez discutable d'un point de vue éthique. Dans ce contexte, l'ADN extracellulaire présent dans le milieu de culture pourrait représenter une alternative efficace et non invasive. **Le deuxième aspect du deuxième objectif de ma thèse était d'évaluer la pertinence de l'utilisation de l'ADN extracellulaire pour détecter des séquences spécifiques de l'embryon dans le milieu de culture, comme la détection du gène *SRY* pour la détermination du genre de l'embryon dans le cas des maladies génétiques associées au sexe, ou l'identification d'embryons porteurs de la mutation responsable de la mucoviscidose chez des couples à risque porteurs de cette mutation.**

Les objectifs spécifiques sont détaillés dans la partie IV « ADN extracellulaire et diagnostic préimplantatoire » du manuscrit.

III. ADN circulant et dépistage du cancer

Après sa découverte en 1948 *(1)*, l'intérêt pour l'ADN circulant a augmenté notamment en oncologie. Le nombre croissant d'études publiées pour valider son utilisation en pratique clinique en témoigne, et indique que l'analyse de l'ADNcir peut avoir un rôle pertinent et décisif pour la prise en charge non invasive des patients atteints de cancer. La détection et la caractérisation des altérations génétiques et épigénétiques de cellules tumorales à partir de l'analyse de l'ADN circulant dans le plasma *(122)*, ainsi que l'étude quantitative et qualitative de cet ADN, peut en effet permettre de diagnostiquer, surveiller la récidive et évaluer la réponse au traitement par une simple prise de sang *(123)*.

A. Caractéristiques de l'ADN circulant tumoral

1. Origine cellulaire

L'ADNcir dans le sang des patients cancéreux peut avoir différentes origines cellulaires. La quantité basale d'ADNcir dans le sang des individus sains a permis de démontrer que les cellules saines (globules blancs et autres) libèrent de l'ADN dans la circulation *(75–77)*. En outre, la détection d'altérations génétiques et épigénétiques dans l'ADNcir libéré par la tumeur *(122)* suggère que les cellules cancéreuses sont une source importante d'ADNcir tumoral. Cependant, la tumeur n'est pas uniquement constituée de cellules malignes, mais également de cellules du microenvironnement tumoral, comme les cellules stromales, les lymphocytes et d'autres cellules immunitaires qui peuvent constituer une source potentielle d'ADNcir *(44)*. Ainsi, l'ADNcir d'un patient atteint de cancer peut provenir des cellules saines, des cellules malignes ou encore des cellules du microenvironnement tumoral **(Figure 8)**.

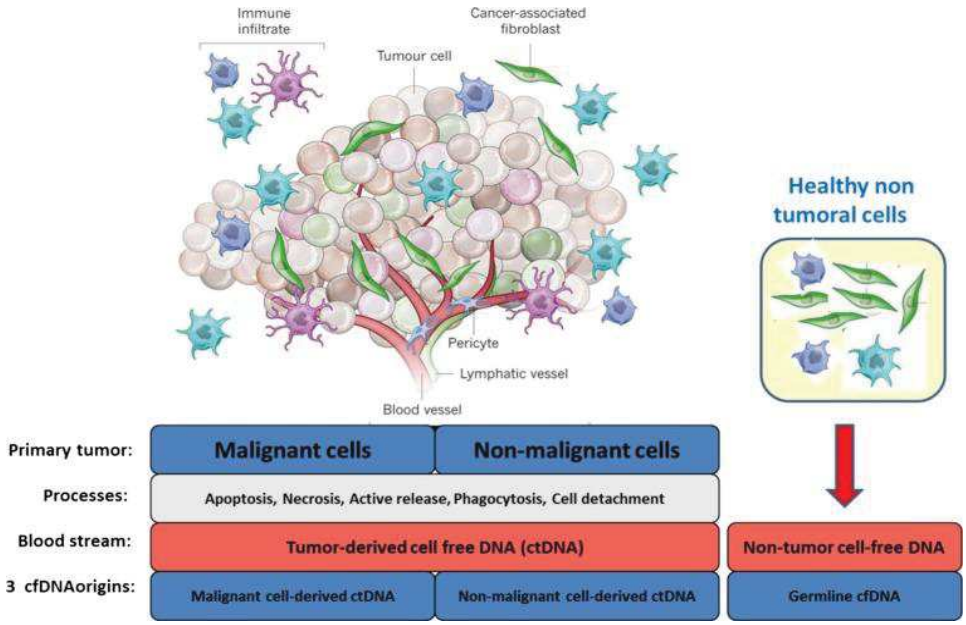


Figure 8: *Origines cellulaires de l'ADNcir dans le sang d'un patient atteint de cancer (44)*

2. Profil de taille et fragmentation de l'ADNcir tumoral

L'étude de la taille des ADN circulants est pertinente, d'une part pour améliorer leur détection, et d'autre part pour mieux comprendre leur structure et leur origine. D'une manière générale la taille de l'ADNcir est de 160-180 pb chez les patients atteints de cancer, ce qui correspond à la taille d'un mononucléosome dégradé et libéré généralement par apoptose *(124, 125)*. Cependant, plusieurs travaux ont montré que la taille de ces ADN circulants était supérieure à 10 000 pb, suggérant une origine nécrotique qui est associée à une dégradation de la chromatine et, de fait, à une libération de fragments de grande taille *(51)*.

En utilisant différentes approches expérimentales, des études récentes ont démontré que l'ADNcir d'origine tumorale était plus court que l'ADNcir sauvage *(126–128)* **(Figure 9)**. L'utilisation d'un modèle de souris xénogreffées par des cellules cancéreuses d'origine humaine *(129)* a permis de discriminer l'ADNcir d'origine non tumorale de l'ADNcir d'origine tumorale, et les résultats ont été confirmés en étudiant les profils de taille de cet ADNcir dans le plasma de patients atteints de cancer colorectal métastasique *(126, 130)*. Les résultats ont en effet montré que ce profil de distribution de taille peut permettre de discriminer des individus sains de patients cancéreux, puisque que la proportion d'ADNcir de taille inférieure

à 100pb est plus élevée chez les patients cancéreux alors que la proportion des fragments de taille supérieure à 400pb est plus élevée chez les individus sains.

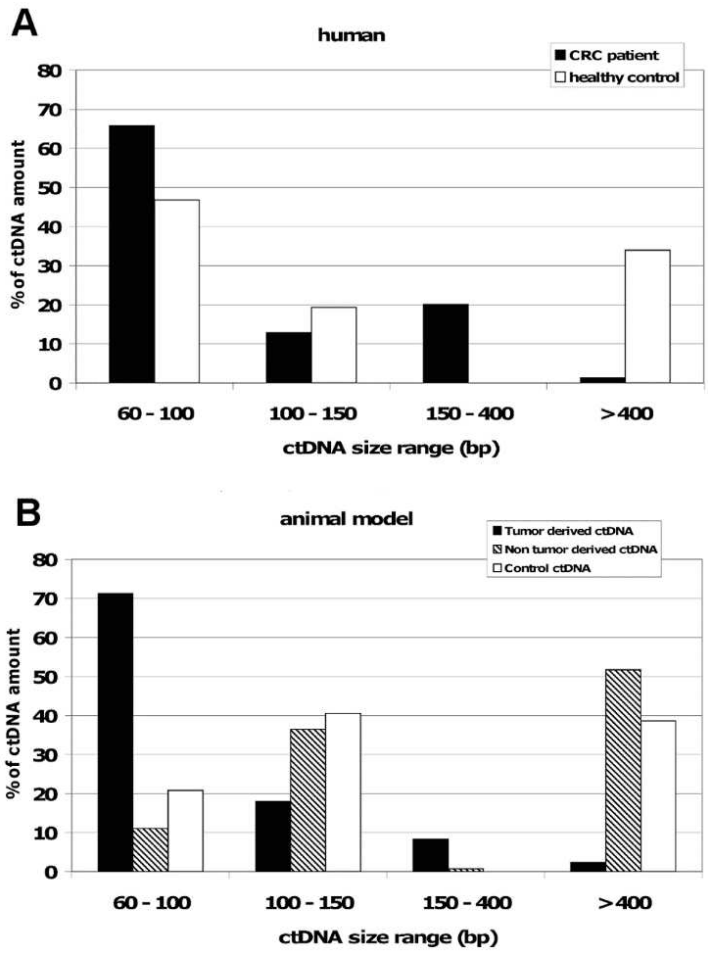


Figure 9: *Comparaison de la distribution de taille des fragments d'ADNcir à partir d'échantillons cliniques d'individus sains et de patients atteints de CCR* ***(A)*** *et de l'ADNcir tumoral et non-tumoral des souris xénogreffées par des cellules cancéreuses humaines et des souris contrôles non-xénogreffées* ***(B)*** *(126)*

Underhill *et al.* ont par ailleurs montré un décalage de la taille des fragments d'ADNcir dans le plasma des patients sains (167pb) par rapport aux patients cancéreux (134-144pb) *(131)*. La cause de ce raccourcissement n'est pas entièrement élucidée, mais les explications possibles incluent des différences dans la compaction des nucléosomes ou le mode d'action des nucléases entre les cellules hématopoïétiques (qui contribuent à la majorité de l'ADNcir chez les individus sains) et d'autres tissus d'origine *(123)*. D'après Sanchez *et al.*, la majorité de l'ADNcir détectable dans le sang de patients cancéreux indique une empreinte nucléosomique

(132). Leurs résultats valident l'utilisation de la Q-PCR ou du séquençage du génome après préparation de librairie simple brin (SSP-S pour Whole-genome sequencing of single-stranded DNA library preparation) pour l'obtention d'un signal analytique optimal de l'ADNcir **(Figure 10)**.

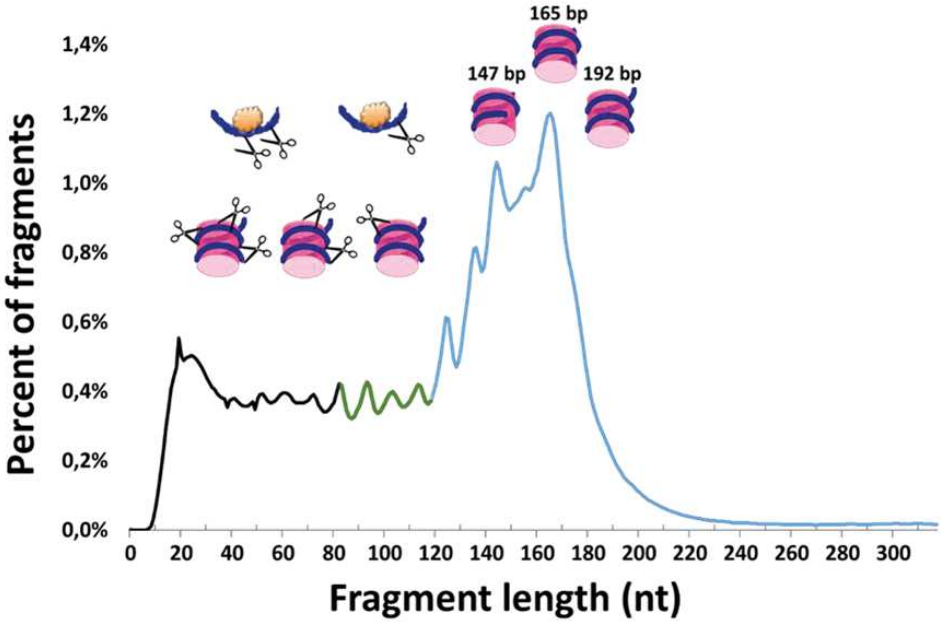


Figure 10: *Structure et fragmentation de l'ADNcir chez un patient cancéreux en ce qui concerne le profil de taille déterminé par SSP-S (132)*

Ainsi, les différences concernant la longueur des fragments d'ADNcir pourraient être exploitées pour augmenter la sensibilité de la détection des acides nucléiques tumoraux *(133)* en sélectionnant, par exemple, les fragments ayant une taille entre 90 et 150pb *(134)*. De plus, une analyse de la distribution de la taille des fragments d'ADNcir, selon le positionnement des nucléosomes, pourrait permettre d'identifier la source de ces acides nucléiques circulants *(77)*.

Des résultats comparables ont été obtenus à partir d'échantillons de sang de femmes enceintes, montrant des différences de profil de taille de l'ADNcir d'origine maternelle ou fœtale *(135, 136)*, ou encore de patients ayant subi une transplantation d'organe, montrant des différences de profil de taille de l'ADNcir provenant du donneur ou du receveur *(137)*.

3. ADNcir mitochondrial en oncologie

- **Mitochondrie et cancer**

Comme évoqué précédemment, l'ADNcir est composé d'ADN d'origine nucléaire (ADNnu) et mitochondriale (ADNmt). Or, des altérations de l'ADN mitochondrial ont été associées à des dysfonctionnements de la mitochondrie et à la promotion de la carcinogenèse *(138)*.

En effet, depuis longtemps, la mitochondrie a été suspectée de jouer un rôle important dans le développement et la progression des cancers, et l'influence fondamentale du métabolisme mitochondrial sur toutes les étapes de l'oncogenèse a finalement été reconnue *(139)*. Ainsi, des mutations somatiques et germinales de l'ADNmt ont été signalées dans une grande variété de cancers *(140, 141)* **(Figure 11)**. De plus, une méta-analyse a révélé que de nombreuses mutations de cet ADN au niveau des cellules cancéreuses, soit inhibent ou altèrent la phosphorylation oxydative et servent à stimuler la transformation néoplasique, soit facilitent l'adaptation des cellules cancéreuses aux changements bioénergétiques de l'environnement.

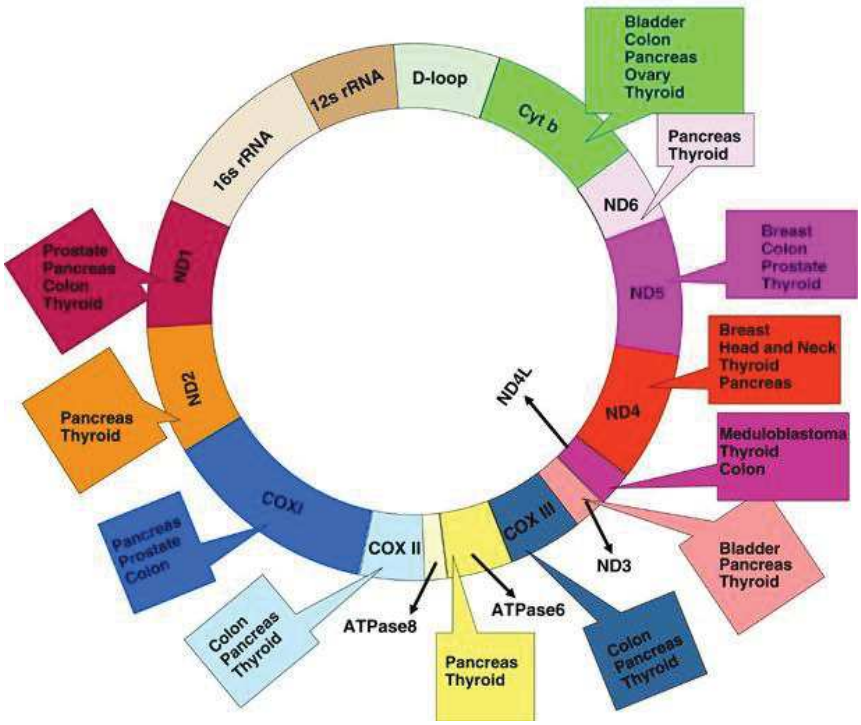


Figure 11: *Les mutations au niveau des gènes de l'ADN mitochondrial dans différents types de cancers (141)*

Ces mutations peuvent être à l'origine d'un dysfonctionnement de la chaine respiratoire mitochondriale, ce qui peut mener à la libération de taux anormalement élevés de ROS et ainsi causer des dommages oxydatifs de l'ADNmt, qui, à leur tour, vont altérer la respiration et générer des taux importants de ROS, provoquant ainsi un stress oxydatif continu, favorable au processus de la cancérogenèse *(87)*.

En plus des différentes mutations qui peuvent s'accumuler au niveaux des régions codantes et non codantes de l'ADNmt, différentes études ont révélé des altérations du nombre de copies d'ADNmt dans un grand nombre de cancers *(87)*. Ces altérations sont soit des augmentations soit des diminutions du nombre de copies, selon le type du cancer *(142)*.

- **ADNcir mitochondrial et cancer**

Différents travaux se sont intéressés à la quantification de l'ADNcir d'origine mitochondriale dans le plasma de patients atteints de plusieurs types de cancer *(89)*. Cette quantification semble avoir une signification pronostique, bien que la tendance de la relation entre taux d'ADNcir mitochondrial et tumeur varie selon le type de cancer.

Certaines études ont, par exemple, mis en évidence une augmentation de la quantité d'ADNcir d'origine mitochondriale chez des patients cancéreux par rapport à des individus sains. Une quantification à partir du sérum, dans le cas de tumeurs urologiques (cancer de la vessie, de la prostate et carcinome des cellules rénales) par exemple *(143)*, a ainsi montré que le taux d'ADNmt circulant augmente significativement chez les patients cancéreux par rapport aux individus sains. De plus, l'intégrité de l'ADNmt circulant était plus élevée dans les cas de cancer du rein et de la vessie en comparaison du groupe contrôle et cancer de la prostate. Des résultats similaires ont été observés dans le cas du cancer du testicule *(144)*, bien que l'intégrité de l'ADNmt était, dans ce cas, similaire chez les patients atteints de cancers et les individus sains. En revanche, dans le cas du cancer épithélial de l'ovaire, les taux d'ADNcir mitochondrial étaient significativement plus élevés chez les patients cancéreux alors que l'intégrité était diminuée *(145)*. Enfin, une autre étude a quantifié le nombre de copies d'ADNmt dans le sang total et le plasma (ADNmt encapsulé ou non dans des exosomes) *(146)*. Une augmentation du nombre de copies a été détectée chez les patients de stade avancé par

rapport aux individus sains, mais cette augmentation n'était pas significative pour l'ADNmt libre circulant.

D'autres travaux, comme ceux concernant le sarcome d'Ewing (un cancer des os affectant principalement les enfants et les adolescents) *(147)*, ou le carcinome hépatocellulaire lié au virus de l'hépatite B *(148)* ont montré que la concentration d'ADNmt circulant était significativement inférieure chez les patients cancéreux en comparaison des individus sains .

En outre, des études comparatives entre le taux d'ADNcir nucléaire et mitochondrial ont été réalisées pour différents types de cancers. Une étude menée par Kohler *et al. (149)*, a montré que le taux d'ADNcir d'origine nucléaire est significativement plus élevé chez les patients cancéreux (*vs* des contrôles sains) et celui d'ADNmt circulant est trouvé supérieur chez les individus sains. Meddeb *et al*. ont publié des résultats similaires dans le cas du cancer colorectal *(150)*. En revanche, les travaux de Lu *et al. (151)* et Jiang *et al. (128)* ont montré que les concentrations d'ADNcir nucléaire et mitochondrial étaient plus élevées dans des cancers rénaux et hépatocellulaires respectivement par rapport aux individus sains.

Finalement, l'ensemble de ces résultats contradictoires et variables sur la quantification de l'ADNmt circulant dans le plasma des patients cancéreux ne permettent pas de conclure quant au potentiel de l'ADNmt circulant comme biomarqueur.

B. Utilité clinique de l'analyse de l'ADN circulant en oncologie

L'analyse de l'ADNcir a été évaluée dans diverses études pour étudier son application et son intérêt clinique en oncologie *(123, 152)*. Il s'est avéré que l'ADNcir peut être un biomarqueur utile à toutes les étapes de la prise en charge des patients atteints de cancer, depuis le diagnostic et la détection précoce, le pronostic, la détection de la maladie minimale résiduelle, le choix du traitement, jusqu'à la surveillance de la réponse au traitement ainsi que l'évolution clonale et l'apparition de résistance **(Figure 12)**.

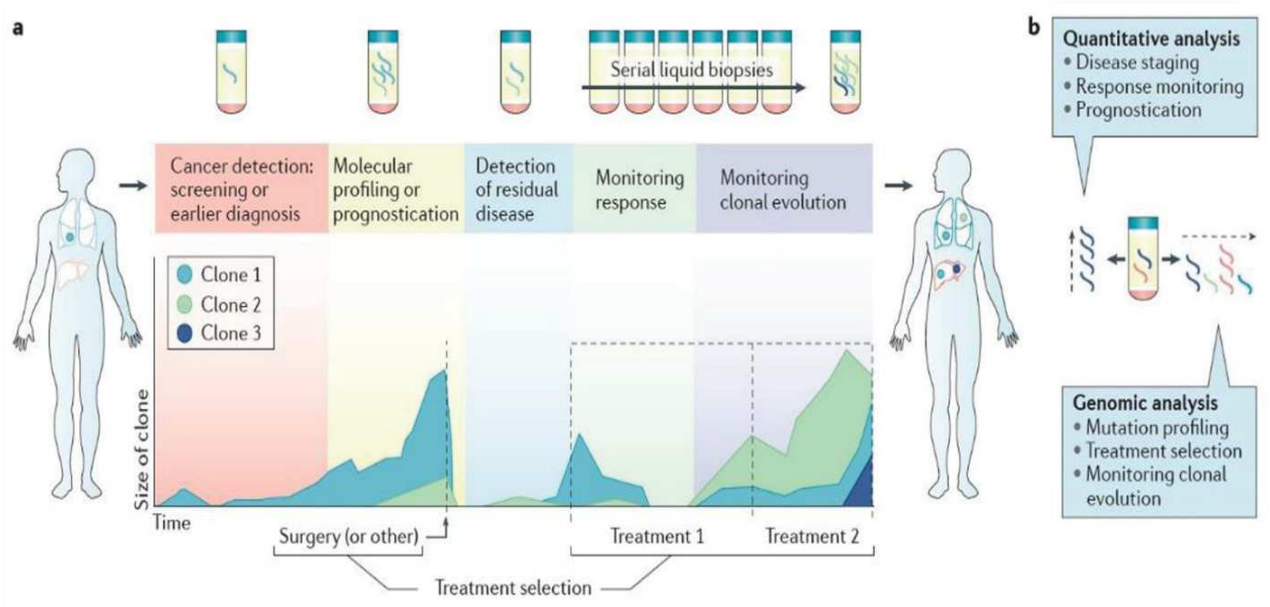


Figure 12: *Les applications potentielles de l'ADN circulant durant la prise en charge d'un patient cancéreux (152)*

Actuellement, l'analyse moléculaire est réalisée à partir du tissu tumoral pour apporter des informations sur la présence de facteurs génétiques prédictifs. Cette méthode nécessite du temps, est coûteuse, et présente certaines limitations comme le manque de tissu ou le faible pourcentage de cellules tumorales prélevées pour l'analyse. De plus, les résultats sont fonction de la pièce du tissu prélevée qui ne fournit pas d'idée globale de la tumeur étant donné l'hétérogénéité clonale intra et inter-tumorale entre tumeur primaire et métastases appariées. L'analyse de l'ADN circulant ou analyse de la biopsie liquide permet de s'affranchir de ces limitations liées à l'analyse du tissu et pourrait même la remplacer, tout en étant rapide,

facile à mettre en œuvre, non invasive, sensible, spécifique, reproductible et peu coûteuse *(12)*. De plus, cette approche permet de détecter l'ensemble des clones, de palier au problème d'hétérogénéité clonale et de refléter ainsi l'hétérogénéité de la tumeur dans sa globalité.

Enfin, parallèlement à l'analyse qualitative, l'analyse quantitative de l'ADN circulant peut également apporter des informations de valeur diagnostique et pronostique.

1. Stade de la tumeur et pronostic

La classification ou le sous-typage du cancer semble possible par analyse de l'ADNcir. En effet, la concentration en ADNcir corrèle avec la taille de la tumeur *(153, 154)* et le stade de la maladie *(155, 156)*. Cette corrélation indique une valeur pronostique potentielle de l'ADNcir.

Ainsi, la fréquence allélique de mutants du gène *TP53* a été calculée dans le plasma de patientes atteintes de carcinome ovarien séreux de haut grade *(153)*. Cette fréquence corrélait avec le volume tumoral. De même, Bettegowda *et al. (155)* ont montré que de l'ADNcir tumoral muté était détectable à des concentrations relativement élevées chez plus de 80% de patients atteints de différents types de cancers métastatiques, et à des concentrations plus faibles mais détectables chez une fraction non négligeable de patients atteints de cancers localisés. Des résultats similaires ont été publiés par Newman *et al.* dans le cas du cancer du poumon non à petites cellules *(156)*, où l'ADNcir tumoral a été détecté chez 100% de patients de stades II à IV et 50 % de stade I de la maladie avec une bonne corrélation entre le taux de cet ADN et la taille de la tumeur.

La valeur pronostique de l'ADNcir a également été évaluée pour différents types de cancer. Une concentration élevée d'ADNcir global et d'ADNcir tumoral, ainsi qu'une charge mutationnelle plus élevée et un indice de fragmentation plus réduit dans le plasma de patients atteints de cancer colorectal, corrélaient avec une survie globale plus courte *(157)* **(Figure 13)**.

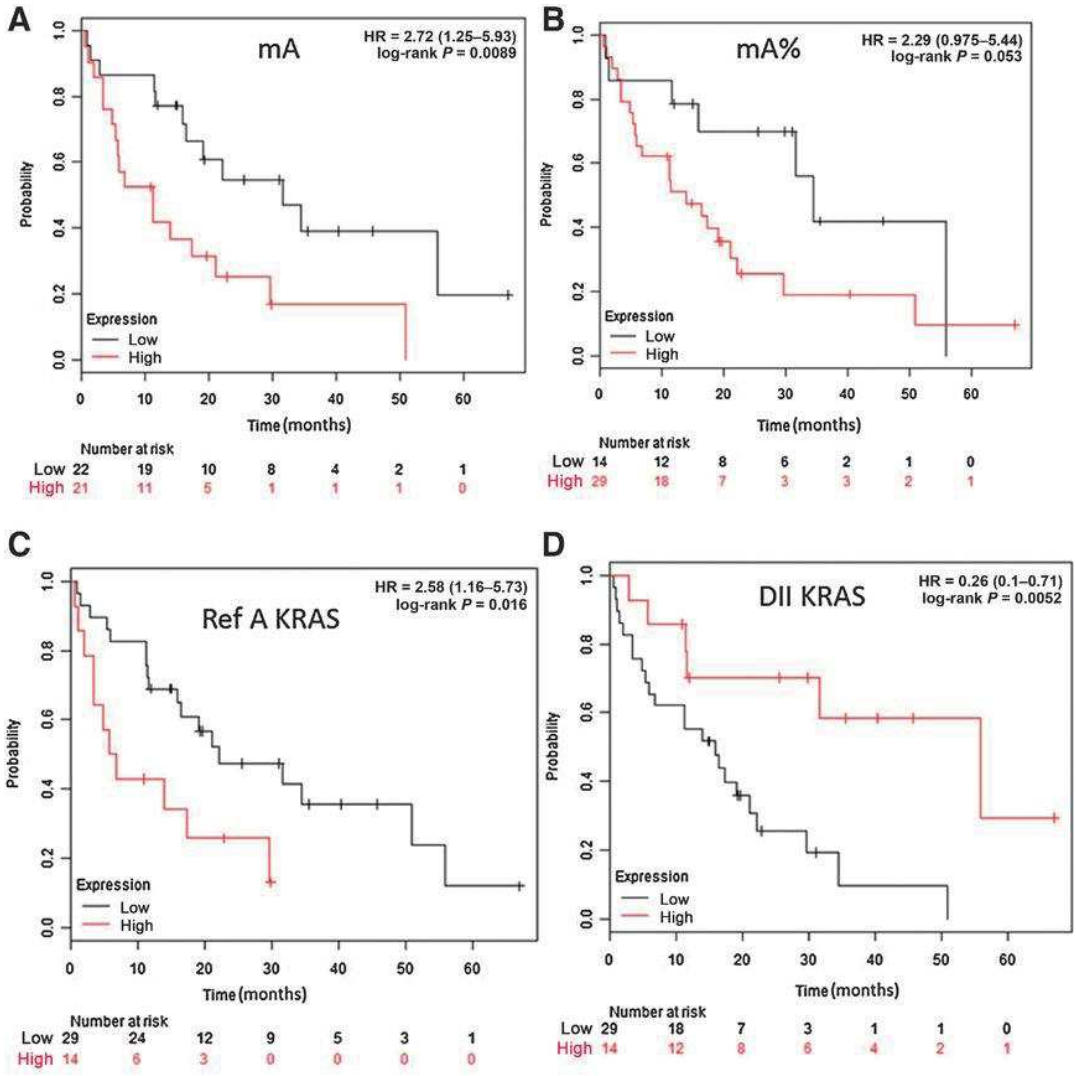


Figure 13: *L'effet de la concentration d'ADNcir tumoral (mA)* ***(A)****, la charge mutationnelle (mA%)* ***(B)****, la concentration totale d'ADNcir (Ref A KRAS)* ***(C)****, et de l'indice de fragmentation de l'ADNcir* ***(D)*** *sur la survie globale des patients mutés pour les gènes KRAS ou BRAF (157)*

De plus, la présence d'ADN tumoral circulant dans le plasma de patients atteints d'adénocarcinome du pancréas a été associée à une survie sans progression et une survie globale plus courtes *(158)*. Cet ADN a été identifié par une approche de séquençage de nouvelle génération (next-generation sequencing) et par PCR digitale (droplet digital PCR) en ciblant des mutations caractéristiques de ce type de cancer. La valeur pronostique de l'ADNcir a également été mise en évidence dans le cas du lymphome folliculaire *(159)*, du mélanome *(160)*, et d'autres types de cancer.

2. Surveillance de la réponse au traitement

La découverte de biomarqueurs spécifiques et l'identification de facteurs génétiques permettant de prédire la réponse au traitement font partie des objectifs dans le domaine de la recherche en oncologie. Un exemple est l'identification des mutations du gène *KRAS* qui reflètent une résistance aux thérapies ciblées utilisant les anticorps monoclonaux anti-EGFR (cetuximab, panitumumab) dans le cancer colorectal métastatique *(161, 162)*. Thierry *et al.* ont montré, dans ce contexte, que la détection de mutations du gène *KRAS* par l'analyse de l'ADNcir dans le plasma de patients atteints de cancer colorectal présentait une spécificité de 98% et une sensibilité de 92%, ainsi qu'une concordance de 96% avec l'analyse du tissu tumoral, et que ce test devrait pouvoir remplacer cette dernière *(12)*.

Des études de suivi des patients pendant le traitement ont également montré qu'un taux d'ADNcir plus faible est corrélé à une réponse positive au traitement dans divers types de cancers, alors qu'un taux plus élevé est corrélé à une réponse faible. Par ailleurs, le taux d'ADNcir tumoral diminue en fonction de la réponse au traitement chez des patients atteints de mélanome *(163)*, et augmente avec la progression de la maladie. Une diminution supérieure à 60% de la fréquence allélique du mutant du gène *TP53* dans le cas du cancer de l'ovaire de haut grade, après un cycle de chimiothérapie, a également été associée à un temps de progression plus court de la maladie *(153)*. La quantification de l'ADNcir tumoral a, par ailleurs, été utilisée pour surveiller, de manière fiable et plus sensible que les méthodes de détection traditionnelles ou alternatives, la dynamique tumorale chez des sujets atteints de cancer colorectal *(164, 165)*, ou de cancer du sein *(166)* opérés ou ayant subi une chimiothérapie. La surveillance quantitative de l'ADN tumoral circulant s'est aussi révélée prometteuse pour le suivi et l'évaluation de la réponse tumorale chez des patients atteints de cancer du poumon non à petites cellules et ayant reçus une immunothérapie anti-PD1 *(167)* ou anti-PDL1 *(168)* par exemple. Enfin, la quantification dans le plasma des mutations du gène *EGFR* à l'origine d'une résistance au traitement par les inhibiteurs de la tyrosine kinase dans ce type de cancer peut présenter un caractère prédictif précoce de la réponse clinique *(169)*.

Finalement, en plus de l'analyse des mutations, la méthylation s'est révélée être un marqueur de substitution pour le suivi des patients atteints de cancer colorectal métastatique *(170)*,

compte tenu du fait qu'elle présente l'avantage supplémentaire de ne pas nécessiter de connaissance préalable de mutations. Ainsi, Garlan *et al.* ont suggéré que les changements précoces de la concentration d'ADNcir tumoral calculée en se basant sur la détection de mutations ou d'hyperméthylation de certains gènes, sont un marqueur de l'efficacité du traitement chez des patients atteints de cancer colorectal métastatique *(171)*.

3. Evolution clonale et résistance

L'application de thérapies adaptées à l'émergence d'une résistance au traitement est un défi en oncologie. Cette résistance résulte de l'acquisition d'altérations moléculaires dans les gènes ou les voies impliquées dans le(s) mécanisme(s) d'action du médicament, et conduit à l'émergence de clones résistants *(172)*. L'analyse de l'ADNcir est un outil adapté à la détection des résistances précoces survenant au cours du traitement anticancéreux. L'analyse de cet ADN dans des échantillons de plasma obtenus avant et après le traitement peut en effet fournir un suivi global de l'évolution génétique et moléculaire de la tumeur. Les mutations responsables du phénomène de résistance peuvent ainsi être identifiées de manière prospective pour adapter le traitement en temps réel *(152)* **(Figure 14)**, en utilisant, des combinaisons de médicaments capables d'inhiber l'expansion des clones résistants, ou encore, en adoptant d'autres traitements avant que la résistance clinique ne se manifeste.

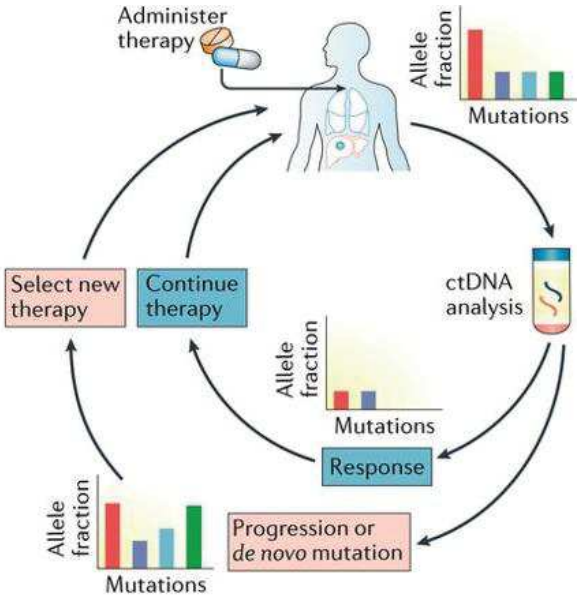


Figure 14: *Adaptation du traitement en temps réel selon les résultats de l'ADNcir (152)*

L'analyse de l'ADNcir dans le suivi de la dynamique clonale et la détection des mutations pendant le traitement, a été particulièrement utilisée dans le cas du cancer colorectal pour la détection de l'émergence de mutations du gène *KRAS*, médiatrices d'une résistance au traitement avec les anticorps monoclonaux anti-EGFR comme le cetuximab et le panitumumab. Diaz *et al*. ont ainsi été parmi les premiers à montrer que ces mutations peuvent être détectées par une approche non-invasive *(173)*. L'équipe de Bardelli a également mis en évidence, par analyse de l'ADN circulant, que 60% des patients métastatiques ayant développé une résistance au cetuximab ou au panitumumab ont acquis des mutations secondaires du gène *KRAS (174)*, et que les patients qui répondent plus longtemps aux anti-EGFR développent préférentiellement des mutations du domaine extracellulaire de l'EGFR *(175)*. Les mutations *RAS* apparaissent en revanche plus fréquemment chez les patients présentant une réduction plus petite du volume tumoral et une survie sans progression plus courte. Thierry *et al.* ont, quant à eux, mis en évidence le fait que des mutations secondaires des gènes RAS et BRAF détectées dans l'ADNcir apparaissent chez presque tous les patients, et qu'une fraction significative de ces mutations présente une fréquence allélique très faible soit avant, soit pendant le traitement *(13)*. Les résultats ont également révélé que pour des limites de détection de 1% ou 0.1%, 64% et 21% des mutations ne seraient pas détectées respectivement.

Enfin, il a été démontré que les marqueurs génétiques de résistance peuvent être suivis de façon non invasive tout au long du traitement dans d'autres cancers que le CCR comme les cancers du sein, de l'ovaire et du poumon *(176)*.

4. Maladie minimale résiduelle et récidive

L'analyse de l'ADNcir permet de détecter la maladie minimale résiduelle, y compris en l'absence de tout signe clinique, et de surveiller les récidives, ce qui pourrait améliorer la prise en charge des patients *(177)*. Tie *et al*. ont utilisé une technique de séquençage de haut débit et ont identifié des mutations somatiques marqueurs de l'ADNcir pour détecter la maladie minimale résiduelle et surveiller l'évolution de la tumeur pendant une période de suivi de 3 ans chez des patients ayant subi une résection du cancer du côlon de stade 2 *(178)*. Une

analyse longitudinale d'échantillons de 27 patients a ainsi révélé la présence d'ADNcir tumoral (muté) chez tous les patients ayant rechuté, mais pas chez les patients en rémission. Une autre étude a montré que, chez des patients atteints de cancer colorectal subissant une résection des métastases hépatiques, avec un suivi minimum d'un an, la sensibilité de détection de l'ADNcir tumoral postopératoire pour la détection de la maladie minimale résiduelle était de 58% avec une spécificité de 100% *(179)*.

Chen *et al.* ont testé l'utilité du séquençage de l'ADNcir tumoral comme biomarqueur de prédiction de la rechute chez des patientes atteintes d'un cancer du sein triple négatif présentant une maladie résiduelle après chimiothérapie néoadjuvante *(180)*. Ils ont montré que cette approche peut prédire une récidive avec une spécificité élevée (100%), mais une sensibilité faible (30%), et que la récidive était rapide lorsque l'ADNcir tumoral était détectable. Enfin, Abbosh *et al.* ont identifié un panel de mutations spécifiques comprenant 12 à 30 variantes mononucléotidiques (SNV) pour le profilage longitudinal de l'ADNcir de patients atteints de cancer du poumon non à petites cellules *(154)*. La détection des SNV par l'analyse de l'ADNcir semblait corréler avec les signes cliniques d'une rechute, avec un délai médian de 70 jours avant une détection clinique et radiologique.

Finalement, cette capacité de détection de la maladie minimale résiduelle et d'identification des patients à haut risque de rechute par recherche de l'ADN tumoral circulant a été validée dans d'autres types de cancer et notamment le mélanome *(181, 182)*.

C. Le dépistage et la détection précoce du cancer

Le cancer est l'une des principales causes de décès dans le monde. Un dépistage précoce, avant l'apparition des symptômes et lorsque le traitement est plus susceptible d'être efficace, pourrait augmenter les chances de survie. Plusieurs biomarqueurs sont utilisés en ce moment pour la détection de différents types de cancers, mais leur efficacité s'est avérée être réduite du fait de leur faible sensibilité ou spécificité. Pour cette raison, la recherche de nouveaux biomarqueurs plus efficaces est indispensable.

De nombreux efforts ont été déployés afin d'évaluer le potentiel de l'ADN circulant dans le dépistage précoce du cancer via une analyse des altérations qualitatives et quantitatives de ce biomarqueur *(114, 115)*. La caractérisation de l'ADNcir chez les individus avant un diagnostic de cancer et chez les individus pré-symptomatiques suggère qu'il pourrait être utilisé comme outil de détection et de dépistage précoce *(152)*.

Le principe du dépistage du cancer, l'importance clinique de la biopsie liquide dans ce domaine, ainsi que les biomarqueurs et tests de dépistage conventionnels utilisés sont développés dans ces deux revues :

- **Rita TANOS**, Alain R. THIERRY, **Clinical relevance of liquid biopsy for cancer screening**, *Translational Cancer Research* **7**, S105–S129 (2018).

- Alain R. THIERRY, **Rita TANOS**, **La biopsie liquide une voie possible pour le dépistage du cancer**, *médecine/sciences* **34**, 824–832 (2018).

Review Article

Clinical relevance of liquid biopsy for cancer screening

Rita Tanos[1,2,3,4], **Alain R. Thierry**[1,2,3,4]

[1]IRCM, Institut de Recherche en Cancérologie de Montpellier, Montpellier, France; [2]INSERM, U1194, Montpellier, France; [3]Université de Montpellier, Montpellier, France; [4]Institut régional du Cancer de Montpellier, Montpellier, France

Contributions: (I) Conception and design: AR Thierry; (II) Administrative support: AR Thierry; (III) Provision of study materials or patients: AR Thierry; (IV) Collection and assembly of data: R Tanos; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Alain R. Thierry. INSERM, U1194, Institut de Recherche en Cancérologie de Montpellier, Campus Val d'Aurelle, 208 avenue des Apothicaires, 34298 Montpellier, cedex 5, France. Email: alain.thierry@inserm.fr.

Abstract: Curative therapies for cancer are often successful when it is early detected and treated. Cancer screening aims at detecting cancer at early stages, before symptoms appear, and when the treatment is most likely to be effective. But most cancer types lack well-established biomarkers allowing the identification of the disease. Non-invasive tests, such as liquid biopsy, hold promise for screening people for cancer and could help advance cancer early detection. Transposing liquid biopsy into clinic to improve survival rates of patients with various types of cancers has been attracting much attention, due to the great potential of blood-based biomarkers for early diagnosis. The present review focuses on the clinical relevance of liquid biopsy for cancer screening, and summarizes the different studies conducted on several circulating biomarkers including circulating cell-free DNA (cfDNA), circulating tumor cells (CTCs), circulating microRNAs (miRNAs) and others, for the development of tests for early cancer detection.

Keywords: Cancer; screening; liquid biopsy; early detection; circulating DNA; circulating tumor cell (CTC)

Submitted Jan 03, 2018. Accepted for publication Jan 26, 2018.
doi: 10.21037/tcr.2018.01.31
View this article at: http://dx.doi.org/10.21037/tcr.2018.01.31

Cancer screening

The progression of cancer to late stages without the appearance of symptoms is one of the main reasons for being among the leading causes of death worldwide. The development of an effective screening test that identifies asymptomatic individuals to assess their likelihood having the disease has a major objective to reduce morbidity or mortality in the screened population by early detection, when treatment is more successful (1). Therefore, an early detection of cancer, before a person shows any signs of illness, would increase the chances of recovery and patients' overall survival and might help reduce cancer-related mortality.

A screening test differs from a diagnostic test by the fact that the second is used when a subject shows signs or symptoms, to determine the presence or absence of a disease, and is usually performed after a positive screening test to establish a definitive diagnosis.

The development of effective screening techniques for early detection does not exist for many types of cancers, and many have not proven effective in reducing cancer mortality. That was due to various reasons, in particular reduced sensitivity and specificity of the tests, inter and intra-tumoral heterogeneity (2-4) or epidemiological factors…

How to evaluate a screening test for cancer

In 1968, the World Health Organization (WHO) published guidelines on the principles and practice of screening for disease, which are often referred to as Wilson's criteria (5). But with the emergence of new genomic technologies, the WHO modified these guidelines in 2008 with the new understanding as follows (6): "*The screening program should respond to a recognized need, and its objectives should be defined*

at the outset. There should be a defined target population and scientific evidence of the screening program's effectiveness. The program should integrate education, testing, clinical services and program management, along with quality assurance and mechanisms to minimize potential risks of screening. It should ensure informed choice, confidentiality and respect for autonomy, and promote equity and access to screening for the entire target population. The evaluation should be planned from the outset, and the overall benefits of screening should outweigh the harm".

An ideal screening test for cancer would be able to perfectly discriminate between individuals who have or do not have the disease (1). In practice, screening tests could exhibit false positives and false negatives (7). The consequences of these results need to be carefully considered when evaluating the advantages and disadvantages of the test, and so taking into account its benefits on one hand and its risks on the other (8). When a new screening test is developed, it is regularly compared to the gold standard test, the best test available, which usually consists of a diagnostic test considered as definitive, like a biopsy for instance. However, the latter is often invasive, expensive, unpleasant, too late, or impractical to be used widely as a screening test (1). The new test would be less expensive or noninvasive for example. Its validity is translated by a high sensitivity, which is the ability of the test to identify correctly those who have that disease, and a high specificity reflecting its capacity to identify correctly those who do not have the disease (9).

In order to fully evaluate the performance of a screening test, particular attention must be paid to the tested cohorts. Indeed, a blind study, and an association of already diagnosed individuals and populations at risk are a necessity.

Conventional biomarkers and screening tests

Different screening tests are currently being used for various types of cancer. The pap smear test, for example, is used for cervical dysplasia or cervical cancer but it does not show a high sensitivity, however, it has a high specificity (10,11). Mammography is the most common test used for breast cancer screening (12,13), it has played a key role in reducing breast cancer mortality nevertheless it presents a limited sensitivity along with excessive false-positive results and the potential of overdiagnosis. Assays for serum markers, such as the tumor antigen CA 15-3 (cancer antigen 15-3) was also applied in patients with breast cancer, but showed an average sensitivity (55,6% sensitivity for a 98%specificity) (14). As a result, this marker is preferentially used for treatment response monitoring than screening and early diagnosis. The prostate-specific antigen (PSA) test could help find prostate cancer before symptoms appear (15), although a high PSA level does not always result from the presence of cancer. Low-dose computed tomographic (CT) screening is usually recommended for people with a high risk of developing lung cancer (16) but some cancers might be missed at screening and others might develop between screening and detection. For colorectal cancer (CRC), the two most common serum-based glycoprotein CRC markers, the cancer embryonic antigen (CEA) and the carbohydrate antigen 199 (CA199) are not appropriate for CRC screening due to their low sensitivity and the lack of specificity, especially for early-stage CRC [for CEA: sensitivity of 40.9–51.8% and specificity of 85.2–95% (17-19)] and are more appropriate to be used in monitoring the CRC recurrence or patients' response to surgical or systemic therapy. In addition, stool based screening tests for CRC were developed. The Hemoccult fecal occult blood test (FOBT) has been used for a long time to aid physicians in detecting hidden blood in stool specimen as an early indication of CRC, with a sensitivity varying between 12.9% and 79.4% and a specificity of 86.7–97.7% (20). However, since this test has many drawbacks in CRC screening, the fecal immunochemical test (FIT) is more commonly used in current CRC screening thanks to its low cost with an overall sensitivity of 0.79 and overall specificity of 0.94 (20). Another is the fecal DNA test (21). This multitarget assay detected invasive cancers and adenomas with high-grade dysplasia with 40.8% sensitivity and 94.4% specificity (22). Cologuard is the first commercial Food and Drug Administration (FDA) approved [2014] fecal DNA test presenting a higher sensitivity than the FIT in CRC and polyps but a lower specificity. However, this test is somewhat expensive (23).

What is liquid biopsy?

Upon the National Cancer Institute (NCI) Dictionary of Cancer Terms liquid biopsy is "*A test done on a sample of blood to look for cancer cells from a tumor that are circulating in the blood or for pieces of DNA from tumor cells that are in the blood*" (24). To our point of view this definition is rather imperfect and one can be uncomfortable to associate a so complex entity such as cells regrouping molecules, intermolecular associations, organelles, compartmentalization, a closed concentration of factors or enzymes, and programming, which are the smallest unit that

can live on its own and that makes up all living organisms and the tissues of the body, with cellular components such as macromolecules (DNA, RNA or microRNA) either strongly associated with proteins or encapsulated in micro-vesicles (25). The improbable association of such different biological entities may only rely on their circulating property and on the potential clinical use of the information provided by both biological sources. As indicated in the NCI Dictionary of Cancer Terms: "*A liquid biopsy may be used to help find cancer at an early stage. It may also be used to help plan treatment or to find out how well treatment is working or if cancer has come back. Being able to take multiple samples of blood over time may also help doctors understand what kinds of molecular changes are taking place in a tumor.*" Despite incoherence in the term, we will use in this review the term liquid biopsy in accordance with the NCI Dictionary terminology and conventionally in the literature. The term liquid biopsy therefore applies mainly in oncology mirroring the biopsy of the tumor tissue. In addition, this terminology cannot be used for circulating DNAs that are analysed in the field of prenatal diagnosis, severe/acute inflammation (sepsis), transplantation, or sports, for instance.

Liquid biopsies are not limited to the blood, though this is greatly where the research is focused. Urine, saliva or cervical fluid may also be used, as genetic information is also present in these fluids.

Circulating cell-free DNA (cfDNA)

cfDNA has emerged as a potential biomarker especially in cancer and is being widely investigated in translational and clinical research (25-27). It may present the opportunity to diagnose, monitor recurrence, and evaluate response to therapy solely through a non-invasive blood draw. Several efforts are being made in order to assess the potential of this biomarker for early cancer screening and qualitative as well as quantitative cfDNA alterations have been examined (26). But despite intensive research, few cfDNA-based tests have been translated to clinical practice. For instance, conflicting data regarding total nuclear cfDNA concentration made it hard for cfDNA-based tests to be developed and used in clinic: plasma cfDNA concentrations in cancer patients range from a few ng/mL to several thousand ng/mL, which overlaps with the concentration range for healthy individuals (27-29).

A cfDNA-based screening test must be able to distinguish between signals from non-cancer and pre-cancerous processes and the invasive malignancy in order to achieve high clinical sensitivity (30). It would also help if the test could provide information on the tissue origin which might be possible through circulating tumor DNA (ctDNA), given the distinct differences in the patterns of somatic alterations between different tumor types.

Until now, several groups worked on developing tests for the early screening of different types of cancer from a single blood analysis. We listed in *Table 1* the most useful and efficient screening tests. For CRC for example, many groups have studied the screening or diagnostic relevance of different cfDNA parameters, and several reports were focused on the detection of methylated Septin9 in the plasma which was found to be significantly higher in patients with CRC than in patients with no evidence of disease (17), making it a potential biomarker for this type of cancer (33,34). The Epi proColon® (Epigenomics AG Corporation, Berlin, Germany), based on a real-time polymerase chain reaction (PCR) detection of methylated Septin9 from blood, is the only commercially available blood based DNA hypermethylation screening test for CRC (32) and is so far the best among the commercial blood-based cancer detection assays. The test discriminated between patients with CRC and healthy controls with a sensitivity of 75–81% and a specificity of 96–99%. Other studies were intended to evaluate, in the plasma, different hypermethylated DNA promoter regions and genes (37) previously found to be CRC specific. The highest area under the curves (AUCs) achieved were 0.85 for a test combining seven promotors along with age and gender (35). In addition, an age-adjusted panel of four cell-free nucleosomes was developed by Volition and it provided an AUC of 0.97 (0.87 if not age-adjusted) for the discrimination between CRC patients and healthy controls (*Table 1*). It showed high sensitivity for early stages (75 and 86 at 90% specificity for stages I and II, respectively). A second combination of four cf-nucleosome biomarkers provided an AUC of 0.72 for the discrimination of polyps from the healthy group (31).

The diagnostic potential of cfDNA was also examined in other types of cancer (43), and cfDNA levels were studied (38). For instance, quantitative analysis in lung cancer (*Table 1*) showed increased levels of cfDNA in cancer patients than in healthy individuals with approximately a value of 0.88 for areas under the summary receiver operating characteristic curves (44-47). Similar results were observed for breast cancer with 78% sensitivity and 83% specificity (53), and ovarian cancer (54) with a sensitivity of 70% and a specificity of 90%.

Circulating DNA consists not only of nuclear but

Table 1 cfDNA and cancer screening

Article	Journal/date	Summary/results
Circulating nucleosomes as new blood-based biomarkers for detection of colorectal cancer (31)	*Clinical Epigenetics*, May 2017	❖ The levels of 12 epigenetic cell-free nucleosome epitopes were measured in the sera of 58 individuals referred for endoscopic screening for CRC ❖ For the discrimination of CRC patients from healthy individuals, they developed an age-adjusted panel of four cell-free nucleosomes. The AUC was 0.97 (0.87 if not age-adjusted) with a high sensitivity at early stages (sensitivity of 75 and 86 at 90% specificity for stages I and II, respectively). A second combination of four cf-nucleosome biomarkers provided an AUC of 0.72 for the discrimination of polyps from the healthy group
Epi proColon® 2.0 CE: A Blood-Based Screening Test for Colorectal Cancer (32)	*Molecular Diagnosis and Therapy*, April 2017	❖ Epi proColon® 2.0 CE consists of the Epi proColon® Plasma Quick Kit (M5-02-001), the Epi proColon® Sensitive PCR Kit (M5-02-002) and the Epi proColon® Control Kit (M5-02-003) → sufficient quantities of reagents for processing up to 32 samples including controls, divided between a maximum of four independent test runs ❖ It consists of a real-time PCR detection of methylated Septin9 from blood derived DNA in bisulfite converted DNA (bisDNA) from 3.5 mL human plasma samples ❖ It is the only commercially available blood based DNA hypermethylation screening test for CRC ❖ The test results do not confirm the presence nor absence of colorectal disease and it must be evaluated together with other clinical parameters, and positive test results should be verified by colonoscopy or sigmoidoscopy ❖ Epi proColon® 2.0 CE has been determined to have an estimated 95% limit of detection (LoD) of 14 pg/mL (95% CI, 9–19 pg/mL), assessed using the Applied Biosystems 7500 Fast Dx with SDS v1.4 ❖ Across studies, Epi proColon® 2.0 CE discriminated between patients with colorectal cancer and healthy controls with a sensitivity of 75–81%; its specificity for colorectal cancer versus healthy individuals was 96–99%
Performance of a second-generation methylated SEPT9 test in detecting colorectal neoplasm (33)	*Journal of Gastroenterology and Hepatology*, April 2015	❖ Peripheral blood samples of 135 patients with CRC, 169 with adenomatous polyps, 81 with hyperplastic polyps, and 91 healthy controls were taken for SEPT9 testing using Epi proColon 2.0 test ❖ The sensitivity and specificity of SEPT9 for CRC were 74.8% and 87.4%, respectively, but the sensitivity for advanced adenomas was 27.4% ❖ SEPT9 was positive in 66.7% of stage I, 82.6% of stage II, 84.1% of stage III, and 100% of stage IV CRCs ❖ For 177 patients, both SEPT9 and FIT (fecal immunochemical test) were performed and the sensitivity and specificity of FIT for CRC was 58.0% and 82.4%, respectively ❖ SEPT9 showed better performance in CRC detection than FIT, but similar results were found for advanced adenomas
Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer (34)	*Gut*, February 2014	❖ This study assessed the accuracy of circulating methylated SEPT9 DNA (mSEPT9) for detecting CRC in a screening population ❖ Blood plasma samples of asymptomatic individuals ≥50 years old scheduled for screening colonoscopy [7,941 men (45%) and women (55%), mean age =60 years] were tested using the first generation of the commercially available Epi proColon Assay ❖ Results from 53 CRC cases and from 1,457 subjects without CRC showed a standardised sensitivity of 48.2%; for CRC stages I–IV, values were 35.0%, 63.0%, 46.0% and 77.4%, respectively for a 91.5% specificity ❖ The sensitivity for advanced adenomas was low (11.2%)

Table 1 *(continued)*

Table 1 *(continued)*

Article	Journal/date	Summary/results
Hypermethylated DNA, a circulating biomarker for colorectal cancer detection (35)	*PloS One*, July 2017	❖ The article presents a cross-sectional case-control study of 193 CRC patients and 102 colonoscopy-verified healthy controls ❖ Thirty DNA promoter regions previously found to be CRC specific were evaluated using methylation specific polymerase chain reaction ❖ Individual DNA promoter regions could not provide an overall sensitivity above 30% at a reasonable specificity showing that individual hypermethylated DNA promoter regions have limited value as CRC screening markers ❖ However, the combination of seven hypermethylated promoter regions (ALX4, BMP3, NPTX2, RARB, SDC2, SEPT9, and VIM) along with the covariates sex and age showed an optimism corrected AUC of 0.86 for all stage CRC and 0.85 for early stage CRC. The overall sensitivity was 90.7% at 72.5% specificity using a cut point value of 0.5
Diagnostic and prognostic role of cell-free DNA testing for colorectal cancer patients (36)	*International Journal of Cancer*, January 2017	❖ Using an *ALU*-based Q-PCR method, the presence and integrity of cfDNA was assessed in a large cohort of CRC patients (n=114) in comparison to healthy subjects (n=56) and patients with adenomatous lesions (n=22) ❖ cfDNA concentration and integrity index were increased in CRC patients, and cfDNA was significantly higher in advanced histopathological stage ❖ The discriminative capacity between CRC patients on one hand and controls or adenoma patients on the other hand was moderate for the *ALU83* and *ALU244* fragment dosage. For *ALU83* ROC curves' AUCs were 0.7105 and 0.77083, respectively, and *ALU244* showed AUCs of 0.7205 and 0.7636, respectively ❖ The methylation profile of the promoters of *OSMR* and *SFRP1* genes was also evaluated in the cohort and was compared for 25 CRC patients in matched tissue and plasma. Only three mismatched cases were observed ❖ A lower methylation quantification was observed in cfDNA than in tissue DNA, but cfDNA methylation frequency was statistically different in controls, adenoma and CRC patients and this frequency increased with the histopathological stage of tumor ❖ The adenoma and CRC patients' methylated cfDNA showed a higher quantity of *ALU83* and *ALU244* ❖ An approach combining the detection of *ALU* fragments and cancer type-specific epigenetic alteration, might improve the diagnostic efficiency for CRC
Aberrant Methylation of APC, MGMT, RASSF2A, and Wif-1 Genes in Plasma as a Biomarker for Early Detection of Colorectal Cancer (37)	*Clinical Cancer Research*, October 2009	❖ The study consists of a retrospective analysis of the methylation status of 10 genes in fresh-frozen tissues and corresponding plasma samples from 243 patients with stage I and II sporadic colorectal cancer, 276 healthy individuals, and plasma from 64 colorectal adenoma patients using methylation-specific PCR ❖ In order to find molecular markers with high sensitivity and specificity, the methylation score (M score) was used ❖ Of the 243 colorectal cancer tissues, methylation was detected in 18% for *p14*, 34% for *p16*, 27% for *APC*, 34% for *DAPK*, 32% for *HLTF*, 21% for *hMLH1*, 39% for *MGMT*, 24% for *RARβ2*, 58% for *RASSF2A*, and 74% for *Wif-1* ❖ The plasma analysis for cancerous patients and healthy individuals showed that the M score of any single gene had a sensitivity less than 40% after controlling for age, sex, and tumor location, but in a model including *APC*, *MGMT*, *RASSF2A*, and *Wif-1* genes, the M score had 86.5% sensitivity and 92.1% specificity when 1.6 was used as a cutoff (AUC =0.927) with a positive predictive value of 90.6% and a negative predictive value of 88.8% ❖ In the plasma of colorectal adenoma patients, the overall M score of the model (*APC*, *MGMT*, *RASSF2A*, and *Wif-1 genes*) was found to show a sensitivity of 74.6% and a specificity of 91.3% (AUC =0.864), when 1.8 was used as the cutoff value, after adjusting for age and sex. The positive predictive value was 71.6% and the negative predictive value was 93.9%

Table 1 *(continued)*

Table 1 *(continued)*

Article	Journal/date	Summary/results
Liquid Biopsies for Cancer: Coming to a Patient near You (38)	*Journal of Clinical Medicine*, January 2017	❖ Circulating DNA size profiling might distinguish early from late malignancies, that's why it's being examined for inclusion in a screening blood test for cancer ❖ The detection of tumor-specific DNA methylation through a liquid biopsy is another feasible approach for the development of diagnostic tests for early-stage cancer: • Differential methylation levels of three promoters, RASSF1A, CALCA, and EP300, in the cell-free plasma could detect ovarian cancer from healthy controls with a sensitivity of 90% and a specificity of 86.7% in a 30-patient cohort study (39) • The methylation of the promoter region of the thrombomodulin gene (*THBD*) could differentiate colorectal cancer and control blood samples with a sensitivity of 71% and a specificity of 80% (40) • Several clinical studies have demonstrated the utility of ctDNA-based biomarkers relative to protein biomarkers • The quantification of ctDNA mutants and the detection of their presence/absence in colon cancer patients after surgery and chemotherapy proved to be more clinically useful than the cancer embryonic antigen (CEA) test (41) • In metastatic breast cancer patients, a study demonstrated an improved sensitivity for cancer detection of ctDNA over CA 15-3: of 85% *vs.* 59% (42)
Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies (43)	*Science Translational Medicine*, February 2014	❖ They evaluated the ability of ctDNA to detect tumors in 640 patients with various cancer types using digital polymerase chain reaction-based technologies ❖ In more than 75% of patients with advanced pancreatic, ovarian, colorectal, bladder, gastroesophageal, breast, melanoma, hepatocellular, and head and neck cancers, ctDNA was detected but in less than 50% of primary brain, renal, prostate, or thyroid cancers ❖ For patients with localized tumors, ctDNA was detected in 73%, 57%, 48%, and 50% of patients with colorectal cancer, gastroesophageal cancer, pancreatic cancer, and breast adenocarcinoma, respectively ❖ ctDNA was often present in patients without detectable circulating tumor cells, suggesting that these two biomarkers are distinct entities ❖ In a separate panel of 206 patients with metastatic colorectal cancers, the sensitivity of ctDNA for the detection of clinically relevant KRAS gene mutations was 87.2% and its specificity was 99.2%
Circulating DNA: diagnostic tool and predictive marker for overall survival of NSCLC patients (44)	*PloS One*, 2012	❖ The study aimed to assess the discriminative capacity and the prognostic value of the amounts of circulating DNA (cDNA) between NSCLC patients and healthy individuals ❖ Plasma of 309 individuals (104 cancer patients and 205 healthy controls) were analysed and the cDNA levels were assessed through a real-time PCR method targeting the hTERT single copy gene ❖ Increased cDNA levels in NSCLC patients compared to control individuals were observed and the area under the ROC curve was 0.88 (95% CI, 0.84–0.92; P<0.0001) ❖ Lower cut-off values increased the sensitivity of the assay but at the cost of specificity and vice versa: with a threshold of 20 ng/mL, there is a probability of illness of 71% when the test is positive (PPV). A DNA cut-off level of >20 ng/mL differentiated between lung cancer patients and controls with a specificity of 83% and sensitivity of 79% ❖ A decreased overall survival time was observed in patients presenting high cDNA levels, when compared to lower cDNA concentrations

Table 1 *(continued)*

Table 1 *(continued)*

Article	Journal/date	Summary/results
Value of quantitative analysis of circulating cell free DNA as a screening tool for lung cancer: a meta-analysis (45)	*Lung Cancer*, August 2010	❖ It consists of a meta-analysis of 10 studies were including 752 lung cancer patients and 635 healthy controls ❖ Sensitivity, specificity, and other measures of accuracy of circulating DNA assay in the diagnosis of lung cancer were pooled using random-effects models and summary ROC curves were used to summarize overall test performance ❖ For quantitative analysis of circulating cell-free DNA in lung cancer screening, the summary estimates were a sensitivity of 0.80 (95% CI, 0.77–0.83); a specificity of 0.77 (95% CI, 0.74–0.80); a positive likelihood ratio of 4.54 (95% CI, 2.66–7.76); a negative likelihood ratio of 0.28 (95% CI, 0.19–0.40); and a diagnostic odds ratio of 20.33 (95% CI, 10.12–40.86) ❖ The AUC was 0.89 (weighted AUC, 0.88), indicating a high level of overall accuracy
The diagnostic value of circulating cell free DNA quantification in non-small cell lung cancer: A systematic review with meta-analysis (46)	*Lung Cancer*, October 2016	❖ The diagnostic value of cfDNA quantification for non-small cell lung cancer (NSCLC) was estimated in 15 studies with a total of 1,193 patients with lung cancer and 1,059 controls ❖ Pooled results showed 81% sensitivity (95% CI, 76–84%); 85% specificity (95% CI, 77–91%); 23.87 diagnostic odds ratio (95% CI, 13.37–42.61); and 0.89 for areas under the summary receiver operating characteristic curves (95% CI, 0.86–0.92)
The Emerging Role of "Liquid Biopsies," Circulating Tumor Cells, and Circulating Cell-Free Tumor DNA in Lung Cancer Diagnosis and Identification of Resistance Mutations (47)	*Current Oncology Reports*, January 2017	❖ Endorsed screening strategies including low-dose CT scans have a low sensitivity and high false positive rates of >90% as well as low adoption as a practiced standard of care ❖ The significant difference in DNA concentration in the serum/plasma of lung cancer patients with healthy controls or patients with benign diseases opened up the possibility of the use of this biomarker in screening assays (48-51)
Identification of Circulating Tumor DNA for the Early Detection of Small-cell Lung Cancer (52)	*EBioMedecine*, August 2016	❖ Plasma of 51 small cell lung cancer (SCLC) and 123 controls were assessed for the presence of TP53 mutations ❖ Thirty-one TP53 mutations were detected in the cfDNA of 49% SCLC patients (35.7% early-stage and 54.1% late-stage) and 18 mutations in 11.4% of non-cancer controls ❖ The results were replicated in an independent series of 102 non-cancer controls and showed a comparable proportion of TP53 mutated samples (10.8%) which suggests that somatic mutations occur in cfDNA among individuals without cancer diagnosis and causes a serious challenge for ctDNA screening tests development ❖ Allelic fractions of the TP53 mutations were significantly higher in cases than in controls (P=0.0004)
Value of circulating cell-free DNA analysis as a diagnostic tool for breast cancer: a meta-analysis (53)	*Oncotarget*, February 2017	❖ The objective of this study was to systematically evaluate the diagnostic value of cfDNA for breast cancer ❖ Twenty-five studies with relevant diagnostic screening were included with 15 quantitative analysis studies and 10 qualitative analysis studies and a total number of 1,705 histologically diagnosed breast cancer patients, 1,079 healthy controls, and 234 patients with benign breast diseases ❖ The mean sensitivity, specificity and AUC of the Summary ROC plots for 24 studies that distinguished breast cancer patients from healthy controls were 0.70, 0.87, and 0.9314, yielding a DOR of 32.31 ❖ For the 14 quantitative analyses of cfDNA for breast cancer diagnosis, the estimates of sensitivity and specificity were 0.78 and 0.83, respectively. The value for PLR was 4.83, and NLR was 0.22. The DOR value was 24.40 and AUC 0.9116 ❖ The 10 qualitative studies produced 0.50, 0.98, 0.9919, and 68.45, sensitivity, specificity, AUC and DOR respectively ❖ The specificity, sensitivity, AUC and DOR for eight studies that distinguished malignant breast cancer from benign diseases were 0.75, 0.79, 0.8213, and 9.49

Table 1 *(continued)*

Table 1 *(continued)*

Article	Journal/date	Summary/results
Circulating Cell Free DNA as the Diagnostic Marker for Ovarian Cancer: A Systematic Review and Meta-Analysis (54)	*PloS One*, June 2016	❖ A meta-analysis of nine diagnostic studies published from 2001 to 2014, including 462 ovarian cancer patients and 407 controls was conducted ❖ The summary estimates for quantitative analysis of circulating cfDNA in ovarian cancer screening showed a sensitivity of 0.70, a specificity of 0.90, a positive likelihood ratio of 6.60, a negative likelihood ratio of 0.34, a diagnostic odds ratio of 26.05, and an AUC of 0.89
Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers (55)	*Proceedings of the National Academy of Sciences*, September 2017	❖ The objective of this study was to combine blood tests for KRAS gene mutations with carefully thresholded protein biomarkers to determine whether the combination of these markers was superior to any single marker ❖ Two hundred and twenty-one patients with resectable pancreatic ductal adenocarcinomas and 182 control patients without known cancer were tested ❖ KRAS mutations were detected in the plasma of 30% of the patients, and with a 100% concordance with the mutations found in the patient's primary tumor ❖ The combination of KRAS with four thresholded protein biomarkers (CA19-9, CEA, HGF, and OPN) increased the sensitivity to 64%. Only one of the 182 plasma samples from the control cohort was positive for any of the DNA or protein biomarkers (99.5% specificity)
Direct detection of early-stage cancers using circulating tumor DNA (56)	*Science Translational Medicine*, August 2017	❖ They developed an approach called targeted error correction sequencing (TEC-Seq) that allows ultrasensitive direct evaluation of sequence changes in circulating cell-free DNA using massively parallel sequencing, and used it to examine 58 cancer-related genes encompassing 81 kb ❖ The analysis of plasma from 44 healthy individuals identified genomic changes related to clonal hematopoiesis in 16% of asymptomatic individuals but no alterations in driver genes related to solid cancers were found ❖ Plasma samples from 194 patients with breast cancer (n=45), colorectal cancer (n=42), lung cancer (n=65), and ovarian cancer (n=42) were analyzed and the concentration of cfDNA in plasma from cancer patients (12 ng/mL) was significantly higher than that observed in healthy individuals (average of 7 ng/mL; P=0.001) ❖ The evaluation of 200 patients with colorectal, breast, lung, or ovarian cancer detected somatic mutations in the plasma of 71%, 59%, 59%, and 68%, respectively, of patients with stage I or II disease ❖ Of the 194 patients analyzed, more than 3/4 of colorectal cancer patients, 2/3 of ovarian cancer patients, and most of the lung and breast cancer patients had detectable alterations in driver genes ❖ The analysis of mutations in the circulation revealed high concordance with alterations in the tumors of these patients

Table 1 *(continued)*

Table 1 *(continued)*

Article	Journal/date	Summary/results
Next-Generation Sequencing of Circulating Tumor DNA for Early Cancer Detection (30)	*Cell*, February 2017	❖ For successful cancer screening, a platform that provides direct, sensitive, and specific measures of cancer and its attributes is needed ❖ The fraction of tumors that shed detectable levels of ctDNA, by tumor type and stage, is not well studied. The studies to date have small numbers of samples and use a variety of measurement techniques that are often not comparable ❖ ctDNA detection has the potential to be more specific to the presence of the tumor than other measurements of proteins and metabolites. However, the implementation of such a test would be technically challenging, since many genes would have to be simultaneously queried for alterations in order to cover enough of the known diversity in cancer genomes ❖ While next-generation DNA sequencing technology does enable high degrees of target multiplexing, the depth of sequencing would also have to be very high to sample enough ctDNA molecules to reliably measure them in a background of mostly non-tumor-derived cfDNA ❖ To achieve high clinical specificity, a ctDNA-based screening test must be capable of distinguishing between the background signal originating from such non-cancer or pre-cancerous processes and the invasive malignancy of real interest ❖ An ideal non-invasive screening test would also provide information on the tissue of origin to streamline the downstream workup, including imaging and tissue diagnosis. This information may be possible through ctDNA, given the distinct differences in the patterns of somatic alterations between different tumor types, at least at a population level ❖ GRAIL is a recently created company, formed to develop ctDNA-based cancer screening tests. It is conducting a 10,000-plus subject study, called the Circulating Cell-Free Genome Atlas (CCGA) to create a reference library of the cancer mutations in the blood for the most common cancers and the background mutations found in matched healthy subjects. It will be the largest database on mutations found in the blood of cancer patients. It intends to apply a next-generation sequencing approach, combining sequencing depth and breadth of genomic coverage, as well as machine learning, to develop models based on cell-free DNA for the accurate classification of subjects with and without cancer ❖ The Cancer Moonshot Initiative recently announced the Blood Profiling Atlas Project, which also aims to compile data on cancer signals in the blood

CRC, colorectal cancer; AUC, area under the curve; PCR, polymerase chain reaction; Q-PCR, quantitative polymerase chain reaction; ROC, receiver operating characteristic; DOR, diagnostic odds ratio; PLR, positive likelihood ratio; NLR, negative likelihood ratio.

mitochondrial DNA (mtDNA). Other studies have been published on the clinical significance of mtDNA levels and integrity in the peripheral blood in different types of cancer (*Table 2*) such as lung (57,58), breast (59), colorectal (60,61), non-Hodgkin lymphoma (62), and others (63-70). At this time, published data are discordant and it is impossible to draw any conclusion. The lack of pre-analytical and analytical studies on circulating cell-free mtDNA could explain in part this discordance, since it is poorly characterized and little is known about its structural properties.

Circulating tumor cells (CTCs)

The discovery of cells released in the bloodstream or escaping from the tumor is of primary importance and has led to intense research for about 20 years. CTCs are incredibly hard to isolate and do not always indicate genetically cancerous cells. The value of CTCs in diagnosing different types of cancers has been also assessed in several studies (*Table 3*) (71,72). In lung cancer for example (78), Tanaka *et al.* showed that CTC enumerations

Table 2 Mitochondrial DNA and cancer screening

Article	Journal/date	Summary/results
Mitochondrial DNA copy number and lung cancer risk in a prospective cohort study (57)	*Carcinogenesis*, May 2010	❖ The association of mtDNA copy number and lung cancer risk was assessed in 227 prospectively collected cases and 227 matched controls ❖ There was evidence that the risk of lung cancer increased in a dose-dependent manner with mtDNA copy number (ptrend 5 0.008) ❖ The association between mtDNA copy number and lung cancer risk was evident among heavy smokers (≥20 cigarettes per day), but not light smokers (<20 cigarettes per day), however, the interaction between mtDNA copy number and smoking was not significant
Pre-diagnostic leukocyte mitochondrial DNA copy number and risk of lung cancer (58)	*Oncotarget*, March 2016	❖ This study consists of a prospective investigation, using a Q-PCR based assay, of the relationship between mitochondrial DNA copy number (mtCN) and the risk of lung cancer in 463 case-control pairs from the the Nurses' Health Study (NHS) (285 cases and 285 controls) and the Health Professionals Follow-Up Study (HPFS) (178 cases and 178 controls) ❖ Current heavy smokers (>24 cigarettes/day) had significantly lower mtCN compared with never smokers (P=0.05) ❖ No overall association was observed between mtCN and lung cancer risk ❖ Compared to the high log_mtCN group, the risk of lung cancer was 1.29 (95% CI, 0.89–1.87) for the median group, and 1.11 (95% CI, 0.75–1.64) for the low group ❖ Among current smokers, compared to participants with high levels of log_mtCN, those with median levels had a significantly higher risk of lung cancer (OR =2.09; 95% CI, 1.12–3.90), but not those with low levels (OR =1.37; 95% CI, 0.75–2.48) ❖ The interaction between mtCN and smoking status on lung cancer risk was not significant
Mitochondrial DNA Copy Number Is Associated with Breast Cancer Risk (59)	*PloS One*, June 2013	❖ The association between mtDNA copy number in peripheral blood and breast cancer risk was studied in 183 breast cancer cases with pre-diagnostic blood samples and 529 individually matched controls ❖ The relative quantification of mtDNA copy number to nuclear DNA, was positively associated with breast cancer risk overall (P for trend =0.01) ❖ Relative mtDNA copy number was associated with breast cancer risk only among those women from whom a blood sample was collected within 3 years of breast cancer diagnosis ❖ No association was observed between mtDNA copy number and breast cancer risk among women who donated a blood sample ≥3 years prior to breast cancer diagnosis (P for trend =0.41) ❖ MtDNA copy number was negatively correlated with time to breast cancer diagnosis (r=20.15; P=0.048)
Mitochondrial Copy Number Is Associated with Colorectal Cancer Risk (60)	*Cancer Epidemiology, Biomarkers & Prevention*, September 2012	❖ The association between mtDNA copy number in peripheral blood and colorectal cancer risk was studied in 422 colorectal cancer cases (168 cases with pre-diagnostic blood and 254 cases with post-diagnostic blood) and 874 controls who were free of colorectal cancer ❖ After measuring the relative mtDNA to nuclear DNA copy number using real-time PCR, a U-shaped relationship between the relative mtDNA copy number and colorectal cancer risk was observed ❖ Compared with the 2nd quartile, the OR (95% CI) for subjects in the lowest and highest quartiles of relative mtDNA copy numbers were 1.81 (1.13–2.89) and 3.40 (2.15–5.36), respectively (P curvilinearity <0.0001) ❖ This U-shaped relationship was present in both men and women, similar for colon cancer and rectal cancer, and independent of the timing of blood draw with regard to cancer diagnosis

Table 2 *(continued)*

Table 2 *(continued)*

Article	Journal/date	Summary/results
Association between mitochondrial DNA content in leukocytes and colorectal cancer risk (61)	*Cancer*, July 2011	❖ mtDNA content was measured in peripheral blood lymphocytes of 320 CRC patients and 320 controls by Q-PCR ❖ mtDNA content was significantly higher in cancer patients than in controls, and high mtDNA content was associated with a significantly increased CRC risk
A prospective study of mitochondrial DNA copy number and risk of non-Hodgkin lymphoma (62)	*Blood*, November 2008	❖ mtDNA copy number was analysed in peripheral white blood cells of 104 males with non-Hodgkin lymphoma (NHL) and 104 control ❖ The results showed that a dose-response relationship exists between mtDNA copy number and NHL risk, with the most pronounced effect for chronic lymphocytic leukemia (CLL)/small lymphocytic lymphoma subtype (SLL)
Diagnostic and prognostic potential of circulating cell-free genomic and mitochondrial DNA fragments in clear cell renal cell carcinoma patients (63)	*Clinical Chimica Acta*, January 2016	❖ CfDNA was extracted from EDTA plasma of healthy people (n=40), non-metastatic (n=145) and metastatic (n=84) clear cell renal cell cancer (RCC) patients using the QIAamp Circulating Nucleic Acid Kit ❖ Genomic and mitochondrial cfDNA concentrations were determined using qPCR of different cfDNA fragments (67–306 bp) (target: *APP* for nuclear DNA) ❖ Genomic cfDNA fragments of APP with 67 bp (*APP-1*) and 180 bp (*APP-2*) as well as of Alu sequences with 79 bp (*SINE-1*) and 248 bp (*SINE-2*) were not different between the controls and non-metastatic RCC patients, but metastatic RCC patients showed lower concentrations of the long 306 bp *APP-3* fragment compared to the controls ❖ Significantly higher concentrations of the short *APP-1* in comparison to *APP*-2 and *APP-3* were found in the RCC groups ❖ Increased mitochondrial cfDNA concentrations in metastatic RCC in comparison to controls and non-metastatic RCC were observed with a decreased integrity index ❖ The cfDNA integrity indices decreased from controls to metastatic patients ❖ An AUC >0.75 was observed for predicting recurrence-free survival and overall survival with concordance indices >0.80
Circulating Mitochondrial DNA Level, a Noninvasive Biomarker for the Early Detection of Gastric Cancer (64)	*Cancer Epidemiology, Biomarkers & Prevention*, November 2014	❖ MtDNA in peripheral leukocytes of 28 patients with non-atrophic gastritis (NAG), 74 patients with gastric cancer, and 48 matched asymptomatic controls was measured by quantitative real-time PCR assay. In parallel, the serologic level of IL8 was determined ❖ Mean mtDNA level was higher in patients with gastric cancer (P=0.0095) than in controls, with values >8.46 significantly associated with gastric cancer (OR =3.93) ❖ Three ranges of mtDNA values were identified: interval I <2.0, interval II 2.0–20, interval III >20 ❖ Interval I included mainly NAG cases, and few gastric cancer samples and interval III corresponded almost exclusively to patients with gastric cancer. All controls fell in interval II, together with some NAG and gastric cancer cases ❖ IL8 levels were significantly higher in patients with gastric cancer (P<0.05), with levels > 50 pg/mL observed exclusively in patients with gastric cancer, allowing to distinguish them within interval II ❖ mMtDNA results validated in a second cohort of patients: mtDNA was significantly higher in gastric cancer than in patients with preneoplasia

Table 2 *(continued)*

Table 2 *(continued)*

Article	Journal/date	Summary/results
Cell-free Circulating Mitochondrial DNA in the Serum: A Potential Non-invasive Biomarker for Ewing's Sarcoma (65)	*Archives of Medical Research*, July 2012	❖ ccf-mtDNA copy number in serum samples obtained from 25 patients with Ewing's sarcoma (EWS) as well as 20 age-matched individuals were detected by quantitative real-time PCR assays using mtDNA 16s RNA-specific primers to amplify a 79-bp fragment ❖ The quantification of ccf nuclear DNA was determined by amplifying a 97-bp fragment of the house-keeping gene *GAPDH* ❖ Levels of ccf-mtDNA in the serum of EWS patients were significantly lower than in healthy controls with a sensitivity =76.1%, a specificity =68.4% and an AUC =0.708 ❖ Serum levels of ccf-mtDNA were associated with the status of tumor metastasis
Cell-free circulating mitochondrial DNA content and risk of hepatocellular carcinoma in patients with chronic HBV infection (66)	*Scientific Reports*, April 2016	❖ This study aimed to determine circulating mtDNA content in serum samples from 116 HBV related hepatocellular carcinoma (HCC) cases and 232 frequency-matched cancer-free HBV controls, and evaluate the retrospective association between mtDNA content and HCC risk ❖ The relative mtDNA content was measured by qRT-PCR in which the ratio of the copy number for mitochondrial ND1 gene to the copy of a human single copy gene 36B4 was used to determine the relative mtDNA content ❖ HCC cases had significantly lower circulating mtDNA content than controls (1.06 versus 2.47, P=0.000017) ❖ Patients with a lower level of serum mtDNA content (≤2.47) exhibited a significantly increased HCC risk with a crude OR of 2.22 (95% CI, 1.39–3.56; P=0.00087) in univariate analysis and an adjusted OR of 2.19 (95% CI, 1.28–3.72, P=0.004) in multivariate analysis adjusting for age, gender, smoking status, drinking status, family history of cancer, and cirrhosis, compared to those with a higher mtDNA content (>2.47) ❖ Using the patients with the highest level of mtDNA content as reference, patients with lower levels of mtDNA content showed significantly increased HCC risk in a dose-dependent manner in both univariate and multivariate analyses (P for trend =0.00016, and 0.001, respectively) ❖ When mtDNA content was added to multivariate analysis adjusting for age, gender, smoking status, drinking status, family history of cancer, and cirrhosis, the AUC of the ROC curve significantly increased from 0.7133 to 0.7511 (P=0.046) ❖ The AUC was 0.8020 in the multivariate model including demographic variables plus AFP, and the AUC significantly increased to 0.8498 after adding mtDNA to the model (P=0.032). Similar results were found in the models including each of the liver enzymes which showed that mtDNA provided additional diagnostic value when jointly used with AFP or common liver enzymes
A prospective study of mitochondrial DNA copy number and the risk of prostate cancer (67)	*Cancer Causes & Control*, June 2017	❖ Seven hundred and ninety-three cases and 790 men control were assessed to evaluate the association between pre-diagnosis mtDNA copy number, measured in peripheral blood leukocytes, and the risk of prostate cancer (PCa) ❖ Overall, no significant difference of the median mtDNA copy number between cases and controls ❖ When the results were stratified by disease aggressiveness, a positive association was found with increasing mtDNA copy number for non-aggressive disease (OR =1.29, P=0.044) but not aggressive PCa (OR =1.02, P=0.933), though a Wald test for heterogeneity of the coefficients for mtDNA copy number was not statistically significant (P=0.334) ❖ Among controls, higher mitochondrial DNA copy number was associated with an increased PSA level (P=0.014) ❖ Increasing mitochondrial DNA copy number was associated with an increased risk of non-aggressive prostate cancer with high (≥4.0 ng/mL) PSA at diagnosis (OR =1.32, P=0.037), but not low PSA at diagnosis (OR =1.16, P=0.527)

Table 2 *(continued)*

Table 2 *(continued)*

Article	Journal/date	Summary/results
Lower mitochondrial DNA copy number in peripheral blood leukocytes increases the risk of endometrial cancer (68)	*Molecular Carcinogenesis*, June 2016	❖ MtDNA copy number was measured in peripheral blood leukocytes (PBLs) from 139 endometrial cancer patients and 139 age-matched controls to determine the association of mtDNA copy number with the risk of endometrial cancer ❖ The normalized mtDNA copy number was significantly lower in endometrial cancer cases (median, 0.84; range, 0.24–2.00) than in controls (median, 1.06; range, 0.64–1.96) (P<0.001) ❖ Dichotomized into high and low groups based on the median mtDNA copy number value in the controls, individuals with low mtDNA copy number had a significantly increased risk of endometrial cancer (adjusted OR, 5.59; 95% CI, 3.05–10.25; P<0.001) compared to those with high mtDNA copy number
Circulating mitochondrial DNA in the serum of patients with testicular germ cell cancer as a novel noninvasive diagnostic biomarker (69)	*BJU International*, July 2009	❖ The diagnostic and prognostic value of the quantification and the integrity of cell-free mtDNA was studied in the serum of 74 patients with testicular cancer and 35 healthy individuals ❖ mtDNA levels were significantly higher in cancer patients than in controls with a distinguishing sensitivity of 59.5%, a specificity of 94.3% and an AUC of 0.787 ❖ No difference in the mtDNA integrity was observed between patients and healthy individuals
Circulating mitochondrial DNA in serum: a universal diagnostic biomarker for patients with urological malignancies (70)	*Urologic Oncology: Seminars and Original Investigations*, July 2012	❖ The serum of 84 bladder cancer, 33 renal cell carcinoma, 23 prostate cancer patient and 79 healthy individual, was analyzed for cell-free circulating mtDNA levels and integrity ❖ The results showed a significant increase in circulating mtDNA levels in cancer patients compared to healthy controls (84% sensitivity and 97% specificity) ❖ mtDNA integrity was increased in renal cell carcinoma and bladder cancer compared to healthy individuals and prostate cancer patients ❖ A correlation was also observed between mtDNA integrity and pathological stage in renal cell cancer on one hand, and tumor grade in bladder cancer on the other hand

Q-PCR, quantitative polymerase chain reaction; CRC, colorectal cancer; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; OR, odds ratio; AUC, area under the curve; PSA, prostate-specific antigen.

had an inadequate discriminating potential between patients with lung cancer and nonmalignant disease [AUC =0.598 (P=0.122)] (73). But on the other hand, other groups showed that a CTC count of more than 25 had a high sensitivity (89%) and specificity (100%) for the differentiation between benign and malignant disease (74), and a cut-off threshold of 8.7 folate receptor-positive-CTC units between the control group and patients with lung cancer presented an AUC of 0.7956 (sensitivity =77.7% and specificity =89.5%) (75). CTCs were also detected in patients with chronic obstructive pulmonary disease (COPD) (*Table 3*), a risk factor for lung cancer, without clinically detected lung cancer (76), in addition CTCs number was higher in patients with stage IV NSCLC compared with patients with stage IIIB (77).

While technology to capture and profile CTCs has advanced rapidly, the complexity and the weak analytical signal may limit clinical utility relative to ctDNA-based methods (38). Initial studies, such as that performed by Diaz *et al.*, suggest that when both ctDNA and CTCs were present, ctDNA fragments outnumbered CTCs by 50 to 1 (79) providing a much higher analytical signal. Nonetheless, CTCs do not have the disadvantage of the necessity to measure very small amounts of mutated fragments in the plasma due to the important release of wild type cfDNA in some patients whose tumors are invaded by a tumor microenvironment in a large proportion (>90% of the cells). In a recent trial of lung cancer patients, ctDNA outperformed CTCs for detection of the KRAS mutation, revealing sensitivities of 96% and 52%, respectively (80). Very recently, there has been a certain enthusiasm for single cell analysis which might be of importance for screening, since this approach is technically feasible. The results

Table 3 CTCs and cancer screening

Article	Journal/date	Summary/results
Circulating tumour cells as a biomarker for diagnosis and staging in pancreatic cancer (71)	*British Journal of Cancer*, June 2016	❖ Blood was collected prospectively from 100 pre-treated patients [28 with non-adenocarcinoma diagnosis and 72 with pancreatic ductal adenocarcinoma (PDAC)], and samples were evaluated for the presence and number of CTCs using the microfluidic NanoVelcro CTC chip ❖ KRAS mutation analysis was used to compare the CTCs with primary tumor tissue ❖ In five patients tested: 100% concordance for KRAS mutation subtype between primary tumor and CTCs ❖ The presence of CTCs was observed in 54/72 patients with confirmed PDAC with a sensitivity =75.0%, a specificity =96.4% and an AUC =0.867) ❖ Using a cut-off of ≥3, CTCs in 4 ml of blood were able to discriminate between local/regional and metastatic disease (AUC =0.885)
The prognostic and diagnostic value of circulating tumor cells in bladder cancer and upper tract urothelial carcinoma: a meta-analysis of 30 published studies (72)	*Oncotarget*, June 2017	❖ Based on the published results of 30 different studies with a total of 2,161 urothelial cancer patients, the prognostic and diagnostic value of CTCs in urothelial cancer was assessed ❖ Concerning the diagnostic accuracy of CTC detection, the overall sensitivity and specificity were 0.35 and 0.97 respectively with significant heterogeneity (I^2=89.40% and 89.71%), with a pooled positive likelihood ratio (PLR) and negative likelihood ratio (NLR) of 11.2 and 0.67 respectively. The diagnostic odds ratio (DOR) was 17 and the summary ROC curve (sROC) for the included studies which reflects the global summary of test's performance showed an AUC of 0.70, so a moderate accuracy of the diagnostic test
Circulating tumor cell as a diagnostic marker in primary lung cancer (73)	*Clinical Cancer Research*, November 2009	❖ The role of CTC counts in the discernment between primary lung cancer and nonmalignant diseases was examined in a cohort of 150 patients clinically suspected to have or with a diagnosis of primary lung cancer (125 primary lung cancer and 25 with nonmalignant disease) ❖ Thirty point six percent of lung cancer patients and 12% of patients with nonmalignant disease had detectable CTCs ❖ CTC enumerations were higher in lung cancer patients, but ROC curve analysis demonstrated an inadequate potential of the CTC counts to discriminate between patients with lung cancer and nonmalignant disease [AUC =0.598 (P=0.122)]
Circulating tumor cells in diagnosing lung cancer: clinical and morphologic analysis (74)	*The Annals of Thoracic Surgery*, June 2015	❖ CTCs were evaluated from potential lung cancer patients to predict the malignancy of lung lesions ❖ CTCs were isolated by size method from peripheral blood of 77 patients with malignant (n=60) and benign (n=17) lung lesions. They were morphologically classified as cells with malignant feature, cells with uncertain malignant feature, and cells with benign feature, then statistically correlated with clinicocytopathologic characteristics of corresponding lung lesion ❖ A CTC count of >25 had high sensitivity and specificity for the differentiation between benign and malignant disease (sensitivity =89% and specificity =100%) ❖ Isolated CTCs shared similar histology and morphological features (72%) with biopsy samples ❖ In tested stage I patients (42%), the numbers of CTCs correlated with tumor size (P=0.001)

Table 3 *(continued)*

Table 3 *(continued)*

Article	Journal/date	Summary/results
Clinical Significance of Folate Receptor-positive Circulating Tumor Cells Detected by Ligand-targeted Polymerase Chain Reaction in Lung Cancer (75)	*Journal of Cancer*, 2017	❖ Folate receptor (FR)-positive circulating tumor cells (FR±CTCs) were detected by a novel ligand-targeted polymerase chain reaction (LT-PCR) detection technique ❖ FR±CTC levels of patients with lung cancer were significantly higher than controls (patients with benign lung diseases and healthy controls) ❖ A cut-off threshold of 8.7 CTC units was established between control group and patients with lung cancer with an AUC =0.7956, a sensitivity =77.7% and a specificity =89.5% ❖ Compared with established clinical biomarkers [CEA, cytokeratin 19 fragment (CYFRA21-1), and neuron-specific enolase (NSE)], FR±CTC showed the highest diagnostic efficiency (highest AUC), so a combination of FR±CTC, CEA, NSE, and CYFRA21-1 could significantly improve the diagnostic efficacy in differentiating patients with lung cancer from benign lung disease
"Sentinel" circulating tumor cells allow early diagnosis of lung cancer in patients with chronic obstructive pulmonary disease (76)	*PloS One*, October 2014	❖ Chronic obstructive pulmonary disease (COPD) is a risk factor for lung cancer; ❖ This study aimed to examine the presence of CTCs, in complement to CT-scan, in COPD patients without clinically detectable lung cancer as a first step to identify a new marker for early lung cancer diagnosis ❖ The presence of CTCs was examined by an ISET filtration-enrichment technique, for 245 subjects without cancer, including 168 (68.6%) COPD patients, and 77 subjects without COPD (31.4%), including 42 control smokers and 35 non-smoking healthy individuals ❖ The presence of CTCs in 5 out of 168 COPD patients predicted the appearance of lung nodules 1–4 years after initial detection of CTCs ❖ No CTCs were detected in control smoking and non-smoking healthy individuals ❖ CTCs can be detected in patients with COPD without clinically detectable lung cancer ❖ Monitoring "sentinel" CTC-positive COPD patients may allow early diagnosis of lung cancer
Evaluation and Prognostic Significance of Circulating Tumor Cells in Patients With Non–Small-Cell Lung Cancer (77)	*Journal of Clinical Oncology*, April 2011	❖ The detection as well as the prognostic significance of CTCs were assessed in 101 stage III or IV NSCLC patients ❖ The CTCs number was higher in patients with stage IV NSCLC (n=60; range, 0–146) compared with patients with stage IIIB (n=27; range, 0–3) or IIIA disease (n=14; no CTCs detected) ❖ In univariate analysis, progression-free survival was 6.8 *vs.* 2.4 months with P<0.001, and overall survival (OS) was 8.1 *vs.* 4.3 months with P<0.001 for patients with fewer than five CTCs compared with five or more CTCs before chemotherapy, respectively ❖ In multivariate analysis, CTC number was the strongest predictor of OS [hazard ratio (HR), 7.92; 95% CI, 2.85–22.01; P<0.001], and the point estimate of the HR was increased with incorporation of a second CTC sample that was taken after one cycle of chemotherapy (HR, 15.65; 95% CI, 3.63–67.53; P<0.001)

CTC, circulating tumor cell; AUC, area under the curve; NSCLC, non-small cell lung cancer.

on this subject are, for the moment, little discussed or not convincing (81). In addition to the paucity of CTCs' number in blood, one of the major drawbacks of CTC analysis would be the necessity of an immediate processing (within a half day), while it would be up to 5 days for cfDNA analysis with full blood stabilizing tubes (*Table 3*). Nevertheless, CTCs are more related to the liquid biopsy terminology and intrinsically more powerful since determination of cellular markers may be combined to the genetic information in the same blood sample, and they are rather of relevance for real-time diagnosis of cancer progression.

Other molecular circulating biomarkers

Other circulating biomarkers were investigated for the early detection of cancer (*Table 4*). The diagnostic relevance of circulating cell-free microRNAs (miRNAs) was studied in the blood of patients with different types of cancer (96). A study showed that tumor-associated circulating miRNAs are elevated in the blood of breast cancer patients and associated with tumor progression (82). A multivariable signature of nine circulating miRNAs was validated and it provided a high discrimination between breast cancer patients and healthy controls with a corresponding AUC of 0.665 (83). Other circulating miRNA signatures were identified for the early diagnosis of lung malignancies (84).

Another study suggested, for breast cancer, that the presence of circulating cancer-associated macrophage like cells might have a utility as a screening tool and may differentiate patients with malignant disease, benign breast conditions, and healthy individuals (92). Tumor educated platelets (TEPs) are another studied circulating biomarker and it was shown that TEPS mRNA profiles can be used to distinguish between healthy donors and cancer patients (93,94) (*Table 4*).

New avenues for molecular cancer screening tests based on cfDNA analysis

Genetic alteration profile

Based on the assumption that early-detection coupled with early treatment would be key to saving lives, liquid biopsies also have the potential to allow physicians to identify patients whose tumors have specific mutations in the least invasive way possible. Several attempts were made towards this goal especially with the use of sophisticated Next-Gen sequencing methods applied on circulating DNA. Thus, the group of Velculescu recently evaluated this strategy on 138 patients with early tumors and it successfully identified the early-stage cancer in more than half of the patients using targeted error correction sequencing (TEC-Seq) (56). cfDNA analysis was used to detect the return of cancers after treatment. Authors noted that 58 genes are typically associated with breast, lung, ovarian cancer and CRC. Of the 138 cancers, they could detect 86 stage I and stage II cancers. The genes were sequenced in 100 patients and 82 of them showed the same mutations in blood samples as well as in the tumor tissue samples. None of 44 tested healthy patients as control group have cancer-derived mutations. The limitations of the study/technological strategy is the difficulty in identifying the rare DNA from cancers and in showing up results from other types of genetic alterations or mutations that a person is born with or develops during his life.

Another study showed that TP53 mutations were also detected in the plasma of 49% small cell lung cancer (SCLC) patients with significantly higher allelic fractions in cases than in controls (52).

Virus genome detection

Dennis Lo's group very recently described an elegant study in which the strategy is to detect the Epstein-Barr virus which is involved in most nasopharyngeal cancer cases, and to hunt for viral DNA that tumors shed into the blood in large quantities, rather than rare bits of cancer cells themselves (97). Viral DNA was found in 1,112 or 5.6% of a cohort of 20,000 men. Of those, 309 also had the DNA on confirmatory tests a month later; and, 34 turned out to have cancer following endoscopy and MRI examinations. More cases were found at the earliest stage. Only one person who tested negative on screening developed nasopharyngeal cancer within a year. Clearly this approach is promising and prescription appears warranted.

Circulating DNA fragmentation

Our team first observed that (I) shorter circulating DNA molecules were more abundant in the plasma of CRC patients relative to healthy individuals; (II) the quantity of short circulating DNA fragments <145 bp is directly correlated with ctDNA concentration (98); (III) and that mutant cfDNA derived from malignant cells is highly fragmented compared to non-mutant cfDNA (99). Optimal detection by quantitative PCR (Q-PCR) of ctDNA is

Table 4 Other molecular circulating biomarkers for cancer screening

Article	Journal/date	Summary/results
Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer (82)	*Breast Cancer Research*, November 2010	❖ In the serum of 59 primary breast cancer patients, 30 metastatic patients and 29 healthy women, the relative concentration of total RNA and of breast cancer-associated miR10b, miR141 and miR155 were measured ❖ The relative concentrations of total RNA (P=0.0001) and miR155 (P=0.0001) in serum significantly discriminated primary breast cancer patients from healthy women ❖ miR10b (P=0.005), miR34a (P=0.001) and miR155 (P=0.008) allowed the discrimination between metastatic patients and healthy controls ❖ The presence of metastases correlated with the levels of total RNA (P=0.0001), miR10b (P=0.01), miR34a (P=0.003) and miR155 (P=0.002) ❖ For patients with primary breast cancer, individuals with an advanced tumor stage (pT3 to 4) had significantly more total RNA (P=0.0001) and miR34a (P=0.01) in their blood than patients at early tumor stages (pT1 to 2)
Novel circulating microRNA signature as a potential non-invasive multi-marker test in ER-positive early-stage breast cancer: A case control study (83)	*Molecular Oncology*, July 2014	❖ Serum from 48 patients with ER-positive early-stage breast cancer obtained at diagnosis (24 lymph node-positive and 24 lymph node-negative) and 24 age-matched healthy controls underwent Global miRNA analysis using LNA-based quantitative real-time PCR (qRT-PCR) ❖ A signature of miRNAs was subsequently validated in an independent set of 111 serum samples from 60 patients with early-stage breast cancer and 51 healthy controls and further tested for reproducibility in three independent data sets from the GEO Database ❖ A multivariable signature, that provided considerable discrimination between breast cancer patients and healthy controls with P=0.012 and a corresponding AUC =0.665, was identified. It consisted of nine miRNAs (miR-15a, miR-18a, miR-107, miR-133a, miR-139-5p, miR-143, miR-145, miR-365, and miR-425) ❖ No association between miRNA expression and tumor grade, tumor size, menopausal or lymph node status was observed ❖ The signature was also successfully validated in a previously published independent data set of circulating miRNAs in early-stage breast cancer (P=0.024)
Circulating epigenetic biomarkers in lung malignancies: From early diagnosis to therapy (84)	*Lung Cancer*, May 2017	❖ The hypermethylation of tumor suppressor genes is frequently observed in cancers, and such epigenetic changes are potential markers for detecting and monitoring tumors ❖ The presence of methylated DNA in the serum or plasma of patients was revealed for various types of malignancy, including lung cancer (29,85) ❖ Methylated tumor suppressor genes, such as *p16INK4A*, *RARB2*, and *RASSF1A* were found in the blood of lung cancer patients (86-88) ❖ A profile of 10-serum miRNAs (miR-20a, miR-24, miR-25, mir-145, miR-152, miR-199a-5p, miR-221, miR-222, miR-223, miR-320) has been identified by analyzing serum miRNAs from a sample set including 400 NSCLC cases and 220 controls (89) ❖ The miR-183 family (miR-96, miR-182, and miR-183), a group of onco miRs, has been found to be overexpressed in lung tumors and serum of NSCLC patients (90) ❖ Based on the use of miRNAratios, a mirRNA signature was developed with a potential for general clinical use. A miRNA signature classified (MSC) algorithm was defined by using 24-miRNA ratios, for prediction, diagnosis, and prognosis of lung cancer. It was able to reduce false-positive rate of low-dose computer tomography (LDCT), thus improving the efficacy of LC screening (91)

Table 4 *(continued)*

Table 4 *(continued)*

Article	Journal/date	Summary/results
Circulating Cancer-Associated Macrophage-Like Cells Differentiate Malignant Breast Cancer and Benign Breast Conditions (92)	*Cancer Epidemiology, Biomarkers & Prevention*, July 2016	The article consists of two related but separate studies: ❖ In the first study, circulating cancer-associated macrophage-like cells (CAML) were isolated from blood samples of patients with known malignant disease (n=41) using CellSieve microfilters. The prevalence and specificity were compared against 16 healthy volunteers ❖ A follow-up double-blind pilot study was conducted on 41 women undergoing core-needle biopsy to diagnose suspicious breast masses. CAMLs were found in 93% of known malignant patients (average 19.4 cell/sample), but none in the healthy controls. In subjects undergoing core biopsy for initial diagnosis, CAMLs were found in 88% of subjects with invasive carcinoma and 26% with benign breast conditions ◆ Comparing subjects with benign conditions (n=19, excluding the high-risk noninvasive lesions) to those with invasive carcinoma (n=17) results in an ROC curve with an AUC of 0.78 (95% CI, 0.63–0.92), with a threshold of 1CAML as a positive finding, a sensitivity =88%, a specificity =74%, a PPV =75%, and a NPV =88% ◆ The results showed that all breast cancer subtypes based on ER, PR, and HER2 status produce detectable levels of CAMLs and that there is not a pronounced effect of tumor stage or nodal status on the presence of these cells ◆ These preliminary pilot studies suggest that the presence of CAMLs may differentiate patients with malignant disease, benign breast conditions, and healthy individuals, and therefore have a utility as a screening tool
RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics (93)	*Cancer Cell*, November 2015	❖ MRNA sequencing of 283 platelet samples [healthy donors (n=55) and both treated and untreated patients with early, localized (n=39) or advanced, metastatic cancer (n=189)] was conducted to determine the diagnostic potential of Tumor-educated blood platelets (TEPs) ❖ The study includes six tumor types: non-small cell lung carcinoma (NSCLC, n=60), colorectal cancer (CRC, n=41), glioblastoma (GBM, n=39), pancreatic cancer (PAAD, n=35), hepatobiliary cancer (HBC, n=14), and breast cancer (BrCa, n=39) ❖ Tumor-educated platelets (TEPs) are implicated as central players in the systemic and local responses to tumor growth, thereby altering their RNA profile ❖ MRNA profiles of tumor-educated platelets are distinct from platelets of healthy individuals, and they were able to distinguish 228 patients with localized and metastasized tumors from 55 healthy individuals with a 96% accuracy ❖ Across six different tumor types, the location of the primary tumor was correctly identified with 71% accuracy ❖ MET or HER2-positive, and mutant KRAS, EGFR, or PIK3CA tumors were accurately distinguished using surrogate TEP mRNA profiles

Table 4 *(continued)*

Table 4 *(continued)*

Article	Journal/date	Summary/results
Platelet RNA signatures for the detection of cancer (94)	*Cancer and Metastasis Reviews*, July 2017	❖ TEP (tumor-educated platelets): Tumor-associated biomolecules are transferred to platelets resulting in their "education". External stimuli, such as activation of platelet surface receptors and lipopolysaccharide-mediated platelet activation, induce specific splicing of pre-messenger RNAs (mRNAs) in circulating TEPs. TEPs may also undergo queue-specific splice events in response to signals released by cancer cells and the tumor microenvironment such as by stromal and immune cells ❖ Platelet mRNA profiles can be used to distinguish between healthy donors and cancer patients ❖ Platelets can sequester extracellular vesicles from cancer cells harboring tumor-specific RNA ❖ EGFRvIII, a deletion mutant of the epidermal growth factor receptor (EGFR), is such a specific tumor RNA which is considered to be present in 30% of glioblastoma tumors. Traces of this very malignant tumor of the central nervous system could be detected by RT-PCR of platelets from these patients. The EGFRvIII RNA transcript was detected with a sensitivity of 80% (4 out of 5 EGFRvIII-positive tumors were detected), and a specificity of 96% (25 out of 26 EGFRvIII-negative tumors were scored as negative). In addition, microarray analysis discovered an RNA signature that could distinguish between glioblastoma patients (n=8) and healthy controls (n=12) (95) ❖ mRNA sequencing of tumor-educated platelets distinguishes cancer patients from healthy individuals with 96% accuracy (93)

RNA, ribonucleic acid; ER, estrogen-receptor; LNA, locked nucleic acid; GEO, Gene Expression Omnibus; AUC, area under the curve; NSCLC, non-small cell lung cancer; NPV, negative predictive value; PPV, positive predictive value; PR, progesterone receptor; LC, lung cancer.

obtained with amplicons <100 bp (100) and Atomic Force Microscopy analysis showed that cfDNA fragments from cancer patient plasma are mostly averaging 135 bp (28). High discrimination between stage IV CRCs and healthy individuals was reported when targeting a short amplicon (63 bp) (28). In another report, we revealed that mutant cfDNA fragment proportion was much higher than non-mutant cfDNA below 145 bp size range (99). These observations were later confirmed by Leszinski *et al.*, who showed that DNA integrity was significantly higher in patients with CRC when compared with healthy controls and with individuals with benign colorectal diseases (P=0.005 and 0.006, respectively) (101); and by Jiang *et al.* using massively parallel sequencing to study the size profiles of hepatocellular carcinoma patient plasma DNA samples at a single-base resolution in a genome-wide manner (102). Based upon these observations, various DNA integrity indexes were evaluated with various efficacies in discriminating healthy and cancer patients due to the lack of readily standard operating procedures (69,70), and sufficient tested patient number (98,99). Only the recent study of Tanos *et al.* reported statistically evaluated screening power of a specific DNA integrity index as determined by a Q-PCR method (103).

This screening strategy is based upon a differential between cfDNA structure deriving from malignant and healthy cells rather than on the cfDNA sequence. This would lead to easier implementation, and to lower screening test cost. Works on circulating DNA size profiling are ongoing to set Q-PCR and sequencing approaches toward its inclusion in a screening blood test for cancer.

A test based on circulating mtDNA: the MiTest

We previously showed that the amount of cfDNA may be a discriminatory factor between healthy subjects and CRC patients (28) and the receiver operating characteristic (ROC) curve analysis revealed an AUC of 0.91 (*Figure 1*). We also showed that quantifying and associating circulating mtDNA and nuclear DNA content enables to distinguish cancer subjects from healthy individuals (103). We determined an index based on the detection of particular sequences in the nuclear and mitochondrial genomes. When applied to cell culture supernatant, a significant difference was observed between normal and cancer cell lines (*Figure 2A*), and in

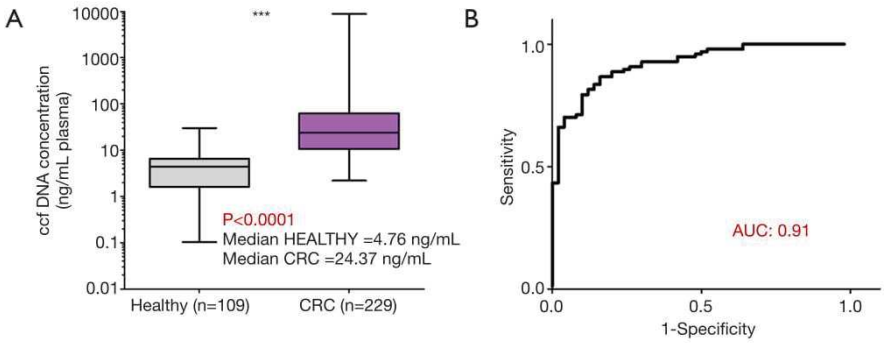


Figure 1 Diagnostic performance of cfDNA concentration in discriminating healthy individuals and cancer patients. (A) Comparison of the quantification of plasma cfDNA (in ng/mL) from healthy individuals (n=109) and colorectal cancer (CRC) patients (stage I–II, n=15; stage III, n=25; stage IV, n=189; and total CRC, n=229). The concentration observed in CRC patients is significantly greater than those of healthy individuals (P<0.0001); (B) diagnosis predictive capacity of total cfDNA concentration to distinguish plasma from CRC patients and healthy subjects. ROC curve representation deriving from the univariate logistic analysis corresponding to the total cfDNA (AUC =0.91) (28). ***, P value <0.0001. cfDNA, circulating cell-free DNA; ROC, receiver operating characteristic; AUC, area under the curve.

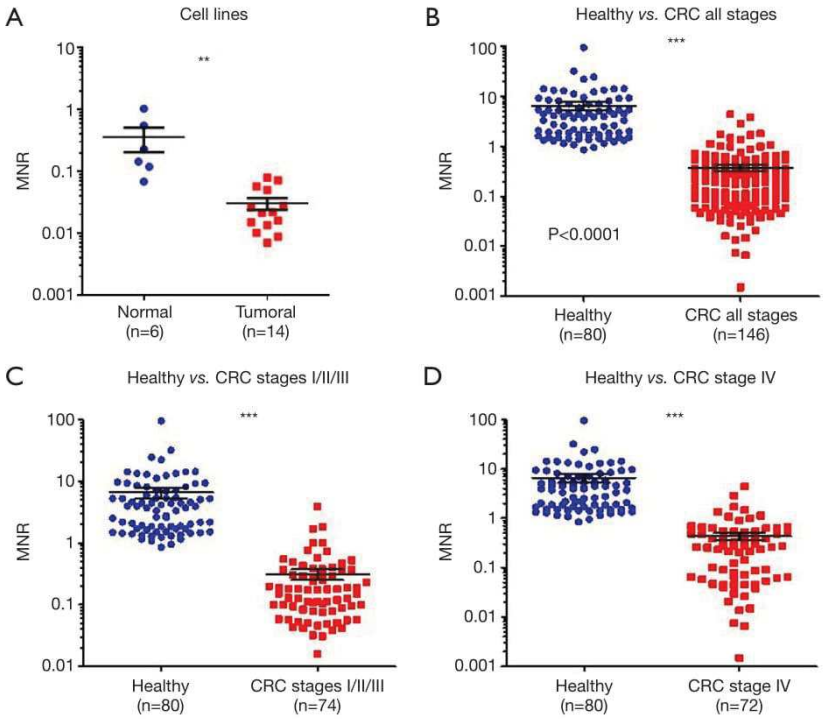


Figure 2 Screening capacity of the MNR test (103). (A) Comparison between the MNR in the supernatant of various tumor cell lines (n=14) and normal cell lines (n=6); (B) dot plot of the MNR in the blood of healthy individuals (n=80) and in CRC patients from stages I to IV (n=146); (C) dot plot of the MNR in blood of healthy individuals (n=80) and in stage I/II/III CRC patients (n=74); (D) dot plot of the MNR in blood of healthy individuals (n=80) and in stage IV CRC patients (n=72). **, P value <0.01; ***, P value <0.001. CRC, colorectal cancer; MNR, multi-normalized ratio.

plasma, this index is statistically lower in CRC patients than that of healthy subjects (*Figure 2B,C,D*). Our findings suggest that the MiTest consists of a powerful screening test for early cancer detection, and studies are ongoing to combine this marker with other parameters in order to increase the discriminative potential of the test in large cohorts of patients and healthy individuals.

Conclusions

Scientists discovered that tumors shed cells and nucleic acids into the blood circulation more than a century and 70 years ago, respectively (104,105). These molecules and cells were more recently found to reveal some of the same information that tissue biopsies provide (25,106,107). As termed here, liquid biopsy research has expanded in the last decade, generating a rapidly growing area of interest in oncology. Both academic and industry researchers from diverse areas of expertise are working on many fronts to develop, refine, and establish clinical uses for liquid biopsy tests (108).

The minimally invasive nature of liquid biopsy for malignancy without the delay, cost, and risk associated with tissue biopsy, potentially at a microscopic stage before radiologic detectability are promising advantages for cancer screening (38). Several circulating biomarkers are being investigated, from cfDNA, CTCs, circulating miRNAs and others, for the development of tests for early cancer detection. Exosomes, containing certain proteins and nucleic acids, could also be a source of multiple markers of malignancy which the analysis might be promising for the development of screening methods (109,110). But a few of these biomarkers were validated towards clinical practice. cfDNA of nuclear and mitochondrial origin seems to have an advantage in cancer screening compared to other biomarkers by showing better efficiency, and at this time, it appears to possess the characteristics to be more rapidly implemented. Combining various analysis from blood sample such as the detection by sequencing of selected mutations and genes and of protein biomarkers might be an attractive approach as very recently reported by Cohen *et al.* (111). While high specificity level (99%) and an overall AUC of 0.91 were observed, their data showed a moderate sensitivity (varying from 30% to 99% upon cancer types), and the cost of this multi-parametric analysis could hinder its routine use as a massive screening test. This approach should, at least, be considered for populations at risk or for specific malignant diseases.

It is to be feared or hoped that the worldwide use of a screening test will be distinguished in two ways: (I) with a test approved by public health administrations and reimbursed, followed by a statistically long and rigorous study; and (II) privately/individually (e.g., pregnancy test) with a moderate level of performance or evaluation proposed in the near future.

Standardization of the pre-analytical parameters and better knowledge on the exact origin and structure of cfDNA would provide the additional step for the implementation of its analysis. Advancement on sophisticated Q-PCR methods or Next-Gen sequencing will inevitably improve reliability of the analytical performance of the future tests. As indicated by Dennis Lo of The Chinese University of Hong Kong: "*We are brick by brick putting that technology into place*". Looking forward, we may consider that liquid biopsies could add a new dimension to the cancer screening and diagnosis role of the primary care physician prior to oncology referral (38). At least, investigation of liquid biopsy screening power in tandem with other tests, such as a magnetic resonance imaging (MRI) is warranted. We envision that liquid biopsy tests may be used to screen for early-stage cancer in high-risk individuals, such as those with hereditary cancer syndromes. Nevertheless, it is crucial to further investigate these emerging biomarkers. In addition, a combined use of multiple markers may be a way to achieve more significance in early cancer detection, and increase the sensitivity and specificity of the tests. The years to come seem to be exciting, while universal screening, that constitutes the "holy grail" in oncology, appears to be accessible.

Acknowledgements

We are grateful to Safia El Messaoudi, Armèle Bonnet-Kerrache and Amaelle Otandault for their help. We would also like to thank Marc Ychou, Denis Pezet and Muriel Mathonnet for providing the blood samples.

Funding: This work was supported by the INSERM (Institut National de la Santé et de la Recherche Médicale), Lilly (France), and the SIRIC Montpellier Grant (INCa-DGOS-Inserm 6045), France.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Maxim LD, Niebo R, Utell MJ. Screening tests: a review with examples. Inhal Toxicol 2014;26:811-28.
2. Parsons DW, Jones S, Zhang X, et al. An Integrated Genomic Analysis of Human Glioblastoma Multiforme. Science 2008;321:1807-12.
3. Navin N, Kendall J, Troge J, et al. Tumour evolution inferred by single-cell sequencing. Nature 2011;472:90-4.
4. Boesch M, Zeimet AG, Reimer D, et al. The side population of ovarian cancer cells defines a heterogeneous compartment exhibiting stem cell characteristics. Oncotarget 2014;5:7027-39.
5. Wilson JM, Jungner G. Principles and practice of screening for disease. Geneva: World Health Organization, 1968.
6. Andermann A, Blancquaert I, Beauchamp S, et al. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. Bull World Health Organ 2008;86:317-9.
7. Petticrew MP, Sowden AJ, Lister-Sharp D, et al. False-negative results in screening programmes: systematic review of impact and implications. Health Technol Assess 2000;4:1-120.
8. PubMed Health. Benefits and risks of screening tests. 2016. Available online: https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0072602
9. Hakama M, Coleman MP, Alexe DM, et al. Cancer screening: Evidence and practice in Europe 2008. Eur J Cancer 2008;44:1404-13.
10. Arbyn M, Sankaranarayanan R, Muwonge R, et al. Pooled analysis of the accuracy of five cervical cancer screening tests assessed in eleven studies in Africa and India. Int J Cancer 2008;123:153-60.
11. Mayrand MH, Duarte-Franco E, Rodrigues I, et al. Human Papillomavirus DNA versus Papanicolaou Screening Tests for Cervical Cancer. N Engl J Med 2007;357:1579-88.
12. Friedewald SM, Rafferty EA, Rose SL, et al. Breast Cancer Screening Using Tomosynthesis in Combination With Digital Mammography. JAMA 2014;311:2499-507.
13. Rafferty EA, Park JM, Philpotts LE, et al. Assessing Radiologist Performance Using Combined Digital Mammography and Breast Tomosynthesis Compared with Digital Mammography Alone: Results of a Multicenter, Multireader Trial. Radiology 2013;266:104-13.
14. Stieber P, Nagel D, Blankenburg I, et al. Diagnostic efficacy of CA 15-3 and CEA in the early detection of metastatic breast cancer—A retrospective analysis of kinetics on 743 breast cancer patients. Clin Chim Acta 2015;448:228-31.
15. Catalona WJ, Smith DS, Ratliff TL, et al. Measurement of Prostate-Specific Antigen in Serum as a Screening Test for Prostate Cancer. N Engl J Med 1991;324:1156-61.
16. Horeweg N, Scholten ET, de Jong PA, et al. Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers. Lancet Oncol 2014;15:1342-50.
17. Tóth K, Wasserkort R, Sipos F, et al. Detection of Methylated Septin 9 in Tissue and Plasma of Colorectal Patients with Neoplasia and the Relationship to the Amount of Circulating Cell-Free DNA. PLoS One 2014;9:e115415.
18. Wild N, Andres H, Rollinger W, et al. A Combination of Serum Markers for the Early Detection of Colorectal Cancer. Clin Cancer Res 2010;16:6111-21.
19. Chen JS, Chen KT, Fan WC, et al. Combined analysis of survivin autoantibody and carcinoembryonic antigen biomarkers for improved detection of colorectal cancer. Clin Chem Lab Med 2010;48:719-25.
20. Song LL, Li YM. Current noninvasive tests for colorectal cancer screening: An overview of colorectal cancer screening tests. World J Gastrointest Oncol 2016;8:793-800.
21. Dhaliwal A, Vlachostergios PJ, Oikonomou KG, et al. Fecal DNA testing for colorectal cancer screening: Molecular targets and perspectives. World J Gastrointest Oncol 2015;7:178-83.
22. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Fecal DNA versus Fecal Occult Blood for Colorectal-Cancer Screening in an Average-Risk Population. N Engl J Med 2004;351:2704-14.
23. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget Stool DNA Testing for Colorectal-Cancer Screening. N Engl J Med 2014;370:1287-97.
24. National Cancer Institute. NCI Dictionary of Cancer Terms. Available online: https://www.cancer.gov/publications/dictionaries/cancer-terms
25. Thierry AR, El Messaoudi S, Gahan PB, et al. Origins, structures, and functions of circulating DNA in oncology. Cancer Metastasis Rev 2016;35:347-76.
26. Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer—A survey. Biochim Biophys Acta 2007;1775:181-232.
27. Perakis S, Auer M, Belic J, et al. Chapter Three - Advances in Circulating Tumor DNA Analysis. In: Makowski GS. Advances in Clinical Chemistry. Philadelphia: Elsevier, 2017:73-153.

28. Mouliere F, El Messaoudi S, Pang D, et al. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. Mol Oncol 2014;8:927-41.
29. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer 2011;11:426-37.
30. Aravanis AM, Lee M, Klausner RD. Next-Generation Sequencing of Circulating Tumor DNA for Early Cancer Detection. Cell 2017;168:571-4.
31. Rahier JF, Druez A, Faugeras L, et al. Circulating nucleosomes as new blood-based biomarkers for detection of colorectal cancer. Clin Epigenetics 2017;9:53.
32. Lamb YN, Dhillon S. Epi proColon® 2.0 CE: A Blood-Based Screening Test for Colorectal Cancer. Mol Diagn Ther 2017;21:225-32.
33. Jin P, Kang Q, Wang X, et al. Performance of a second-generation methylated SEPT9 test in detecting colorectal neoplasm. J Gastroenterol Hepatol 2015;30:830-3.
34. Church TR, Wandell M, Lofton-Day C, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut 2014;63:317-25.
35. Rasmussen SL, Krarup HB, Sunesen KG, et al. Hypermethylated DNA, a circulating biomarker for colorectal cancer detection. PLoS One 2017;12:e0180809.
36. Bedin C, Enzo MV, Del Bianco P, et al. Diagnostic and prognostic role of cell-free DNA testing for colorectal cancer patients. Int J Cancer 2017;140:1888-98.
37. Lee BB, Lee EJ, Jung EH, et al. Aberrant Methylation of APC, MGMT, RASSF2A, and Wif-1 Genes in Plasma as a Biomarker for Early Detection of Colorectal Cancer. Clin Cancer Res 2009;15:6185-91.
38. Krishnamurthy N, Spencer E, Torkamani A, et al. Liquid Biopsies for Cancer: Coming to a Patient near You. J Clin Med 2017;6:3.
39. Liggett TE, Melnikov A, Yi Q, et al. Distinctive DNA methylation patterns of cell-free plasma DNA in women with malignant ovarian tumors. Gynecol Oncol 2011;120:113-20.
40. Lange CP, Campan M, Hinoue T, et al. Genome-Scale Discovery of DNA-Methylation Biomarkers for Blood-Based Detection of Colorectal Cancer. PLoS One 2012;7:e50266.
41. Diehl F, Li M, Dressman D, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci U S A 2005;102:16368-73.
42. Shaw JA, Stebbing J. Circulating free DNA in the management of breast cancer. Ann Transl Med 2014;2:3.
43. Bettegowda C, Sausen M, Leary RJ, et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. Sci Transl Med 2014;6:224ra24.
44. Catarino R, Coelho A, Araújo A, et al. Circulating DNA: Diagnostic Tool and Predictive Marker for Overall Survival of NSCLC Patients. PLoS One 2012;7:e38559.
45. Zhang R, Shao F, Wu X, et al. Value of quantitative analysis of circulating cell free DNA as a screening tool for lung cancer: A meta-analysis. Lung Cancer 2010;69:225-31.
46. Jiang T, Zhai C, Su C, et al. The diagnostic value of circulating cell free DNA quantification in non-small cell lung cancer: A systematic review with meta-analysis. Lung Cancer 2016;100:63-70.
47. Esposito A, Criscitiello C, Trapani D, et al. The Emerging Role of "Liquid Biopsies," Circulating Tumor Cells, and Circulating Cell-Free Tumor DNA in Lung Cancer Diagnosis and Identification of Resistance Mutations. Curr Oncol Rep 2017;19:1.
48. Gautschi O, Bigosch C, Huegli B, et al. Circulating Deoxyribonucleic Acid As Prognostic Marker in Non–Small-Cell Lung Cancer Patients Undergoing Chemotherapy. J Clin Oncol 2004;22:4157-64.
49. Sozzi G, Conte D, Mariani L, et al. Analysis of Circulating Tumor DNA in Plasma at Diagnosis and during Follow-Up of Lung Cancer Patients. Cancer Res 2001;61:4675-8.
50. Sozzi G, Conte D, Leon M, et al. Quantification of Free Circulating DNA As a Diagnostic Marker in Lung Cancer. J Clin Oncol 2003;21:3902-8.
51. Paci M, Maramotti S, Bellesia E, et al. Circulating plasma DNA as diagnostic biomarker in non-small cell lung cancer. Lung Cancer 2009;64:92-7.
52. Fernandez-Cuesta L, Perdomo S, Avogbe PH, et al. Identification of Circulating Tumor DNA for the Early Detection of Small-cell Lung Cancer. EBioMedicine 2016;10:117-23.
53. Lin Z, Neiswender J, Fang B, et al. Value of circulating cell-free DNA analysis as a diagnostic tool for breast cancer: a meta-analysis. Oncotarget 2017;8:26625-36.
54. Zhou Q, Li W, Leng B, et al. Circulating Cell Free DNA as the Diagnostic Marker for Ovarian Cancer: A Systematic Review and Meta-Analysis. PLoS One 2016;11:e0155495.
55. Cohen JD, Javed AA, Thoburn C, et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. Proc Natl Acad Sci 2017;114:10202-7.
56. Phallen J, Sausen M, Adleff V, et al. Direct detection of early-stage cancers using circulating tumor DNA. Sci

Transl Med 2017;9. pii: eaan2415.
57. Hosgood HD, Liu CS, Rothman N, et al. Mitochondrial DNA copy number and lung cancer risk in a prospective cohort study. Carcinogenesis 2010;31:847-9.
58. Meng S, De Vivo I, Liang L, et al. Pre-diagnostic leukocyte mitochondrial DNA copy number and risk of lung cancer. Oncotarget 2016;7:27307-12.
59. Thyagarajan B, Wang R, Nelson H, et al. Mitochondrial DNA Copy Number Is Associated with Breast Cancer Risk. PLoS One 2013;8:e65968.
60. Thyagarajan B, Wang R, Barcelo H, et al. Mitochondrial Copy Number is Associated with Colorectal Cancer Risk. Cancer Epidemiol Biomarkers Prev 2012;21:1574-81.
61. Qu F, Liu X, Zhou F, et al. Association between mitochondrial DNA content in leukocytes and colorectal cancer risk. Cancer 2011;117:3148-55.
62. Lan Q, Lim U, Liu CS, et al. A prospective study of mitochondrial DNA copy number and risk of non-Hodgkin lymphoma. Blood 2008;112:4247-9.
63. Lu H, Busch J, Jung M, et al. Diagnostic and prognostic potential of circulating cell-free genomic and mitochondrial DNA fragments in clear cell renal cell carcinoma patients. Clin Chim Acta 2016;452:109-19.
64. Fernandes J, Michel V, Camorlinga-Ponce M, et al. Circulating Mitochondrial DNA Level, a Noninvasive Biomarker for the Early Detection of Gastric Cancer. Cancer Epidemiol Biomarkers Prev 2014;23:2430-8.
65. Yu M, Wan YF, Zou QH. Cell-free Circulating Mitochondrial DNA in the Serum: A Potential Non-invasive Biomarker for Ewing's Sarcoma. Arch Med Res 2012;43:389-94.
66. Li L, Hann HW, Wan S, et al. Cell-free circulating mitochondrial DNA content and risk of hepatocellular carcinoma in patients with chronic HBV infection. Sci Rep 2016;6:23992.
67. Moore A, Lan Q, Hofmann JN, et al. A prospective study of mitochondrial DNA copy number and the risk of prostate cancer. Cancer Causes Control 2017;28:529-38.
68. Sun Y, Zhang L, Ho SS, et al. Lower mitochondrial DNA copy number in peripheral blood leukocytes increases the risk of endometrial cancer. Mol Carcinog 2016;55:1111-7.
69. Ellinger J, Albers P, Müller SC, et al. Circulating mitochondrial DNA in the serum of patients with testicular germ cell cancer as a novel noninvasive diagnostic biomarker. BJU Int 2009;104:48-52.
70. Ellinger J, Müller DC, Müller SC, et al. Circulating mitochondrial DNA in serum: A universal diagnostic biomarker for patients with urological malignancies. Urol Oncol 2012;30:509-15.
71. Ankeny JS, Court CM, Hou S, et al. Circulating tumour cells as a biomarker for diagnosis and staging in pancreatic cancer. Br J Cancer 2016;114:1367-75.
72. Zhang Z, Fan W, Deng Q, et al. The prognostic and diagnostic value of circulating tumor cells in bladder cancer and upper tract urothelial carcinoma: a meta-analysis of 30 published studies. Oncotarget 2017;8:59527-38.
73. Tanaka F, Yoneda K, Kondo N, et al. Circulating Tumor Cell as a Diagnostic Marker in Primary Lung Cancer. Clin Cancer Res 2009;15:6980-6.
74. Fiorelli A, Accardo M, Carelli E, et al. Circulating Tumor Cells in Diagnosing Lung Cancer: Clinical and Morphologic Analysis. Ann Thorac Surg 2015;99:1899-905.
75. Wang L, Wu C, Qiao L, et al. Clinical Significance of Folate Receptor-positive Circulating Tumor Cells Detected by Ligand-targeted Polymerase Chain Reaction in Lung Cancer. J Cancer 2017;8:104-10.
76. Ilie M, Hofman V, Long-Mira E, et al. "Sentinel" Circulating Tumor Cells Allow Early Diagnosis of Lung Cancer in Patients with Chronic Obstructive Pulmonary Disease. PLoS One 2014;9:e111597.
77. Krebs MG, Sloane R, Priest L, et al. Evaluation and Prognostic Significance of Circulating Tumor Cells in Patients With Non–Small-Cell Lung Cancer. J Clin Oncol 2011;29:1556-63.
78. Hofman P. Liquid biopsy for early detection of lung cancer. Curr Opin Oncol 2017;29:73-8.
79. Diaz LA, Bardelli A. Liquid Biopsies: Genotyping Circulating Tumor DNA. J Clin Oncol 2014;32:579-86.
80. Freidin MB, Freydina DV, Leung M, et al. Circulating Tumor DNA Outperforms Circulating Tumor Cells for KRAS Mutation Detection in Thoracic Malignancies. Clin Chem 2015;61:1299-304.
81. Alix-Panabières C, Pantel K. Characterization of single circulating tumor cells. FEBS Lett 2017;591:2241-50.
82. Roth C, Rack B, Müller V, et al. Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. Breast Cancer Res 2010;12:R90.
83. Kodahl AR, Lyng MB, Binder H, et al. Novel circulating microRNA signature as a potential non-invasive multi-marker test in ER-positive early-stage breast cancer: A case control study. Mol Oncol 2014;8:874-83.
84. Tomasetti M, Amati M, Neuzil J, et al. Circulating epigenetic biomarkers in lung malignancies: From early diagnosis to therapy. Lung Cancer 2017;107:65-72.
85. Jiang T, Ren S, Zhou C. Role of circulating-tumor DNA analysis in non-small cell lung cancer. Lung Cancer

2015;90:128-34.
86. Nikolaidis G, Raji OY, Markopoulou S, et al. DNA Methylation Biomarkers Offer Improved Diagnostic Efficiency in Lung Cancer. Cancer Res 2012;72:5692-701.
87. Bearzatto A, Conte D, Frattini M, et al. p16INK4A Hypermethylation Detected by Fluorescent Methylation-specific PCR in Plasmas from Non-Small Cell Lung Cancer. Clin Cancer Res 2002;8:3782-7.
88. Ponomaryova AA, Rykova EY, Cherdyntseva NV, et al. Potentialities of aberrantly methylated circulating DNA for diagnostics and post-treatment follow-up of lung cancer patients. Lung Cancer 2013;81:397-403.
89. Volinia S, Calin GA, Liu CG, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A 2006;103:2257-61.
90. Castro D, Moreira M, Gouveia AM, et al. MicroRNAs in lung cancer. Oncotarget 2017;8:81679-85.
91. Sozzi G, Boeri M, Rossi M, et al. Clinical Utility of a Plasma-Based miRNA Signature Classifier Within Computed Tomography Lung Cancer Screening: A Correlative MILD Trial Study. J Clin Oncol 2014;32:768-73.
92. Adams DL, Adams DK, Alpaugh RK, et al. Circulating Cancer-Associated Macrophage-Like Cells Differentiate Malignant Breast Cancer and Benign Breast Conditions. Cancer Epidemiol Biomarkers Prev 2016;25:1037-42.
93. Best MG, Sol N, Kooi I, et al. RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. Cancer Cell 2015;28:666-76.
94. Sol N, Wurdinger T. Platelet RNA signatures for the detection of cancer. Cancer Metastasis Rev 2017;36:263-72.
95. Nilsson RJ, Balaj L, Hulleman E, et al. Blood platelets contain tumor-derived RNA biomarkers. Blood 2011;118:3680-3.
96. Schwarzenbach H. Diagnostic relevance of circulating cell-free and exosomal microRNAs and long non-coding RNAs in blood of cancer patients. LaboratoriumsMedizin. 2016;40:345-53.
97. Chan KC, Woo JK, King A, et al. Analysis of Plasma Epstein–Barr Virus DNA to Screen for Nasopharyngeal Cancer. N Engl J Med 2017;377:513-22.
98. Mouliere F, Robert B, Peyrotte EA, et al. High Fragmentation Characterizes Tumour-Derived Circulating DNA. PLoS One 2011;6:e23418.
99. Mouliere F, El Messaoudi S, Gongora C, et al. Circulating Cell-Free DNA from Colorectal Cancer Patients May Reveal High KRAS or BRAF Mutation Load. Transl Oncol 2013;6:319-28.
100. Thierry AR, Molina F. Analytical methods for cell free nucleic acids and applications, WO/2012/028746. 2012. Available online: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2012028746&recNum=281&docAn=EP2011065333&queryString=(%2520&
101. Leszinski G, Lehner J, Gezer U, et al. Increased DNA Integrity in Colorectal Cancer. In Vivo 2014;28:299-303.
102. Jiang P, Chan CWM, Chan KC, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc Natl Acad Sci 2015;112:E1317-25.
103. Thierry A, El Messaoudi S. Methods for screening a subject for cancer, WO/2016/063122. 2016. Available online: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2016063122
104. Mandel P, Metais P. Les acides nucléiques du plasma sanguin chez l'homme. C R Seances Soc Biol Fil 1948;142:241-3.
105. Ashworth TR. A case of cancer in which cells similar to those in the tumours were seen in the blood after death. Australas Med J 1869;14:146-9.
106. Stroun M, Anker P, Lyautey J, et al. Isolation and characterization of DNA from the plasma of cancer patients. Eur J Cancer Clin Oncol 1987;23:707-12.
107. Leon SA, Shapiro B, Sklaroff DM, et al. Free DNA in the Serum of Cancer Patients and the Effect of Therapy. Cancer Res 1977;37:646-50.
108. National Cancer Institute. Liquid Biopsy: Using Tumor DNA in Blood to Aid Cancer Care. 2017.
109. Vlassov AV, Magdaleno S, Setterquist R, et al. Exosomes: Current knowledge of their composition, biological functions, and diagnostic and therapeutic potentials. Biochim Biophys Acta 2012;1820:940-8.
110. Wang M, Ji S, Shao G, et al. Effect of exosome biomarkers for diagnosis and prognosis of breast cancer patients. Clin Transl Oncol 2017. [Epub ahead of print].
111. Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science 2018;359:926-30.

Cite this article as: Tanos R, Thierry AR. Clinical relevance of liquid biopsy for cancer screening. Transl Cancer Res 2018;7(Suppl 2):S105-S129. doi: 10.21037/tcr.2018.01.31

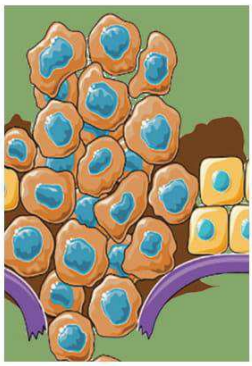
médecine/sciences 2018 ; 34 : 824-32

La biopsie liquide

Une voie possible pour le dépistage du cancer

Alain R. Thierry[1-4], Rita Tanos[1-4]

> La biopsie liquide est apparue comme une voie prometteuse pour le dépistage du cancer. En effet, plusieurs biomarqueurs comme les ADN circulants, les cellules tumorales circulantes, les micro-ARN circulants etc. se sont révélés prometteurs pour le théragnostic ou le suivi du patient. La détection précoce peut aider à réduire la mortalité associée au cancer et augmenter la survie globale des patients. La plupart des types de cancer manquent de biomarqueurs spécifiques et le développement de techniques de dépistage efficaces appliquées en clinique a été limité malgré des efforts intenses dans ce domaine. La nature non invasive de la biopsie liquide lui donne un avantage vis-à-vis d'autres méthodes, notamment pour le développement de tests de dépistage du cancer. Les différentes études fondées sur l'analyse de la biopsie liquide dans le but de développer des tests de dépistage et de détection précoce du cancer sont présentées dans cette revue. Bien qu'actuellement aucun test développé à partir de la biopsie liquide s'avère à la fois assez spécifique et sensible pour être utilisé comme test universel de dépistage, le potentiel de cette nouvelle approche apparaît de plus en plus crédible, eu égard aux récents développements de méthodes sophistiquées, notamment multiparamétriques. <



[1]IRCM, Institut de recherche en cancérologie de Montpellier, 208, avenue des Apothicaires, Montpellier, F-34298, France.
[2]Inserm U1194, Montpellier, F-34298, France.
[3]Université de Montpellier, Montpellier, F-34090, France.
[4]Institut régional du cancer de Montpellier, Montpellier, F-34298, France.
alain.thierry@inserm.fr

Le dépistage du cancer

La progression du cancer à des stades tardifs sans apparition de symptômes est l'une des principales raisons pour lesquelles il figure parmi les premières causes de décès dans le monde. Le dépistage par sa détection précoce, avant qu'une personne présente des signes de la maladie et lorsque le traitement est plus efficace, augmenterait les chances de rétablissement et la survie globale des patients. Il pourrait ainsi aider à réduire la mortalité liée au cancer [1].

Un test diagnostique diffère d'un test de dépistage par le fait qu'il est utilisé lorsqu'un sujet présente des signes ou des symptômes afin de déterminer la présence ou l'absence d'une maladie et qu'il est habituellement pratiqué après un test de dépistage positif pour établir un diagnostic définitif. Le développement de techniques de dépistage pour la détection précoce ne s'est pas avéré efficace pour de nombreux types de cancers, notamment en ce qui concerne la prévention de la mortalité liée à la maladie. Cela est dû à diverses raisons techniques et cliniques, la relative faible performance des tests en termes de sensibilité et de spécificité, l'hétérogénéité inter et intra-tumorale, ou à des facteurs épidémiologiques.

Comment évaluer un test de dépistage du cancer

En 1968, l'Organisation mondiale de la santé (OMS) a publié des directives, appelées souvent critères de Wilson [2], sur les principes et la pratique du dépistage de la maladie. En 2008, avec l'émergence de nouvelles technologies génomiques, l'Organisation a modifié ces directives avec la définition suivante [3] : « *Le programme de dépistage doit répondre à un besoin reconnu, et ses objectifs doivent être définis dès le départ. Il devrait y avoir une population cible définie et des preuves scientifiques de l'efficacité du programme de dépistage. Le programme devrait intégrer l'éducation, les tests, les services cliniques et la gestion de programme, ainsi que l'assurance de la qualité et des mécanismes pour minimiser les risques potentiels de dépistage. Il devrait garantir un choix informé, la confidentialité et le respect de l'autonomie, et promouvoir l'équité et l'accès au dépistage pour l'ensemble de la population cible. L'évaluation devrait être planifiée dès le départ et les avantages globaux du dépistage devraient l'emporter sur les dommages* ».

Vignette (Photo © Daniel Birnbaum).

m/s n° 10, vol. 34, octobre 2018
https://doi.org/10.1051/medsci/2018208

Un test de dépistage idéal du cancer permettrait donc d'établir une distinction parfaite entre les personnes atteintes et indemnes [1]. En pratique, un test de dépistage présente un certain pourcentage de faux positifs (des personnes détectées positives par le test alors qu'elles ne sont pas malades) et de faux négatifs (des personnes malades non détectées). La détermination de ces valeurs est soigneusement prise en compte lors de l'évaluation des avantages et des inconvénients du test, en considérant le rapport bénéfices/risques. Une détection précoce ne conduit pas toujours à de meilleurs résultats. Elle révèle parfois des faux positifs, ce qui crée des inquiétudes chez les personnes détectées et peut entraîner un traitement inutile. Le sur-diagnostic est ainsi remis en question, notamment pour le dépistage du cancer du sein. Le but du dépistage, quel que soit sa performance, doit consister à proposer uniquement un avantage pour la santé de l'individu.

Lorsqu'un nouveau test est développé, il est régulièrement comparé au test de référence, c'est-à-dire « le meilleur test disponible » qui consiste généralement en un test de diagnostic considéré comme définitif, comme une biopsie par exemple. Cette dernière est souvent coûteuse, invasive et potentiellement dangereuse, trop tardive ou peu pratique, pour être largement utilisée comme test de dépistage [1]. Un nouveau test devrait donc être non invasif, ou moins coûteux ou plus bénéfique, le rapport coûts/bénéfices étant au centre de l'évaluation par les administrations de santé publique.

La validité d'un test se traduit par une sensibilité (Se) élevée, qui est la capacité du test à identifier correctement les personnes qui ont la maladie, et une haute spécificité (Sp), reflétant sa capacité à identifier correctement celles qui ne sont pas atteintes. La courbe ROC (*receiver operating characteristic curve*) combine ces deux paramètres en représentant la Se en ordonnée et le complément de la Sp (ou 1-Sp) en abscisse. Elle met donc en rapport la proportion de vrais positifs (parmi les malades) avec la proportion de faux positifs (parmi les non-malades) pour toutes les valeurs-seuil du test [4]. L'aire sous la courbe ROC (AUC pour *area under curve*) est un indice simple et quantitatif de la performance du test et révèle sa capacité à discriminer entre malades et non-malades. Un test parfait présente une AUC égale à 1 ; pour un test sans valeur discriminante, l'aire sous la courbe sera de 0,5.

Une attention toute particulière doit être apportée aux cohortes des individus testés afin d'évaluer les performances d'un test de dépistage. Une étude en aveugle et une analyse intégrant des individus déjà diagnostiqués, ainsi que des personnes avant diagnostic dans des populations à risques, sont une nécessité.

Biomarqueurs et tests de dépistage conventionnels

Différents tests de dépistage et biomarqueurs sont utilisés en pratique pour identifier divers types de cancer *(Tableau I)*. Le test de frottis (*pap smear test*) est par exemple utilisé pour la dysplasie cervicale, ou le cancer du col de l'utérus. Il a une haute spécificité mais ne montre pas une sensibilité élevée [5]. La mammographie est le test le plus couramment utilisé pour le dépistage du cancer du sein. Ce test a joué un rôle clé dans la réduction de la mortalité par ce cancer, mais il présente une sensibilité limitée ainsi que de nombreux faux positifs et la possibilité de sur-diagnostic, d'où des opinions récentes assez contradictoires quant à son utilisation à une échelle de masse. Le dosage des marqueurs sériques, comme l'antigène tumoral CA 15-3 (*cancer antigen 15-3*), a également été appliqué pour le cancer du sein. Bien que très spécifique (Sp de 98 %), il a cependant montré une sensibilité moyenne (Se de 55,6 %)[6] ; il est donc préférentiellement utilisé pour le suivi de la réponse au traitement plutôt que pour le dépistage et le diagnostic précoce. Le test de détection de l'antigène prostatique spécifique (*prostate specific antigen,* PSA) aide à révéler un cancer de la prostate avant l'apparition des symptômes [7]. Un taux élevé de PSA ne résulte pas toujours de la présence d'un cancer et une polémique autour du dosage du PSA et le dépistage du cancer de la prostate est apparue en 2009. Deux études, une américaine [8] et une européenne [9], publiées dans le *New England Journal of Medicine* ont montré des résultats contradictoires sur la capacité de ce test de réduire le taux de mortalité ; cependant, en 2017, une ré-évaluation statistique des résultats de ces deux études a confirmé la diminution de la mortalité dans le groupe avec dépistage [10]. La tomodensitométrie à faible dose (*low-dose CT scan*) est généralement recommandée pour les personnes à risque élevé de développer un cancer du poumon, mais certaines tumeurs pourraient ne pas être détectées et d'autres pourraient se développer entre le dépistage et la détection. Pour le cancer colorectal (CCR), les deux glycoprotéines sériques les plus communément utilisées comme biomarqueurs, le CEA (*carcinoembryonic antigen*) et le CA19-9 (*carbohydrate antigen 19-9*), ne sont pas appropriées pour son dépistage en raison de leur faible sensibilité et spécificité, surtout pour les stades précoces (Se entre 40,9 % et 51,8 %, Sp entre 85,2 % et 95 %, pour le CEA [11]). Leurs mesures sont en fait plus appropriées pour le suivi de la récurrence ou de la réponse au traitement. D'autres tests de dépistage du CCR, fondés sur des analyses de selles par exemple, ont donc été développés. Le *hemoccult fecal occult blood test (FOBT)* est utilisé depuis longtemps pour détecter du sang dans les échantillons de selles comme indicateur précoce du CCR avec une sensibilité variant entre 12,9 % et 79,4 % et une spécificité de 86,7 % à 97,7 % [12]. Ce test présentant certains inconvénients, le *fecal immunochemical test (FIT)* est plus couramment utilisé en raison de son faible coût et de sa sensibilité globale de 79 % et sa spécificité de 94 % [12]. Le test de l'ADN fécal *(fecal DNA test)* [13] a permis la détection de cancers invasifs et d'adénomes avec une dysplasie de haut grade, avec une sensibilité de 40,8 % et une spécificité

de 94,4 % [14]. *Cologuard* a été le premier test d'ADN fécal commercialisé et approuvé par la *Food and drug administration* (FDA) en 2014 aux États-Unis. Son coût est plus élevé que les autres tests et sa spécificité inférieure, mais il présente une sensibilité plus élevée que le FIT, que ce soit pour le CCR ou la présence de polypes [15].

La biopsie liquide

Selon la définition du dictionnaire du NCI *(National Cancer Institute)*, la biopsie liquide est « *un test effectué sur un échantillon de sang pour la recherche de cellules cancéreuses ou de fragments d'ADN provenant d'une tumeur et circulant dans le sang* » [16]. Cette définition est, selon nous, plutôt imparfaite. Réunir sous le même terme de « biopsie liquide » une entité si complexe que sont des cellules et des macromolécules (ADN, ARN ou microARN) ne peut dépendre que de la nature liquide de leur prélèvement, de leur origine tumorale, et de l'information diagnostique qu'elles peuvent fournir en oncologie clinique. Comme indiqué dans le dictionnaire du NCI, « *une biopsie liquide peut être utilisée pour aider à trouver un cancer à un stade précoce, elle peut également être utilisée pour planifier un traitement ou déterminer si le traitement est efficace ou si le cancer a resurgi. Ainsi que permettre plusieurs prélèvements sanguins au fil du temps peut aussi aider les médecins à comprendre les changements moléculaires qui se produisent dans une tumeur* ». Malgré l'incohérence du terme regroupant ces différentes entités, nous utiliserons dans cette revue le terme « biopsie liquide » selon le dictionnaire du NCI et son utilisation conventionnelle dans la littérature. La terminologie biopsie liquide ne s'applique donc principalement qu'en oncologie, en miroir de la biopsie de tissu tumoral. Elle ne peut s'appliquer aux ADN circulants que l'on analyse dans d'autres domaines : diagnostic prénatal, inflammations sévères/aiguës (sepsis), transplantation, ou sport.

Les biopsies liquides ne sont pas limitées au sang. L'urine, la salive ou le liquide cervical peuvent également être utilisés, puisque l'information génétique est également présente dans ces fluides. La recherche sur les biopsies liquides ne s'est développée qu'au cours de la dernière décennie, générant un intérêt croissant en oncologie, et très récemment pour le dépistage ou la détection précoce du cancer.

L'ADN circulant

Pour la première fois, il y a plus de 70 ans, des chercheurs ont montré que les tumeurs libéraient des acides nucléiques dans les fluides corporels [17]. Plus récemment, les ADN circulants (ou ADNcir, pour *circulating cell free DNA)* se sont avérés porter une partie de la même information moléculaire que les biopsies tissulaires [18, 19]. L'ADNcir est ainsi apparu comme un biomarqueur potentiel particulièrement dans le cancer et il fait l'objet d'études en recherche translationnelle et clinique [19]. Ce biomarqueur permettrait de diagnostiquer, de surveiller la récurrence et d'évaluer la réponse à la thérapie, uniquement par un prélèvement sanguin. Une analyse qualitative de l'ADNcir permettrait, par exemple, la détection non invasive d'altérations géniques liées à différents types de cancer, comme les mutations touchant les gènes *KRAS* (*Kirsten rat sarcoma viral oncogene homolog*) dans le cancer colorectal, et *EGFR* (*epidermal growth factor receptor*) dans le cancer des poumons. Une analyse quantitative de son taux circulant pourrait renseigner sur la présence ou l'absence d'un cancer, suivre le développement de la tumeur ou être le signe d'une éventuelle rechute. Le profil de méthylation de l'ADNcir est également étudié afin de déterminer le tissu dont il est originaire, ainsi que sa taille et sa fragmentation, pour caractériser sa structure et ses mécanismes de libération. Plusieurs limites de l'analyse de l'ADNcir réduisent l'intérêt de son utilisation en clinique. L'une des contraintes majeures de ce type de détection est certainement le manque de procédure standardisée (SOP, pour *standard operating procedure*), notamment en ce qui concerne le contrôle de la phase pré-analytique, en particulier le contrôle des différents paramètres intervenant depuis le prélèvement sanguin jusqu'à l'analyse de l'ADNcir [20]. Une grande majorité des méthodes de détection et d'analyse de l'ADNcir ne se sont pas avérées assez sensibles en raison de la rareté des ADNcir mutés présents dans l'échantillon : une cellule tumorale portant la mutation ne libère qu'un seul génome dans la circulation et donc qu'un (ou très rarement deux) allèle muté ; les ADNcir mutés peuvent être en très grande minorité, compte tenu de la fréquente disproportion entre cellules malignes (mutées) et cellules du microenvironnement tumoral (cellules du stroma, endothéliales et lymphocytaires). Toutefois, les efforts de développement technologique récents permettent d'entrevoir des solutions envisageables cliniquement. Les administrations européennes de santé ont déjà autorisé en oncologie thoracique, l'analyse de l'ADN tumoral circulant (mutations du gène *EGFR*) comme une alternative à la biopsie tissulaire, uniquement lorsque celle-ci est impossible à réaliser ou insuffisante. D'autres tests compagnons aux thérapies ciblées utilisant l'examen de l'ADNcir pourraient donc être autorisés prochainement.

Le potentiel de ce biomarqueur pour le dépistage précoce du cancer et des altérations qualitatives et quantitatives de cet ADN a fait l'objet de nombreuses études [21]. Cependant, malgré cette recherche intense, peu de tests ont été implémentés en pratique clinique. Des données contradictoires concernant la concentration totale d'ADNcir nucléaire ont en effet limité la mise au point et l'utilisation de ces tests en clinique : alors que les concentrations plasmatiques d'ADNcir chez les patients cancéreux varient de quelques ng à plusieurs milliers par millilitre, un chevauchement avec la concentration mesurée chez les individus sains a été retrouvé [21, 22].

Un test de dépistage fondé sur l'ADNcir devrait permettre de distinguer les processus non cancéreux et précancéreux de ceux de la malignité invasive afin d'atteindre une sensibilité clinique élevée. L'utilité d'un tel test résulterait également de sa capacité de fournir des informations sur l'origine tissulaire de l'ADNcir, ce qui devrait être possible étant donné les différences d'altérations somatiques existant entre les différents types de tumeurs.

Jusqu'à présent, plusieurs groupes ont travaillé sur le développement de tests pour le dépistage et le diagnostic précoce de plusieurs types de cancer à partir d'une seule analyse de sang *(Tableau I)*. Ainsi, pour le dépistage du cancer colorectal par exemple, plusieurs études ont porté sur la détection dans le plasma du gène méthylé de la Septine 9, une protéine de la famille des protéines liant le GTP (*GTP-binding protein*) dont le taux est significativement plus élevé chez les patients atteints de cancer colorectal par rapport à celui observé chez les individus sains, ce qui en fait un biomarqueur potentiel de ce cancer [23]. L'*Epi proColon*® (*Epigenomics AG Corporation*, Berlin, Allemagne), fondé sur la détection dans le sang par PCR (*polymerase chain reaction*) en temps réel du gène codant la Septine 9 méthylé est le seul test de dépistage sanguin de l'hyperméthylation d'ADN pour le CCR [24]. Il montre jusqu'à présent les meilleurs résultats parmi les tests commerciaux de détection du cancer par analyse sanguine, permettant de distinguer les patients atteints de CCR des individus sains avec une sensibilité de 75-81 % et une spécificité de 96-99 %.

D'autres gènes ou régions de l'ADNcir ont également été examinés afin de déterminer la pertinence de leur détection pour un diagnostic précoce. Ainsi, différentes régions promotrices hyperméthylées de certains gènes identifiés comme spécifiques au cancer colorectal ont été examinées. Les AUC (aire sous la courbe ROC) les plus élevées (0,85) ont été obtenues pour un test combinant la détection de sept promoteurs, après normalisation de l'âge et du sexe [25]. Une combinaison de quatre marqueurs de nucléosomes circulants a été développée par *Volition* et a fourni une AUC de 0,87 (0,97 pour le même panel avec l'ajustement de l'âge) pour discriminer patients atteints de CCR et témoins sains, avec une sensibilité élevée pour les stades précoces (75 et 86 % respectivement pour les stades I et II, pour une spécificité de 90 %). Une deuxième combinaison de quatre autres biomarqueurs de nucléosomes circulants a fourni une AUC de 0,72 pour discriminer patients présentant des polypes et individus sains [26], montrant ainsi une application potentielle de ce test dans le dépistage du CCR.

Le potentiel diagnostic de l'ADNcir et son taux sanguin ont également été examinés dans d'autres types de cancer. Ainsi, dans le cas du cancer du poumon, une analyse quantitative a montré des taux accrus d'ADNcir chez les patients cancéreux, comparés aux individus sains, avec une valeur d'AUC d'environ 0,88 [27]. Des résultats similaires ont été observés pour le cancer du sein avec 78 % de sensibilité et 83 % de spécificité [28], ainsi que le cancer de l'ovaire [29] avec une sensibilité de 70 % et une spécificité de 90 %.

L'ADN circulant est constitué non seulement d'ADN d'origine nucléaire mais aussi d'ADN mitochondrial (ADNmt), dont certaines altérations ont été associées à un dysfonctionnement mitochondrial et au développement d'un cancer. La signification clinique des niveaux d'ADN mitochondrial dans le sang périphérique et de son intégrité a été notamment étudiée pour différents types de cancers tels que le poumon, le sein, le cancer colorectal, le lymphome non hodgkinien [30, 31]. Les données restent actuellement discordantes et aucune conclusion ne peut être donnée. L'absence d'études pré-analytiques et analytiques concernant l'ADN mitochondrial circulant, qui est mal caractérisé, notamment ses propriétés structurales, pourrait expliquer en partie ces discordances.

Les cellules tumorales circulantes

La remarquable découverte dans la circulation sanguine de cellules libérées ou s'échappant de la tumeur par des chercheurs australiens il y a plus d'un siècle [32] a été à l'origine d'intenses recherches au cours de la dernière décennie. Les cellules tumorales circulantes (CTC) sont cependant difficiles à isoler et ne sont pas toujours le reflet des cellules génétiquement cancéreuses. La valeur de ces CTC a également été évaluée dans plusieurs études pour le dépistage de différents types de cancers *(Tableau I)*. Dans le cancer du poumon, Tanaka *et al.* ont montré que le dénombrement des CTC ne présentait pas un potentiel discriminant adéquat entre patients cancéreux et individus sains (avec une AUC de 0,598) [33]. D'autres groupes ont déterminé en revanche que la détection d'un nombre de CTC supérieur à 25 présentait une sensibilité (89 %) et une spécificité élevées (100 %) pour distinguer maladie bénigne et maligne [34]. Les CTC ont également été détectées chez des patients atteints de bronchopneumopathie chronique obstructive (BPCO), un facteur de risque du cancer du poumon, sans présenter de signes cliniques de cancer [35]. Wang *et al.* ont évalué la valeur des CTC pour le diagnostic précoce du cancer du poumon : une AUC de près de 0,80 (Se = 77,7 % et Sp = 89,5 %) [36] a été mesurée pour un seuil de 8,7 unités de CTC exprimant le récepteur du folate (ou vitamine B9) entre le groupe témoin et les patients atteints.

La technologie permettant d'isoler et de phénotyper les CTC a rapidement progressé. Néanmoins, la complexité de leur analyse et le faible signal analytique limitent leur utilité clinique par rapport aux méthodes fondées sur l'ADNcir tumoral [37]. En effet, lorsque les deux sont présents, les fragments d'ADNcir tumoral sont plus nombreux (d'un facteur de 50) que les CTC [37]. Ils apportent donc un signal analytique bien supérieur. Les CTC ne présentent en revanche pas le désavantage de l'ADNcir tumoral de devoir mesurer des quantités très faibles de fragments mutés dans le plasma, en raison

Catégorie		Test	Cancer	Sensibilité	Spécificité	AUC	Réf
Biomarqueurs et tests conventionnels utilisés		*CEA*	Colorectal	41-52 %	85-95 %	0,63-0,77	[11]
		CA19-9	Colorectal/ Pancréas	23 %	95 %	0,64	[11]
		FOBT	Colorectal	13-79 %	87-98 %	-	[12]
		FIT	Colorectal	79 %	94 %	-	[12]
		Cologuard (test d'ADN fécal)	Colorectal	41 %	94 %	-	[13]
		CA 15-3	Sein	55,6 %	98 %	-	[6]
		PHI (prostate health index)	Prostate	-	-	0,69-0,77	[54]
		PSA	Prostate	-	-	0,55	[7]
Biomarqueurs circulants en cours de développement	ADN circulant	*Epi ProColon* (détection du gène de la Septine 9 méthylé)	Colorectal	75-81 %	96-99 %	-	[23,24]
		7 promoteurs méthylés	Colorectal	91 %	73 %	0,85	[25]
		Volition (marqueurs de nucléosomes circulants)	Colorectal	75 %	86-90 %	0,97	[26]
			Poumon			0,88	[27]
		Concentration totale d'ADNcir nucléaire	Sein	78 %	83 %	0,91	[28]
			Ovaire	70 %	90 %	0,89	[29]
			Colorectal	-	-	0,91	[22]
		Détection du virus d'Epstein-Barr	Nasopharynx	97,1 %	98,6 %	-	[49]
		CancerSeek	Divers	30-99 %	99 %	0,91	[48]
	CTC	Numération	Poumon	30-78 %	88-100 %	0,60-0,86	[33,34,36]
	miARN circulants	Panel	Sein	83 %	41 %	0,67	[41]
			Poumon et mésothélium	73-75 %	54-79 %	-	[42]
	Autres	Combinaison de biomarqueurs protéiques circulants	Sein	-	-	0,70-0,99	[44]

Tableau I. ***Récapitulatif de différents tests de dépistage du cancer et de leurs performances actuelles.*** La sensibilité, la spécificité et l'AUC représentent les paramètres de performance.

de la libération importante d'ADNcir non muté issu majoritairement de l'ADN des cellules hématopoïétiques. Dans un essai récent réalisé sur des patients atteints d'un cancer du poumon, l'analyse de l'ADNcir tumoral a surpassé l'examen des CTC pour la détection de la mutation du gène *KRAS*, révélant des sensibilités respectives de 96 % et de 52 % [38].

Très récemment, un engouement certain s'est porté sur l'analyse de la cellule unique (*single cell analysis*). Techniquement réalisable, cette approche peut s'avérer d'importance pour le dépistage du cancer. Les études sont cependant peu nombreuses et les résultats peu convaincants pour l'instant [39]. Outre la paucité du nombre de CTC dans le sang, l'un des inconvénients majeurs de leur analyse est la nécessité de leur préparation immédiate juste après le prélèvement (dans la demi-journée), alors que le délai pour l'analyse d'ADNcir peut atteindre 5 jours si le sang est prélevé sur stabilisant. Les tests utilisant les CTC sont néanmoins plus liés à la terminologie de la biopsie liquide et, intrinsèquement, plus informatifs puisque la détermination des marqueurs cellulaires peut être combinée à l'information génétique dans un même échantillon.

Les autres biomarqueurs circulants

La pertinence diagnostique des micro-ARN circulants (miARN) dans le sang a été examinée chez des patients atteints de différents types de cancer *(Tableau I)*. Dans le sang de patientes atteintes d'un cancer du sein, les miARN circulants associés aux tumeurs sont élevés et liés à la progression tumorale [40] et une signature multivariée de neuf miARN circulants a été validée et a permis de discriminer les patientes atteintes des femmes saines avec une AUC de 0,665 [41]. D'autres signatures de miARN circulants ont été identifiées pour le diagnostic précoce des tumeurs malignes pulmonaires [42].

La présence de macrophages circulants associés à la tumeur pourrait être utile comme outil de dépistage du cancer du sein et permettrait de différencier les patientes atteintes d'une maladie maligne de celles présentant une affection mammaire bénigne ou saines [43]. Dans ce cancer, la combinaison de plusieurs biomarqueurs circulants protéiques a aussi été évaluée et a montré des valeurs d'AUC variant entre 0,7 et 0,99 selon le sous-type de cancer [44].

Les plaquettes éduquées par la tumeur *(tumor-educated platelets,* TEP*)* sont un autre biomarqueur circulant. Les profils d'ARN messagers de ces TEP permettent en effet de révéler les patients cancéreux [45].

Dépistage du cancer et analyse de l'ADN circulant : vers de nouvelles possibilités

Profil des altérations génétiques

Les biopsies liquides ont le potentiel de permettre d'identifier les patients dont les tumeurs présentent des mutations spécifiques. Velculescu *et al.* ont récemment évalué les méthodes de séquençage de nouvelle génération *(Next-Gen sequencing)* appliquées à l'ADN circulant sur 138 patients atteints de tumeurs précoces. L'analyse a aussi été utilisée pour détecter la rechute après traitement. En utilisant un séquençage de type *TEC-Seq (targeted error correction sequencing)*, le type de cancer a pu être identifié chez plus de la moitié des patients [46] : cinquante-huit gènes sont typiquement associés aux cancers du sein, du poumon, de l'ovaire et colorectal. Sur les 138 cancers examinés, 86 cancers de stade I et II ont été détectés et l'analyse des gènes, séquencés chez 100 patients, a montré les mêmes mutations dans les échantillons de sang et dans les échantillons de tissus tumoraux chez 82 patients. Aucun des patients sains n'a montré de mutations. Cette étude présente quelques limites avec la difficulté d'identifier l'ADN rare provenant de la tumeur et de révéler d'autres altérations génétiques ou mutations héréditaires ou somatiques.

Des mutations du gène *TP53* ont été retrouvées dans les échantillons sanguins de 49 % des patients atteints du cancer de poumon à petites cellules *(small cell lung cancer,* SCLC*)* avec des formes alléliques significativement plus élevées chez les patients cancéreux [47].

Association du profil des altérations génétiques avec des biomarqueurs protéiques

En janvier 2018, un nouveau test sanguin pour le dépistage du cancer, appelé *CancerSEEK*, a reçu beaucoup d'attention [48]. L'un des éléments les plus innovants de l'étude réalisée sur le plasma de 1 005 patients ayant reçu un diagnostic de cancer sans propagation au moment de l'inclusion et de 812 individus sains, était l'utilisation d'une combinaison de marqueurs génétiques et protéiques. *CancerSEEK* concernait le séquençage ciblé de seize gènes, associé à huit biomarqueurs protéiques caractéristiques d'un type spécifique de cancer, qui non seulement identifient la présence de cancers relativement précoces mais aussi contribuent à localiser l'organe d'origine de ces cancers. *CancerSEEK* a permis de détecter 98 % des cancers de l'ovaire et du foie, entre 60 % et 70 % des cancers de l'estomac, du pancréas, de l'œsophage, du côlon et du poumon, et 33 % des cancers du sein. Bien que la performance de ce test sanguin reste actuellement inégalée, il présente des limitations dans un contexte de dépistage universel du cancer. *CancerSEEK* apparaît comme une confirmation d'un diagnostic au stade du pré-dépistage. Pour le qualifier de test de dépistage, il aurait dû être réalisé sur une population saine, avec potentiellement des patients asymptomatiques (➜). Néanmoins, *CancerSEEK* se montre intéressant pour la détection précoce et le suivi des tumeurs à partir d'un échantillon de sang. Son développement est clairement un pas de plus vers un test de dépistage universel.

(➜) Voir la Chronique de B. Jordan, *m/s* n° 4, avril 2018, page 363

Détection de génome viral

Plutôt que d'analyser les cellules cancéreuses elles-mêmes, Lo *et al.* ont récemment décrit une étude visant à détecter le virus d'Epstein-Barr, impliqué dans la plupart des cas de cancer du nasopharynx, par la recherche de l'ADN viral libéré dans le sang par les cellules tumorales [49]. L'ADN viral a été retrouvé dans le sérum de 1 112 individus d'une cohorte de 20 000 hommes (soit

5,5 %) ; 309 patients ont vu la confirmation de leur diagnostic un mois plus tard, et pour 34 d'entre eux, un cancer a été révélé après examen endoscopique et imagerie par résonance magnétique (IRM) ; une seule personne détectée négative au dépistage a développé un cancer en moins d'un an. Le test a détecté plus de cas positifs pour les stades les plus précoces de la maladie.

Fragmentation de l'ADN circulant

Nous avons observé que les fragments d'ADN courts étaient plus abondants dans le plasma de patients atteints de cancer colorectal que chez des individus sains. La quantité de fragments d'ADN circulants de moins de 145 paires de bases (pb) est également directement liée à la concentration d'ADN tumoral circulant [50]. L'ADNcir muté provenant des cellules tumorales est en fait très fragmenté comparé à l'ADN cirulant non muté [22] : la détection de l'ADN tumoral circulant repose sur des amplicons inférieurs à 100 pb et l'analyse par microscopie à force atomique révèle des fragments d'ADNcir en moyenne de 135 pb chez les patients cancéreux [22]. Une forte discrimination entre cancers colorectaux de stade IV et individus sains a pu être observée en ciblant un petit amplicon (de 63 pb) [22]. Ces observations ont été confirmées par Leszinski *et al.* qui ont montré que l'ADNcir était significativement plus fragmenté chez les patients atteints de cancer colorectal que chez les individus sains ou atteints de maladies colorectales bénignes [51]. Plusieurs indices d'intégrité de l'ADN ont été évalués, avec de très diverses efficacités pour la discrimination des patients sains et cancéreux sans doute en raison du manque de procédures opératoires standards et du nombre de patients testés. Un pouvoir de dépistage avancé pour un indice d'intégrité d'ADN spécifique, déterminé par PCR (*polymerase chain reaction*) quantitative, a pu être obtenu. La stratégie utilisée est fondée sur la différence entre structure de l'ADNcir dérivant de cellules malignes et celle de l'ADNcir issu de cellules saines. Elle facilite la mise en œuvre de l'analyse et réduit les coûts des tests de dépistage. Des travaux sur le profil de taille de l'ADNcir sont en cours pour définir des approches de q-PCR et de séquençage massif en vue de son inclusion dans un test sanguin de dépistage du cancer.

Un test fondé sur l'ADN circulant mitochondrial : le *MiTest*

La quantité d'ADNcir dans le sang peut être un facteur discriminatif entre sujets sains et patients atteints de CCR [22], avec une AUC de 0,91. La quantification et l'association de l'ADN mitochondrial circulant et nucléaire circulant permettent une telle distinction [52]. Nous avons déterminé un indice fondé sur la détection de séquences particulières dans le génome nucléaire et mitochondrial qui, appliqué à des surnageants de cultures cellulaires, montre une différence significative entre lignées cellulaires normales et cancéreuses. Dans le plasma, cet indice a une valeur statistiquement plus faible chez les patients atteints de cancer colorectal que chez les sujets sains. La mesure de cet indice, qui constitue ce que nous avons nommé le *Mitest*, constitue donc un test de dépistage puissant pour la détection précoce du cancer. Des études sont en cours afin de combiner ce nouveau marqueur avec d'autres paramètres pour augmenter le potentiel discriminant du test au sein de larges cohortes de patients et d'individus sains.

Conclusion

La biopsie liquide présente des avantages prometteurs pour le dépistage du cancer : sa nature minimalement invasive, sans le délai, le coût et les risques associés à la biopsie tissulaire, ainsi que la possibilité de détection à un stade précoce avant la détectabilité radiologique. Plusieurs biomarqueurs circulants sont actuellement évalués pour être validés en ce qui concerne leur utilité clinique (l'ADN circulant, les cellules tumorales circulantes, les microARN circulants et autres) pour développer des tests de détection précoce du cancer. Les exosomes, qui contiennent des acides nucléiques et des protéines, pourraient également être une source de multiples marqueurs de malignité dont l'analyse pourrait être prometteuse pour le développement de méthodes de dépistage [53]. Mais peu de ces biomarqueurs ont été à ce jour validés pour une utilisation clinique. L'ADNcir d'origine nucléaire et mitochondriale semble avoir un avantage dans le dépistage du cancer comparativement à d'autres biomarqueurs en montrant une meilleure efficacité et semble posséder les caractéristiques pour que sa détection soit mise en œuvre plus rapidement.

Il est à craindre ou à souhaiter que l'utilisation, à travers le monde, d'un test de dépistage se distingue de deux façons : (1) qu'il soit approuvé par les administrations de santé publique et remboursé, à la suite d'une étude statistique longue et rigoureuse, et (2) qu'il soit mis à disposition individuellement (par exemple, test de grossesse) avec un niveau de performance ou d'évaluation modéré et qu'il soit proposé dans un proche futur.

La standardisation des paramètres pré-analytiques et une meilleure connaissance de l'origine et de la structure exacte de l'ADNcir constitueraient l'étape supplémentaire de sa mise en œuvre. L'avancement des méthodes sophistiquées de q-PCR ou de séquençage *Next-Gen* améliorera inévitablement la fiabilité de la performance analytique des futurs tests. Comme indiqué par Dennis Lo de l'Université de Hong Kong : « *Brique par brique, nous sommes en train de mettre en place cette technologie* ». À l'avenir, il semble, au vu des résultats prometteurs, que les biopsies liquides pourraient ajouter une nouvelle dimension au rôle du médecin eu égard au dépistage et au diagnostic. Pour le moins, l'étude de l'évaluation de la performance de dépistage de la biopsie liquide en tandem avec d'autres tests, comme l'IRM, semble justifiée. Par ailleurs, il est envisageable que des tests de biopsie liquide puissent être utilisés pour dépister le cancer à un stade précoce chez les individus à haut risque, tels que ceux atteints de cancers familiaux. Ainsi, il est crucial de poursuivre

l'étude de ces biomarqueurs émergents. L'utilisation combinée de plusieurs marqueurs de l'ADN circulant ou bien de biomarqueurs conventionnels apparait être un moyen pour augmenter la sensibilité et la spécificité des tests et donc obtenir des résultats plus significatifs dans la détection précoce des cancers. Les années à venir semblent être excitantes alors que le « sacré graal » en cancérologie que constitue le dépistage universel apparaît être désormais accessible. ◊

SUMMARY

Liquid biopsy: a possible approach for cancer screening

Liquid biopsy has emerged as a promising avenue for cancer screening. Several circulating biomarkers such as circulating DNA, circulating tumor cells, circulating microRNAs and others have shown promise for theragnostics and patient's monitoring. Early detection may help reduce cancer-related mortality and increase overall patient survival. Most cancer types lack specific biomarkers and despite intensive efforts in this area, the development of effective clinical screening techniques has been limited. The noninvasive nature of liquid biopsy represents an advantage over other approaches to define cancer biomarkers, particularly for the development of cancer screening tests. This review presents the various studies based on the analysis of liquid biopsy aiming to develop tests for cancer screening and early detection. So far, no test developed from liquid biopsy proves to be both specific and sensitive enough to be used as a universal screening test. However, the potential of this new approach appears more and more credible, given the recent developments of sophisticated multi-parametric methods. ◊

REMERCIEMENTS

Nous remercions Safia El Messaoudi, Sandrine Bonizec et Amaelle Otandault pour leur aide. Nous remercions également Marc Ychou, Denis Pezet et Muriel Mathonnet pour avoir fourni les échantillons sanguins. Ce travail a été soutenu par l'Inserm, Lilly (France) et la bourse « INCa-DGOS-Inserm 6045 » du SIRIC de Montpellier, France.

LIENS D'INTÉRÊT

A.R. Thierry possède des actions de DiaDx SAS. R. Tanos déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

1. Maxim LD, Niebo R, Utell MJ. Screening tests: a review with examples. *Inhal Toxicol* 2014 ; 26 : 811-28.
2. Wilson JMG, Jungner G. *Principles and practice of screening for disease*. Geneva : World Health Organisation, 1968 : 164 p.
3. Andermann A, Blancquaert I, Beauchamp S, *et al.* Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ* 2008 ; 86 : 317-19.
4. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Casp J Intern Med* 2013 ; 4 : 627-35.
5. Mayrand MH, Duarte-Franco E, Rodrigues I, *et al.* Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. *N Engl J Med* 2007 ; 357 : 1579-88.
6. Stieber P, Nagel D, Blankenburg I, *et al.* Diagnostic efficacy of CA 15-3 and CEA in the early detection of metastatic breast cancer: a retrospective analysis of kinetics on 743 breast cancer patients. *Clin Chim Acta* 2015 ; 448 : 228-31.
7. Catalona WJ, Smith DS, Ratliff TL, *et al.* Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N Engl J Med* 1991 ; 324 : 1156-61.
8. Andriole GL, Crawford ED, Grubb RL, *et al.* Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* 2009 ; 360 : 1310-19.
9. Schröder FH, Hugosson J, Roobol MJ, *et al.* Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 2009 ; 360 : 1320-28.
10. Tsodikov A, Gulati R, Heijnsdijk EAM, *et al.* Reconciling the effects of screening on prostate cancer mortality in the ERSPC and PLCO trials. *Ann Intern Med* 2017 ; 167 : 449-55.
11. Chen JS, Chen KT, Fan WC, *et al.* Combined analysis of survivin autoantibody and carcinoembryonic antigen biomarkers for improved detection of colorectal cancer. *Clin Chem Lab Med* 2010 ; 48 : 719-25.
12. Song LL, Li YM. Current noninvasive tests for colorectal cancer screening: an overview of colorectal cancer screening tests. *World J Gastrointest Oncol* 2016 ; 8 : 793-800.
13. Dhaliwal A, Vlachostergios PJ, Oikonomou KG, *et al.* Fecal DNA testing for colorectal cancer screening: molecular targets and perspectives. *World J Gastrointest Oncol* 2015 ; 7 : 178-83.
14. Imperiale TF, Ransohoff DF, Itzkowitz SH, *et al.* Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med* 2004 ; 351 : 2704-14.
15. Imperiale TF, Ransohoff DF, Itzkowitz SH, *et al.* Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med* 2014 ; 370 : 1287-97.
16. NCI dictionary of cancer terms *Natl Cancer Inst* nd. https://www.cancer.gov/publications/dictionaries/cancer-terms
17. Mandel P, Metais P. Les acides nucléiques du plasma sanguin chez l'homme. *CR Seances Soc Biol Fil* 1948 ; 142 : 241-3.
18. Stroun M, Anker P, Lyautey J, *et al.* Isolation and characterization of DNA from the plasma of cancer patients. *Eur J Cancer Clin Oncol* 1987 ; 23 : 707-12.
19. Thierry AR, Messaoudi S El, Gahan PB, *et al.* Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev* 2016 ; 35 : 347-76.
20. Ilie M, Hofman V, Long E, *et al.* Current challenges for detection of circulating tumor cells and cell-free circulating nucleic acids, and their characterization in non-small cell lung carcinoma patients. What is the best blood substrate for personalized medicine? *Ann Transl Med* 2014 ; 2 : 107.
21. Wan JCM, Massie C, Garcia-Corbacho J, *et al.* Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* 2017 ; 17 : 223-38.
22. Mouliere F, Messaoudi S El, Pang D, *et al.* Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol Oncol* 2014 ; 8 : 927-41.
23. Church TR, Wandell M, Lofton-Day C, *et al.* Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut* 2014 ; 63 : 317-25.
24. Lamb YN, Dhillon S. Epi proColon® 2.0 CE: a blood-based screening test for colorectal cancer. *Mol Diagn Ther* 2017 ; 21 : 225-32.
25. Rasmussen SL, Krarup HB, Sunesen KG, *et al.* Hypermethylated DNA, a circulating biomarker for colorectal cancer detection. *PLoS One* 2017 ; 12 : e0180809.
26. Rahier JF, Druez A, Faugeras L, *et al.* Circulating nucleosomes as new blood-based biomarkers for detection of colorectal cancer. *Clin Epigenetics* 2017 ; 9 : 53.
27. Jiang T, Zhai C, Su C, *et al.* The diagnostic value of circulating cell free DNA quantification in non-small cell lung cancer: a systematic review with meta-analysis. *Lung Cancer* 2016 ; 100 : 63-70.
28. Lin Z, Neiswender J, Fang B, *et al.* Value of circulating cell-free DNA analysis as a diagnostic tool for breast cancer: a meta-analysis. *Oncotarget* 2017 ; 8 : 26625-36.
29. Zhou Q, Li W, Leng B, *et al.* Circulating cell free DNA as the diagnostic marker for ovarian cancer: a systematic review and meta-analysis. *PLoS One* 2016 ; 11 : e0155495.
30. Hosgood HD, Liu CS, Rothman N, *et al.* Mitochondrial DNA copy number and lung cancer risk in a prospective cohort study. *Carcinogenesis* 2010 ; 31 : 847-49.
31. Thyagarajan B, Wang R, Barcelo H, *et al.* Mitochondrial copy number is associated with colorectal cancer risk. *Cancer Epidemiol Prev Biomarkers* 2012 ; 21 : 1574-81.
32. Ashworth TR. A case of cancer in which cells similar to those in the tumours were seen in the blood after death. *Australasian Med J* 1869 ; 14 : 146-7.
33. Tanaka F, Yoneda K, Kondo N, *et al.* Circulating tumor cell as a diagnostic marker in primary lung cancer. *Clin Cancer Res* 2009 ; 15 : 6980-86.
34. Fiorelli A, Accardo M, Carelli E, *et al.* Circulating tumor cells in diagnosing lung cancer: clinical and morphologic analysis. *Ann Thorac Surg* 2015 ; 99 : 1899-905.
35. Ilie M, Hofman V, Long-Mira E, *et al.* Sentinel circulating tumor cells allow early diagnosis of lung cancer in patients with chronic obstructive pulmonary disease. *PLoS One* 2014 ; 9 : e111597.

RÉFÉRENCES

36. Wang L, Wu C, Qiao L, *et al.* Clinical significance of folate receptor-positive circulating tumor cells detected by ligand-targeted polymerase chain reaction in lung cancer. *J Cancer* 2017 ; 8 : 104-10.

37. Diaz LA, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol* 2014 ; 32 : 579-86.

38. Freidin MB, Freydina D V, Leung M, *et al.* Circulating tumor DNA outperforms circulating tumor cells for KRAS mutation detection in thoracic malignancies. *Clin Chem* 2015 ; 61 : 1299-304.

39. Alix-Panabières C, Pantel K. Characterization of single circulating tumor cells. *FEBS Lett* 2017 ; 591 : 2241-50.

40. Roth C, Rack B, Müller V, *et al.* Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. *Breast Cancer Res* 2010 ; 12 : R90.

41. Kodahl AR, Lyng MB, Binder H, *et al.* Novel circulating microRNA signature as a potential non-invasive multi-marker test in ER-positive early-stage breast cancer: a case control study. *Mol Oncol* 2014 ; 8 : 874-83.

42. Tomasetti M, Amati M, Neuzil J, *et al.* Circulating epigenetic biomarkers in lung malignancies: from early diagnosis to therapy. *Lung Cancer* 2017 ; 107 : 65-72.

43. Adams DL, Adams DK, Alpaugh RK, *et al.* Circulating cancer-associated macrophage-like cells differentiate malignant breast cancer and benign breast conditions. *Cancer Epidemiol Prev Biomarkers* 2016 ; 25 : 1037-42.

44. Gonzalez RM, Daly DS, Tan R, *et al.* Plasma biomarker profiles differ depending on breast cancer subtype but RANTES is consistently increased. *Cancer Epidemiol Biomarkers Prev* 2011 ; 20 : 1543-51.

45. Sol N, Wurdinger T. Platelet RNA signatures for the detection of cancer. *Cancer Metastasis Rev* 2017 ; 1-10.

46. Phallen J, Sausen M, Adleff V, *et al.* Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med* 2017 ; 9 : eaan2415.

47. Fernandez-Cuesta L, Perdomo S, Avogbe PH, *et al.* Identification of circulating tumor DNA for the early detection of small-cell lung cancer. *EBioMedicine* 2016 ; 10 : 117-23.

48. Cohen JD, Li L, Wang Y, *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 2018 ; eaar3247.

49. Chan KCA, Woo JKS, King A, *et al.* Analysis of plasma Epstein–Barr virus DNA to screen for nasopharyngeal cancer. *N Engl J Med* 2017 ; 377 : 513-22.

50. Mouliere F, Robert B, Peyrotte EA, *et al.* High fragmentation characterizes tumour-derived circulating DNA. *PLoS One* 2011 ; 6 : e23418.

51. Leszinski G, Lehner J, Gezer U, *et al.* Increased DNA integrity in colorectal cancer. *In Vivo (Brooklyn)* 2014 ; 28 : 299-303.

52. Thierry A, Messaoudi SE. *Methods for screening a subject for cancer.* WO2016/063122.

53. Wang M, Ji S, Shao G, *et al.* Effect of exosome biomarkers for diagnosis and prognosis of breast cancer patients. *Clin Transl Oncol* 2017 ; 1-6.

54. Loeb S, Catalona WJ. The prostate health index: a new test for the detection of prostate cancer. *Ther Adv Urol* 2014 ; 6 : 74-7.

55. Jordan B. Dépister les cancers asymptotiques ? *Med Sci (Paris)* 2018 ; 34 : 363-5.

TIRÉS À PART

A.R. Thierry

D. Projet de thèse partie 1 : une nouvelle approche pour le dépistage et la détection précoce du cancer

Résumé

L'ADN circulant (ADNcir) est un nouveau biomarqueur en oncologie et plusieurs études ont récemment cherché à comprendre son rôle et son intérêt pour le dépistage et la détection précoce du cancer. Nous avons identifié deux paramètres quantitatifs de l'ADNcir, le Ref A 67 (concentration totale d'ADNcir nucléaire) et le MNR (Rapport de la concentration entre l'ADNcir mitochondrial et nucléaiare), déterminés par q-PCR en ciblant des séquences nucléaires et mitochondriales. Ces paramètres ont été validés dans trois modèles expérimentaux : modèle murin de souris xénogreffées, le milieu de culture des cellules et dans le plasma d'une petite cohorte d'individus sains et cancéreux. Ces deux paramètres, ainsi que d'autres paramètres quantitatifs et structurels de l'ADNcir ont été évalués dans une large cohorte rétrospective de 289 individus sains, 99 individus à risque de cancer colorectal (CCR) et 983 patients atteints de CCR (n = 791), de cancer du sein (n = 169) et d'autres cancers (hépatocellulaire, pancréatique, ovarien et lymphome) (n = 23), et ont été combinés, après un réajustement en fonction de l'âge, selon une approche d'apprentissage automatique. Nous avons mis en place un modèle de prédiction basé sur des arbres de décision pour la détection et la classification des patients sains et cancéreux. Cette approche a fourni des résultats très encourageants, en particulier concernant les stades précoces. Par conséquent, nous avons validé la preuve de concept de l'utilisation des biomarqueurs quantitatifs et structurels et de l'apprentissage automatique que nous avons proposée. L'ajout de ces biomarqueurs quantitatifs à la fragmentation, la méthylation ou la détection d'altérations génétiques pourrait être synergique.

Cet article est soumis à *Science Translational Medicine*.

Science Translational Medicine AAAS

Submitted Manuscript: Confidential

Title: Machine learning study for setting and evaluating circulating DNA quantitative analysis towards cancer screening

Authors: Rita Tanos [1,2,3], Guillaume Tosato [1,2,3], Amaelle Otandault [1,2,3], Zahra Al Amir Dache [1,2,3], Laurence Pique Lasorsa [1,2,3], Geoffroy Tousch [1,2,3], Safia El Messaoudi [1,2,3], Mona Diab Assaf [4], Marc Ychou [1,2,3], Stanislas Du Manoir[1,2,3], Denis Pezet [5], Pierre-Emmanuel Colombo[2], William Jacot[2], Eric Assénat [6], Marie Dupuy [7], José Maria Sayagués [8], Caroline Mollevi[1,2,3], Jacques Collinge[1,2,3], Alain R. Thierry[1,2,3]*

Affiliations :

[1] IRCM, Institut de Recherche en Cancérologie de Montpellier, INSERM U1194, Montpellier, France.

[2] Institut régional du Cancer de Montpellier, Montpellier, F-34298, France.

[3] Université de Montpellier, Montpellier, France.

[4] Lebanese University, Beirut, Lebanon.

[5] CHU Clermont-Fd, Clermont-Ferrand, France.

[6] CHU Saint-Eloi, Montpellier, France.

[7] CHU Lapeyronnie, Montpellier, France.

[8] University Hospital, Salamanca, Spain.

*To whom correspondence should be addressed: alain.thierry@inserm.fr

One Sentence Summary: We implemented a machine learning decision tree prediction model based on cfDNA quantitative parameters that might provide a potential approach for noninvasive detection of patients with early-stage cancers.

Abstract:

Circulating cell-free DNA (cfDNA) is an emerging biomarker in oncology and several studies have recently investigated its role and utility for cancer screening and early detection. We identified two quantitative cfDNA candidate markers: the Ref A 67 (total nuclear cfDNA concentration) and the MNR (Mitochondrial to Nuclear Ratio) determined by specific q-PCR while targeting both nuclear and mitochondrial sequences. First, we independently validated them in three experimental models: cell culture media, tumor xenografted mice, and plasma from an exploratory cohort of healthy and cancer subjects. Subsequently, these two variables, as well as other quantitative and structural cfDNA parameters were evaluated in a large retrospective cohort including 289 healthy individuals, 99 individuals at risk for colorectal cancer (CRC), and 983 patients with CRC (n = 791), breast (n = 169), and other types of cancers (hepatocellular, pancreatic, ovarian and lymphoma) (n = 23), and were combined in a machine learning approach after age resampling. We implemented a decision tree prediction model for the detection and classification of healthy and cancer patients. This approach showed very encouraging results especially for early stages. Consequently, we demonstrated the proof of concept of using both the quantitative and structural biomarkers and the machine learning method we proposed. The addition of such quantitative biomarkers to fragmentomics, methylation or the detection of genetic alterations might be synergistic. The evaluation of this multi-analyte strategy with using this machine learning method is warranted.

Introduction:

Cancer is among the leading causes of death worldwide, coming second after heart diseases. The whole mechanisms leading to its progression are still unknown, but an early detection could greatly increase the chances for a successful treatment. Hence, in order to decrease its global mortality, tests should be used for screening and early detection, rather than waiting for the first symptoms to appear (1). Unfortunately, effective screening tests for early detection do not exist for many types of cancers due to limitations in sensitivity and specificity, and they are preferentially used for monitoring the response to treatment.

Recently, a few groups looked to develop strategies including several biomarkers for empowering screening test performance. Since most of these approaches include the combination of several biomarkers, it seems powerful to use the emerging machine learning. Many teams, from biomedical and bioinformatics fields, have studied the application of machine learning (ML) methods in this domain, such as Artificial Neural Networks (ANNs), Bayesian Networks (BNs), Support Vector Machines (SVMs) and Decision Trees (DTs), for the development of predictive models (2). Several studies based on different strategies have been reported for the detection of different cancer types (3, 4), by identifying, for example, a subset of three SNPs as key discriminators between controls and breast cancer patients (5). ANNs models trained on a large prospectively collected dataset of consecutive mammography findings demonstrated superior discrimination between benign and malignant breast cancer compared with radiologists (6). In addition, a study supported the utility of Bayesian classification to develop a predictive model that could support surgical decisions in patients with colon carcinoma (7).

Circulating cell-free DNA (cfDNA) has gained prominence and has become an emerging biomarker especially in cancer. It is being widely investigated in translational and clinical research (8, 9), since it may present the opportunity to diagnose, monitor recurrence, and evaluate response to therapy solely through a non-invasive blood draw. The use of cfDNA was recently proposed by our team as well as others (10, 11) towards cancer screening and early detection, and work is ongoing by several groups to develop tests for early cancer detection from a single blood analysis. Different approaches were adopted, from the detection of the viral DNA of the Epstein-Barr virus in nasopharyngeal cancer cases (12), to the analysis of the methylome and cfDNA methylation patterns in different cancer types (13, 14). The diagnostic potential was also examined by studying cfDNA levels (15) or detecting genetic alterations (16, 17). Others even associated mutation detection and protein markers (10), but bearing in mind

that the same mutation can be associated with different types of cancers. However, despite intensive research, the reliability of cfDNA-based tests was impacted by issues regarding their sensitivity and specificity, especially for early cancer stages.

We worked for more than a decade on quantitative and structural characteristics of cfDNA (18–22) that may, among others, show a capacity in discriminating cancer and healthy subjects . We believe that cfDNA based-attempts towards a pan-cancer test should be multi-analyte, and that the combination of qualitative (i.e. genetic or epigenetic alterations) and quantitative candidates would be a solution. In this work, we first identified different quantitative and structural cfDNA parameters determined by specific q-PCR by targeting both nuclear and mitochondrial sequences in a murine model, in cell culture supernatants as well as in the plasma of healthy and cancer individuals. We assessed their independent potential in distinguishing between healthy and cancer state in a large cohort of healthy (n= 289) and cancer individuals (n= 983). In order to optimally classify the data, we proposed a machine learning method based on a predictive decision tree model as a proof of concept for evaluating the combination of different quantitative parameters to rigorously estimate its potential before, eventually, including them in a multi-analyte cancer early detection test.

Results:

Determination of cfDNA quantitative parameters in a xenografted mouse model

In order to determine the potential of different quantitative cfDNA parameters, we used a murine model of nude mice xenografted with SW620 colorectal cancer human cells previously established by the team (19). This model allowed us to differentiate tumor human cfDNA from normal murine cfDNA in the same mouse. We quantified, by q-PCR, nuclear and mitochondrial cfDNA of human and murine origin in 14 xenografted mice. We found a significant increase of nuclear total DNA concentrations (ng/ml), that we called Ref A 67, between normal and tumor cfDNA, and the receiver operating curve (ROC) analysis showed an AUC (area under curve) of 0.91 (0.8125-1.014, 95% CI; confidence interval) **(Fig. 1A and B)** (Sensitivity Se= 0.79; Specificity Sp=0.86). The results were also significant when analyzing mitochondrial DNA concentrations (Ref M 67) **(Fig. 1C and D)** and the mitochondrial to nuclear ration (MNR) (Fig. 1E and F). These two parameters decreased significantly and showed higher AUCs of 0.99 (0.963-1.016, 95% CI) and 0.96 (0.9047-1.024, 95% CI) respectively with respective sensitivities of 0.93 and 1, and specificities of 1 and 0.86, which suggested the high potential of mitochondrial DNA concentration and the MNR to separate between cfDNA of normal and tumor origin.

Evaluation of the Ref A 67 and MNR in cell culture supernatants

The Ref A 67 and the MNR were then tested in the supernatant of cells in culture to assess their ability to discriminate between normal and cancer cells. The supernatant of 14 cancer cell lines and 5 normal cell lines were tested **(Fig. 2)**. The Ref A 67 showed a significant increase with a high AUC (0.901-1.042, 95% CI) of 0.97 (Se= 0.86, Sp= 1), while the MNR had a higher discriminative potential of 100% between normal and cancer cell lines with a 100% sensitivity and a specificity.

Quantitative parameters' evaluation in an exploratory cohort of healthy and cancer subjects

These quantitative parameters were studied in the plasma of a small independent exploratory cohort of 80 healthy individuals and 146 colorectal cancer (CRC) patients of stages I to IV. The CRC cohort consisted of 50% patients of stage IV and 50% of stages I to III. The Ref A 67 showed a significant increase in CRC patients with an AUC of 0.82 **(Fig. S1 A)**. As for the

MNR, a significant decrease was observed in cancer patients with a high AUC of 0.98 between CRC and healthy subjects **(Fig. S1 B)**.

Validation of the cfDNA parameters in a large retrospective cohort

The discriminative value of these two cfDNA quantitative parameters was then evaluated in a large retrospective cohort of 289 healthy individuals, 99 individuals at risk for CRC, and 983 patients with CRC (n = 791), breast (n = 169), or other types of cancers (hepatocellular, pancreatic, ovarian and lymphoma) (n = 23) **(Table 1)**. The Ref A 67 (copy nb/ml of plasma), or the total nuclear cfDNA concentration, showed a significant increase with a p-value < 0.0001 between healthy individuals and the different cancer types (CRC, breast and others) **(Fig. S2 A)**. It was able to significantly discriminate the healthy group from CRC patients no matter their stage (stage 0/I/II, stage III, stage IV) and from non-cancerous patients at risk for CRC showing benign tumors or dysplasia **(Fig. S2 B)**. An AUC of 0.83 (0.81-0.86, 95% CI) was observed for the CRC group when compared to healthy subjects **(Fig. S3 A)**, with a high value of 0.79 (0.76-0.83, 95% CI) for early stages 0, I and II **(Fig. S3 C)**, and 0.76 (0.70-0.83, 95% CI) for patients at risk for CRC **(Fig. S3 D)**. The Ref A 67 showed an AUC of 0.7 (0.74-0.82, 95% CI) for the discrimination of the breast cancer group from the healthy group **(Fig. S3 B)**.

We then assessed the potential of the MNR or the mitochondrial to nuclear ratio. This parameter significantly decreased between healthy patients and all cancer types **(Fig. S2 C)**. It showed a good discriminative potential for CRC patients (AUC= 0.78; 0.75-0.82, 95% CI) and breast cancer patients (AUC= 0.82; 0.78-0.86, 95% CI) **(Fig. S4 A and B)**. A significant decrease and a high discriminative potential was observed as well between the healthy group and early stage CRC patients (AUC= 0.79; 0.76-0.82, 95% CI) **(Fig. S2 D and Fig. S4 C)**, and patients at risk for CRC (AUC= 0.83; 0.78-0.88, 95% CI) **(Fig. S2 D and Fig. S4 D)**.

Cohort age adjustment and cfDNA parameters' individual performance for CRC detection

In order to achieve the most discriminative power, we wanted to assess the potential of a machine learning approach by evaluating the combination of different cfDNA variables. We first focused on the CRC group compared to healthy subjects. A bias was observed in the age distribution of the studied cohort **(Fig. 3A)**; the fraction of patients with an age < 30 years old was constituted of healthy individuals only, and the fraction of patients > 70 years old of CRC patients. No gender-based bias was observed **(Fig. S5)**. For this reason, subjects between 30

and 70 years old were resampled in 5 categories: ≤ 45 years old, 45 – 50, 50 – 55, 55 – 60, and > 60 years old **(Fig. 3B)**. The tendency of all the parameters was then appreciated on a heatmap **(Fig. 3C)** clustering patients (by columns) and parameters (by rows). The importance score of each of the parameters was then extracted after recursive partitioning was applied on 500 different resampling **(Fig. 3D)**. The descriptive statistics of the CRC cohort are presented in **Table S1**. The different quantitative cfDNA parameters were assessed: Ref A 67 (copy nb/ml) or total nuclear cfDNA concentration, Ref M 67 (copy nb/ml) or total mitochondrial cfDNA concentration, and the MNR. The age-adjusted estimation of the AUC for the Ref A 67 (copy nb/ml) parameter slightly decreased between healthy and CRC patients compared to the non-adjusted value: 0.79 (0.73-0.84, 95% CI) for the age-adjusted AUC with 0.81 Sp (0.72-0.89, 95% CI) and 0.70 Se (0.62-0.78, 95% CI) vs 0.83 (0.81-0.86, 95% CI) AUC, 0.82 (0.77-0.87, 95% CI) Sp and 0.74 (0.69-0.78, 95% CI) Se for the non-adjusted cohort **(Fig. 4A and Table 2)**. The MNR showed an AUC of 0.75 (0.69-0.81, 95% CI) after age sampling with 0.71 (0.61-0.781, 95% CI) Sp and 0.7 (0.62-0.78, 95% CI) Se **(Fig. 4B and Table2)**. The total mitochondrial cfDNA concentration (Ref M 67 ng/ml plasma) did not have a very high discriminative potential with a relatively low AUC value after resampling of 0.61 (0.53-0.67, 95% CI); Sp= 0.60 (0.45-0.69, 95% CI), Se= 0.62 (0.52-0.78, 95% CI) **(Fig. S6 and Table1)**.

Two other cfDNA parameters were studied in the plasma of the majority of the tested cohort to be added to the quantitative variables. These parameters correspond to the nuclear cfDNA concentration of the fragments with a size ≥ 145 base pairs (bp) (Ref A 145 (copy nb/ml plasma)) or ≥ 320 bp (Ref A 320 (copy nb/ml plasma)), and provide information regarding the size distribution of cfDNA in the healthy and the cancer group. The Ref A 145 (ng/ml plasma) was significantly higher in CRC patients than healthy individuals with a high discriminative value (age-adjusted AUC = 0.79 (0.74-0.87, 95% CI); Sp= 0.79 (0.67-0.91, 95% CI), Se= 0.72 (0.62-0.84, 95% CI) **(Fig. S7 and Table1)**. The AUC of the Ref A 320 (ng/ml) was shown to be lower than that of the Ref A 145 age-adjusted AUC = 0.63 (0.53-0.73, 95% CI) with 0.76 Sp (0.67-0.92, 95% CI) and 0.63 Se (0.53-0.73, 95% CI) **(Fig. S8 and Table 1)**.

Decision tree construction and prediction value for CRC detection

We implemented a decision tree prediction model by including our 5 different cfDNA quantitative and structural parameters obtained from q-PCR analysis after age resampling, to examine whether their association would further increase the discrimination potential **(Fig. 5 and Table S2)**. 527 CRC and healthy patients were used for the construction of the tree, and

showed 0.87 (0.79-0.92, 95% CI) specificity and 0.76 (0.67-0.85, 95% CI) sensitivity **(Table 3)**. The tree's performance characteristics were validated by cross-validation of 2000 rounds of resampling of a total of 622 patients that were not used for the tree construction. This validation estimated a specificity of 0.89 (0.84-0.94, 95% CI) and a sensitivity of 0.77 (0.73-0.80, 95% CI) for CRC and healthy patients' classification **(Table 3)**. When applied to early stages CRC, a relatively high Sp and Se of 0.87 (0.83-0.91, 95% CI) and 0.72 (0.67-0.76, 95% CI) were observed respectively **(Table 3)**. The discrimination between cancer patients and patients at risk for CRC was not possible with a specificity of 0.24 **(Table 3)** since more patients were predicted as cancerous rather than healthy (chi-square test p-value: 1.86E-05). No difference regarding the clinical parameters was observed between patients at risk classified as cancer or healthy individuals **(Table S3)**.

Breast cancer and other cancer types

The distribution and discriminative potential of each of the five individual cfDNA parameters were tested in stages II and III breast cancer patients (n=169) vs healthy individuals, and an unpaired t-test was adopted for continuous variables comparison among the populations studied **(Fig. S9 and Table S4)**. The MNR and the Ref A 67 showed the highest potential with an AUC of 0.81 (0.76-0.87, 95% CI) and 0.77 (0.71-0.83, 95% CI), a sensitivity of 0.76 (0.66-0.85, 95% CI) and 0.75 (0.66-0.86, 95% CI) and a specificity of 0.75 (0.66-0.85, 95% CI) and 0.77 (0.61-0.81, 95% CI) respectively **(Fig. S10 and Table S5)**. The other three parameters were less effective for breast cancer vs healthy discrimination (Ref M 67: AUC = 0.72 (0.66-0.88, 95% CI), Se = 0.67 (0.50-0.77, 95% CI), Sp = 0.71 (0.59-0.85, 95% CI) ; Ref A 145: AUC = 0.69 (0.62-0.76, 95% CI), Se = 0.61 (0.44-0.76, 95% CI), Sp = 0.71 (0.53-0.89, 95% CI) ; Ref A 320: AUC = 0.60 (0.53-0.70, 95% CI), Se = 0.55 (0.34-0.65, 95% CI), Sp = 0.66 (0.52-0.86, 95% CI)) **(Fig. S10 and Table S5)**. The same pattern adopted for the decision tree construction for the CRC cohort was followed for the breast cancer cohort **(Fig. S11 and Table S2)**. The MNR parameter appeared at the first node, followed by the Ref A 67 and the Ref A 145. 236 patients in total were used for the construction of the tree with a specificity of 0.72 (0.65-0.84, 95% CI) and a sensitivity of 0.86 (0.80-0.91, 95% CI) **(Table 3)**. The cross-validation using the 46 patients left showed a high specificity of 0.80 (0.64-0.95, 95% CI) and a high sensitivity of 0.95 (0.85-1, 95% CI) **(Table 3)**. The predictive tree established on a cohort was then used to predict the results of the other. The CRC decision tree had a 0.90 (0.84-0.95, 95% CI) specificity and 0.58 (0.50-0.65, 95% CI) sensitivity for breast cancer patients' classification, and the breast

cancer decision tree showed a Sp of 0.66 (0.62-0.72, 95% CI) and a Se of 0.85 (0.82-0.87, 95% CI) for CRC patients prediction **(Table 3)**. The ability of these two decision trees for the classification of 23 patients with different types of cancers (pancreatic, hepatocellular, lymphoma and ovarian) was assessed. The CRC and the breast prediction tree respectively **(Table S6)** classified 21 and 22 individuals out of the 23 as cancer patients.

Discussion:

Since the discovery of cfDNA (23) and especially the presence of higher concentrations in the plasma of cancer patients, quantitative determination of cfDNA was conceived as a tool for cancer screening (24, 25). However, despite significant statistical difference, the large overlapping in both cancer and healthy subjects severely weakens this approach. Alternatively, the detection of genetic alterations from cfDNA appeared as a valuable strategy (10, 17, 26, 27), and several recent attempts were implemented, but the false positive rate observed precludes it, mainly due to (i), the detection of white blood cells derived mutations among cfDNA variants especially with age, and (ii), the very low cfDNA tumor fraction in a significant fraction of cancer patient plasma. This suggests that quantitative and structural indicators from total cfDNA, which are not totally and directly associated with malignancy, may compensate the eventual deficiencies of qualitative indicators. We believe that the combination of both types of candidates in a multi-analyte approach might obey to the stringent evaluation for the validation of a screening or early detection test. This study aims at providing the proof of concept of using cfDNA quantitative biomarkers, as well as proposing a specific machine learning method we selected towards this goal.

Independent evaluation of the performance of various cfDNA parameters

The biomarker candidates were identified from our earlier works (18–22) based on the observation that nuclear total cfDNA analysis showed significantly higher plasma concentration (18, 28) (AUC = 0.91, N= 109 healthy and 229 CRC patients (21)), that cfDNA fragment size from cancer patients are lower than those of healthy individuals (18, 20, 22, 29) and on this study of the comparison of the respective amounts of mitochondrial and nuclear cfDNA as revealed here.

First, we used the previously established xenograft mouse model (19), and compared different quantitative cfDNA parameters of murine (normal) and human (tumor) origin. Each of the studied parameters, from the total nuclear (Ref A 67) or mitochondrial (Ref M 67) cfDNA concentration showed a significant difference between murine normal DNA and human tumor DNA, with the highest discriminative power to the mitochondrial to nuclear ratio (MNR). Second, these three quantitative cfDNA parameters were then validated in the supernatant of cells in culture to determine their capacity to discriminate between normal and cancer cells. In vitro data revealed that the MNR show the strongest potential to discriminate cell culture

supernatant of normal to cancer cell lines (AUC = 1). Third, the evaluation on human plasma in a retrospective study on a small exploratory cohort confirmed that the total nuclear cfDNA concentration and the MNR ratio might be discriminative factors (AUC = 0.82 and 0.98 respectively). Next, these variables were assessed in the plasma of a large cohort of 289 healthy individuals and 982 cancer patients including 791 CRC, 169 breast cancer and 23 patients with other types of cancer. Early stages 0/I/II constituted about 54% of the CRC cohort. Each of these parameters showed a good potential for discriminating cancer and healthy patients, but the AUC, specificity and sensitivity were lower than what we previously observed from the small exploratory cohort due to the large number of individuals and the high percentage of cancer patients with early stage disease.

Parameters relative to the cfDNA structure were also studied. It has been shown by our team (20, 21), as well as others (30–33) that cfDNA fragmentation and size profile differ between healthy and cancer patients. For this reason, we added two parameters to our study the Ref A 145 (nuclear cfDNA concentration of the fragments with a size ≥ 145 base pairs) and the Ref A 320 (≥ 320 bp) which provide information regarding the size distribution of cfDNA. The identification of these parameters originated from previous reports showing that the examination of size profile and fragmentomics (34) may help in discriminating healthy to cancer individuals (20, 29, 35). The Ref A 145 showed a high performance for CRC patients, and a lower potential for breast cancer patients. A lower discrimination potential was observed for both cohorts of cancer patients, when evaluating the Ref A 320.

Machine learning selection

The aim of the decision algorithm was to serve as a visual tool or method for clinical application. The easiest way to do so was the use of ROC curves and optimal thresholds (36) for the different cfDNA parameters which produced good results but could sometimes give opposite prediction if used independently. The combination of both these methods to predict cancer status could have been done in several ways like Support Vector Machine, Random Forest and Logistic regression as top algorithms used in a comparative study for the selection of a risk prediction model for CRC (37). However, an association between age and the measured cfDNA seems to exist since some of these parameters tend to have a non-linear interaction showed using logistic regression with colorectal cancer status as a predictive variable **(Table S7)**. A non-parametric method as recursive regression trees (38) is adapted to this kind of situation where some of the parameters are inter-independent, like the MNR that is based on

the Ref A 67 and the Ref M 67. The visual representation of this method tends to be more adapted for a potential clinical application instead of Random Forest or Support Vector Machine techniques that need to apply computational tools. This method allows as well the use of most of the available parameters and the assessment of their potential combination, as it solves some of the problems regarding the interactions between our available variables.

These decision tree methods have already shown good performances in regards to other machine learning techniques as artificial neural networks or support vector machine used for mortality prognostication in the case of glioma for example (39). However, the use of casual methods like those used for the discrimination of head and neck squamous cell carcinoma patients and healthy adults (40), with cross-validation -2/3 of the values as a learning-set and validation with bootstrap estimation using the 1/3 left-, gave unstable results. When applying a similar method to our cohort, large confidence intervals were observed for the sensitivity and specificity **(Table S8 and Fig. S12)** and the decision trees obtained were highly dependent on the patients selected for tree construction. For that reason, the resampling of aged paired patients was used and a global tree was built based on the most recurrent variable at each node. The iterative procedure we used was also applied by other teams in the field of suicide attempt prediction (41) or in the detection of drug usage (42). The tree was then qualified on the rest of the patients not used for learning, which gave us more stable results with shorter confident intervals that we hope would be more robust if applied on an extended population.

Test performance by machine learning decision tree classification

Patients with early stages CRC (stages 0, I and II) were mostly predicted as of cancer status by the CRC decision tree with a sensitivity of 0.72 and a specificity of 0.87. The breast cancer cohort studied consisted of patients with stage II or III of the disease, and the breast cancer decision tree cross-validation showed a 0.80 sensitivity and a 0.95 specificity. This tends to encourage us on the ability of this approach for the early detection of cancer. In fact, Cohen et al developed a test called CancerSeek for the noninvasive detection and localization of eight types of cancer (stages I to III) by combining the detection of tumor-specific mutations in cfDNA with eight circulating protein markers (10). Even though the test showed a high specificity of 0.99 with a median sensitivity of 0.70 (highest sensitivity = 0.98 for ovarian cancer), the sensitivity was 0.33 for breast cancer and 0.65 for CRC. Regarding early stages, the sensitivity was 0.43 (ranging from 0.2 for esophageal cancer to 1 for liver cancer) for stage I patients and 0.73 for stage II patients. Other cfDNA based methods, like the methylome

analysis by the recovery of circulating tumor DNA-associated methylation profile could enable sensitive detection, classification and monitoring of cancer across a range of cancer types (13). The study of the methylation landscape of cfDNA isolated from 100 metastatic breast and colorectal cancer patients and 45 healthy individuals showed a statistical diagnostic efficacy with 0.84 sensitivity and 0.82 specificity (14), but was not able to detect cancer on a very early stage.

In our study, patients labelled at risk were not discriminated from patients with colorectal cancer (Sp = 0.24), being predicted to be cancer patients rather than healthy individuals, which indicates that they share a similar profile regarding the cfDNA measured parameters. In addition, no specific profile in the clinical parameters of this sub-population was found to correlate with the classification of these patients as cancerous or healthy (Table S3). However, the stratification of these individuals as CRC patients could be beneficial given that patients with these profiles are advised to be regularly followed.

Test performance vs cancer type

Breast cancer patients were poorly predicted by the decision tree determined based on the CRC cohort (Se = 0.58) despite the high specificity found (0.90). This high specificity could be explained by the fact that the same healthy patients were used in both algorithms. Alternatively, the use of the breast cancer tree to classify colorectal cancer patients was quite efficient (Se = 0.85) but the loss of specificity (0.66) was significant. This could be due to the fact that the breast cancer cohort only concerns women, and that a probable difference between healthy males and females regarding the cfDNA measured parameters could exist, especially for mitochondrial cfDNA (28). This may explain as well the importance of the MNR, which appears in first node of the breast cancer decision tree, for the classification of breast cancer patients. These algorithms were then applied on a small cohort of patients with other cancer types (n=23) showing good predictions considering the sensitivity, which tends to indicate that these trees are not cancer-type specific.

Conclusion, study limitations and perspectives

The selected machine learning method seems adapted towards classifying and testing the combination of the quantitative biomarkers, promoting its further use when combining quantitative as well as qualitative biomarkers. The selected ones evaluated in this study showed

some screening power, especially in early cancer detection, with the ability to detect two or three quarters of patients with early-stage cancer, representing a higher level than that found with the CancerSeek test (10) or the analysis of the methylation profile of cfDNA (13). Use of biomarkers based on the total cfDNA analysis are also known to vary in conditions and pathologies other than cancer like autoimmune diseases (43, 44), sepsis (45, 46), myocardial infraction (47, 48), exercise (49) and others. Therefore, it is crucial to investigate this approach in respect to confounding conditions such as inflammation. However, the studied biomarkers cannot determine the tumor tissue of origin as potentially reported with CancerSeek or methylation analysis (10, 13). Although the quantitative multi-analyte test performance as a blood test is noteworthy, the study presents critical limitations in the context of a universal cancer screening. The studied cohort is retrospective and composed of healthy controls devoid of asymptomatic patients. It rather appears as the confirmation of a diagnosis at the pre-screening stage, whereas, to properly call it screening, the test ought to have been performed on an asymptomatic population ostensibly in good health, with positive results being confirmed by other testing techniques.

Protein markers have about a >10% sensitivity and a >99% specificity as well as the capacity of detecting one or more cancer types. They, at this time, do not comply with the requirements for multi-cancer tests used at the population scale. Primary requirements would be test performance (high sensitivity without sacrifying specificity; >90% and >99%, respectively), reproducibility and robustness. Secondary requirements are generalizability to population, and ability to identify anatomic location to direct appropriate diagnostic work-up. In addition, cancer screening test validation should ultimately be subject to large scale population studies with people with no known diagnosis, inclusion of potentially confounding conditions to ensure specificity, and multiple study sites for demographic diversity. The detection of the genetic alterations such as mutations, aneuploidy, or translocation/re-arrangements would have the perfect screening parameters, but white blood cells derived mutations among cfDNA variants especially with age, and the very low cfDNA tumor fraction occurring in a significant part of cancer patient plasma precludes their use as a single strategy. We believe that a blood test based on cfDNA analysis might have the capacity towards this goal, given especially its minimal invasiveness, high compliance, affordability and scalability. A cfDNA based cancer screening test would be multi-analyte. Although the test performance of the combination of quantitative and structural parameters as reported here are not good enough, it may reveal synergic power, especially in respect to early cancer stage detection, when associated with other cfDNA

parameters that are from different origins such as genetic alterations. Examination of genome wide methylation profiling was found to be specific to cancer and very informative of cancer type while showing high sensitivity and specificity, and could be a candidate partner of choice. In addition, analysis of the cfDNA fragment size profile may be another powerful candidate as we previously demonstrated (29, 35) and recently confirmed (31, 50). The evaluation of the combination of the quantitative biomarkers tested here with methylation and fragmentomics analysis is ongoing with using of the machine learning method we proposed here.

Materials and methods:

Study design

This study presents an observational analysis of different quantitative and structural cfDNA parameters calculated using an ultrasensitive q-PCR method to quantify wild type nuclear and mitochondrial cfDNA sequences in the plasma of cancer patients and healthy individuals, as well as a proposition of a machine learning method of a predictive decision tree model combining these different parameters for early cancer detection.

The plasma of a small exploratory cohort of 80 healthy subjects and 146 colorectal cancer (CRC) patients (stages I to IV) were first analyzed. We then analyzed 1371 plasma samples (Table 1) of 289 healthy individuals, 99 individuals at risk for CRC, and 983 patients CRC (n = 791), breast (n = 169), or other cancer types (hepatocellular, pancreatic, ovarian and lymphoma) (n = 23) over a range of stages. Early stages CRC patients (stages 0/ 1/ 2) represent more than 50% of the total CRC cohort tested.

We evaluated the utility of each of the cfDNA parameters tested to discriminate between healthy individuals and cancer patients, especially for early stages. We then developed a decision tree that combined five different parameters and established its sensitivity and specificity to assess whether it could potentially serve for early cancer detection.

Patient and sample characteristics

Blood samples from healthy individuals were obtained from healthy donors, from the Etablissement Français du Sang (E.F.S), the blood transfusion center of Montpellier (Convention EFS-PM N° 21PLER2015-0013). They were analyzed (virology, serology, immunology, blood numeration) and ruled out if an abnormality was detected.

Plasma samples from patients with colorectal cancer were provided by the hospital centers of Clermont-Ferrand and Limoges in France, and of Salamanca in Spain, as well as from the “BCB colon” cohort from Montpellier in France and the KPLEX I and KPLEX II studies previously conducted by the team. Plasma samples from individuals with breast cancer were obtained from the “IDEA Sein” cohort from Montpellier. 18 Hepatocellular carcinoma plasma samples were provided by Eric Assenat and Marie Dupuy from Montpellier, and 3 pancreatic adenocarcinomas, 1 lymphoma and 1 ovarian cancer samples were obtained from the hospital center of Salamanca in Spain. All plasma samples from cancer patients were obtained at the

time of diagnosis, or at least 45 days after treatment interruption to eliminate any bias due to the liberation of cfDNA by the treatment.

Written informed consent was obtained from all participants prior to the onset of the study. Clinical data for all patients included are listed in **Table 1**.

Cell lines

We analyzed the supernatant of 14 tumor cell lines in culture: 3 colorectal cancer (SW620, SW480 and CaCo2), 5 prostate cancer (VCAP, 22rV1, DU145, LNCAP and PC3), 3 breast cancer (SUM159, MDA 468 and R2sh P53), 2 lymphoma (RAMOS and BJAB), and 1 lung cancer cell line (H1975); and 5 normal cell lines of various origins: human foreskin fibroblasts (HFF), skin fibroblasts (CDC45K), mammary fibroblasts (R2), lung fibroblasts (IMR-90 A), and hepatocytes (LWFD). Genomic DNA of the DIFI human colorectal cancer cell line grown in RPMI 1640 and supplemented with 10% FCS and ATB, was used as a standard for human nuclear DNA quantification. Most cell lines were obtained from the American type culture collection (ATCC) (Manassas, VA, USA). The R2sh P53 cell line was obtain from the Charles Theillet/ Claude Sardet team. Cell culture medium supernatant was collected when cells were at 70% confluence, and underwent two centrifugations, the first at 1200g and the second at 16000g at 4°C for 10min each. DNA was then extracted using the QIAamp DNA Blood Mini Kit (Qiagen, Courtaboeuf, France) from 200µL of the supernatant and eluted in a final volume of 80µl.

Mice models

An experimental mouse xenograft model previously developed (19), allowing to distinguish in the same subject non-tumor derived cfDNA (of murine origin) and tumor derived cfDNA (of human origin) was used. Peripheral blood was drawn into EDTA pre-coated tubes of 14 female athymic nude mice (6–8 weeks old) xenograted with SW620 human cancer cells, and was used for plasma preparation within one hour. All experiments complied with the current national and institutional regulations and ethical guidelines and were performed by an accredited person.

Plasma isolation and cfDNA extraction

Human and murine blood samples were handled according to the preanalytical guideline we established (20, 51). Blood was drawn in EDTA tubes and was centrifuged at 1200 g at 4°C in a Heraeus Multifuge LR centrifuge for 10 min. The supernatants were isolated and centrifuged at 16000 g at 4°C for 10 min. Subsequently, the plasma was either immediately handled for DNA extraction or stored at -80°C. CfDNA was extracted from 200µL of plasma and eluted in a final volume of 80µl using the QIAamp DNA Blood Mini Kit (Qiagen, Courtaboeuf, France) according to our detailed protocol(52). DNA samples were kept at-20°C until use. Freeze-thaw cycles were avoided in order to reduce the phenomenon of cfDNA fragmentation.

DNA quantification by Q-PCR and copy number calculation

Q-PCR amplifications were carried out at least in triplicate in a 25 µl reaction volume on a CFX96 touchTM Real-Time PCR detection system (Bio-Rad) instrument, using the CFX manager software. Each PCR reaction mixture was composed of 12.5µl of SsoAdvancedTM Universal SYBR® Green Supermix (Bio-Rad, Marnes-la-Coquette, France), 2.5µl of each amplification primer (3 pmol/µl), 2.5µl of Nuclease free water (Qiagen), and 5µl of DNA extract. Thermal cycling consisted of three repeated steps: a 3-min Hot-start Polymerase activation-denaturation step at 95°C, followed by 40 repeated cycles at 90°C for 10s, and then at 60°C for 30s. Melting curves were obtained by increasing the temperature from 55°C to 105°C with a plate reading every 0.2°C. The concentration was calculated from the Cq detected by Q-PCR. A triplicate of non-template negative control was included in each run for each couple of primers.

Human DNA quantification

For human nuclear DNA, serial dilutions of genomic DNA from Difi cells were used as a standard for nuclear DNA quantification. The initial concentration and purety were assessed by optic density at λ=260 nm, 230 nm and 280nm, with an Eppendorf BioPhotometer® D30.

A control standard curve using a 3382-bp human ORF vector with a 786-bp MT-CO3 insert obtained from ABM good® (accession no.YP_003024032) of known concentration was used for human mitochondrial DNA quantification in cell lines and human samples. Initial vector

solution concentration and purity were determined by measuring the optic density at λ=260 nm, 230 nm and 280 nm, with an Eppendorf BioPhotometer® D30.

For each human plasma sample and cell line supernatant extract, a nuclear wild type KRAS amplicon and a MT-CO3 mitochondrial amplicon of 67 bp each were separately quantified. A wild-type BRAF nuclear amplicon of 105 bp was also targeted for several human plasma samples to validate the total nuclear DNA quantification calculated by targeting the KRAS 67 bp amplicon. Amplicons of 145 and 320 bp of the KRAS gene were targeted as well for cfDNA size profile analysis.

Nuclear cfDNA copy number calculation:

Nuclear cfDNA copy number per milliliter of plasma/supernatant was determined with the following calculation:

$$Qnuclear = \left(\frac{c}{3{,}3}\right) * \left(\frac{Velution}{Vplasma/supernatant}\right)$$

Qnuclear is the NcfDNA copy number per milliliter, c is the NcfDNA concentration (pg/µl) determined by Q-PCR targeting the nuclear KRAS gene sequence and 3.3 pg is the human haploid genome mass. Velution is the volume of cfDNA extract (µl) and Vplasma/supernatant is the volume of plasma or supernatant used for the extraction (ml).

Mitochondrial cfDNA copy number calculation:

Mitochondrial cfDNA copy number per milliliter of plasma/supernatant was determined with the following calculation:

$$Qmito = \left(\frac{c * Na}{2 * MW * Lvector}\right) * \left(\frac{Velution}{Vplasma/supernatant}\right)$$

Qmito is the McfDNA copy number per milliliter of plasma/supernatant, ‘c’ is the McfDNA mass concentration (g/µl) determined by Q-PCR targeting the mitochondrial MT-CO3 gene. NA is Avogadro’s number (6.02*1023 molecules per mole), Lvector is the plasmid length (nucleotides), and MW is the molecular weight of one nucleotide (g/mol). Velution is the elution volume of cfDNA extract (µl) and Vplasma is the volume of plasma or supernatant used for the extraction (ml).

Murine DNA quantification

For murine samples, murine nuclear DNA was quantified using as reference a serial dilution of genomic murine DNA (Promega), and human nuclear DNA of tumor origin was quantified using a standard curve of genomic DNA of the human Difi cell line. The relative amount of mitochondrial DNA to nuclear DNA was determined using the equation 2^{-dCq}, where dCq = (Cq mito – Cq nuc). Human cfDNA (of tumor origin) and murine cfDNA (of non-tumor origin) were determined in the same murine plasma samples using adequate primer sets for each amplification and expressed in nanograms per milliliter of plasma (ng/ml). Human KRAS nuclear amplicon and MT-CO3 mitochondrial amplicon of 67 bp each were quantified to assess the human cfDNA of tumor origin. On the other hand, murine cfDNA of non-tumor origin was analysed by targeting a KRAS murine nuclear amplicon of 63 bp and a MT-CO1 murine mitochondrial amplicon of 114 bp.

Oligonucleotides

Oligonucleotides were synthesized and purified on HPLC by Integrated DNA Technologies (Coralville, Iowa) and quality control of the oligonucleotides was performed by MALDI TOF. They were tested for specificity and sensitivity before use. The sequences and characteristics of the selected primers are presented in the **Table S9**.

Ref A and MNR calculation

Human samples and cell culture supernatant extracts

The mean nuclear DNA copy nb/ml value when targeting the KRAS wild type 67bp amplicon, is considered as the Ref A 67 (copy nb/ml of plasma or supernatant) and corresponds to the total concentration of nuclear cfDNA. Plasma samples with a Ref A 67 below 450 copies/ml were excluded from the study. The RefA 145 (copy nb/ml) and the RefA 320 (copy nb/ml) correspond to the concentration of cfDNA fragments >145 bp and >320 bp respectively, and are calculated by targeting the KRAS 145 bp or 320 bp amplicon.

The mean mitochondrial DNA copy nb/ml value when targeting the MT-CO3 wild type 67bp amplicon, is considered as the Ref M 67 (copy nb/ml of plasma or supernatant) and corresponds to the total concentration of mitochondrial cfDNA.

The mitochondrial to nuclear DNA ratio (MNR) was expressed as the ratio of the mean of mitochondrial DNA copy nb/ml of plasma (or supernatant) value to the Ref A 67 (copy nb/ml of plasma (or supernatant)) value of the experiments (MNR = Ref M 67/Ref A 67).

Murine samples

For murine samples, Human tumor MNR was calculated by the ratio between human mcfDNA relative concentration to human ncfDNA concentration (ng/ml). Murine non tumor MNR was calculated by the ratio between murine mcfDNA relative concentration to murine ncfDNA concentration (ng/ml).

Statistical analysis

Murine and cell culture supernatant data. Data were expressed as mean ± SD. The non-parametric Wilcoxon-Mann-Whitney test was used for comparison of medians for murine and cell culture supernatant data. Receiver operating characteristic (ROC) curves are presented and the area under the ROC curve (AUC) calculated. A probability of less than 0.05 was considered to be statistically significant; *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$.

Human plasma samples. The analysis was conducted using R 3.6.0 and the stats package.

Colorectal cancer cohort. Patients with missing values on parameters used for tree regression were removed which concerned 152 patients (129 CRC patients, 2 healthy patients and 21 at risk patients). An unpaired t-test was adopted for continuous variables comparisons, rather a two samples proportions test with continuity correction for binary variables comparison among the populations studied (Table S1).

Breast cancer cohort. Patients with missing values were removed from the analysis, which concerned 2 patients diagnosed with a breast cancer. An unpaired t-test was adopted for continuous variables comparison among the populations studied (Table S3).

Age pairing. Due to different age distribution among healthy and colorectal patients (Fig. 3A), we split the age in 5 different categories in order to pair patients: less than 45 years old, 45 to 50 years old, 50 to 55 years old, 55 to 60 years old and more than 60 years old (Fig. 3B). This pairing reduced number of patients included to 527. Moreover, interaction effect between age and some measured parameters was observed using logistic modelling of the colorectal cancer status probability (Table S7) which confirmed the utility to remove this bias in order to

appreciate, if present, the interest of the age in a non-parametric algorithm. This age distribution was not so different between breast cancer patients and healthy individuals and resulted only in a high number of young Healthy patients. Accordingly, a window of patients between 25 and 65 years was created.

Univariate first analysis. For each measured parameter, ROC curves were produced using the pROC package and an optimal threshold corresponding to the best prediction was estimated by minimizing the distance between the ROC curve and the ideal point (Se=1, 1-Sp=0) (53). Sensitivity, specificity, area under the curve and the optimal threshold were estimated with their respective confidence intervals using an empirical bootstrap method already shown useful for quantitative diagnostic tests (54). First, we applied 2000 resampling -with replacement- and we estimated the punctual values for each step. Then the difference between each step estimation and the targeted value were calculated, the 2.5th and 97.5th distance percentiles were obtained and added to the targeted punctual value. Secondly, we applied the same method on 2000 resampling of 155 CRC and 155 healthy patients -paired on the age categories previously made.Thirdly, we applied the same method on patients from the breast cancer cohort with 2000 resampling of 236 patients. Thirdly, we applied the same method on patients from Breast cancer cohort with 2000 resampling of 236 patients.

Bioinformatics analysis. The cancer prediction algorithm was conducted by applying recursive partitioning for regression tree (55) using the rpart package. First, a global view of all the measured parameters was generated in order to appreciate the tendency of each of them. The resulting heatmap was built using the ComplexHeatmap package **(Fig. 3C)** clustering patients -by columns- and parameters -by rows- based on Euclidian distance and using ward.D2 agglomeration method. The values were reduced among the samples after a log10 transformation to appreciate the variability in an easier color scale. Secondly, recursive partitioning was applied on 500 different resampling to extract the importance score of each variable introduced in the procedure **(Fig. 3D)**. That allowed us to exclude some new transformed variables using these parameters.

To build the global tree, a stepwise procedure was performed on each node using the resampling method described in univariate first analysis part to keep the most frequently selected variable. It was confirmed by applying a homogeneity chi-square test on the frequency table, and a proportion test with continuity correction was applied on the two most frequent variables in order to discriminate them. Once the selection confirmed, we fixed the parameter of interest at the node and we estimated its conditional threshold on 2000 resamples –with replacement-. The

confidence interval was obtained based on the distribution of the difference to the mean, we kept the median as the condition applied at the node. The same method was applied on breast cancer cohort in order to obtain and compare a specific tree on this cohort **(Table 3)**. We iterated this procedure to acquire the full tree for colorectal cancer cohort **(Fig. 5)**. Each node contains the variable with its estimated threshold and the proportions of CRC/Healthy Patients present at this step. Each leaf contains the concluded status –CRC or healthy-, the proportion of CRC|healthy patients and the global proportion of patients ending in it. The same pattern was followed for breast cancer cohort **(Fig. S11)**. For each tree obtained, the performance was qualified by using the predictive values given by the tree and the real patient's diagnosis. We punctually estimated sensitivity and specificity and then we obtained the 95% confidence interval by empirical bootstrap as previously described in univariate first analysis. Different groups of patients were used. First we calculated the parameters with the patients used for tree building (n=527 for colorectal cancer cohort, n=236 for breast cancer). Secondly, we applied the tree prediction on the patients excluded of the tree building. We also applied these predictions on early colorectal cancer stages (0-1-2) and patients at risk to develop colorectal cancer.

Other cancer types. Our parameters were measured on 23 patients with different cancer types. The results were predicted using both breast cancer cohort tree and colorectal cancer cohort tree (Table S5).

Acknowledgments:

We would like to thank Philippe BLACHE, Corinne PREVOSTEL and Vanessa GUILLAUMON for their help, as well as Laura MUINELO ROMAY (MD PhD from Santiago de Compostela) and Sandrine BONIZEC.

Funding: This work was funded by the "SIRIC Montpellier Cancer Grant INCa_Inserm_DGOS_12553" and MSDAvenir through the MSD-Mitest grant.

References:

1. L. D. Maxim, R. Niebo, M. J. Utell, Screening tests: a review with examples, Inhal. Toxicol. 26, 811–28 (2014).

2. K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, D. I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Comput. Struct. Biotechnol. J. 13, 8–17 (2015).

3. L. Hussain, A. Ahmed, S. Saeed, S. Rathore, I. A. Awan, S. A. Shah, A. Majid, A. Idris, A. A. Awan, Prostate cancer detection using machine learning techniques by employing combination of features extracting strategies, Cancer Biomarkers 21, 393–413 (2018).

4. S. Sharma, A. Aggarwal, T. Choudhury, in 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), (IEEE, 2018), pp. 114–118.

5. J. Listgarten, S. Damaraju, B. Poulin, L. Cook, J. Dufour, A. Driga, J. Mackey, D. Wishart, R. Greiner, B. Zanke, Predictive Models for Breast Cancer Susceptibility from Multiple Single Nucleotide Polymorphisms, Clin. Cancer Res. 10, 2725–2737 (2004).

6. T. Ayer, O. Alagoz, J. Chhatwal, J. W. Shavlik, C. E. Kahn, E. S. Burnside, Breast cancer risk estimation with artificial neural networks revisited, Cancer 116, 3310–3321 (2010).

7. A. Stojadinovic, A. Nissan, J. Eberhardt, T. C. Chua, J. O. W. Pelz, J. Esquivel, Development of a Bayesian Belief Network Model for Personalized Prognostic Risk Assessment in Colon Carcinomatosis, .

8. A. R. Thierry, S. El Messaoudi, P. B. Gahan, P. Anker, M. Stroun, Origins, structures, and functions of circulating DNA in oncology, Cancer Metastasis Rev. 35, 347–76 (2016).

9. M. Fleischhacker, B. Schmidt, Circulating nucleic acids (CNAs) and cancer—A survey, Biochim. Biophys. Acta - Rev. Cancer 1775, 181–232 (2007).

10. J. D. Cohen, L. Li, Y. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A. A. Javed, F. Wong, A. Mattox, R. H. Hruban, C. L. Wolfgang, M. G. Goggins, M. Dal Molin, T.-L. Wang, R. Roden, A. P. Klein, J. Ptak, L. Dobbyn, J. Schaefer, N. Silliman, M. Popoli, J. T. Vogelstein, J. D. Browne, R. E. Schoen, R. E. Brand, J. Tie, P. Gibbs, H.-L. Wong, A. S. Mansfield, J. Jen, S. M. Hanash, M. Falconi, P. J. Allen, S. Zhou, C. Bettegowda, L. A. Diaz, C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. M. Lennon, N. Papadopoulos, Detection and localization of surgically resectable cancers with a multi-analyte blood test., Science 359, 926–930 (2018).

11. R. Tanos, A. R. Thierry, Clinical relevance of liquid biopsy for cancer screening, Transl. Cancer Res. 7, S105–S129 (2018).

12. K. C. A. Chan, J. K. S. Woo, A. King, B. C. Y. Zee, W. K. J. Lam, S. L. Chan, S. W. I. Chu, C. Mak, I. O. L. Tse, S. Y. M. Leung, G. Chan, E. P. Hui, B. B. Y. Ma, R. W. K. Chiu, S.-F. Leung, A. C. van Hasselt, A. T. C. Chan, Y. M. D. Lo, Analysis of Plasma Epstein–Barr Virus DNA to Screen for Nasopharyngeal Cancer, N. Engl. J. Med. 377, 513–22 (2017).

13. S. Y. Shen, R. Singhania, G. Fehringer, A. Chakravarthy, M. H. A. Roehrl, D. Chadwick, P. C. Zuzarte, A. Borgida, T. T. Wang, T. Li, O. Kis, Z. Zhao, A. Spreafico, T. da S. Medina, Y. Wang, D. Roulois, I. Ettayebi, Z. Chen, S. Chow, T. Murphy, A. Arruda, G. M. O'Kane, J. Liu, M. Mansour, J. D. McPherson, C. O'Brien, N. Leighl, P. L. Bedard, N. Fleshner, G. Liu, M. D. Minden, S. Gallinger, A. Goldenberg, T. J. Pugh, M. M. Hoffman, S. V. Bratman, R. J. Hung, D. D. De Carvalho, Sensitive tumour detection and classification using plasma cell-free DNA methylomes, Nature 563, 579–583 (2018).

14. A. A. I. Sina, L. G. Carrascosa, Z. Liang, Y. S. Grewal, A. Wardiana, M. J. A. Shiddiky, R. A. Gardiner, H. Samaratunga, M. K. Gandhi, R. J. Scott, D. Korbie, M. Trau, Epigenetically reprogrammed methylation landscape drives the DNA self-assembly and serves as a universal cancer biomarker, Nat. Commun. 9, 4915 (2018).

15. N. Krishnamurthy, E. Spencer, A. Torkamani, L. Nicholson, Liquid Biopsies for Cancer: Coming to a Patient near You, J. Clin. Med. 6, 3 (2017).

16. L. Fernandez-Cuesta, S. Perdomo, P. H. Avogbe, N. Leblay, T. M. Delhomme, V. Gaborieau, B. Abedi-Ardekani, E. Chanudet, M. Olivier, D. Zaridze, A. Mukeria, M. Vilensky, I. Holcatova, J. Polesel, L. Simonato, C. Canova, P. Lagiou, C. Brambilla, E. Brambilla, G. Byrnes, G. Scelo, F. Le Calvez-Kelm, M. Foll, J. D. McKay, P. Brennan, Identification of Circulating Tumor DNA for the Early Detection of Small-cell Lung Cancer, EBioMedicine 10, 117–23 (2016).

17. J. Phallen, M. Sausen, V. Adleff, A. Leal, C. Hruban, J. White, V. Anagnostou, J. Fiksel, S. Cristiano, E. Papp, S. Speir, T. Reinert, M.-B. W. Orntoft, B. D. Woodward, D. Murphy, S. Parpart-Li, D. Riley, M. Nesselbush, N. Sengamalay, A. Georgiadis, Q. K. Li, M. R. Madsen, F. V. Mortensen, J. Huiskens, C. Punt, N. van Grieken, R. Fijneman, G. Meijer, H. Husain, R. B. Scharpf, L. A. Diaz, S. Jones, S. Angiuoli, T. Ørntoft, H. J. Nielsen, C. L. Andersen, V. E. Velculescu, Direct detection of early-stage cancers using circulating tumor DNA, Sci. Transl. Med. 9, eaan2415 (2017).

18. A. Thierry, S. El Messaoudi, Methods for screening a subject for cancer (2016).

19. A. R. Thierry, F. Mouliere, C. Gongora, J. Ollier, B. Robert, M. Ychou, M. Del Rio, F. Molina, Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts, Nucleic Acids Res. 38, 6159–6175 (2010).

20. F. Mouliere, B. Robert, E. A. Peyrotte, M. Del Rio, M. Ychou, F. Molina, C. Gongora, A. R. Thierry, High Fragmentation Characterizes Tumour-Derived Circulating DNA, PLoS One 6, e23418 (2011).

21. F. Mouliere, S. El Messaoudi, D. Pang, A. Dritschilo, A. R. Thierry, Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer, Mol. Oncol. 8, 927–41 (2014).

22. C. Sanchez, M. W. Snyder, R. Tanos, J. Shendure, A. R. Thierry, New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA analysis, npj Genomic Med. 3, 31 (2018).

23. P. Mandel, P. Metais, Les acides nucléiques du plasma sanguin chez l'homme., C. R. Seances Soc. Biol. Fil. 142, 241–43 (1948).

24. C. Rago, D. L. Huso, F. Diehl, B. Karim, G. Liu, N. Papadopoulos, Y. Samuels, V. E. Velculescu, B. Vogelstein, K. W. Kinzler, L. A. Diaz, Serial Assessment of Human Tumor Burdens in Mice by the Analysis of Circulating DNA, Cancer Res. 40, 2281–2287 (2007).

25. M. Van Der Vaart, P. J. Pretorius, Circulating DNA, Ann. N. Y. Acad. Sci. 1137, 18–26 (2008).

26. A. M. Aravanis, M. Lee, R. D. Klausner, Next-Generation Sequencing of Circulating Tumor DNA for Early Cancer Detection, Cell 168, 571–574 (2017).

27. A. M. Newman, S. V Bratman, J. To, J. F. Wynne, N. C. W. Eclov, L. A. Modlin, C. L. Liu, J. W. Neal, H. A. Wakelee, R. E. Merritt, J. B. Shrager, B. W. Loo Jr, A. A. Alizadeh, M. Diehn, An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage, Nat. Med. 20, 548–554 (2014).

28. R. Meddeb, Z. A. A. Dache, S. Thezenas, A. Otandault, R. Tanos, B. Pastor, C. Sanchez, J. Azzi, G. Tousch, S. Azan, C. Mollevi, A. Adenis, S. El Messaoudi, P. Blache, A. R. Thierry, Quantifying circulating cell-free DNA in humans, Sci. Reports 2019 91 9, 5220 (2019).

29. A. R. Thierry, F. Molina, Analytical methods for cell free nucleic acids and applications (2012).

30. G. Leszinski, J. Lehner, U. Gezer, S. Holdenrieder, Increased DNA Integrity in Colorectal Cancer, In Vivo (Brooklyn). 28, 299–303 (2014).

31. P. Jiang, C. W. M. Chan, K. C. A. Chan, S. H. Cheng, J. Wong, V. W.-S. Wong, G. L. H. Wong, S. L. Chan, T. S. K. Mok, H. L. Y. Chan, P. B. S. Lai, R. W. K. Chiu, Y. M. D. Lo, Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients, Proc. Natl. Acad. Sci. 112, E1317–E1325 (2015).

32. H. R. Underhill, J. O. Kitzman, S. Hellwig, N. C. Welker, R. Daza, D. N. Baker, K. M. Gligorich, R. C. Rostomily, M. P. Bronner, J. Shendure, D. J. Kwiatkowski, Ed. Fragment Length of Circulating Tumor DNA, PLOS Genet. 12, e1006162 (2016).

33. F. Mouliere, D. Chandrananda, A. M. Piskorz, E. K. Moore, J. Morris, L. B. Ahlborn, R. Mair, T. Goranova, F. Marass, K. Heider, J. C. M. Wan, A. Supernat, I. Hudecova, I. Gounaris, S. Ros, M. Jimenez-Linan, J. Garcia-Corbacho, K. Patel, O. Østrup, S. Murphy, M. D. Eldridge, D. Gale, G. D. Stewart, J. Burge, W. N. Cooper, M. S. van der Heijden, C. E. Massie, C. Watts, P. Corrie, S. Pacey, K. M. Brindle, R. D. Baird, M. Mau-Sørensen, C. A. Parkinson, C. G. Smith, J. D. Brenton, N. Rosenfeld, Enhanced detection of circulating tumor DNA by fragment size analysis., Sci. Transl. Med. 10, eaat4921 (2018).

34. M. Ivanov, A. Baranova, T. Butler, P. Spellman, V. Mileyko, Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation, BMC Genomics 16, S1 (2015).

35. A. R. Thierry, C. Sanchez, Method for screening a subject for a cancer (2017).

36. P. Charoentong, F. Finotello, M. Angelova, C. Mayer, M. Efremova, D. Rieder, H. Hackl, Z. Trajanoski, Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade, Cell Rep. 18, 248–262 (2017).

37. N. Cueto-López, M. T. García-Ordás, V. Dávila-Batista, V. Moreno, N. Aragonés, R. Alaiz-Rodríguez, A comparative study on feature selection for a risk prediction model for colorectal cancer, Comput. Methods Programs Biomed. 177, 219–229 (2019).

38. H. Tang, E. T. Donnell, Application of a model-based recursive partitioning algorithm to predict crash frequency, Accid. Anal. Prev. 132, 105274 (2019).

39. S. S. Panesar, R. N. D'Souza, F.-C. Yeh, J. C. Fernandez-Miranda, Machine Learning Versus Logistic Regression Methods for 2-Year Mortality Prognostication in a Small, Heterogeneous Glioma Database, World Neurosurg. X 2, 100012 (2019).

40. G. Wichmann, C. Gaede, S. Melzer, J. Bocsi, S. Henger, C. Engel, K. Wirkner, J. R. Wenning, T. Wald, J. Freitag, M. Willner, M. Kolb, S. Wiegand, M. Löffler, A. Dietz, A. Tárnok, Discrimination of Head and Neck Squamous Cell Carcinoma Patients and Healthy Adults by 10-Color Flow Cytometry: Development of a Score Based on Leukocyte Subsets, Cancers (Basel). 11, 814 (2019).

41. J. T. Jordan, D. E. McNiel, Characteristics of a suicide attempt predict who makes another attempt after hospital discharge: A decision-tree investigation, Psychiatry Res. 268, 317–322 (2018).

42. Q. Q. Tiet, Y. E. Leyva, R. H. Moos, S. M. Frayne, L. Osterberg, B. Smith, Screen of Drug Use: Diagnostic Accuracy of a New Breif Tool for Primary Care, JAMA Intern. Med. 175, 1371 (2015).

43. Y. Xu, Y. Song, J. Chang, X. Zhou, Q. Qi, X. Tian, M. Li, X. Zeng, M. Xu, W. Zhang, D. S. Cram, J. Liu, High levels of circulating cell-free DNA are a biomarker of active SLE, Eur. J. Clin. Invest. 48, e13015 (2018).

44. A. Truszewska, B. Foroncewicz, L. Pączek, The role and diagnostic value of cell-free DNA in systemic lupus erythematosus., Clin. Exp. Rheumatol. 35, 330–336 (2016).

45. D. J. Dwivedi, L. J. Toltl, L. L. Swystun, J. Pogue, K.-L. Liaw, J. I. Weitz, D. J. Cook, A. E. Fox-Robichaud, P. C. Liaw, Prognostic utility and characterization of cell-free DNA in patients with severe sepsis, Crit. Care 16, R151 (2012).

46. T. Huang, Z. Yang, S. Chen, J. Chen, [Predictive value of plasma cell-free DNA for prognosis of sepsis]., Zhonghua Wei Zhong Bing Ji Jiu Yi Xue 30, 925–928 (2018).

47. L. Wang, L. Xie, Q. Zhang, X. Cai, Y. Tang, L. Wang, T. Hang, J. Liu, J. Gong, Plasma nuclear and mitochondrial DNA levels in acute myocardial infarction patients., Coron. Artery Dis. 26, 296–300 (2015).

48. K. V. Glebova, N. N. Veiko, A. A. Nikonov, L. N. Porokhovnik, S. V. Kostuyk, Cell-free DNA as a biomarker in stroke: Current status, problems and perspectives, Crit. Rev. Clin. Lab. Sci. 55, 55–70 (2018).

49. S. Tug, S. Helmig, E. R. Deichmann, A. Schmeier-Jürchott, E. Wagner, T. Zimmermann, M. Radsak, M. Giacca, P. Simon, Exercise-induced increases in cell free DNA in human plasma originate predominantly from cells of the haematopoietic lineage., Exerc. Immunol. Rev. 21, 164–73 (2015).

50. S. Cristiano, A. Leal, J. Phallen, J. Fiksel, V. Adleff, D. C. Bruhm, S. Ø. Jensen, J. E. Medina, C. Hruban, J. R. White, D. N. Palsgrove, N. Niknafs, V. Anagnostou, P. Forde, J. Naidoo, K. Marrone, J. Brahmer, B. D. Woodward, H. Husain, K. L. van Rooijen, M.-B. W. Ørntoft, A. H. Madsen, C. J. H. van de Velde, M. Verheij, A. Cats, C. J. A. Punt, G. R. Vink, N. C. T. van Grieken, M. Koopman, R. J. A. Fijneman, J. S. Johansen, H. J. Nielsen, G. A. Meijer, C. L. Andersen, R. B. Scharpf, V. E. Velculescu, Genome-wide cell-free DNA fragmentation in patients with cancer, Nature 570, 385–389 (2019).

51. R. Meddeb, E. Pisareva, A. R. Thierry, Guidelines for the Preanalytical Conditions for Analyzing Circulating Cell-Free DNA., Clin. Chem. , clinchem.2018.298323 (2019).

52. S. El Messaoudi, F. Rolet, F. Mouliere, A. R. Thierry, Circulating cell free DNA: Preanalytical considerations, Clin. Chim. Acta 424, 222–230 (2013).

53. C. E. Metz, Basic principles of ROC analysis, Semin. Nucl. Med. 8, 283–298 (1978).

54. R. W. Platt, J. A. Hanley, H. Yang, Bootstrap confidence intervals for the sensitivity of a quantitative diagnostic test, Stat. Med. 19, 313–322 (2000).

55. L. Breiman, Classification and Regression Trees (Routledge, New York, 2017).

Figures:

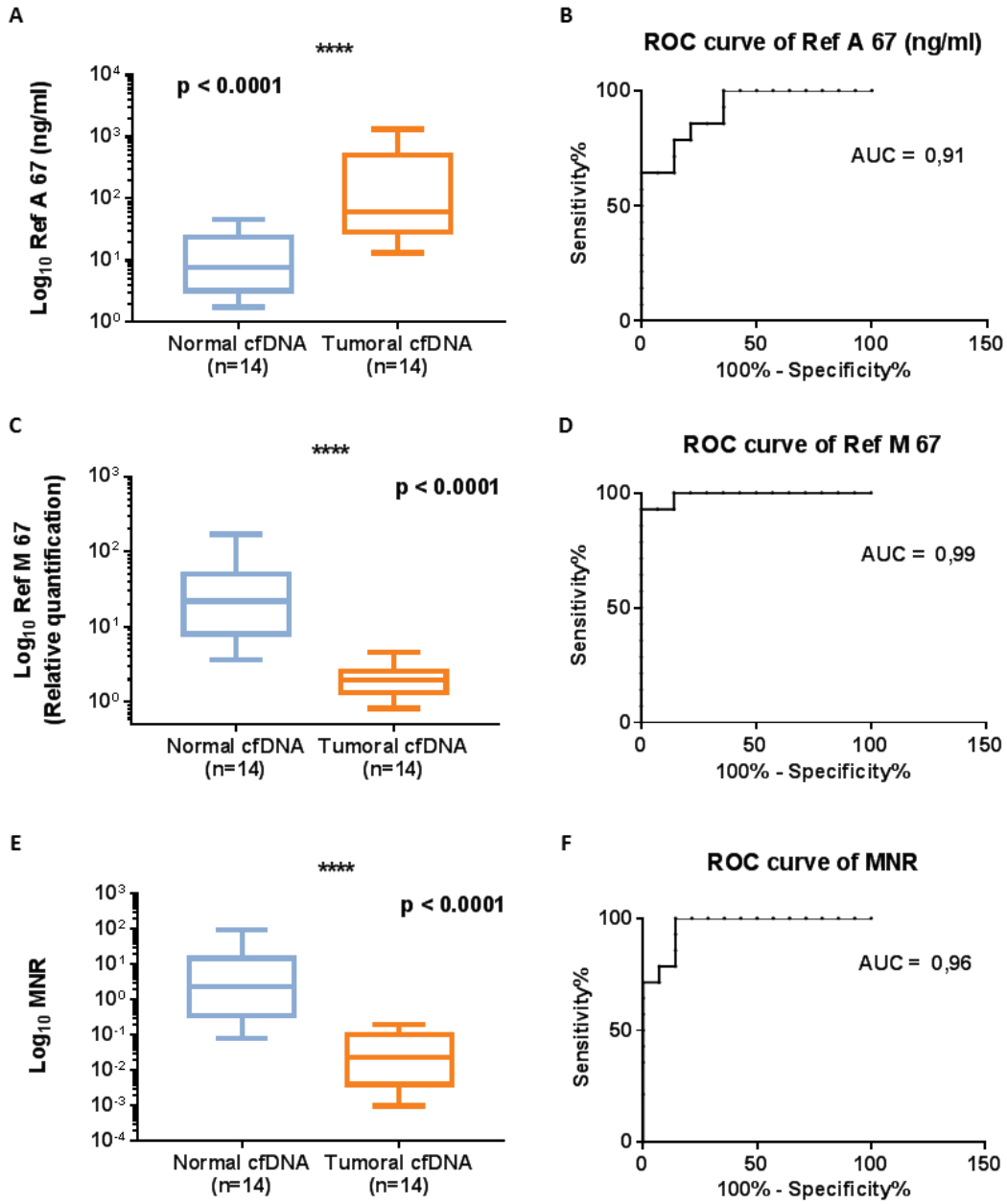
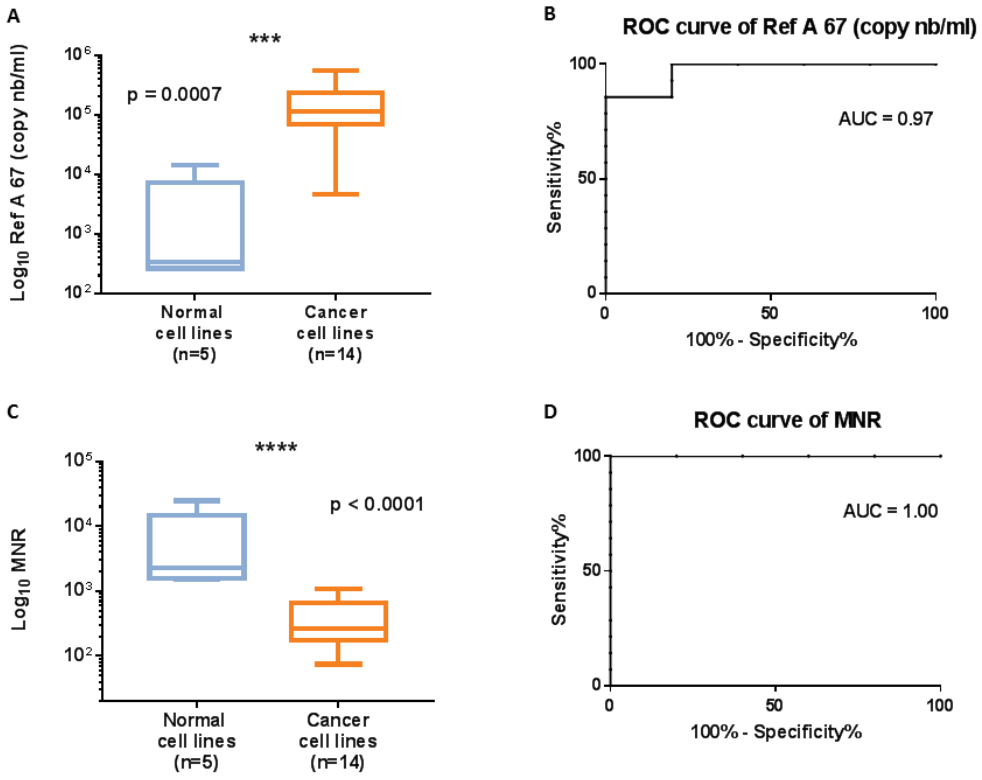


Fig. 1. Use of the xenografted mouse model for independently evaluating candidate performance. **A.** Total nuclear cfDNA amount (Ref A 67; ng/ml of plasma) from normal (murine) or tumor (human) origine in 14 SW620 xenografted nude mice. **B.** ROC curve of tumor vs non tumor Ref A 67 (ng/ml). C. Relative quantification of mitochondrial cfDNA (Ref M 67) from normal or tumor origin. **C.** ROC curve of tumor vs non tumon Ref M 67. **D.** Mitochondrial to nuclear cfDNA ratio (MNR) as determined in normal and tumor cfDNA. E. ROC curve from normal and tumor MNR. (AUC: Area under curve; MNR: mitochondrial to nuclear ratio).



E

	Type	Cell line
Cancer cell lines (n=14)	Colon	SW620
		SW480
		CaCo2
	Burkitt lymphoma	Ramos
		BJAB
	Lung	H1975
	Prostate	VCAP
		22rV1
		DU145
		LNCAP
		PC3
	Breast	SUM159
		MDA 468
		R2shP53

F

	Type	Cell line
Normal cell lines (n=5)	human foreskin fibroblasts	HFF
	human skin fibroblasts	CDC45K
	human mammary fibroblasts	R2
	human lung fibroblasts	IMR 90 A
	human normal hepatocytes	LWFD

Fig. 2. Use of the cell culture model for independently evaluating candidate performance.

A. Box plot of the total nuclear cfDNA concentration (Ref A 67 copy number/ml of supernatant) for normal and cancer cell lines. **B.** Receiver operating characteristics (ROC) curve for Ref A 67 between normal and cancer cell lines. **C.** Box Plot of the mitochondrial to nuclear DNA concentration ratio (MNR) for normal and cancer cell lines. **D.** ROC curve for MNR between normal and cancer cells. **E.** The different cancer cell lines tested. **F.** The different normal cell lines tested. (The box plot whiskers represent the minimal and maximal value; p: p-value, probability value; AUC: Area under curve).

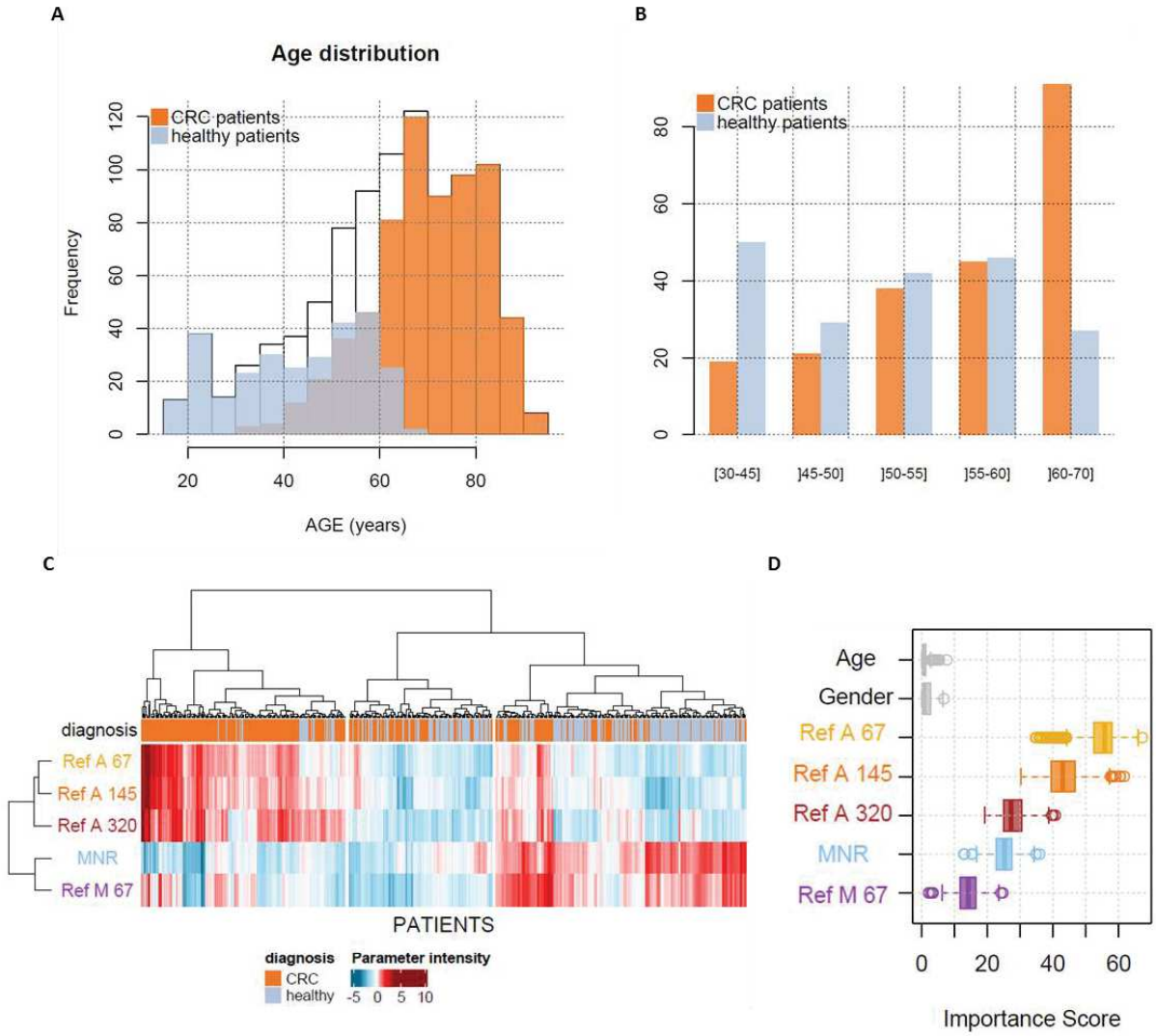
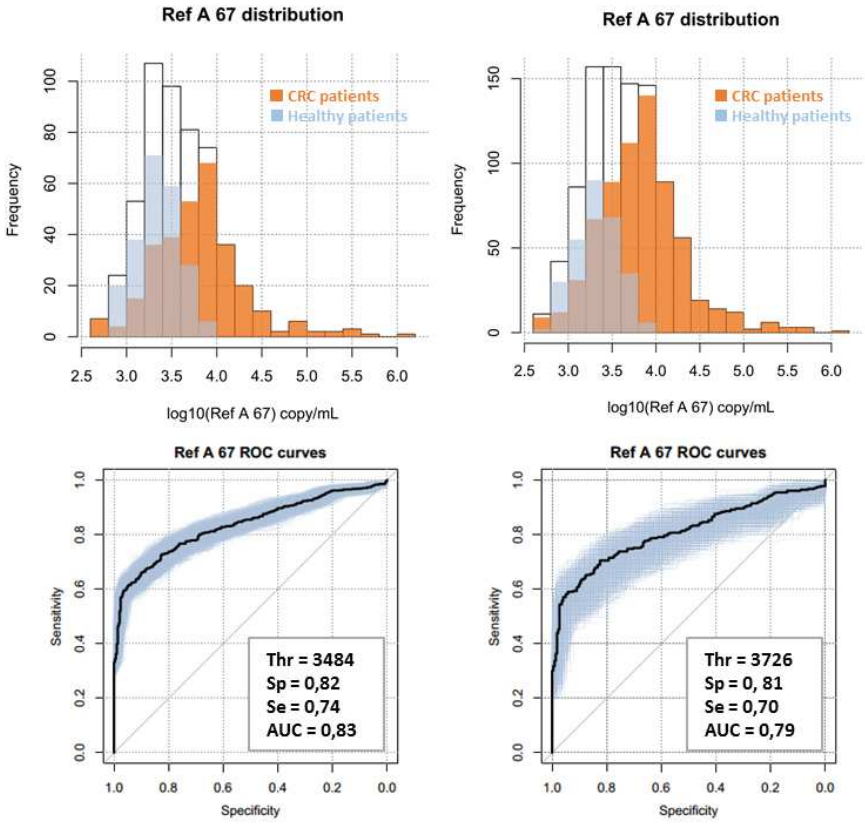


Fig. 3. Representation of patients used for boostraping and age correction. **A.** Age distribution among the two populations: patients with colorectal cancer (orange) and healthy patients (grey). **B.** Categories made based on the patients' age in order to select 1 patient with colorectal cancer with 1 healthy patient. **C.** Ascending hierarchical clustering performed on all patients used for resampling and all parameters available. **D.** Boxplot of each parameter importance score obtained after 500 recursive partitioning regression trees and resampling.

A **Ref A 67** **CRC Vs. Healthy** **Age adjusted**



B **MNR** **CRC Vs. Healthy** **Age adjusted**

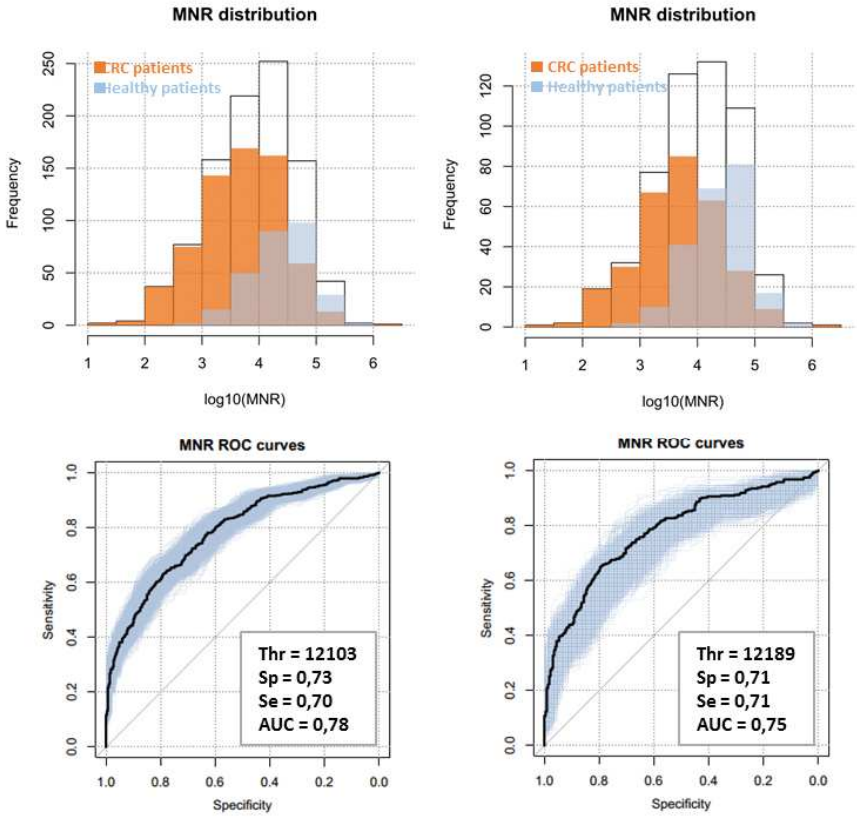


Fig. 4. Distribution and ROC curves of the Ref A 67 (copy nb/ml) (A) and MNR (B) parameters in CRC patients (orange) and healthy individuals (blue) before and after age resampling.

Ref A 67: Total nuclear cfDNA concentration; MNR: Mitochondrial to nuclear ratio; ROC: receiver operating characteristics; Thr: Threshold; Sp: Specificity; Se: Sensitivity; AUC: Area under ROC curve.

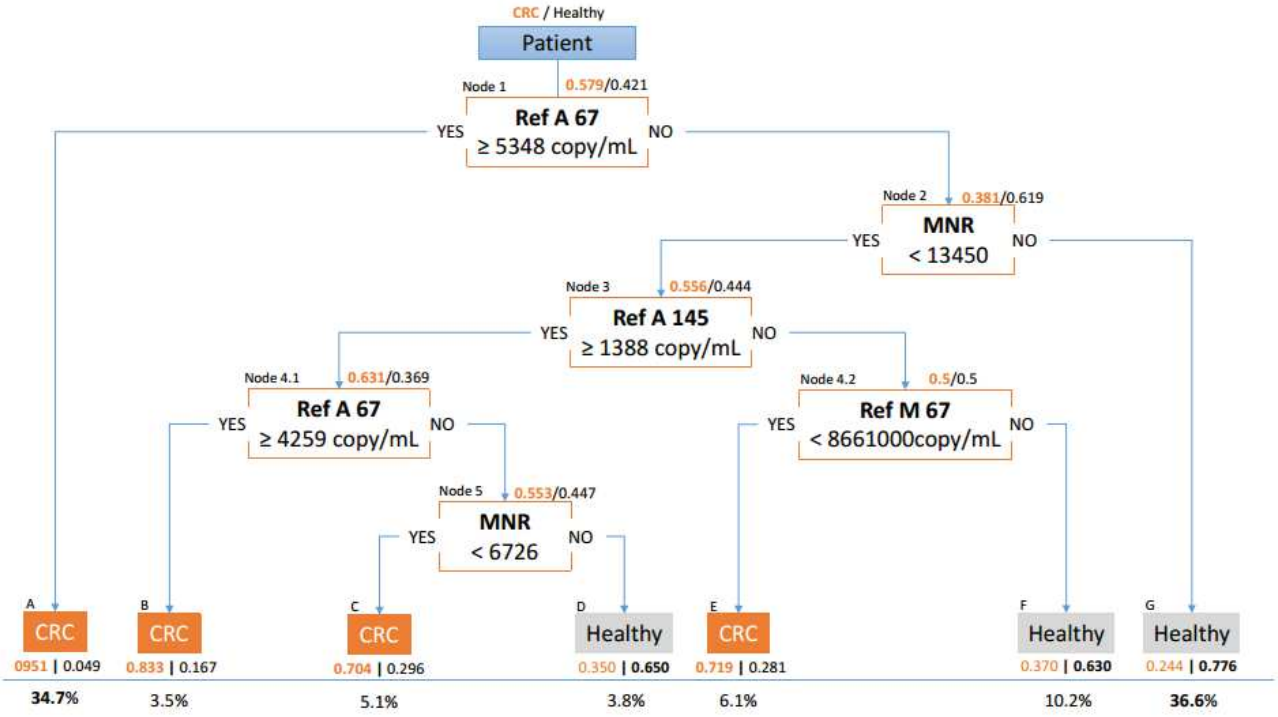


Fig. 5. Global decision tree obtained from the colorectal cancer cohort.

Global decision tree obtained from the colorectal cancer cohort by iterating a recursive partitioning regression tree after resampling patients at each step. Proportions of patients with colorectal cancer (orange) and healthy patients (grey) are represented at each node. Each final leaf contains the name of the most representative population inside, with their relative proportions. The proportion of patients that ended in each leaf is presented below the corresponding leaf. One node features the parameter of interest, the condition on the parameter (below) and the direction of the answer (YES/NO).

Table 1. Patients' characteristics.

Nb: number; NA: not available; HCC: Hepatocellular carcinoma

Diagnosis group		Nb	Gender			Age	
			Male	Female	NA	Mean	Median
Colorectal cancer	All Stages	791	452 (57%)	331 (42%)	8 (1%)	70	70
	Stages 0/I/II	425	252 (59,3%)	172 (40,5%)	1 (0,2%)	72	72
	Stage III	180	108 (60%)	72 (40%)	-	70	71
	Stage IV	186	92 (49%)	87 (47%)	7 (4%)	65	66
At risk for CRC		99	58 (%)	41 (%)	-	65	66
Breast Cancer	Stages 2 or 3	169	-	169 (100%)	-	49	48
Other cancers		23	21 (91%)	2 (9%)	-	63	66
	HCC	18	18 (100%)	-	-	62	64
	Pancreatic	3	2 (67%)	1 (33%)	-	78	78
	Ovarian	1	-	1 (100%)	-	44	44
	Lymphoma	1	1 (100%)	-	-	69	69
All cancers		983	473	502	8	66	67
Healthy		289	179 (62%)	110 (38%)	-	43	46

Table 2. Parameters estimation for each variable in the CRC cohort.

	Threshold		Specificity		Sensitivity		AUC	
Parameter	Value	IC95%	Value	IC95%	Value	IC95%	Value	IC95%
Ref A 67	**3484**	[3076-4857]	**0.82**	[0.77-0.87]	**0.74**	[0.70-0.78]	**0.83**	[0.81-0.86]
MNR	**12103**	[6945-15360]	**0.73**	[0.65-0.82]	**0.70**	[0.61-0.78]	**0.78**	[0.75-0.82]
Ref A 145	**1491**	[1121-1639]	**0.79**	[0.70-0.86]	**0.74**	[0.69-0.82]	**0.83**	[0.80-0.86]
Ref A 320	**425**	[286-509]	**0.77**	[0.61-0.83]	**0.67**	[0.62-0.74]	**0.76**	[0.73-0.79]
Ref M 67	**4.41E+07**	[2.78E+07-6.04E+07]	**0.60**	[0.46-0.69]	**0.57**	[0.50-0.69]	**0.59**	[0.56-0.63]
Patients paired by age category								
Ref A 67	**3726**	[2921-4511]	**0.81**	[0.72-0.89]	**0.70**	[0.62-0.78]	**0.79**	[0.73-0.84]
MNR	**12189**	[4295-16225]	**0.71**	[0.61-0.84]	**0.71**	[0.59-0.82]	**0.75**	[0.69-0.81]
Ref A 145	**1613**	[1267-1874]	**0.79**	[0.67-0.91]	**0.72**	[0.62-0.84]	**0.79**	[0.74-0.87]
Ref A 320	**617**	[100-899]	**0.76**	[0.67-0.92]	**0.63**	[0.53-0.73]	**0.63**	[0.53-0.73]
Ref M 67	**4.78E+07**	[3.54E+07-6.87E+07]	**0.60**	[0.45-0.69]	**0.62**	[0.52-0.78]	**0.61**	[0.53-0.67]

Parameters estimation for each variable based on minimum distance between corresponding ROC curve and the point (Se=1, 1-Sp=0) with the use of an empirical bootstrap to build their respective confidence interval. These estimations were obtained after patient age pairing. Se = Sensitivity, Sp = Specificity, AUC = Area under ROC curve.

Table 3. Estimation of tree performances over several groups of patients obtained on the colorectal and breast cancer cohort.

Specificity		Sensitivity	
Estimation	95% CI	Estimation	95% CI
CRC Decision Tree			
Patients used for tree building n=527 (CRC + healthy)			
0.87	[0.79-0.92]	**0.76**	[0.67-0.85]
Patients not used for tree building n=622 (CRC + healthy)			
0.89	[0.84-0.94]	**0.77**	[0.73-0.80]
CRC vs At risk n = 749			
0.24	[0.21-0.27]	**0.74**	[0.63-0.83]
CRC stages 0/I/II vs healthy			
0.87	[0.83-0.91]	**0.72**	[0.67-0.76]
Breast Cancer patients n= 278 vs healthy			
0.90	[0.84-0.95]	**0.58**	[0.50-0.65]
Breast Cancer Decision Tree			
Patients used for tree building n= 236 (breast cancer + healthy)			
0.72	[0.65-0.84]	**0.86**	[0.80-0.91]
Patients not used for tree building n=46 (breast cancer + healthy)			
0.80	[0.64-0.95]	**0.95**	[0.85-1]
CRC patients n=951 vs healthy			
0.66	[0.62-0.72]	**0.85**	[0.82-0.87]

Confidence intervals were obtained by using empirical bootstrap method.

Supplementary Materials:

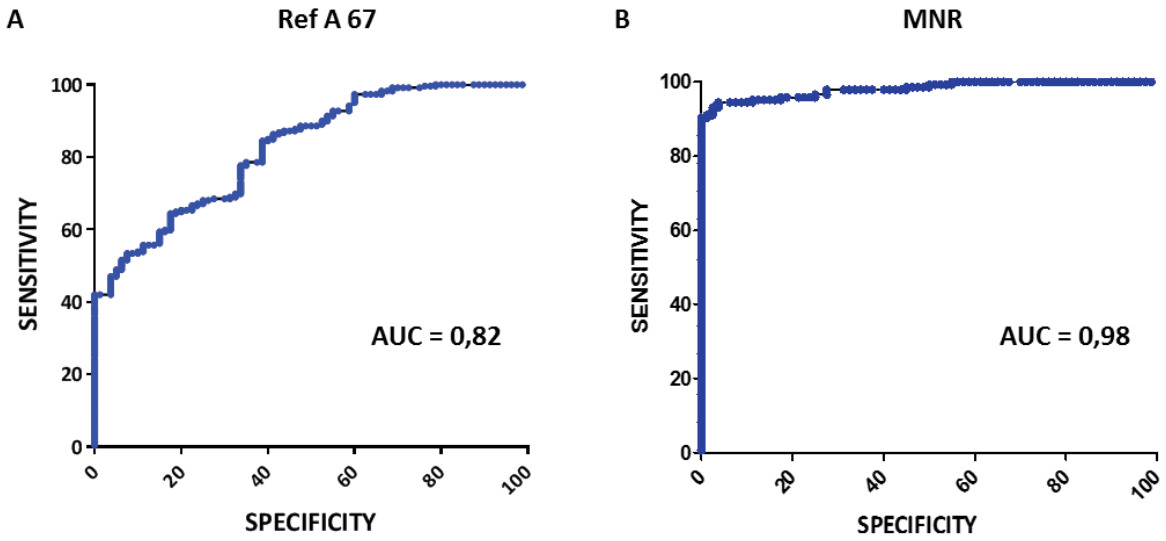


Fig. S1. ROC curves of the Ref A 67 (copy nb/ml plasma) **(A)** and the MNR parameter **(B)** for 80 healthy individuals vs 146 CRC patients of all stages. AUC: Area under the ROC curve; ROC curve: Receiver operating characteristic curve

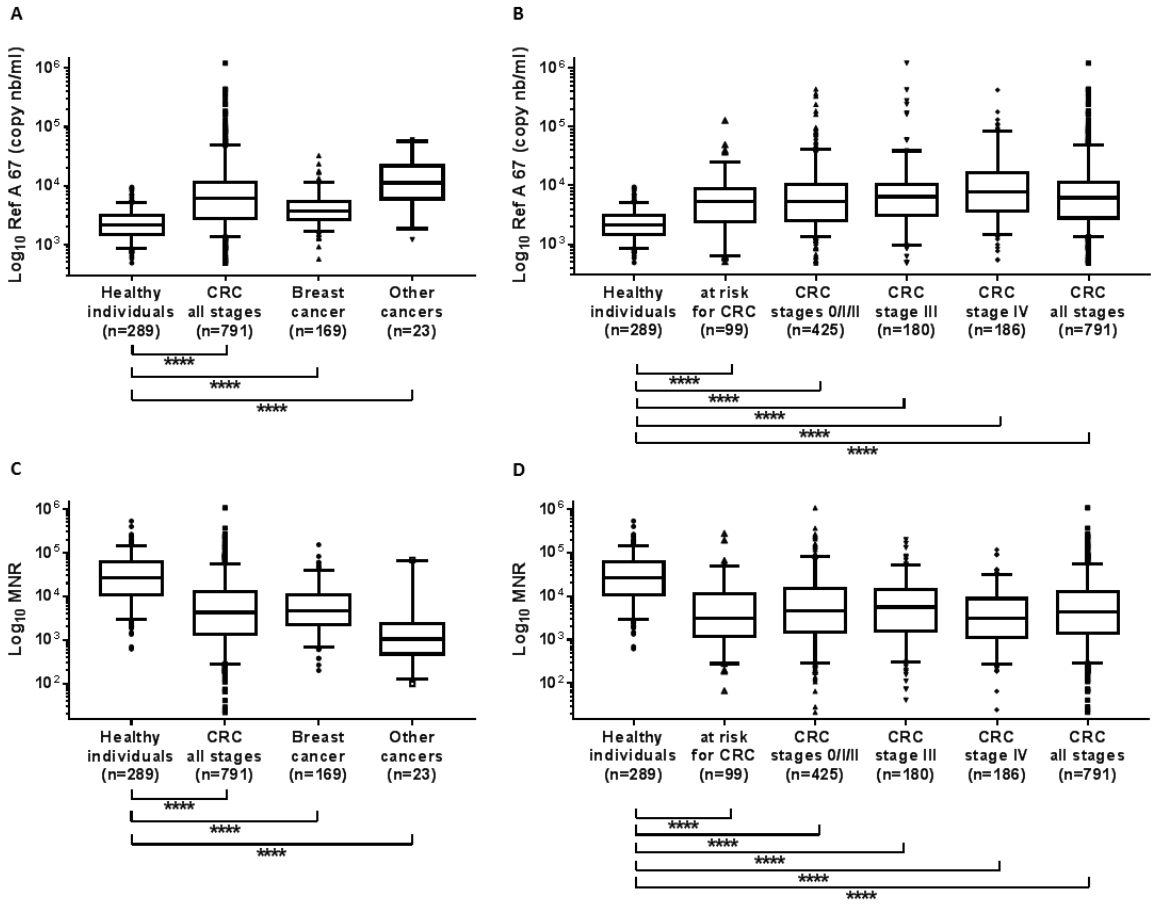


Fig. S2. Ref A 67 and MNR evaluation in the plasma of healthy individuals and cancer patients.

A. Box plot of the total nuclear cfDNA concentration (Ref A 67 copy number/ml of plasma) for healthy individuals and patients with different cancer types. **B.** Box plot of Ref A 67 for different stages of CRC. **C.** Box Plot of the mitochondrial to nuclear DNA concentration ratio (MNR) for healthy individuals and patients with different cancer types. **D.** Box plot of MNR for different stages of CRC. (The box plot whiskers represent the 5th and 95th percentiles; **** correspond to a p-value < 0,0001).

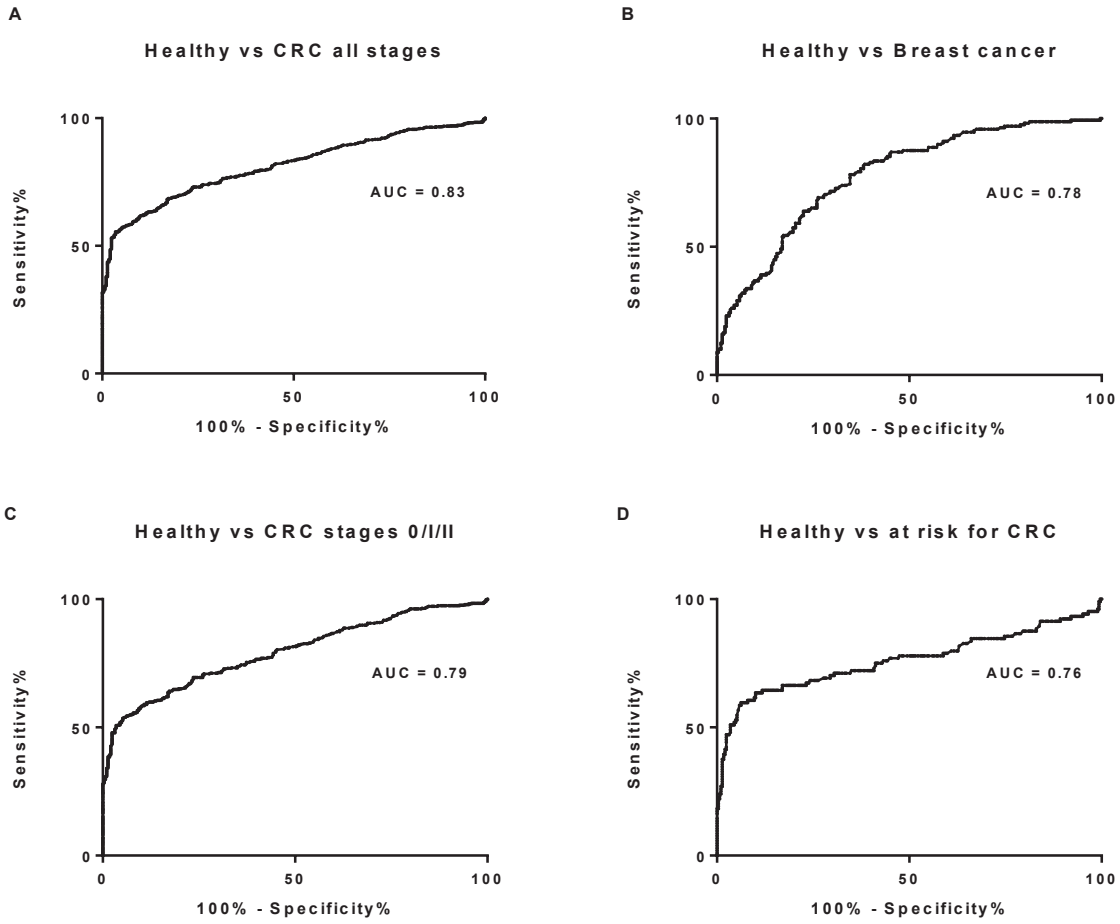


Fig. S3. ROC curve of the Ref A 67 (copy nb/ml plasma) parameter for healthy individuals vs **A.** CRC patients of all stages, **B.** brest cancer patients, **C.** Early stages CRC patients (0/I/II) and **D.** patients at risk for CRC. AUC: Area under the ROC curve; ROC curve: Receiver operating characteristic curve

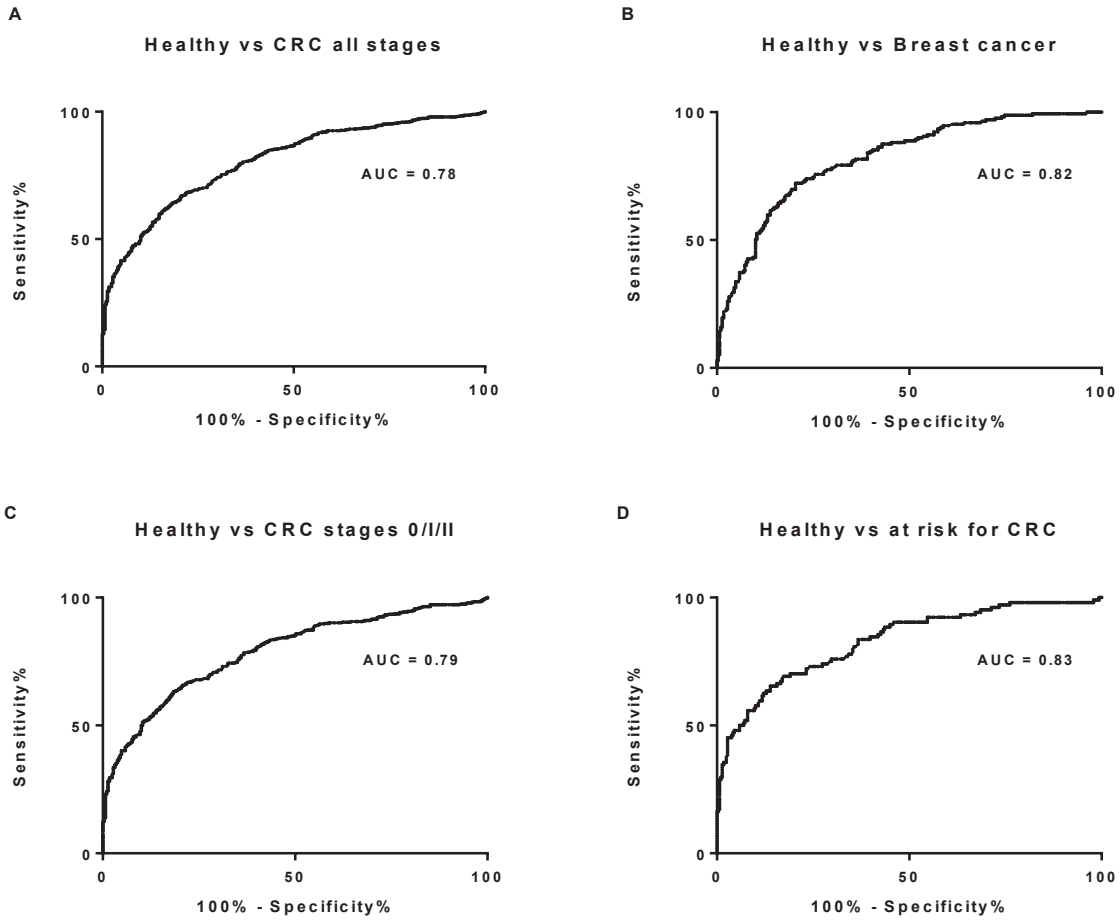


Fig. S4. ROC curve of the MNR parameter for healthy individuals vs **A.** CRC patients all stages, **B.** Brest cancer patients, **C.** early stages CRC patients (0/I/II) and **D.** patients at risk for CRC. MNR: mitochondrial to nuclear ratio; AUC: Area under the ROC curve; ROC curve: Receiver operating characteristic curve

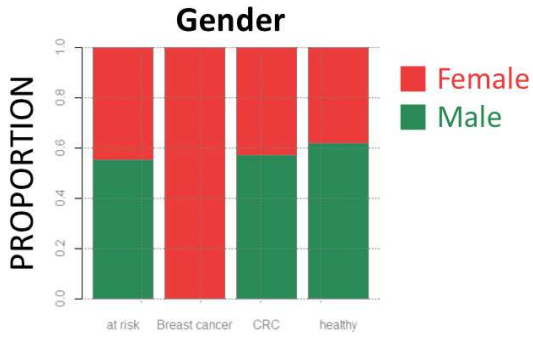


Fig. S5. Gender distribution among healthy individuals and CRC patients.

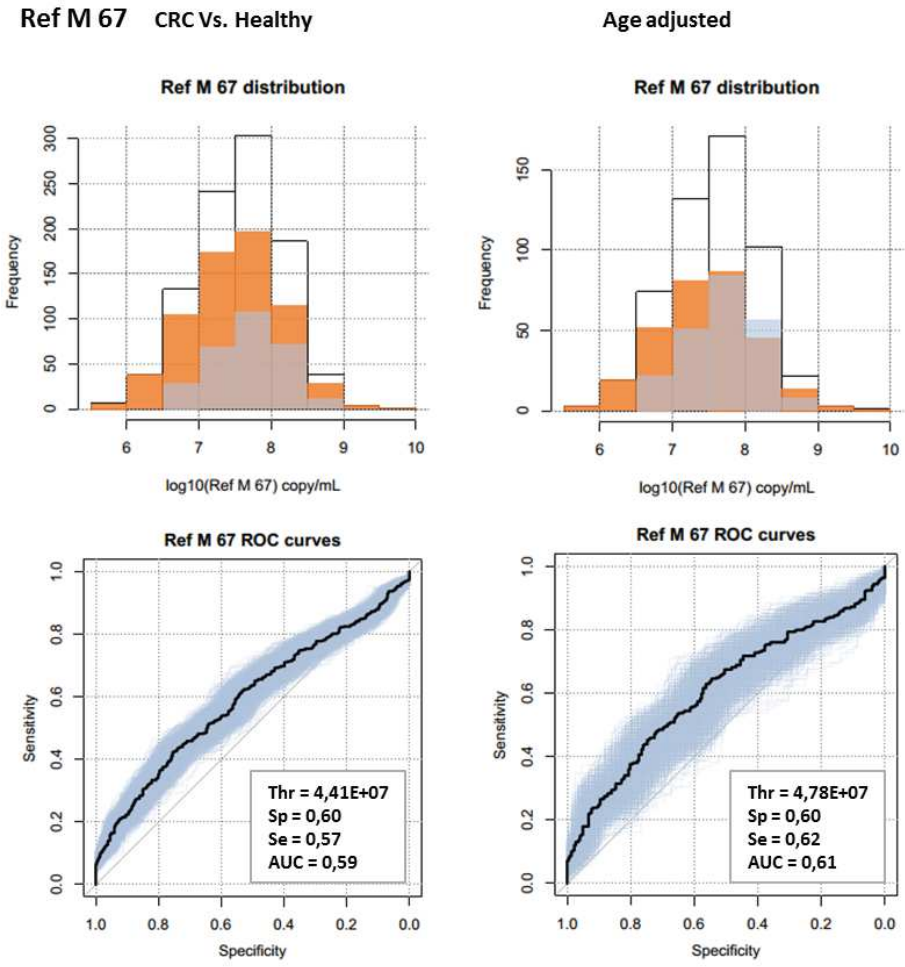


Fig. S6. Distribution and ROC curves of the Ref M 67 (copy nb/ml) parameter in CRC patients (orange) and healthy individuals (blue) before and after age resampling.

Ref M 67: Total mitochondrial cfDNA concentration; ROC: receiver operating characteristics; Thr: Threshold; Sp: Specificity; Se: Sensitivity; AUC: Area under ROC curve.

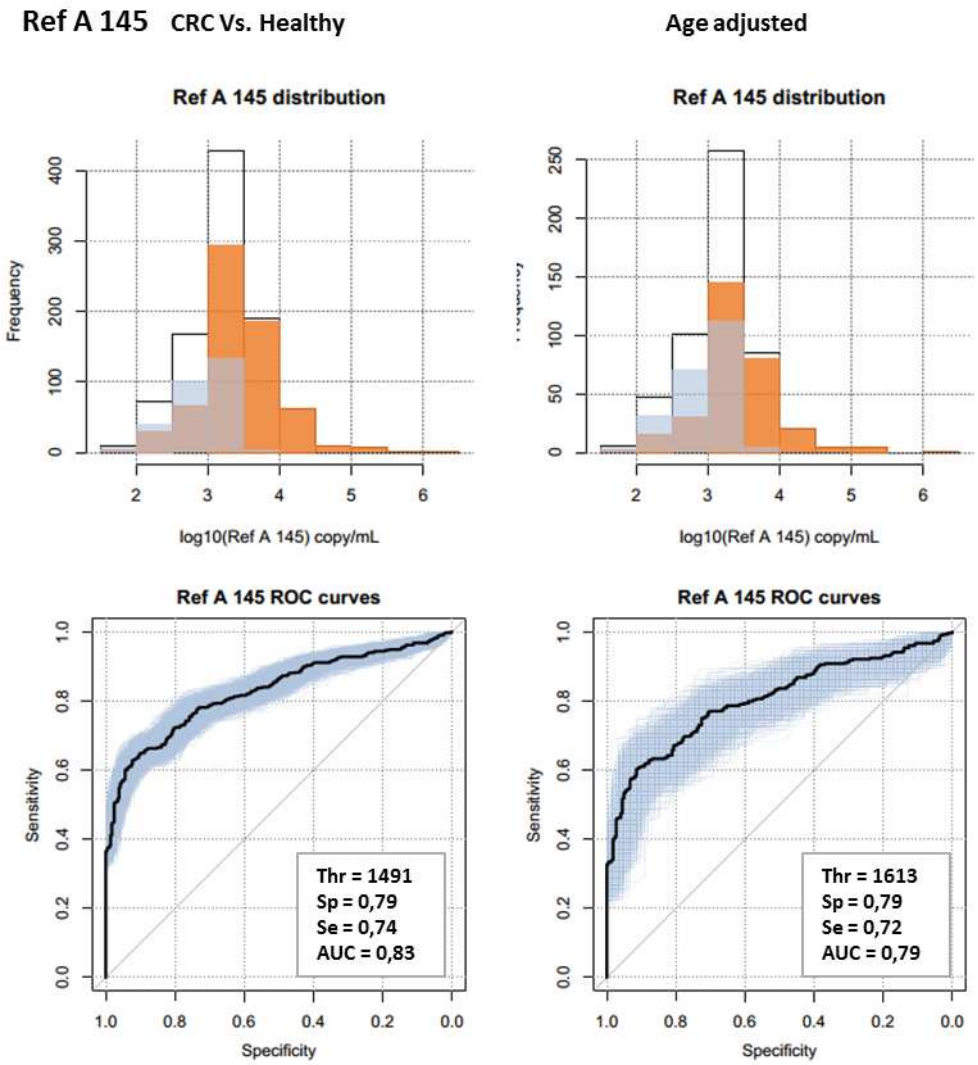


Fig. S7. Distribution and ROC curves of the Ref A 145 (copy nb/ml) parameter in CRC patients (orange) and healthy individuals (blue) before and after age resampling.

Ref A 145: nuclear cfDNA concentration of the fragments with a size ≥ 145 base pairs; ROC: receiver operating characteristics; Thr: Threshold; Sp: Specificity; Se: Sensitivity; AUC: Area under ROC curve.

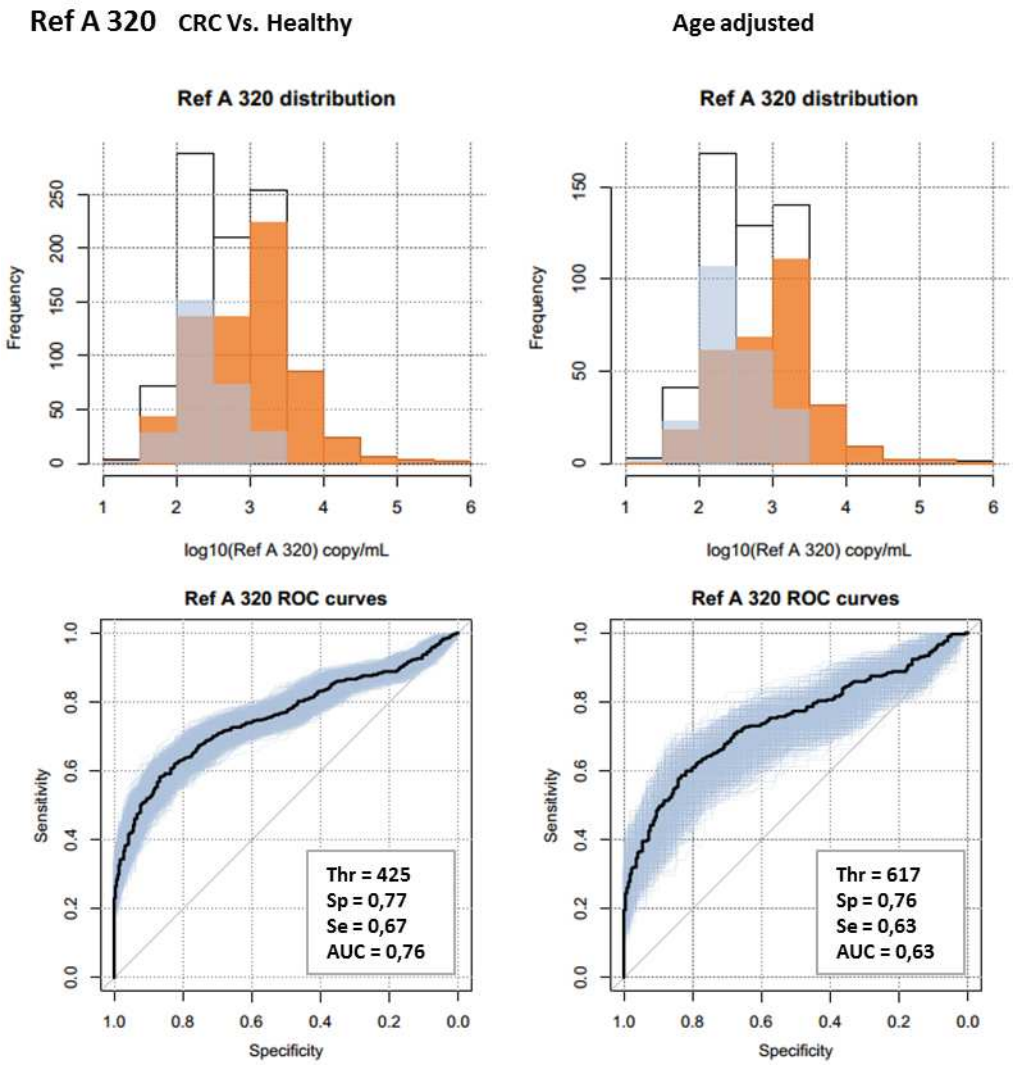


Fig. S8. Distribution and ROC curves of the Ref A 320 (copy nb/ml) parameter in CRC patients (orange) and healthy individuals (blue) before and after age resampling.

Ref A 320: nuclear cfDNA concentration of the fragments with a size ≥ 320 base pairs; ROC: receiver operating characteristics; Thr: Threshold; Sp: Specificity; Se: Sensitivity; AUC: Area under ROC curve.

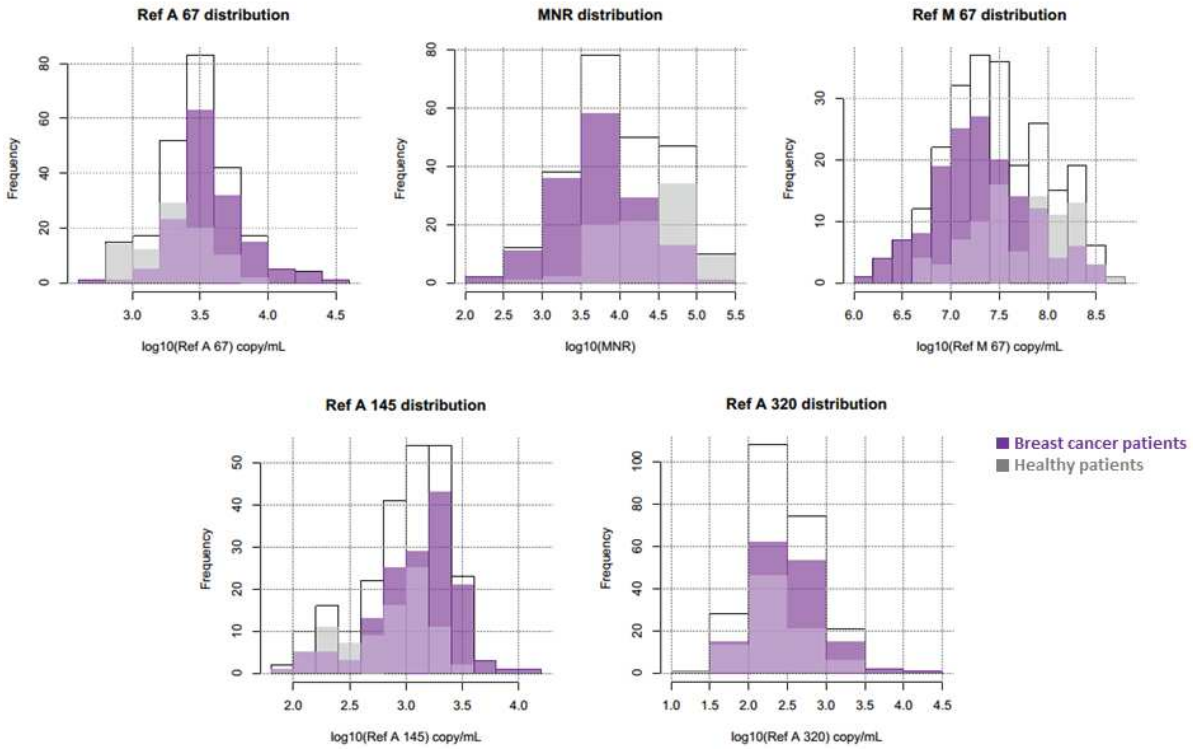


Fig. S9. Distribution of each measured parameter in breast cancer patients (in purple) and healthy individuals (in grey).

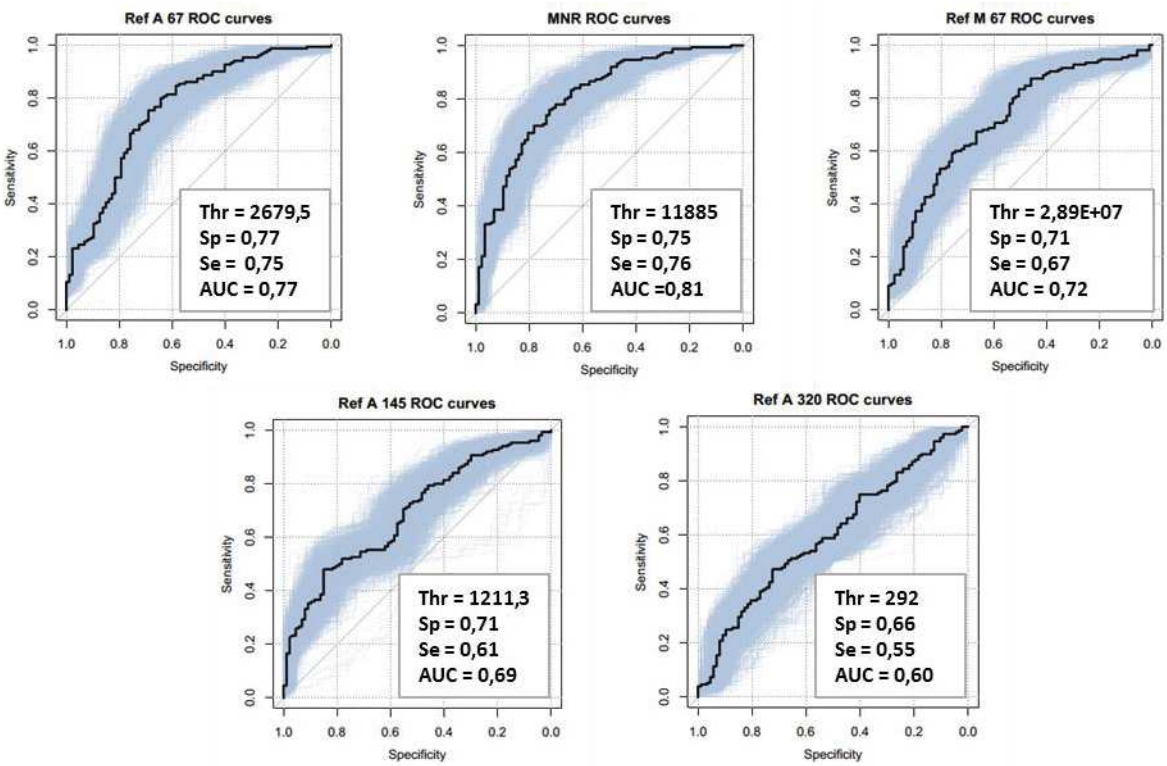


Fig. S10. ROC curves of each measured parameter on the breast cancer cohort after 2000 resampling for empirical bootstrapped estimation to build parameters confidence intervals. The original curve is presented in black.

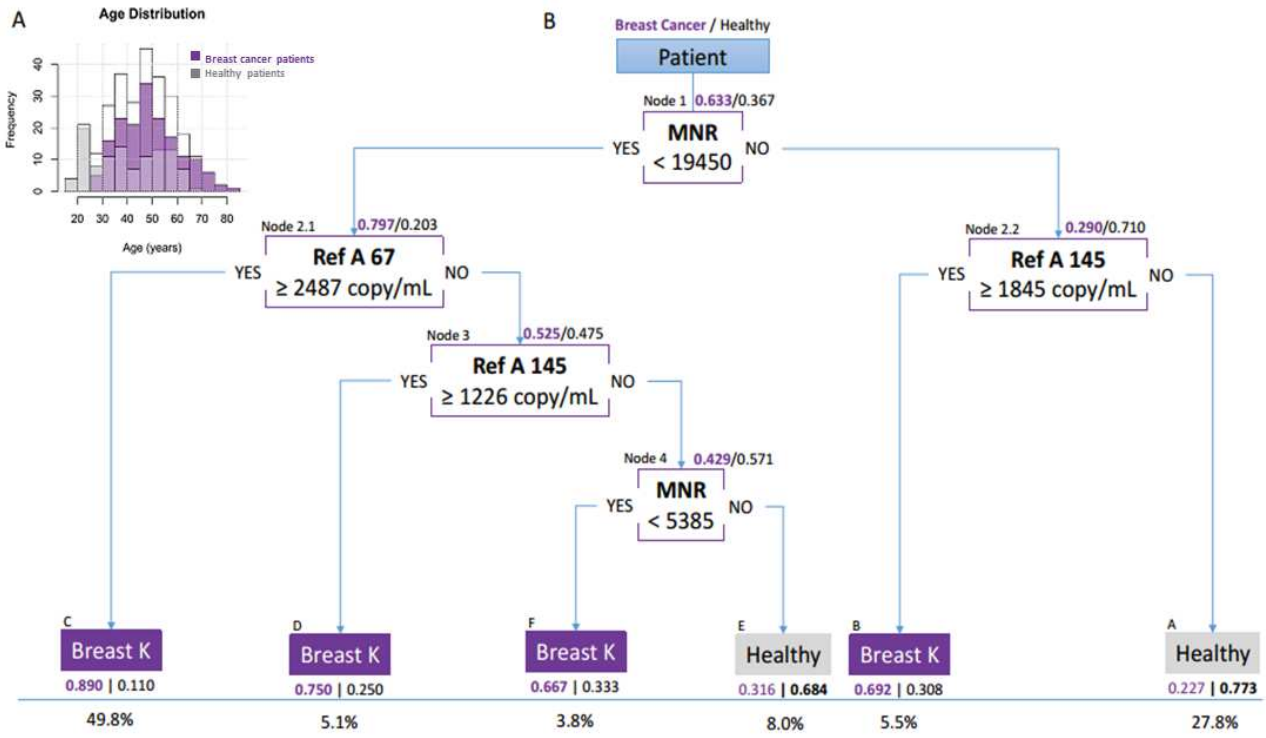


Fig. S11. A. Distribution histogram of the age among patients with breast cancer (purple) and healthy subjects (grey). **B.** Global decision tree obtained from the breast cancer cohort by iterating a recursive partitioning regression tree after resampling patients at each step. Proportions of patients with breast cancer (purple) and healthy patients (grey) are represented at each node. Each final leaf contains the name of the most representative population inside, with their relative proportions. The proportion of patients that ended in each leaf is presented below the corresponding leaf. One node features the parameter of interest, the condition on the parameter (below) and the direction of the answer (YES/NO).

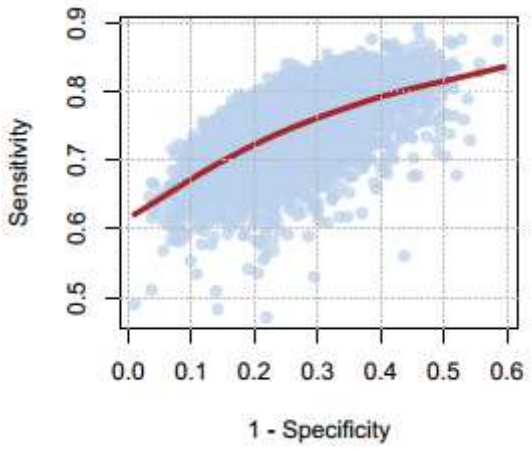


Fig. S12. Plot of sensitivities and specificities obtained by creating multiple trees with resampling and cross-validation and its loess curve (in red).

Table S1. Descriptive statistics of the colorectal cancer cohort.

	CRC Patients		Healthy Patients		
	used	entire cohort	Used	entire cohort	
	665	794	287	289	
Variable	Frequency	Proportion	Frequency	Proportion	p value
Gender - female	283	42.60%	109	40.00%	0.698
	Mean	S.D.	Mean	S.D.	p value
Age (years)	70.4	11.7	43.4	14.0	<0.00001
Ref A 67 (copy/mL)	17456.5	62933.2	2473.4	1476.5	<0.00001
Ref A 145 (copy/mL)	8805.8	49244.8	1021.5	699.0	0.00005
Ref A 320 (copy/mL)	5772.5	40694.0	392.4	431.9	0.00069
Ref M 67 (copy/mL)	90164077.0	218748965.5	85632814.2	87677092.9	0.648
MNR (Ref M 67 / Ref A 67)	16354.5	50162.3	45396.3	57714.4	<0.00001

Descriptive statistics of the colorectal cancer cohort after removing patients with missing values concerning the parameters of interest. Qualitative variables –frequency and proportion- are presented with the p-value resulting of a two samples proportions test with continuity correction. Continuous variables –mean and standard deviation- are presented with the p-value resulting of an unpaired two samples t-test.

Table S2. Estimation of the thresholds used for each node of the decision tree.

Tree localization		Ponctual estimation			95% Confidence interval	
Node	Parameter	**Median**	Mean	Sd	low	high
CRC Decision Tree						
1	Ref A 67	**5348**	5218.1	572.5	3616	5818
2	MNR	**13450**	17440	10032	3020	34560
3	Ref A 145	**1388**	1507.1	388.8	451	2104
4.1	Ref A 67	**4259**	3987.7	505.6	2692	4312
4.2	Ref M 67	**8661000**	9388360	3734777	6807000	21270000
5	MNR	**6726**	5874.5	1844.6	2744	8644
Breast Cancer Decision Tree						
1	MNR	**19450**	22792.5	11088.6	7468	37830
2.1	Ref A 67	**2487**	2455.8	291.5	1855	2928
2.2	Ref A 145	**1845**	1883.9	330	1398	2494
3	Ref A 145	**1226**	1193.2	360.4	410.9	1765
4	MNR	**5385**	6930.5	3017.4	3724	13330

Confidence intervals were obtained using empirical bootstrap method on 2000 independent resampling.

Table S3. At risk patients' predictions and their clinical parameters.

missing	**Healthy**	**CRC**	**p-value**
21 (20.2%)	22 (26.5%)	61 (73.5%)	1.86E-05
Diverticulosis	2 (9.1%)	11 (18%)	0.5175
Fibrosis	0	1	1
Polyp	9 (40.9%)	11 (17.9%)	0.06287
RCH	0	1	1
Adenomous	6 (27.3%)	19 (31.1%)	0.9453
Dysplasia	0	5 (8.2%)	0.3883
Treated	1 (4.5%)	6 (9.8%)	0.7504
Metastasis	2 (9.1%)	4 (6.6%)	1
Nothing	2 (9.1%)	3 (4.9%)	0.8551

Chi-square test was used to compare distribution of predicted cancer status.

Table S4. Descriptive statistics of the breast cancer cohort.

	Breast Cancer Patients n = 169		Healthy Patients n = 109		
	mean	**S.D.**	**mean**	**S.D.**	**p-value**
Age (years)	49.4	11.9	40.7	14.1	<0.00001
Ref A 67 (copy/mL)	4629.7	4040.9	2411.1	1460.9	<0.00001
Ref A 145 (copy/mL)	1664.1	1428.0	954.4	672.8	<0.00001
Ref A 320 (copy/mL)	636.6	1497.1	340.1	374.7	0.02315
Ref M 67 (copy/mL)	40213853.3	62578589.7	85697790.8	85962478.6	0.00003
MNR (Ref M 67 / Ref A 67)	11223.9	17749.4	45204.7	47141.6	<0.00001

Descriptive statistics of the breast cancer cohort after removing patients with missing values concerning the parameters of interest. The variables are presented with their mean, standard deviation- and the p-value resulting of an unpaired two samples t-test between breast cancer and healthy patients.

Table S5. ROC curves results on the breast cancer cohort with the use of empirical bootstrap method to build confidence intervals.

	Threshold		Specificity		Sensitivity		AUC	
Parameter	**Value**	IC95%	**Value**	IC95%	**Value**	IC95%	**Value**	IC95%
Ref A 67	**2679.5**	[2431.3 - 3123.7]	**0.77**	[0.61 - 0.81]	**0.75**	[0.66 - 0.86]	**0.77**	[0.71 - 0.83]
MNR	**11884.8**	[4321.9 - 12869.4]	**0.75**	[0.66 - 0.85]	**0.76**	[0.66 - 0.85]	**0.81**	[0.76 - 0.87]
Ref A 145	**1211.3**	[868.9 - 1593.5]	**0.71**	[0.53 - 0.89]	**0.61**	[0.44 - 0.76]	**0.69**	[0.62 - 0.76]
Ref A 320	**291.6**	[226.8 - 432.2]	**0.66**	[0.52 - 0.86]	**0.55**	[0.34 - 0.65]	**0.60**	[0.53 - 0.70]
Ref M 67	**2.89e+07**	[1.78e+06 – 3.81e+07]	**0.71**	[0.59 - 0.85]	**0.67**	[0.50 - 0.77]	**0.72**	[0.66 - 0.88]

2000 resampling have been done for the estimations. Parameters estimations for each variable on the breast cancer cohort were based on the distance between corresponding ROC curve and the point (Se=1; 1-Sp=0), with the use of an empirical bootstrap method to build their respective confidence interval. These estimations were obtained after patient age pairing. Se: Sensitivity, Sp: Specificity; AUC: Area under the curve.

Table S6. Prediction results over the other cancer types using the tree obtained on the CRC and breast cancer cohort.

	Colorectal Cancer TREE prediction		Breast Cancer TREE prediction	
	Cancer	Not Cancer	Cancer	Not Cancer
ADP	2	1	3	0
HCC	18	0	18	0
Lymphoma	1	0	1	0
Ovarian cancer	0	1	0	1

Contingency table of predictive status after applying the decision trees built using both the colorectal cancer and breast cancer cohort.

Table S7. Logistic model results concerning possible interaction between age and the different parameters.

	Univariate		Adjusted on age			Interaction with age		
Parameters	**Value**	**p-value**	**Value**	**p-value**	**age p-value**	**estimation**	**p-value**	**interaction p-value**
Ref A 67 (copy/mL)	5.19E-04	<0.00001	4.48E-04	<0.00001	<0.00001	1.19E-03	<0.00001	<0.00001
Ref M 67 (copy/mL)	1.43E-10	0.723	6.95E-10	0.212	<0.00001	5.30E-09	0.17	0.2
Ref A 320 (copy/mL)	0.001412	<0.00001	1.16E-03	<0.00001	<0.00001	2.57E-03	0.0123	0.1549
Ref A 145 (copy/mL)	0.0011382	<0.00001	9.96E-04	<0.00001	<0.00001	2.49E-03	0.000655	0.031487
MNR (Ref M 67 / Ref A 67)	-1.64E-05	<0.00001	-5.63E-06	0.026	<0.00001	-2.78E-05	0.134	0.212

Logistic parameters estimation and their associated p-value to evaluate the effect of each measured parameter on colorectal cancer probability in the cohort and the age's effect. First, the effect of the parameter alone was evaluated on a single univariate logistic model, secondly a multivariate model using the age as an adjusted covariable was built. Finally, a multivariate logistic model was applied with the age and the interaction between age and measured parameter as adjusted covariables.

Table S8. Table obtained using regression trees based on age resampling of patients from the CRC cohort.

Parameter	Estimation	IC low	IC high	Best
Sensitivity	**0.742**	0.65	0.851	0.788
Specificity	**0.746**	0.595	0.93	0.835

A training set of 206 patients was used among the 310 obtained at each resampling, the resulting tree was applied on the patients left, and the sensitivity and specificity of the cross-validation results were calculated. Confidence intervals were built by empirical bootstrapping method on 2000 resampling.

Table S9. Characteristics of the selected primers to study nuclear and mitochondrial cfDNA.

HUMAN SEQUENCES				
Gene	**PRIMER NAME**	**Orientation**	**SEQUENCE 5' - 3'**	**AMPLICON SIZE (bp)**
MITOCHONDRIAL DNA				
MT-CO3	MIT MT-CO3 F	SENSE	GACCCACCAATCACATGC	67
	MIT MT-CO3 R 67	ANTISENSE	TGAGAGGGCCCCTGTTAG	
NUCLEAR DNA				
KRAS	KRAS B2 inv k	ANTISENSE	CCCTGACATACTCCCAAGGA	
	KRAS B1 inv k	SENSE	CCTTGGGTTTCAAGTTATATG	67
	KRAS 145 E	SENSE	GATAAAGGTTTCTCTGACCA	145
	KRAS A1 inv k	SENSE	GCCTGCTGAAAATGACTGA	320
BRAF	BRAF A1	SENSE	TTATTGACTCTAAGAGGAAAGATGAA	105
	BRAF A2	ANTISENSE	GAGCAAGCATTATGAAGAGTTTAGG	

MURINE SEQUENCES				
Gene	**PRIMER NAME**	**Orientation**	**SEQUENCE 5' - 3'**	**AMPLICON SIZE (bp)**
MITOCHONDRIAL DNA				
MT-CO1	MUMTCO1 F	SENSE	GTCCCACTAATAATCGGAGC	114
	MUMT CO1 REV C	ANTISENSE	TGCTTCTACTATTGATGATGC	
NUCLEAR DNA				
KRAS	KRAS 63-382 Mf	SENSE	AAGAGTGAAGACCCGTGTGC	63
	KRAS 63 Mr	ANTISENSE	GGAGAACAAGCACCCAACAG	

E. Conclusion, discussion et perspectives

De nombreux efforts ont été déployés pour développer de nouvelles méthodes sensibles et spécifiques de dépistage et de détection précoce du cancer. Nous avons utilisé ici une approche de « machine learning », ou d'apprentissage automatique, en combinant différents paramètres quantitatifs de l'ADNcir obtenus par q-PCR pour traiter cette question. Nous avons développé un modèle de prédiction basé sur des arbres de décision pour la classification des individus sains et des patients atteints de cancer. Cette approche pourrait permettre de détecter précocement et de manière non invasive des patients atteints de cancer.

• Evaluation des paramètres individuels d'ADNcir

A l'aide d'un modèle murin de souris xénogreffées précédemment établi par l'équipe *(129)*, nous avons comparé différents paramètres quantitatifs d'ADNcir d'origine murine (normale) et humaine (tumorale). Chacun des paramètres étudiés, comme la concentration totale d'ADNcir nucléaire (Ref A 67) ou mitochondriale (Ref M 67), a montré une différence significative entre l'ADN murin normal et l'ADN humain tumoral. Le MNR (Mitochondrial to Nuclear Ratio), ou le rapport entre la concentration d'ADNcir mitochondrial et nucléaire, a montré le pouvoir discriminant le plus élevé. Ces paramètres quantitatifs ont été ensuite validés *in vitro* dans le milieu de culture de cellules pour évaluer leur capacité à discriminer des cellules normales et de cellules cancéreuses. Nous avons montré que le MNR avait un potentiel significatif de discrimination entre lignées cellulaires normales et tumorales avec une AUC (aire sous la courbe ROC) égale à 1. L'évaluation dans le plasma d'une petite cohorte exploratoire de 80 individus sains et 147 patients atteint de CRC a confirmé que la concentration totale d'ADNfc nucléaire et le MNR pourraient être des facteurs discriminants (AUC = 0,82 et 0,98 respectivement). Ces variables ont ensuite été évaluées dans le plasma d'une importante cohorte de 289 individus sains et de 982 patients atteints de cancer, dont 791 CCR, 169 cancers du sein et 23 autres types de cancer. Les stades précoces 0/I/II constituaient environ 54 % de la cohorte des CCR. Chacun de ces paramètres présentait un bon potentiel discriminatoire entre patients sains et cancéreux. Cependant, les AUC, les spécificités et les sensibilités étaient inférieures à ce que nous avions observé précédemment

en raison du grand nombre d'individus dans la cohorte et du pourcentage élevé de patients atteints de cancer de stade précoce.

Des paramètres relatifs à la structure de l'ADNcir ont également été étudiés. Il a été démontré par plusieurs équipes *(128, 131, 183)* y compris la nôtre *(126, 130)*, que la fragmentation et le profil de taille de l'ADNcir diffèrent entre les patients sains et les patients cancéreux, et que ces paramètres pourraient constituer un marqueur potentiel de détection du cancer. Pour cette raison, nous avons rajouté deux autres paramètres à notre étude : le Ref A 145 (concentration des fragments d'ADNcir nucléaire ayant une taille ≥ 145 paires de bases) et le Ref A 320 (≥ 320 pb) qui fournissent des informations sur la distribution de taille des ADNcir. Le Ref A 145 s'est révélé très performant pour les patients atteints d'un cancer colorectal mais moins significatif pour les patientes atteintes d'un cancer du sein. Concernant le Ref A 320, un faible potentiel de discrimination a été observé pour les deux cohortes de patients cancéreux.

La plupart des méthodes basées sur l'analyse d'ADNcir utilisées actuellement pour la différenciation entre les patients cancéreux et les individus sains impliquent la détection de mutations somatiques *(116, 156, 184)*. Cependant, ces méthodes pourraient montrer une faible sensibilité chez les patients à un stade précoce étant donné le nombre limité de mutations récurrentes *(185)* et compte tenu du fait que la même mutation peut être associée à différents types de cancers. De plus, l'ADNcir peut provenir non seulement de cellules cancéreuses mais aussi de cellules sanguines susceptibles aux altérations génétiques, notamment avec l'âge. Notre approche permet de contourner cette limitation en étudiant des paramètres quantitatifs et structurels de l'ADNcir, provenant non seulement des cellules cancéreuses mais aussi des cellules saines.

- L'association des différents paramètres selon un arbre de décision

Il semble exister un lien entre l'âge et les différentes variables d'ADNcir calculées. Pour cette raison, les différents paramètres recherchés ont été ajustés en fonction de l'âge avant d'être combinés sous forme d'un arbre de décision pour la classification des individus sains et cancéreux. Le but de cet algorithme était d'élaborer un outil visuel pour une application clinique. Cette approche est adaptée à ce type de situation où certains paramètres sont interdépendants *(186)*, comme le MNR qui est basé sur le Ref A 67 et le Ref M 67. Elle permet

également de combiner la plupart des paramètres disponibles et d'éliminer certains problèmes inhérents aux interactions entre les différentes variables.

Ces approches basées sur des arbres de décision ont déjà montré de bonnes performances par rapport à d'autres techniques d'apprentissage automatique utilisées pour le pronostic de mortalité dans le cas du gliome par exemple *(187)*. Cependant, l'utilisation de méthodes courantes comme celles utilisées pour discriminer des patients atteints d'un carcinome épidermoïde de la tête et du cou *vs* des patients sains *(188)*, avec une validation croisée, en utilisant les 2/3 des valeurs comme ensemble d'apprentissage et le 1/3 restant pour la validation, a donné des résultats variables. En appliquant une méthode de ce type à notre cohorte, de grands intervalles de confiance ont été observés pour la sensibilité et la spécificité, et les arbres de décision obtenus étaient fortement dépendants des patients sélectionnés pour l'étape de construction. Ceci nous a conduit à ré-échantillonner des patients appariés selon l'âge, et à construire un arbre global sur la base de la variable la plus récurrente à chaque nœud. La procédure que nous avons utilisée a également été utilisée par d'autres équipes dans le domaine de la prédiction des tentatives de suicide *(189)* ou pour la détection de l'usage des drogues *(190)*. L'arbre a ensuite été appliqué au reste des patients non utilisés pour l'apprentissage, ce qui nous a donné des résultats plus stables avec des intervalles de confiance plus faibles.

- Stades précoces et patients à risque

Les patients atteints de CCR de stades précoces (stades 0, I et II) ont été pour la plupart prédits comme cancéreux par l'arbre décisionnel construit pour la cohorte de CCR, avec une sensibilité de 0,72 et une spécificité de 0,87. La cohorte du cancer du sein étudiée comprenait des patientes de stade II ou III, et la validation croisée de l'arbre de décision pour cette cohorte a montré une sensibilité de 0,80 et une spécificité de 0,95. Cela tend à nous encourager sur la capacité de cette approche pour la détection précoce du cancer. Cohen *et al.* ont développé un test dénommé CancerSeek pour la détection non invasive et la localisation de huit types de cancer (stades I à III) en combinant la détection de mutations spécifiques de tumeurs dans l'ADNcir avec huit marqueurs protéiques circulants *(116)*. Même si le test a montré une spécificité élevée de 0,99 avec une sensibilité médiane de 0,70 (sensibilité maximale = 0,98

pour le cancer de l'ovaire), la sensibilité était de 0,33 pour le cancer du sein et de 0,65 pour le CCR. En ce qui concerne les stades précoces, une sensibilité moyenne de 0,43 (de 0,2 pour le cancer de l'œsophage à 1 pour le cancer du foie) a été trouvée pour les patients de stade I, et de 0,73 pour ceux de stade II. D'autres méthodes basées sur l'ADNcir, comme l'analyse du méthylome de l'ADNcir provenant des tumeurs, pourraient permettre la détection, la classification et la surveillance du cancer dans plusieurs types de cancer *(117)*. L'étude du profil de méthylation de l'ADNcir isolé chez 100 patients atteints d'un cancer du sein ou d'un CCR métastatique et 45 personnes saines, a montré une efficacité diagnostique avec une sensibilité de 0,84 et une spécificité de 0,82 *(191)*, mais n'a pas permis la détection d'un cancer à un stade précoce.

Dans notre étude, les patients à risque n'ont pas pu être discriminés des patients atteints d'un cancer colorectal (Sp = 0,24), ce qui indique qu'ils partagent un profil similaire concernant les paramètres mesurés d'ADNcir. De plus, aucun profil spécifique concernant les paramètres cliniques de cette sous-population n'était corrélé avec la classification de ces patients comme cancéreux ou sains. Toutefois, la stratification de ces personnes en tant que patients atteints de CCR pourrait être bénéfique étant donné qu'il est conseillé de suivre régulièrement les patients présentant ces profils.

- Spécificité du test concernant le type de cancer

L'une de nos préoccupations était d'étudier la spécificité de ces paramètres en prenant en compte le type de cancer. Nous avons alors utilisé l'arbre prédictif établi sur une cohorte afin de prédire les résultats de l'autre cohorte. La prédiction utilisant l'arbre décisionnel de la cohorte du cancer colorectal s'est avérée mauvaise pour les patientes atteintes d'un cancer du sein (Se = 0,58) y compris lorsque la spécificité était assez élevée (0,90). Cette Sp élevée peut s'expliquer par le fait que les mêmes patients sains ont été utilisés dans les deux algorithmes. Au contraire, l'utilisation de l'arbre du cancer du sein pour classer les patients atteints d'un cancer colorectal était assez efficace (Se = 0,85) mais la perte de spécificité (0,66) était significative. Cela pourrait être dû au fait que la cohorte de patientes atteintes d'un cancer du sein ne concerne que les femmes, et qu'une différence pourrait exister entre les hommes et les femmes sains concernant les paramètres mesurés par l'ADNcir, en particulier

pour l'ADNcir mitochondrial *(150)*. Ceci expliquerait également l'importance du MNR, qui apparaît au premier noeud de l'arbre décisionnel pour le cancer du sein, dans la classification de ces individus. Ces algorithmes ont ensuite été appliqués à une petite cohorte de patients atteints d'autres types de cancer (n=23), et ont montré de bonnes prédictions compte tenu de la sensibilité, ce qui tend à indiquer que ces arbres ne seraient pas spécifiques du type de cancer.

- Limitations de l'étude

Bien que notre méthode puisse s'appliquer à différents types de cancer, elle ne permet pas de déterminer le type et la localisation du cancer comme dans le cas de l'étude CancerSeek *(116)* ou de l'analyse du profil de méthylation de l'ADNcir *(117)*. De plus, notre méthode est principalement basée sur des paramètres quantitatifs de l'ADNcir qui sont également connus pour varier dans des conditions physiologiques et pathologiques autres que le cancer comme les maladies auto-immunes *(28, 29)*, le sepsis *(38, 39)*, l'infarctus du myocarde *(41, 42)*, l'exercice *(192)* et autres. Par conséquent, cette approche devrait être validée pour évaluer sa capacité à faire la distinction entre patients cancéreux et patients atteints d'autres types de troubles comme l'inflammation par exemple. Malgré ces limites, la capacité de détecter la moitié ou les trois quarts des patients atteints d'un cancer à un stade précoce pourrait permettre un dépistage précoce *(185)*.

- Perspectives

Ajout d'autres paramètres

Afin d'augmenter le potentiel de notre méthode, l'ajout d'autres paramètres tels que le profil de méthylation de l'ADNcir ou ou la détection d'altérations génétiques, pourrait être bénéfique. De plus, il a été démontré par notre équipe que l'étude du profil de fragmentation de l'ADNcir, par séquençage de nouvelle génération suite à la préparation d'une librairie à simple ou double brin, pourrait servir de biomarqueur potentiel pour le cancer *(132)*, et pourrait être incluse dans notre approche d'apprentissage automatique.

Confirmation sur individus asymptomatiques

Une validation plus large de cette méthode devrait être envisagée sur une plus grande cohorte de patients à risque ainsi que sur des populations d'individus asymptomatiques pour une éventuelle application dans le dépistage du cancer.

Cause(s) de la variation du MNR entre état normal et état cancéreux

Nous avons montré que le taux d'ADN extracellulaire d'origine nucléaire augmente dans le cas du cancer par rapport au cas normal ou sain, alors que celui de l'ADNcf d'origine mitochondriale diminue, engendrant ainsi une diminution du MNR. Pour expliquer cette différence dans le comportement de ces deux ADN d'origines différentes, nous avons étudié au début l'effet de l'apoptose et de la nécrose sur leur libération. Nous avons induit l'apoptose ou la nécrose au niveau des cellules SW620 et DLD1 en culture en les traitant par des concentrations croissantes d'agents pro-apoptotiques ou pro-nécrotiques comme la staurosporine ou l'H_2O_2 respectivement. L'induction de la mort cellulaire a été validée par cytométrie de flux (FACS) **(Figures 15 et 16)**. L'ADN a été extrait à partir du milieu de culture des cellules, et l'ADNcf nucléaire et mitochondrial a été quantifié. Aucune différence significative de concentrations d'ADN extracellulaire nucléaire (Ref A 67) et mitochondriale (Ref M 67) **(Figure 17)** et du MNR **(Figure 18)** n'a été observée entre les contrôles (cellules non traitées) et les cellules en apoptose ou en nécrose. Ces observations suggèrent que la mort cellulaire par apoptose ou nécrose n'est pas à l'origine de la différence du comportement entre l'ADNcf d'origine différente. D'autres hypothèses, comme la différence du contenu en ADN mitochondrial entre les cellules saines et les cellules cancéreuses et/ou une possible dégradation de l'ADNmt au niveau de la circulation devraient être testées.

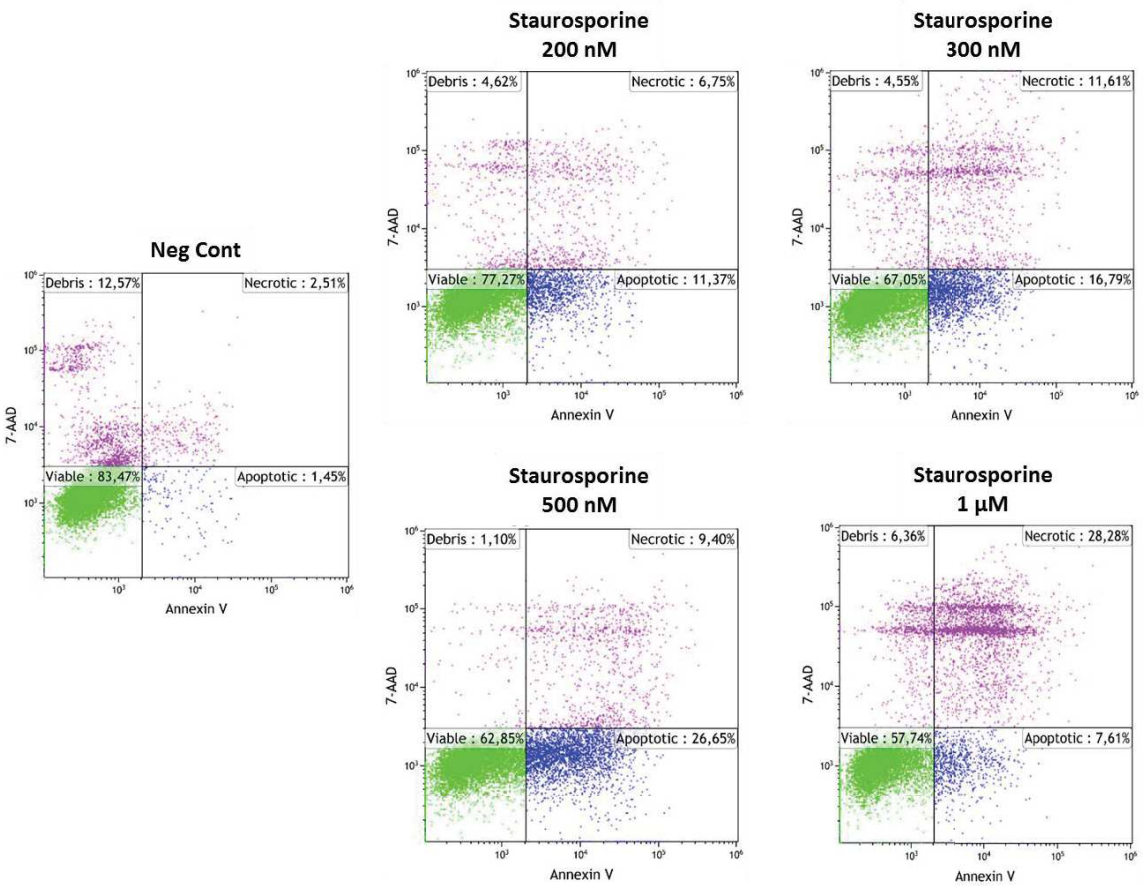


Figure 15: *Validation par FACS de l'induction de l'apoptose au niveau des cellules SW620 suite au traitement avec des concentrations croissantes de staurosporine*

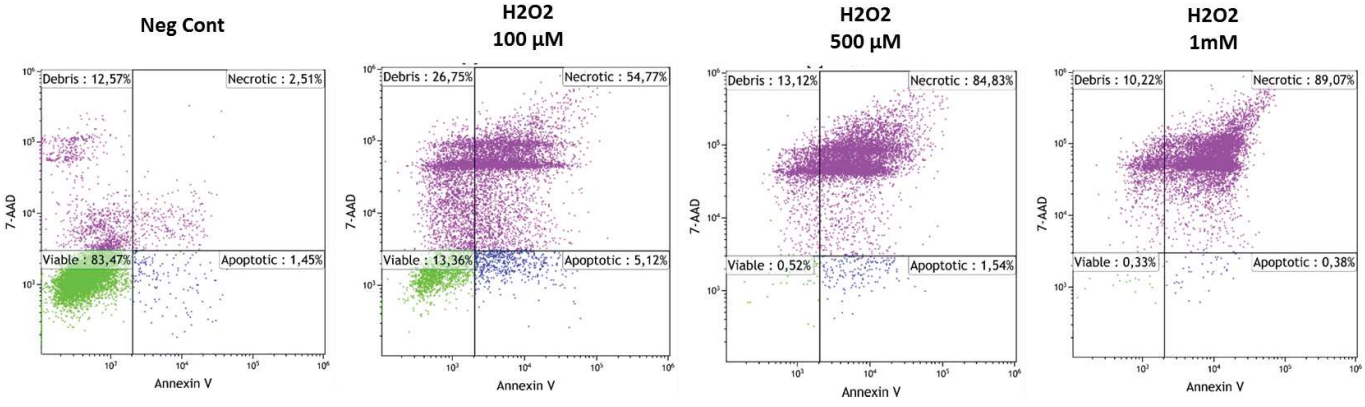


Figure 16: *Validation par FACS de l'induction de la nécrose au niveau des cellules SW620 suite au traitement avec des concentrations croissantes de H_2O_2*

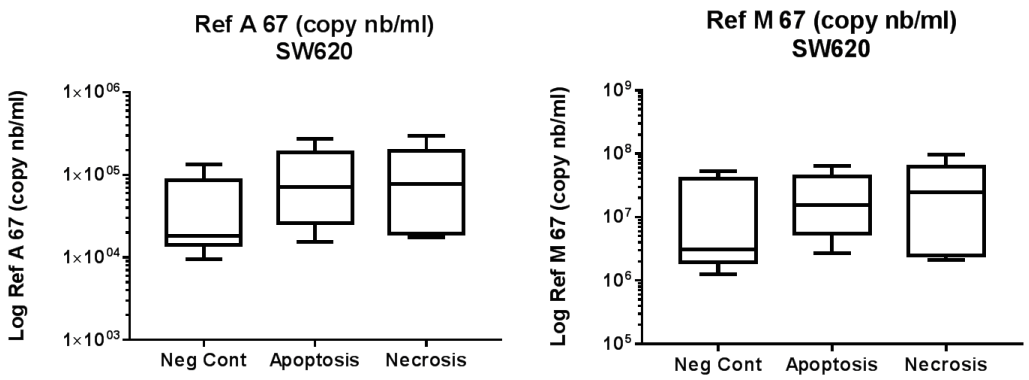


Figure 17: *Concentration d'ADNcf nucléaire (Ref A 67 nb de copies/ml) et mitochondrial (Ref M 67 nb de copies/ml) dans le milieu de culture des cellules SW620 contrôles (non traitées), en apoptose ou en nécrose*

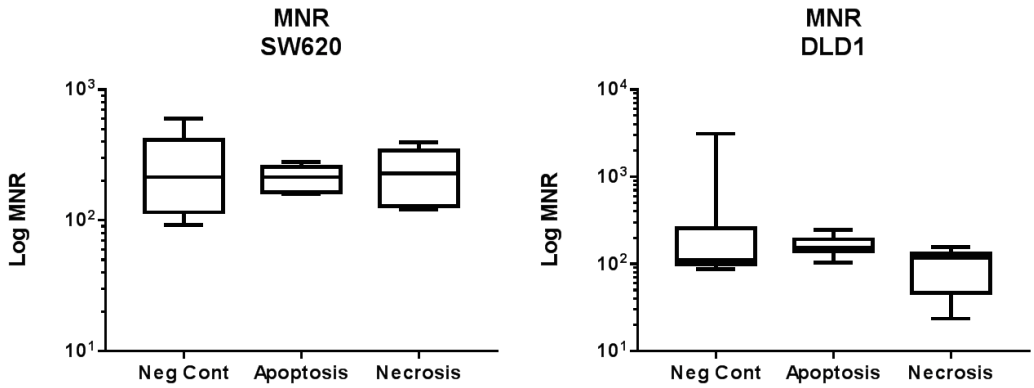


Figure 18: *La différence du MNR (Mitochondrial to Nuclear ratio) dans le milieu de culture des cellules SW620 et DLD1 contrôles (non traitées), en apoptose ou en nécrose*

IV. ADN extracellulaire et test génétique préimplantatoire

Lo *et al.* ont montré que de l'ADN d'origine fœtal circule dans le plasma des femmes enceintes *(6)*, ce qui a conduit à la mise en place d'une approche non invasive pour le diagnostic prénatal *(18)* et au développement de nouveaux tests de diagnostic et de pronostic pour la détection d'aneuploïdie chromosomale chez le fœtus, comme la trisomie 21 et 28 *(22, 23, 193)* par exemple, ou encore l'identification du sexe du fœtus *(24)*, et autres.

La pertinence de l'utilisation de l'ADN extracellulaire (ADNcf) a également été évaluée dans le domaine des techniques de procréation médicalement assistée (ART pour Assisted Reproductive Technology). Les échecs répétés d'implantation constituent en effet un problème réel en ART ; il est donc nécessaire d'identifier de nouveaux biomarqueurs pour évaluer la qualité des embryons et prédire le taux de réussite des implantations. Dans ce but, l'ADNcf a été recherché dans les liquides biologiques et les milieux de culture des embryons **(Figure 19)** afin de tenter de développer des tests génétiques préimplantatoires non invasifs et plus fiables.

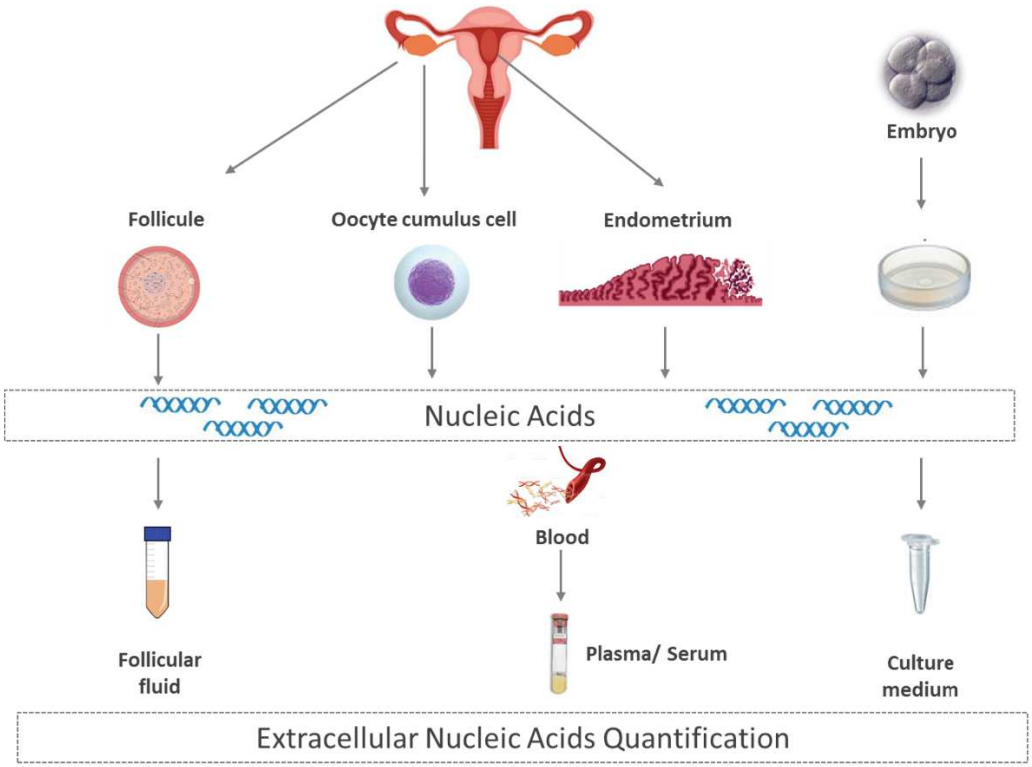


Figure 19: *Différentes sources d'ADN extracellulaire pour des tests non invasifs en diagnostic prénatal et préimplantatoire*

A. ADN circulant et diagnostic prénatal

Pendant la grossesse, des molécules d'ADN fœtal passent dans la circulation maternelle et se mélangent à l'ADN maternel circulant *(6)* **(Figure 20)**. L'ensemble du génome fœtal est représenté dans le plasma maternel *(135)* et constitue 3 à 6% de l'ADN total circulant chez la mère *(194)* . Ces molécules d'ADNcir d'origine fœtale sont rapidement éliminées avec des taux indétectables dès le premier jour après l'accouchement *(195)*. Cet ADN a été utilisé pour le développement de différents tests prénataux non invasifs (TPNI), à partir du sang de la mère, pour évaluer le risque que le fœtus naisse avec certaines anomalies génétiques *(196)*.

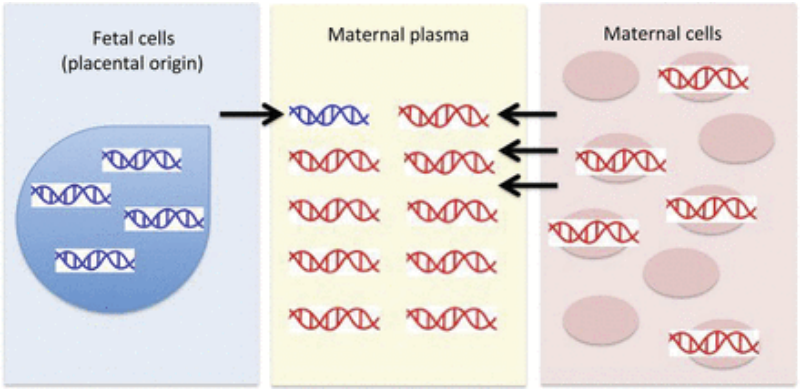


Figure 20: *L'origine de l'ADNcir dans le sang d'une femme enceinte (197)*

L'analyse de cet ADN a permis en premier lieu de déterminer le sexe du fœtus en détectant dans le plasma de la mère des séquences du chromosome Y présentes chez le fœtus masculin et absentes chez la mère *(194, 198)*. Le diagnostic prénatal du sexe est réalisé dans le cadre des maladies récessives liées au sexe *(199)*, pour lesquelles les gènes impliqués sont présents sur le chromosome X, comme dans le cas de l'hémophilie *(200)*. Pour ces pathologies, un fœtus masculin, qui n'a qu'un chromosome X, est plus susceptible de présenter des troubles qu'un fœtus féminin avec ses 2 copies du chromosome X, où une anomalie du gène sur un chromosome X est compensée par le gène sain porté par l'autre chromosome.

Des TPNI pour le génotypage du rhésus *(25, 201)* ont également été utiles chez des femmes enceintes rhésus négatif sensibilisées à la RhD, dont les partenaires sont hétérozygotes pour ce gène. Cela concerne alors les femmes ayant développé des anticorps dirigés contre cet

antigène en raison d'une grossesse antérieure. La détermination du rhésus du fœtus pourrait dans ce cas éviter l'administration répétée à la mère d'immunoglobulines anti-D passifs destinés à inhiber la réponse immunitaire aux antigènes RhD lorsque le fœtus est rhésus négatif.

Les TPNI sont par ailleurs souvent utilisés pour détecter des aneuploïdies chromosomiques chez des femmes à risque élevé pour ces anomalies. Par exemple, en 2008, l'équipe de Lo a utilisé une méthode de séquençage haut débit pour détecter la trisomie 21 dans le plasma maternel *(202)*. Cette technique a montré une sensibilité et une spécificité très élevées *(203)*. Cette nouvelle approche pour le diagnostic prénatal non invasif a été validée par d'autres méthodes, pour différentes aneuploïdies chromosomiques fœtales, comme celles affectant le chromosome 18, le chromosome 13 et les chromosomes sexuels *(193, 204–206)*. À l'heure actuelle, des tests non invasifs pour les trisomies 13, 18 et 21, ainsi que d'autres aneuploïdies, sont disponibles dans plusieurs pays, et des millions de femmes enceintes sont testées chaque année dans le monde *(207)*.

En revanche, le diagnostic prénatal des grossesses à risque de maladies monogéniques utilise toujours des techniques invasives, comme l'amniocentèse ou le prélèvement des villosités choriales **(Figure 21)**, qui causent de l'inconfort, présentent un risque de fausse couche et ne peuvent être pratiquées que pendant certaines périodes de la grossesse.

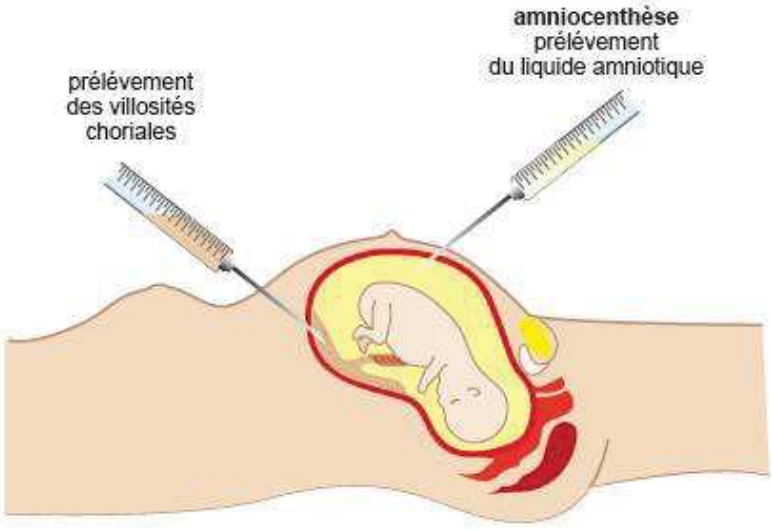


Figure 21: *Les techniques invasives utilisées pour le diagnostic prénatal comme l'amniocentèse et le prélèvement des villosités choriales (208)*

Ainsi, des approches technologiques ont été mises au point pour réaliser des TPNI destinés aux couples porteurs de mutations relatives à la bêta-thalassémie (une forme héréditaire d'anémie) *(209, 210)*, et la mucoviscidose (maladie génétique affectant les épithéliums glandulaires de nombreux organes) *(211, 212)* par exemple, ainsi que pour la recherche de mutations *de novo* responsables de certaines maladies monogéniques chez le fœtus dans le plasma maternel *(19)*.

Toutefois, les TPNI présentent plusieurs limitations et font encore l'objet de défis à relever. Par exemple, une fraction faible d'ADN fœtal dans la circulation maternelle peut réduire la sensibilité du test. Par conséquent, toute approche permettant de détecter des quantités infimes d'ADN ou de mieux distinguer ADN maternel et ADN fœtal améliorerait fortement la validité clinique des TPNI *(213)*. De plus, la présence d'ADN maternel dans la circulation interfère avec l'analyse de l'ADN fœtal et rend le diagnostic prénatal non invasif des mutations héréditaires maternelles impossible, contrairement aux mutations héréditaires paternelles *(214)*. Pour s'affranchir de ces limitations, Yu *et al.* ont montré que les TPNI basés sur la taille des fragments d'ADN dans la circulation et non sur leur dénombrement pourraient être pertinents, compte tenu du fait que l'ADN d'origine fœtale présente une taille plus petite que l'ADN maternel *(136)*. Cette approche pourrait être applicable à toutes les grossesses, indépendamment du genre et du profil génétique du fœtus.

B. Procréation médicalement assistée et test génétique préimplantatoire

1. Les premières étapes du développement embryonnaire

La fécondation est la première étape du développement embryonnaire. Après pénétration de la zone pellucide de l'ovocyte (enveloppe glycoprotéique assurant sa protection et son alimentation) par le spermatozoïde **(Figure 22)**, la membrane plasmique du spermatozoïde fusionne avec celle de l'ovocyte pour induire son activation *(215)*. L'ovocyte entame alors sa deuxième division méiotique pour devenir un ovule haploïde et libère le deuxième globule polaire, cellule haploïde avec une petite quantité de cytoplasme et n'ayant généralement pas la capacité d'être fécondée. Le noyau du spermatozoïde fusionne alors avec celui de l'ovule et la zone pellucide subit une transformation chimique qui la rend imperméable à d'autres spermatozoïdes.

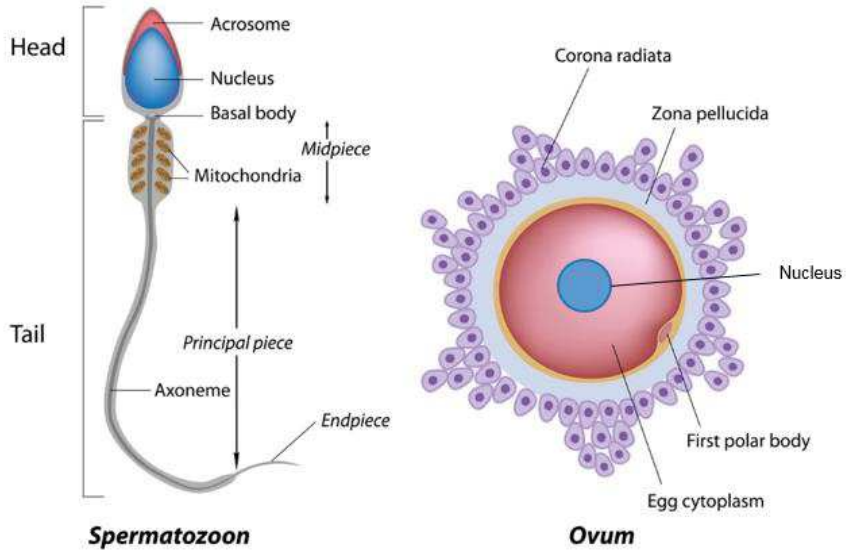


Figure 22: *Structure d'un spermatozoïde et d'un ovule (adaptée de 186)*

Suite à la fécondation de l'ovule par un spermatozoïde au niveau de l'appareil reproducteur féminin ou *ex-vivo* via les techniques de procréation assistée, le zygote est formé *(217)*. Au

premier jour du développement embryonnaire, il est formé d'une seule cellule et a la même taille que celle de l'ovocyte d'origine. Il est toujours entouré par la zone pellucide et contient les deux globules polaires formés pendant les deux divisions méiotiques de l'ovule **(Figure 23)**. Lors du développement embryonnaire, le zygote subit des divisions mitotiques formant 4 cellules symétriques au deuxième jour, et 8 cellules au troisième jour. Les cellules issues des premières divisions du zygote au stade de clivage sont appelées blastomères. Au quatrième jour l'embryon se trouve au stade de morula et comporte 12 à 16 cellules.

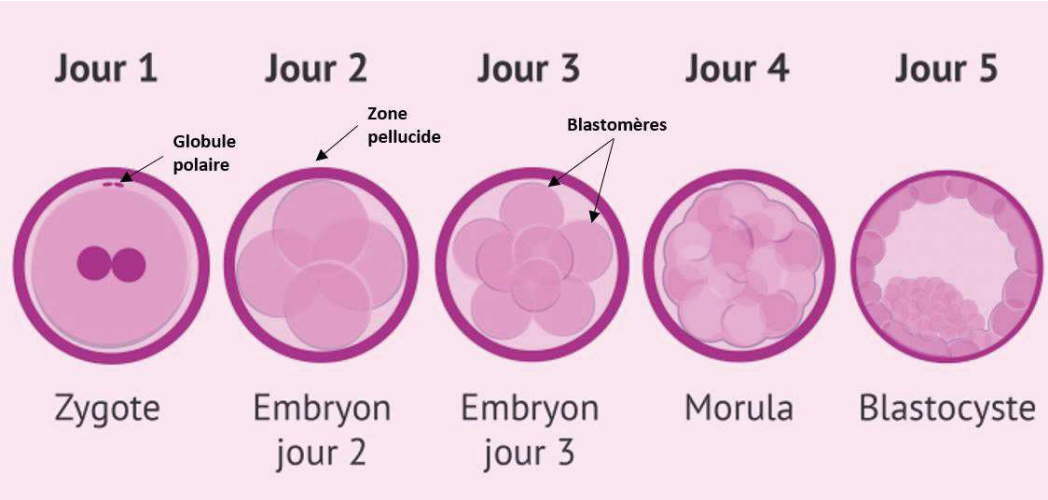


Figure 23: *Développement embryonnaire entre le jour 1 et le jour 5 (adaptée de 188)*

Au cinquième jour, l'embryon se transforme en blastocyste, commençant alors les premiers évènements de différentiation cellulaire **(Figure 24)**. Une cavité remplie de liquide appelée blastocèle se forme accompagnée de deux types de cellules : les cellules du trophectoderme qui vont former le placenta, et la masse cellulaire interne qui formera le fœtus.

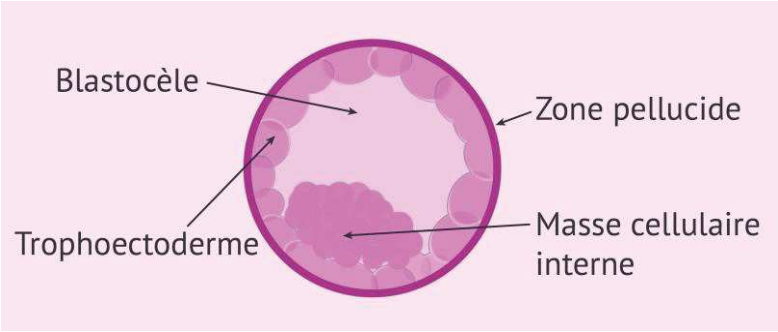


Figure 24: Structure d'un blastocyste *(218)*

2. Technologies de procréation médicalement assistée

Les technologies de reproduction assistée ou de procréation médicalement assistée regroupent l'ensemble des techniques et traitements médicaux destinés à favoriser la grossesse en cas de problèmes de fertilité masculins, féminins ou des deux partenaires **(Figure 25)**. Elles peuvent également être indiquées pour éviter la transmission d'une maladie génétique invalidante ou mortelle à la descendance.

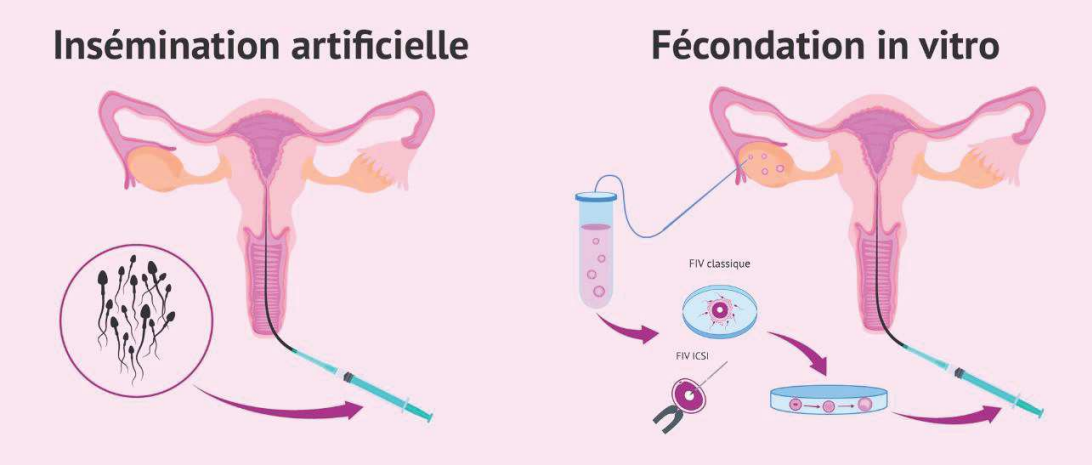


Figure 25: *Les différentes techniques de procréation médicalement assistée (adaptée de 162)*

L'insémination artificielle est la méthode de procréation assistée la plus ancienne. Elle consiste à déposer un échantillon de sperme à l'intérieur de l'utérus de la femme afin d'augmenter les chances de fécondation de l'ovule. Elle respecte au maximum l'environnement naturel des gamètes et est utilisée dans le cas de troubles de l'ovulation ou des problèmes d'infertilité masculine.

La fécondation *in vitro* (FIV) est la technique de reproduction assistée la plus courante et la plus efficace. Elle consiste à stimuler les ovaires par traitement hormonal pour induire une maturation de nombreux follicules *(121)*. Les ovocytes localisés à l'intérieur de ces follicules sont alors prélevés par aspiration du liquide folliculaire et mis au contact de spermatozoïdes *ex-vivo*, dans un milieu favorable à leur survie. L'étape de fécondation peut aussi s'effectuer par **injection intracytoplasmique de spermatozoïde (ICSI** pour intracytoplasmic sperm injection) dans l'ovule. Il s'agit d'une technique relativement récente qui implique l'injection

d'un spermatozoïde dans le cytoplasme d'un ovocyte mature à l'aide d'une micro-pipette. L'ICSI est la méthode la plus souvent utilisée avec un taux de réussite de 30% contre 25% pour la FIV classique *(220)*. Les embryons issus d'ovules fécondés sont alors implantés dans l'utérus de la mère ou éventuellement congelés pour une implantation ultérieure. Ce processus s'effectue normalement au stade du clivage, c'est-à-dire au jour 3 du développement embryonnaire, mais il est de plus en plus fréquemment effectué au jour 5 qui correspond au stade de développement des blastocystes. Cette technique est parfois associée à des grossesses multiples en raison du nombre d'embryons (deux ou trois) implantés dans l'utérus, ce qui augmente les risques d'accouchement prématuré, d'insuffisance pondérale à la naissance et met en danger la mère et l'enfant *(221)*. C'est la raison pour laquelle, les efforts en médecine moderne de la reproduction ont pour objectif de mettre en place de nouvelles stratégies destinées à améliorer le choix de l'embryon à implanter afin de passer progressivement au transfert d'un seul embryon **(voir partie 3 : Comment s'effectue le choix des embryons à transférer lors d'une FIV ?)**

Une FIV peut être associée à un test génétique préimplantatoire. Ce test ressemble au diagnostic prénatal pratiqué pour dépister certaines maladies génétiques avant la naissance, mais il permet surtout de sélectionner des embryons ne portant pas d'anomalies génétiques graves avant implantation dans l'utérus **(voir partie 2 : le test génétique préimplantatoire : généralités)**.

3. Le test génétique préimplantatoire : généralités

Le Test Génétique Préimplantatoire (PGT pour preimplantation genetic testing) permet de détecter, chez l'embryon, des altérations génétiques ou des anomalies chromosomiques à l'origine de maladies graves, et d'éviter ainsi toute transmission à la descendance *(121)* **(Figure 26)**. Il s'adresse donc principalement aux couples à risque porteurs de ces anomalies. Il est basé sur l'analyse génétique d'un embryon obtenu après fécondation *in vitro*, avant son implantation dans l'utérus. Ce test est également proposé dans certains pays pour les couples dont la femme est âgée de plus de 38 ans où un risque élevé d'anomalie chromosomique peut exister chez l'enfant.

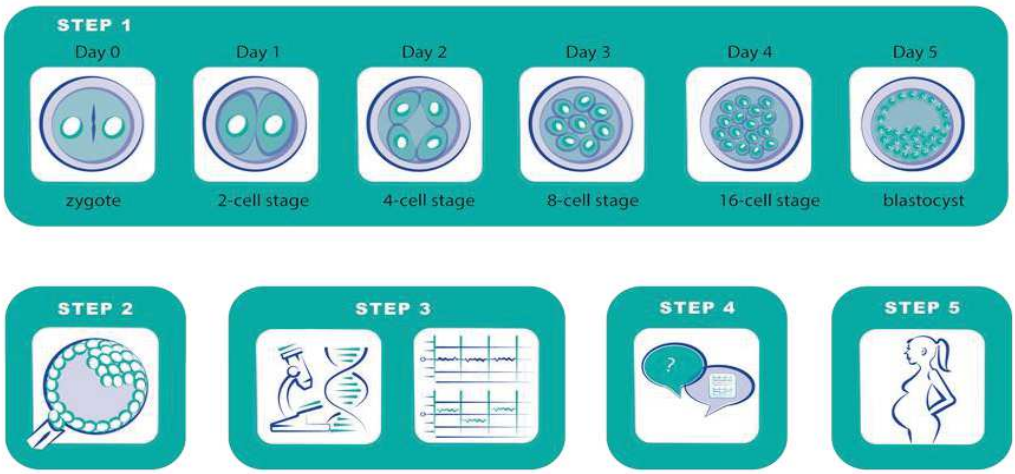


Figure 26: *Les différentes étapes d'un PGT.* ***Etape 1 :*** *FIV ;* ***Etape 2 :*** *Biopsie d'une ou plusieurs cellules de l'embryon ;* ***Etape 3 :*** *Recherche des altérations génétiques et des anomalies chromosomiques ;* ***Etape 4 :*** *Choix de l'embryon exempt de la mutation à transférer dans l'utérus de la mère ;* ***Etape 5 :*** *Implantation de l'embryon choisi (222)*

Il existe trois types différents de PGT *(223)*:

- PGT-M (Preimplantation Genetic Testing for monogenic diseases ou test génétique préimplantatoire pour les maladies monogéniques) : ce test permet de détecter les maladies héréditaires dues à une altération ou une mutation d'un seul gène. Il est donc utilisé pour des couples atteints d'une maladie génétique ou porteurs du gène muté et désirant éviter la naissance d'un enfant malade.

- PGT-A (Preimplantation Genetic Testing for aneuploidy ou test génétique préimplantatoire pour les aneuploïdies) : ce test permet de détecter les aneuploïdies chromosomiques, que ce soit des additions ou des pertes de fragments chromosomiques ou de chromosomes entiers. Parmi quelques exemples on peut citer les syndromes de Down (trisomie 21), de Turner (absence complète ou partielle d'un des deux chromosomes X) et de Klinefelter (mâle avec 2 chromosomes X). Ce test est destiné aux patientes présentant des échecs d'implantation répétés, des fausses couches récurrentes ou un âge maternel avancé.
- PGT-SR (Preimplantation Genetic Testing for structural rearrangements) : ce test permet de détecter les anomalies affectant le réarrangement de la structure d'un ou plusieurs chromosomes, comme les translocations par exemple. Les embryons porteurs de ces anomalies peuvent donner naissance à des enfants potentiellement stériles à l'âge adulte.

Pour effectuer un PGT, une biopsie est nécessaire pour accéder au matériel génétique du fœtus *(224, 225)* **(Figure 27)**. Cette biopsie peut s'effectuer au troisième jour du développement embryonnaire (J3), au stade du clivage, lorsque l'embryon est constitué de 6 à 8 cellules. La zone pellucide est ouverte au moyen d'un laser ou par des moyens mécaniques et chimiques, et 1 ou 2 blastomères sont prélevés pour l'analyse. Une biopsie au jour 5 (J5 : stade blastocyste) peut aussi être pratiquée. Dans ce cas 5 à 6 cellules du trophectoderme sont prélevées par micromanipulation du blastocyste constitué de 80 à 100 cellules. Dans ce cas, la proportion de la masse cellulaire embryonnaire prélevée est plus faible et l'échantillon de cellules ne contient pas la masse cellulaire interne. Il semblerait que la biopsie à J3, même en fournissant des échantillons adéquats, serait nuisible pour 3 embryons sur 5 et pourrait entrainer une perte de capacité implantatoire et une réduction de leur potentiel reproductif de 40% environ *(226)*. En revanche, une biopsie à J5 ne présente pas ces dangers. Une biopsie du globule polaire peut aussi être pratiquée sur l'ovocyte avant la fécondation, bien qu'elle présente l'inconvénient de ne permettre de déceler que les anomalies maternelles dans le cas des maladies monogéniques. Cette technique est plutôt utilisée dans les pays où la biopsie des cellules de l'embryon est interdite.

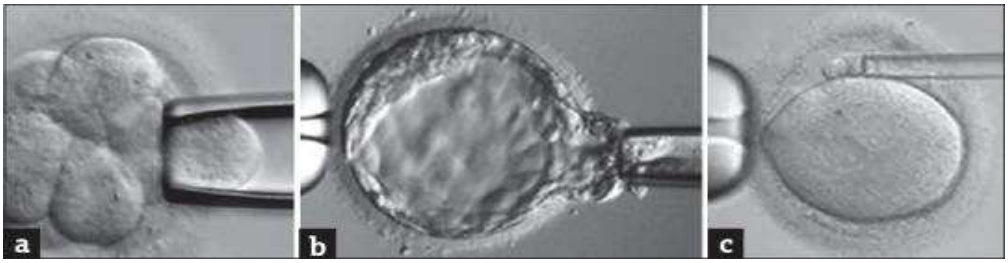


Figure 27: *Les différentes techniques de biopsie:* ***(a)*** *Biopsie d'un blastomère en J3 ;* ***(b)*** *Biopsie des cellules du trophectoderme en J5 ;* ***(c)*** *Biopsie du globule polaire (225)*

Après prélèvement, plusieurs techniques peuvent être utilisées pour effectuer l'analyse génétique de l'ADN embryonnaire *(225)*. L'hybridation *in situ* par fluorescence (FISH pour fluorescence in situ hybridization) a été la première technique utilisée et c'est qu'ainsi en 1993, des embryons mâles, femelles et atteints du syndrome de Turner ont pu être identifiés *(227)*. Cette technique permet de marquer des fragments chromosomiques spécifiques ou des chromosomes entiers par fluorescence et de mettre ainsi en évidence la présence de l'anomalie recherchée. Elle peut être donc utilisée pour les translocations ou inversions chromosomiques et la recherche d'aneuploïdie. Plusieurs autres méthodes comme la PCR, le NGS (next-generation sequencing) et autres sont utilisées de nos jours pour détecter la présence de différentes altérations génétiques *(225)*. Les résultats de cette analyse génétique par l'une ou l'autre de ces techniques permettent de décider quels embryons sont exempts de l'anomalie recherchée, et par suite optimaux pour l'implantation.

Cependant, les tests génétiques préimplantatoires soulèvent de nombreuses questions éthiques *(228)*. La nature invasive du test constitue l'une de ces questions puisque le fait de perturber l'embryon et de prélever une ou plusieurs cellules n'est pas toujours acceptable. De plus, les résultats issus des PGT déterminent si un embryon est implanté ou détruit ce qui pose la question de savoir si la vie commence dès la conception et si l'embryon doit être considéré comme une personne. Une autre question soulevée est celle de l'utilisation de ces tests pour déterminer le profil HLA des embryons et identifier ceux qui sont compatibles avec un frère ou sœur ayant besoin d'une transplantation de cellules souches ou d'une greffe d'organes *(229, 230)*. Les questions éthiques et juridiques relatives à cette pratique sont complexes et controversées. Ainsi le PGT pourrait être considéré comme un outil d'eugénisme, par élimination des caractères indésirables ou sélection des qualités désirées et constitue, de ce fait, une question morale très importante pour le futur.

4. Comment s'effectue le choix des embryons à transférer lors d'une FIV ?

A l'heure actuelle, l'ART a pour objectif d'implanter un embryon unique sain et viable *(231)*. De ce fait, pour optimiser le taux de réussite de la FIV, limiter le nombre d'embryons implantés et ainsi éviter le risque de grossesses multiples, la sélection des embryons les plus viables est essentielle *(232)*. Actuellement, les paramètres les plus utilisés pour le choix des embryons destinés à une FIV sont ceux concernant la morphologie de l'embryon et le taux de développement embryonnaire en culture *(118)*. Ces paramètres prennent en compte le nombre de cellules, la régularité cellulaire, le degré de fragmentation ... En effet, au deuxième jour du développement embryonnaire, l'embryon est en général constitué de quatre cellules (certains en ont deux, trois ou cinq) et les cellules doivent être de taille identique. De plus, des fragmentations de cellules peuvent avoir lieu au cours des divisions cellulaires, et lorsqu'un embryon présente plus de 50% de fragmentation on considère que ses chances de développement et d'implantation sont réduites. Cependant, bien que non invasive, l'observation morphologique des embryons présente des limites et une valeur prédictive relativement faible *(221)*, tout en étant subjective et en exigeant une grande expérience.

La surveillance du développement embryonnaire à intervalles de temps réguliers (time-lapse imaging) peut représenter un meilleur moyen non invasif de sélection d'embryon *in vitro (233)*. Grâce à une fréquence élevée, une image plus complète des processus biologiques permet d'évaluer d'une manière plus précise la morphologie de l'embryon. Différents groupes ont d'ailleurs proposé différents algorithmes potentiellement capables de sélectionner objectivement le ou les embryons à transférer *(234)*. Khosravi *et al.* ont aussi mis en place une approche par intelligence artificielle pour sélectionner les embryons à l'aide d'une vaste collection d'images prises à intervalle de temps régulier (environ 50 000 images) *(235)*. Leur méthode a permis d'obtenir une haute sensibilité lors de l'évaluation de la qualité de l'embryon tout en étant non invasive et entièrement automatisée.

Le test génétique préimplantatoire peut également être considéré comme un outil de sélection des embryons en particulier pour la recherche d'aneuploïdies. Cependant la nécessité d'une manipulation invasive des embryons a limité son utilisation en routine, les

tests non invasifs, ne causant aucun dommage à l'embryon, étant plutôt favorisés afin de ne pas perturber la croissance de l'embryon ou diminuer son potentiel implantatoire.

Diverses méthodes non invasives basées sur l'étude du milieu de culture de l'embryon dans les domaines de la protéomique et de la métabolomique ont par ailleurs été développées pour évaluer la qualité et la viabilité de l'embryon *(236)*. La fraction soluble de la protéine HLA-G (sHLA-G pour soluble human leukocyte antigen-G) dans le milieu de culture de l'embryon a été considérée comme un marqueur prédictif potentiel pour le développement embryonnaire. L'absence de sHLA-G dans le milieu avait une valeur prédictive négative *(237)*, et lorsque des embryons sHLA-G positifs ont été implantés dans l'utérus de la mère, le pourcentage d'implantation et de grossesse était de 44% et 75% respectivement, comparé à 14% et 23% pour des embryons sHLA-G négatifs *(238)*.

La métabolomique peut être une approche alternative en tant qu'outil d'évaluation rapide et non invasif du potentiel reproducteur et de la viabilité des embryons *(239)*. La mesure de l'absorption de pyruvate *(240)*, de l'augmentation de l'absorption du glucose *(241)*, le taux de renouvellement des acides aminés *(242)*, ainsi que la mesure de la fréquence respiratoire *(243)* et de la consommation d'oxygène *(244)* dans le milieu de culture peuvent être liés à l'évaluation de la viabilité des ovocytes et des embryons.

L'une des nouvelles approches permettant d'améliorer les résultats des FIV est l'évaluation du contenu de l'embryon en ADN mitochondrial. Un nombre élevé de copies d'ADNmt au niveau des embryons indique en effet une faible viabilité et est lié à un faible potentiel d'implantation *(245–247)*. De plus, des quantités élevées d'ADNmt dans les blastocystes sont associées à un état d'aneuploïdie *(248)*. Cependant, Victor *et al.* ont montré qu'il n'existe pas de différences significatives dans le taux d'ADNmt quel que soit le potentiel d'implantation et l'état de ploïdie de l'embryon *(249)*.

Toutes ces approches peuvent apparaître dans l'avenir comme une addition à l'évaluation morphologique traditionnelle des embryons basée sur son apparence.

5. L'ADN extracellulaire comme PGT et pour le choix des embryons ?

Le test génétique préimplantatoire (PGT) implique actuellement l'utilisation de biopsies relativement invasives pour les embryons, et le développement d'une approche moins ou non invasive est souhaitable. L'ADN du liquide folliculaire a été étudié en tant que biomarqueur potentiel en FIV *(250)*. Ce liquide est important pour le développement de l'ovocyte et contient des composants plasmatiques et des facteurs sécrétés par les cellules de la granulosa lors de la folliculogenèse. Le taux d'ADN extracellulaire présent au niveau de ce liquide corrèle significativement avec la qualité des embryons *(250)*, et constitue un facteur prédictif du succès ou l'échec de la grossesse *(251)*.

En outre, l'ADN embryonnaire extracellulaire présent dans le fluide du blastocèle ainsi que dans le milieu de culture de l'embryon est recherché comme source d'information génétique intéressante *(252)*.

Il a été rapporté que les embryons, lors d'une fécondation *in vitro*, libèrent de l'ADN dans leur milieu de culture *(119, 120)*. Des études ont également montré que le fluide du blastocèle contient de l'ADNcf *(253, 254)*. Cet ADN peut être libéré par l'embryon lors d'une fragmentation cellulaire au stade de clivage, ou lors d'une lyse cellulaire, apoptose ou nécrose des cellules de la masse cellulaire interne, ou présent dans les corps de fragmentation *(255)* **(Figure 28)**.

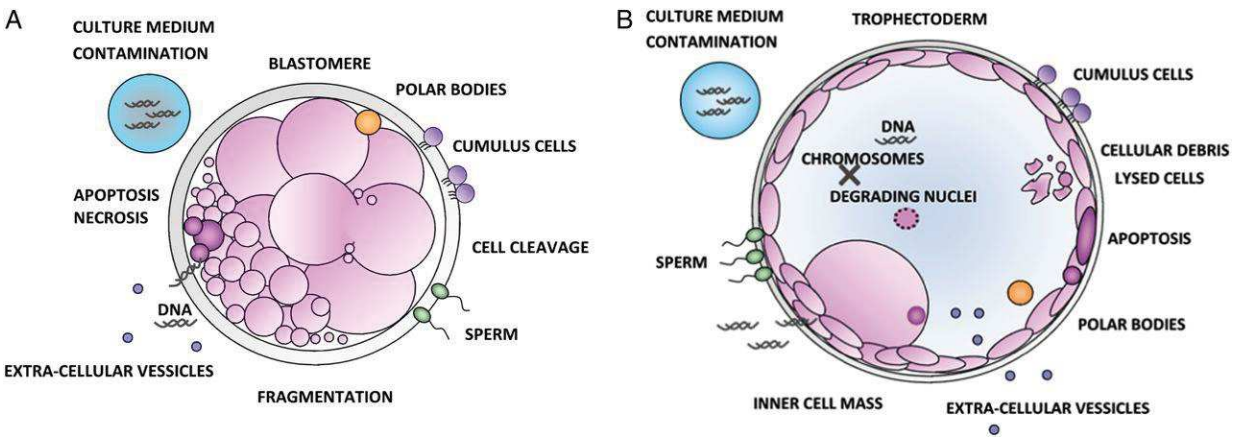


Figure 28: *Les sources potentielles et les mécanismes proposés de libération de l'ADN dans le milieu de culture des embryons au stade de clivage (A) et dans le fluide du blastocèle et le milieu de culture au stade blastocyste (B) (255)*

L'ADN peut être également présent dans des vésicules destinées à la communication entre les cellules du trophectoderme et la masse cellulaire interne. L'ADN peut traverser la zone pellucide et se présenter dans des vésicules extracellulaires. Une contamination par de l'ADN maternel peut avoir lieu et provenir des cellules du cumulus (cellules de la corona radiata), qui restent collées à la zone pellucide ou des corps polaires après leur expulsion de l'ovocyte, et une contamination paternelle peut provenir de spermatozoïdes qui adhèrent à la zone pellucide dans le cas d'une FIV classique.

Bien que l'isolement du fluide du blastocèle puisse représenter une approche peu invasive, l'analyse du milieu de culture des embryons est quant à elle complètement non invasive. L'utilité de cet ADNcf comme PGT a été démontrée dans plusieurs études pour augmenter le succès de l'implantation.

Le milieu de culture et le fluide du blastocèle ont été évalués comme sources d'ADN pour le génotypage de l'embryon *(256)*. Le polymorphisme C677T du gène de la méthylènetétrahydrofolate réductase (*MTHR*) a été recherché avec un taux de détection de 62.5% et 44.4% dans le milieu de culture et le fluide respectivement.

L'ADN embryonnaire extracellulaire a été largement étudié pour la détermination du statut de ploïdie de l'embryon. Une combinaison du milieu de culture du blastocyste et du fluide du blastocèle s'est révélée contenir suffisamment d'ADN embryonnaire pour l'amplification du génome et le dépistage précis de l'aneuploïdie *(257)*. Gianaroli *et al.* ont comparé les résultats obtenus par l'analyse du fluide du blastocèle et des cellules du trophectoderme et ont montré un taux de concordance de 96.6% quant à la ploïdie de l'embryon *(258)*. Une méthode de NGS a été validée pour le séquençage de l'ADN génomique sécrété par l'embryon dans le milieu de culture *(259)*, et a fourni l'information concernant la ploïdie des 24 paires de chromosome. Les résultats ont été validés en les comparant avec la biopsie de l'embryon correspondant testé, et une corrélation élevée a été retrouvée avec une sensibilité de 88% et une spécificité de 84%. D'autres méthodes ont montré également l'utilité de la détection non invasive de l'aneuploïdie par analyse de l'ADNcf dans le milieu de culture de l'embryon *(260–262)* avec des niveaux de concordance élevés avec l'analyse de la biopsie de l'embryon. Cette approche pourra ainsi identifier les anomalies chromosomiques avec des résultats cliniques satisfaisants sans avoir recours à la biopsie.

L'intérêt de l'ADN extracellulaire comme PGT a été étudié dans le cas de la délétion de l'α-thalassemia-^{SEA}, la forme la plus courante et la plus grave de l'α-thalassémie au Sud-Est de l'Asie et au sud de la Chine (SEA pour Southeast Asian deletion) *(263)*. L'efficacité diagnostique de détection de la mutation par l'analyse de l'ADN dans le milieu de culture est significativement plus élevée (88.6%) que celle par l'analyse d'une biopsie. La détermination du sexe de l'embryon constitue une autre utilisation de l'analyse de l'ADNcf, une application utile dans le cas des maladies génétiques liées au sexe *(120)*. Des séquences du gène *SRY* qui se trouve sur le chromosome Y ont été détectées dans le milieu de culture des embryons à partir du troisième jour du développement embryonnaire *(264)***.** La concordance pour la détermination du sexe des embryons entre le milieu de culture en J3 et la biopsie de l'embryon est de 81.3% *(262)*.

Le nombre de copies d'ADN mitochondrial chez l'embryon est corrélé avec sa qualité et sa viabilité *(245)*. C'est la raison pour laquelle l'ADN mitochondrial extracellulaire est recherché dans le milieu de culture *(255)*. La concentration de cet ADN est significativement associée à la qualité embryonnaire et plus précisément au taux de fragmentation de l'embryon avec une quantité d'ADN mitochondrial plus élevée dans le cas d'embryons présentant un clivage de mauvaise qualité *(265)*. Le rapport de la concentration d'ADN mitochondrial sur la concentration d'ADN nucléaire est plus élevé dans le milieu de culture des embryons de bonne qualité ayant atteint le stade de blastocyste complet au jour 5 *(266)*. Ce ratio, associé au classement morphologique, permet de mieux prédire la qualité de l'embryon et de le choisir en fonction de son potentiel de développement *(267)*.

C. Projet de thèse partie 2 : l'ADN extracellulaire pour le choix des embryons et la détection des altérations génétiques

1. Résultats antérieurs

Notre équipe a travaillé, en collaboration avec le Pr. Samir Hamamah du département de la biologie de la reproduction à l'hôpital Arnaud de Villeneuve à Montpellier, sur la mise en place d'une méthode non invasive pour détecter l'ADN extracellulaire (ADNcf) dans le milieu de culture des embryons afin d'évaluer la qualité de l'embryon et sa capacité à permettre une grossesse *(268)*. La méthode vise à déterminer, dans le contexte de la FIV, si un embryon est compétent et/ou porte une anomalie génétique ou une séquence spécifique. Les résultats préliminaires ont été obtenus sur un nombre réduit d'embryons et ont montré la présence de l'ADNcf dans le milieu de culture des embryons en J3 et J5/6 du développement embryonnaire **(Figure 29)** en ciblant un amplicon de 67 pb du gène nucléaire *KRAS*.

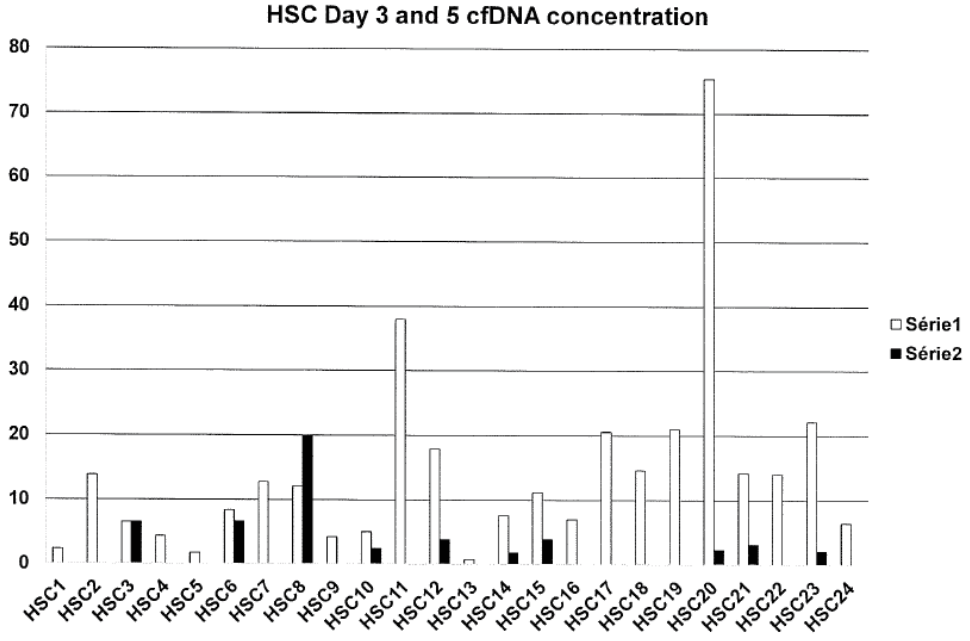


Figure 29: *Concentration de l'ADN extracellulaire (en ng/ml) dans le milieu de culture de différents embryons en J3 (série 1, histogrammes clairs) et en J5/6 (série 1, histogrammes noirs) (268)*

Un grade de 1 à 10 a été attribué à chaque embryon reflétant sa qualité sur la base de ses critères morphologiques (grade 1 pour embryon de mauvaise qualité et grade 10 pour embryon de bonne qualité). La concentration d'ADNcf nucléaire obtenue a été comparée pour onze embryons. Le taux en J5/6 ainsi que la différence de concentration en J5/6 et J3 étaient inversement proportionnels au bon développement de l'embryon **(Figure 30)**, cette différence apparaissant comme un marqueur du développement embryonnaire *in vitro*.

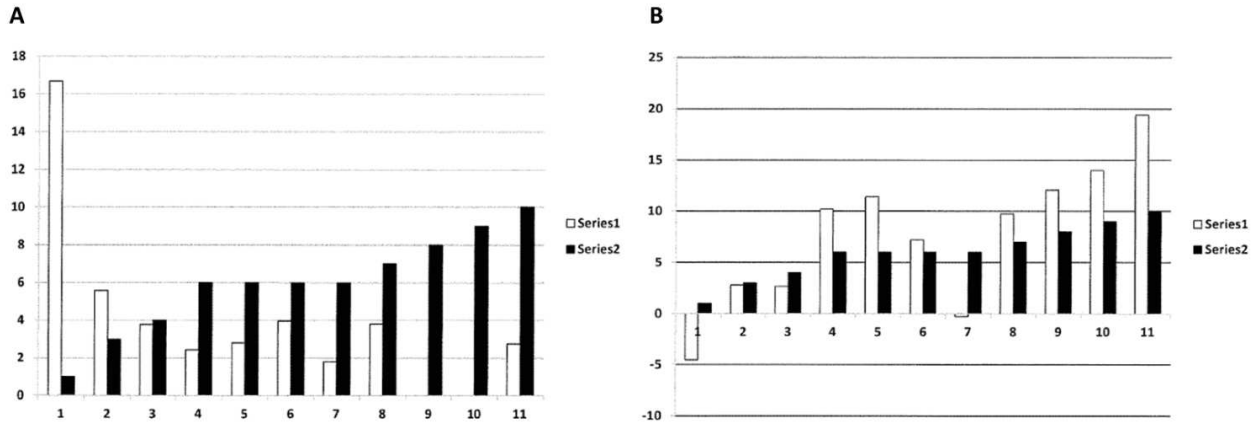


Figure 30: *Concentration de l'ADNcf en J5/6* ***(A)*** *et la différence de la concentration entre J5/6 et J3* ***(B)*** *par rapport au grade de 11 embryons différents. Les histogrammes noirs correspondent au grade de l'embryon. Les Histogrammes clairs représentent la concentration d'ADNcf dans le milieu en J5/6* ***(A)*** *ou la différence de concentration entre J5/6 et J3* ***(B)*** *(adaptée de 238).*

De plus, la quantité d'ADNcf dans le milieu de culture embryonnaire en J5/6 était plus importante chez les patientes pour lesquelles l'implantation embryonnaire a échouée que chez les patientes enceintes **(Figure 31)**.

	Day 5	
Patient	Cf DNA (ng/ml)	Pregnancy
HSC	2.75	Yes
BT	0.998	Yes
AGG	11.4	Yes
GS	18.57	No
MMB	13.16	No
LA	18.98	No

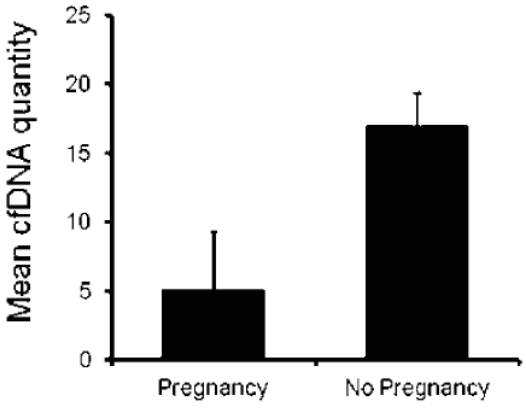


Figure 31: *Relation entre la concentration d'ADNcf dans le milieu de culture de l'embryon et l'issue de la grossesse (268).*

2. Résumé du travail

Un des objectifs de ma thèse était de poursuivre ce travail et de valider les résultats sur une cohorte plus importante. L'objectif était d'étudier la concentration d'ADN extracellulaire d'origine nucléaire et mitochondriale dans le milieu de culture des embryons lors d'une fécondation *in vitro* comme biomarqueur de qualité pour une évaluation plus précise de la viabilité embryonnaire et la prévention des échecs d'implantation.

De plus, notre objectif était d'évaluer la pertinence de l'utilisation de l'ADN extracellulaire comme test génétique préimplantatoire en détectant des séquences spécifiques et des altérations génétiques dans le milieu de culture des embryons pouvant être à l'origine de l'apparition de certaines maladies. En particulier, nous avons recherché le gène *SRY* (Sex-determining Region of Y chromosome) situé sur le chromosome Y pour la détermination du sexe de l'embryon, qui constitue une information utile dans le cas des maladies génétiques et des pathologies liées au sexe. De plus, nous avons tenté d'identifier des embryons porteurs d'altérations génétiques à l'origine de certaines pathologies. La mutation Delta F508, qui correspond à la mutation la plus fréquente responsable de la mucoviscidose, a été analysée chez un couple porteur de cette mutation et à risque de donner naissance à un enfant malade.

D. Matériels et méthodes

- **Procédure de la fécondation *in vitro***

Après une stimulation ovarienne contrôlée (traitement par un agoniste de la GnRH (hormone de libération des gonadotrophines) suivie d'une stimulation par la gonadotrophine ménopausique humaine (hMG) ou la FSH (hormone folliculostimulante) recombinante, les femmes ont subi un prélèvement d'ovocytes par ponction transvaginale et une fécondation par ICSI (intracytoplasmic sperm injection). La fécondation a été confirmée après 16 à 20 h par la présence de deux pronucléi distincts et de deux globules polaires sous le microscope inversé. Les zygotes ont ensuite été placés dans des gouttes individuelles de 30µl de milieu de culture (G1-Plus, Vitolife, Suède) et maintenus dans un incubateur à 37°C et à 5% d'oxygène. Les embryons ont été cultivés dans le milieu de culture G1 du jour 1 (J1) au jour 3 (J3) du développement embryonnaire, puis dans un milieu de culture G2 (G2-Plus, Vitolife, Suède) pour une culture prolongée jusqu'à J5 ou J6. Un grade de 1 à 10 a été attribué à chaque embryon en J3, J5 et J6. Ce grade est fonction de l'observation morphologique de l'embryon et reflète sa qualité. Ainsi, un embryon de mauvaise qualité sera gradé à 1, alors qu'un embryon de bonne qualité sera gradé 10.

- **Prélèvement des milieux de culture et extraction de l'ADN extracellulaire**

69 embryons ont été collectés chez 11 couples subissant une tentative de FIV. Après le retrait de l'embryon, les milieux de culture en J1, J3, J5 et J6 ont été récupérés et immédiatement conservés à -80°C. Une goutte témoin de milieu sans embryon incubée dans les mêmes conditions a été prélevée pour servir de contrôle. Avant l'extraction, les gouttes de milieux ont été décongelées et le volume initial a été complété avec du PBS 1X jusqu'à un volume final de 200µl. L'ADN extracellulaire a été extrait à partir des 200µl d'échantillon et élué dans un volume final de 80µl en utilisant le QIAamp DNA Blood Mini Kit (Qiagen, Courtaboeuf, France) selon notre protocole *(269)*. Les échantillons d'ADN ont été conservés à -20°C jusqu'à leur analyse.

- **Quantification de l'ADN par Q-PCR et calcul du nombre de copies**

Les réactions de Q-PCR ont été réalisées dans un volume réactionnel de 25µl sur un CFX96 touch™ Real-Time PCR detection system (Bio-Rad) en utilisant le logiciel CFX manager software. Chaque mélange réactionnel était composé de 12,5µl de SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad, Marnes-la-Coquette, France), 2,5µl de chaque amorce d'amplification (3 pmol/µl), 2,5µl d'eau sans nucléase (Qiagen), et 5µl d'extrait d'ADN. Le cycle thermique comporte trois étapes répétées : une étape d'activation-dénaturation de la polymérase de 3 minutes à 95°C, suivie de 40 cycles répétés à 90°C pendant 10 secondes, puis à 60°C pendant 30 secondes. Les courbes de fusion ont été obtenues en augmentant la température de 55°C à 105°C avec une lecture de plaque tous les 0,2°C. La concentration a été calculée à partir du Cq détecté par Q-PCR. Chaque échantillon a été analysé en triplicat pour chaque couple d'amorces et l'analyse a été répétée au moins une fois. Un triplicat de contrôle négatif a été inclus à chaque analyse pour chaque couple d'amorces utilisé.

Quantification de l'ADNcf nucléaire :

Des dilutions en série de l'ADN génomique de concentration connue provenant de la lignée cellulaire Difi ont été utilisées comme standard pour la quantification de l'ADN nucléaire. La concentration et la pureté initiales ont été évaluées par la densité optique à λ=260 nm, 230 nm et 280 nm avec un Eppendorf BioPhotometer® D30. Le Ref A 67 (nb de copies/ µl de milieu) correspond à la concentration totale d'ADN nucléaire extracellulaire. Cette valeur est calculée en utilisant la formule suivante :

$$Ref\ A\ 67 = \left(\frac{c}{3{,}3}\right) * \left(\frac{Velution}{Vmedia}\right)$$

« C » correspond à la concentration d'ADN nucléaire en pg/µl déterminée par Q-PCR en ciblant un amplicon de 67pb du gène *KRAS* sauvage, 3.3pg est la masse d'un génome haploïde humain, $V_{elution}$ est le volume final de l'extrait obtenu après extraction en µL et V_{media} le volume du milieu de culture utilisé pour l'extraction (en µl).

Quantification de l'ADNcf mitochondrial :

Une gamme étalon obtenue par dilution en série d'un vecteur ORF humain de 3382 pb avec un insert de 786 pb du gène mitochondrial *MT-CO3* de concentration connue obtenu de chez ABM good® (numéro d'accès YP_003024032) a été utilisée pour quantifier l'ADN mitochondrial extracellulaire. La concentration et la pureté initiales de la solution ont été déterminées en mesurant la densité optique à λ=260 nm, 230 nm et 280 nm, avec un BioPhotomètre Eppendorf® D30. La valeur moyenne du nombre de copies d'ADN mitochondrial extracellulaire ou Ref M 67 (nb de copies/µl de milieu) a été calculée en utilisant la formule suivante :

$$Ref\ M\ 67 = \left(\frac{c * Na}{2 * MW * Lvector}\right) * \left(\frac{Velution}{Vmedia}\right)$$

« C » est la concentration d'ADN mitochondrial (g/µL) déterminée par Q-PCR en ciblant un amplicon sauvage de 67pb du gène *MT-CO3*. NA est le nombre d'Avogadro (6.02 x 1023 molécules par mole), MW est la masse moléculaire (molecular weight) d'un nucléotide (g/mol) et L_{vector} correspond à la longueur du plasmide en nucléotides. $V_{elution}$ est le volume final de l'extrait obtenu après extraction en µL et V_{media} le volume du milieu de culture utilisé pour l'extraction (en µl).

- **Contrôles : analyse des milieux de culture neufs**

Des milieux de culture utilisés pour les différentes étapes du développement embryonnaire lors de la FIV et provenant de différents fournisseurs (Vitrolife, Origio et Irvine Scientific) **(Tableau 1)** ont été analysés pour tester l'hypothèse d'une contamination possible en ADN extracellulaire nucléaire et mitochondrial. L'ADNcf a été extrait à partir de 200µl de milieux de culture neufs dans un volume final de 80µl en utilisant le QIAamp DNA Blood Mini Kit (Qiagen, Courtaboeuf, France) selon notre protocole *(269)* et le Ref A 67 (nb copies/µl de milieu) ainsi que le Ref M 67 (nb de copies/µl de milieu) ont été calculés par Q-PCR.

Fournisseur	*Milieu testé*	
Vitrolife	G-1 Plus	Milieu de culture des embryons aux jours 2 ou 3 du développement embryonnaire
	G-2 Plus	Milieu de culture des embryons du jour 3 jusqu'au stade blastocyste (jour 5/6)
	G-IVF Plus	Milieu utilisé pour la préparation et la manipulation des gamètes pour la FIV
	G-MOPS Plus	Milieu pour la manipulation d'ovocytes et d'embryons dans l'atmosphère ambiante
Origio	Fert	Milieu qui soutient les gamètes pendant la fécondation et favorise la fonction des spermatozoïdes
	Cleav	Milieu qui favorise le développement embryonnaire précoce
	Blast	Milieu conçu pour répondre aux besoins métaboliques après l'activation du génome de l'embryon (blastocyste)
	Flushing Media	Milieu utilisé pour le prélèvement, la conservation et le lavage des ovocytes
Irvine Scientific	Continuous Single Culture Media	Milieu conçu pour la fécondation et la culture embryonnaire jusqu'au jour 5/6 du développement embryonnaire

Tableau 1 : *Liste des différents milieux de culture frais testés.*

- **Détection de séquences spécifiques et d'altérations génétiques**

Détection du gène SRY et détermination du sexe de l'embryon

Le gène *SRY* présent sur le chromosome Y a été recherché dans le milieu de culture des 69 embryons testés en J1, J3, J5 et J6 pour la détermination du sexe. Lorsque ce gène a été détecté en ciblant, par Q-PCR, un amplicon de 72pb avec une courbe de fusion ayant un Tm (température de fusion) de 77.6°C, l'embryon a été considéré mâle.

Recherche de la mutation Delta F508 dans le cas de la mucoviscidose

Le milieu de culture des embryons provenant d'un couple (BES VIR) où les deux parents sont porteurs de la mutation Delta F508 du gène *CFTR* (pour Cystic fibrosis transmembrane conductance regulator) à l'origine de la maladie de la mucoviscidose a été récupéré en J1, J3 et J5/6. L'ADN extracellulaire a été extrait selon la méthodologie déjà décrite. L'ADN a été ensuite concentré par précipitation dans un volume final de 10µl à l'aide du GenElute^{TM}-LPA (Sigma-Aldrich, St. Quentin Fallavier Cedex). La mutation a été recherchée par la méthode Intplex développée par l'équipe *(127)*. Cette méthode combine la détection de la mutation par une Q-PCR allèle-spécifique et l'utilisation d'un oligo-bloqueur porteur à son extrémité 3' d'un groupement phosphate ciblant la séquence sauvage de la mutation et empêchant ainsi une hybridation non spécifique probable du primer muté sur cette séquence. Des contrôles de qualité internes positifs (ADN génomique portant la mutation recherchée) et négatifs (ADN génomique sauvage pour la mutation) ont été intégrés à chaque analyse.

En J3, chaque embryon a subi un test génétique préimplantatoire ; un blastomère a été prélevé par micromanipulation et a été utilisé pour analyser le site F508 du gène *CFTR* dans un séquenceur automatique (ABI 3130XL). Les résultats ont été analysés par le logiciel Gen Mapper 4.0. à l'Institut Universitaire de Recherche Clinique à Montpellier.

- **Oligonucléotides**

Les séquences et les caractéristiques des amorces sélectionnées sont présentées dans le **Tableau 2**. Les amorces ont été conçues à l'aide du logiciel Primer 3 et les séquences ont été vérifiées pour « l'annealing » intra et intermoléculaire par les logiciels mfold et Oligo analyzer 1.2. Des analyses d'alignement par BLAST ont été réalisées pour confirmer la spécificité des amorces. Les oligonucléotides ont été synthétisés et purifiés par Integrated DNA Technologies (Coralville, Iowa) sur chromatographie en phase liquide à haute performance (HPLC) et le contrôle qualité a été effectué par MALDI-TOF.

GENE	PRIMER NAME	ORIENTATION	SEQUENCE 5' - 3'	AMPLICON SIZE (bp)
MITOCHONDRIAL DNA				
MT-CO3	MIT MT-CO3 F	SENSE	GACCCACCAATCACATGC	67
	MIT MT-CO3 R 67	ANTISENSE	TGAGAGGGCCCCTGTTAG	
NUCLEAR DNA				
KRAS	KRAS B1 inv k	SENSE	CCTTGGGTTTCAAGTTATATG	67
	KRAS B2 inv k	ANTISENSE	CCCTGACATACTCCCAAGGA	
SRY	SRY F1	SENSE	AGCTCTTCCTTCCTTTGC	72
	SRY F2	ANTISENSE	GCCTTTACTGTTTTCTCCCG	
CFTR WT	CFTR WT F2	SENSE	ACAGAGTGAGCACTTGGC	85
	CFTR WT R2	ANTISENSE	CACATGATTCAGATAGTTGG	
CFTR Mu	CFTR WT 19 F	SENSE	GTGGAAGAATTTCATTCTG	87
	CFTR mt 19 R	ANTISENSE	CATCATAGGAAACACCAAT	
	Blocker CFTR 23 R	ANTISENSE	GAAACACCAAAGATGATATTTTC-PHO	

Tableau 2: *Les séquences et caractéristiques des oligonucléotides utilisés*

E. Résultats

1. Quantification de l'ADN extracellulaire nucléaire et mitochondrial dans le milieu de culture des embryons

a) Contamination des milieux de culture frais par de l'ADN extracellulaire

Les milieux de culture habituellement utilisés pour les différentes étapes du développement embryonnaire lors d'une fécondation *in vitro* contiennent de façon générale de l'albumine de sérum humain en plus d'autres constituants. Nous avons soupçonné la présence d'ADN extracellulaire dans ces milieux frais étant donné que l'albumine est connue pour interagir avec l'ADN *(270)*. La première étape de notre analyse a consisté à établir des contrôles en évaluant l'ADNcf nucléaire et mitochondrial présents dans les milieux de culture frais. Des milieux provenant de différents fournisseurs (Vitrolife, Origio et Irvine Scientific) utilisés pour les différentes étapes du développement embryonnaire lors de la FIV ont ainsi été testés **(Tableau 1)**.

La concentration totale d'ADN extracellulaire nucléaire (Ref A 67 en nb de copies/µl de milieu) et mitochondrial (Ref M 67 en nb de copies/µl de milieu) a été calculée. De l'ADNnu a été détecté dans la plupart des milieux testés (8 milieux sur 9) avec un taux variant de 1.37 à 3.65 copies/µl (moyenne : 2.59 copies/µl de milieu ; médiane : 2.64 copies/µl de milieu) **(Figure 32)**. Tous les milieux contenaient de l'ADNcf mitochondrial avec un Ref M 67 qui varie entre 42.6 et 1680 copies/µl et une concentration moyenne de 435 copies/µl (médiane : 371 copies/µl de milieu) **(Figure 33)**. Ces résultats indiquent une contamination des milieux frais de culture d'embryons avec de l'ADN extracellulaire d'origine nucléaire et mitochondriale et suggèrent la nécessité d'une normalisation lors de la quantification de l'ADN extracellulaire provenant de l'embryon dans ces milieux de culture.

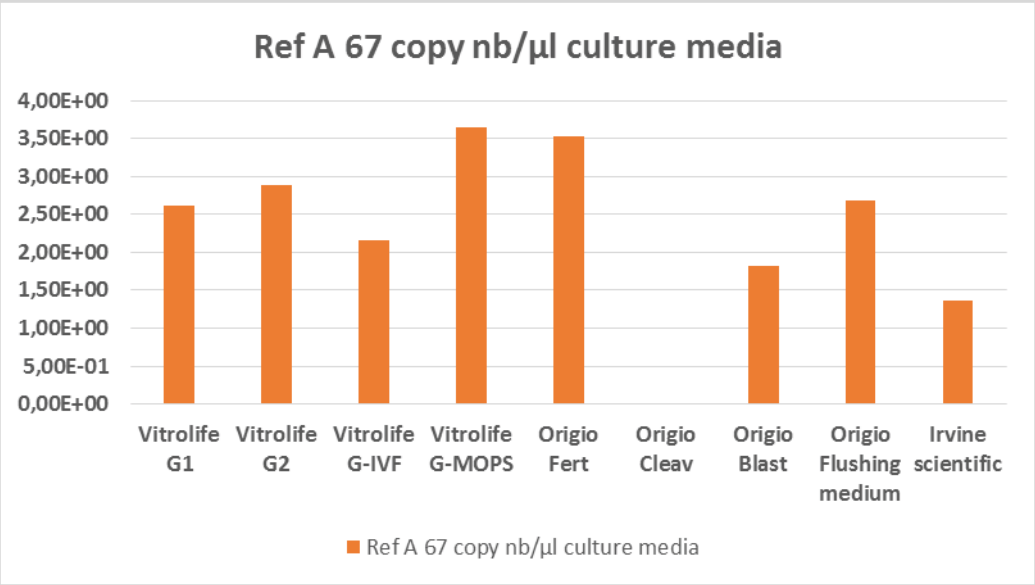


Figure 32: *Concentration totale d'ADNcf nucléaire (Ref A 67 copies/µl de milieu) dans les milieux contrôles testés.*

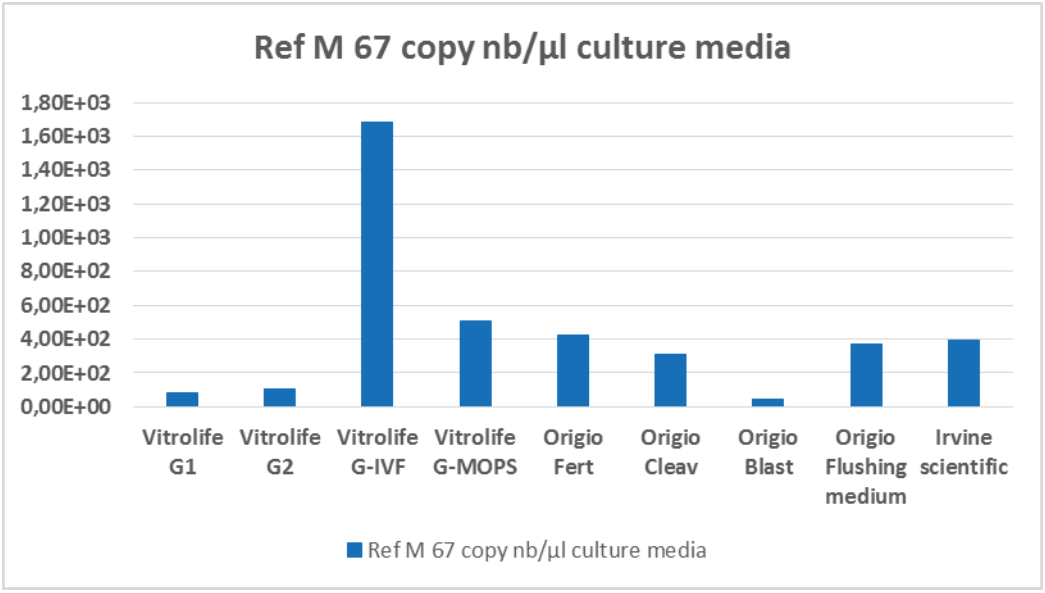


Figure 33: *Concentration totale d'ADNcf mitochondrial (Ref M 67 copies/µl de milieu) dans les milieux contrôles testés.*

b) Quantification de l'ADNcf dans le milieu de culture des embryons

229 échantillons provenant des milieux de culture en J1, J3, J5 et J6 de 69 embryons différents et 60 contrôles ont été évalués. Les embryons correspondent à 11 couples candidats à une fécondation *in vitro*. L'ADNcf a été extrait à partir des milieux de culture récupérés et a été quantifié par Q-PCR en ciblant des séquences spécifiques pour calculer la concentration totale d'ADNcf d'origine nucléaire (Ref A 67 en ng/µl de milieu) et mitochondriale (Ref M 67 en ng/µl de milieu).

Le taux d'ADN provenant de l'embryon a été calculé en le normalisant par rapport à la concentration basale d'ADN détectée au niveau des gouttes contrôles. Après normalisation, l'ADNcf nucléaire a été détecté dans le milieu de culture de 48% (33 sur 69 embryons) des embryons testés quel que soit le jour du développement embryonnaire **(Tableau 3)**. L'ADNmito extracellulaire était quant à lui détectable dans le milieu de culture de 36% des embryons en J1, 77% en J3 et 41% en J5/6 **(Tableau 3)**.

Détection de l'ADN extracellulaire dans le milieu de culture des embryons

	ADNcf nucléaire	**ADNcf mitochondrial**
J1	33 (48%)	25 (36%)
J3	33 (48%)	53 (77%)
J5/6	33 (48%)	28 (41%)

Tableau 3: *La détection d'ADNcf nucléaire et mitochondrial dans les milieux de culture des embryons testés (n=69) au niveau des différents jours du développement embryonnaire*

Le niveau d'ADNcf nucléaire et mitochondrial a ensuite été comparé, pour chaque embryon, en fonction du jour du développement embryonnaire (J3 à J6) et du grade (de 1 à 10) de l'embryon. D'après l'observation morphologique, un embryon de grade 10 était considéré de bonne qualité alors qu'un grade 1 de mauvaise qualité.

Les résultats du couple ANT MYR sont présentés dans la **Figure 34**. Pour les différents embryons testés (embryons 3, 7, 9, 13, 14 et 16), la concentration d'ADNcf mitochondrial (Ref M 67 nb de copies/µl) (en bleu) était proportionnelle au grade de l'embryon ; en effet, une augmentation du grade entre J3 et J5 ou entre J5 et J6 est accompagnée d'une augmentation

du Ref M 67 dans le milieu de culture **(Figure 34)**. La concentration totale d'ADN extracellulaire nucléaire (Ref A 67 nnb de copies/µl) (en orange) a montré des résultats plus divergents. Une augmentation du Ref A 67 dans le milieu de culture est observée pour les différents embryons entre J3 et J5 lorsque le grade de l'embryon diminue. Par contre, entre J5 et J6, lorsque le grade de l'embryon augmente, cette concentration augmente dans certains cas (embryons 3, 7, 13 et 14) et diminue dans d'autres (embryons 9 et 16) **(Figure 34)**.

Des validations supplémentaires sont nécessaires, et la variation du Ref A 67 et du Ref M 67 dans le milieu de culture des embryons correspondant aux autres couples en fonction du grade de l'embryon ainsi que du jour du développement embryonnaire doit être analysée. En outre, l'évaluation des résultats du MNR (Ref M 67/ Ref A 67) pourrait probablement donner des renseignements complémentaires quant à la qualité des embryons.

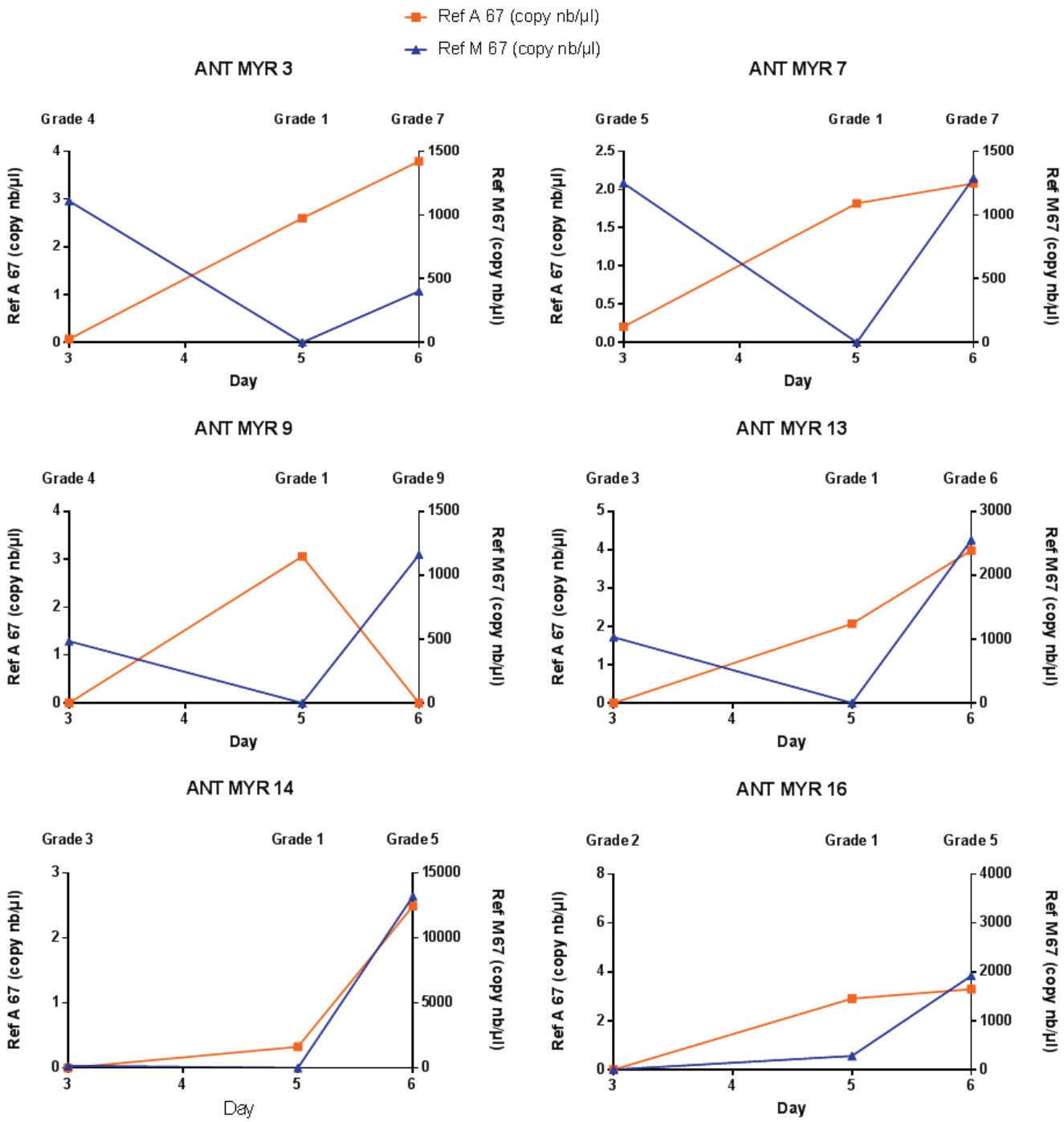


Figure 34: *Variation du Ref A 67 (nb copies/µl) (en orange) et du Ref M 67 (nd de copies/µl) (en bleu) en fonction du grade de l'embryon et du jour de développement embryonnaire pour le couple ANT MYR. Les embryons analysés sont les numéros 3, 7, 9, 13, 14 et 16.*

2. Détection de séquences spécifiques et d'altérations génétiques dans le milieu de culture

a) Gène SRY et détermination du sexe des embryons

Le gène *SRY* présent sur le chromosome Y a été recherché dans le milieu de culture des 69 embryons testés en J1, J3, J5 et J6. La présence de ce gène a été détectée chez 24 embryons (49%), qui ont alors été considérés comme mâles **(Tableau 4)**. La non détection de ce gène dans le milieu signifie que l'embryon est soit de sexe féminin, soit de sexe masculin mais que la quantité d'ADN libérée dans le milieu n'était pas suffisante pour la détection de la séquence recherchée.

Détection du gène SRY dans le milieu de culture des embryons

Embryon	**Jour du développement embryonnaire**				*Embryon*	**Jour du développement embryonnaire**			
	J1	J3	J5	J6		J1	J3	J5	J6
FL 1		+			*AM 2*	+			
FL 3	+		+		*AM 5*			+	
FL 5				+	*AM 9*	+			
Lj 4			+		*AM 12*				+
Lj 5			+		*AM 14*		+		
LL 2	+				*AM 16*		+		+
LL 3		+			*RA 1*	+			
LL 4	+				*RA 4*		+		+
PA 1			+		*RA 7*	+			
PA 3		+	+		*AN 9*		+		
PA 4		+			*CM 6*		+		
PA 5	+				*CM 8*		+		
Total n=24	**8**	**10**	**6**	**4**					

Tableau 4: *Détection du gène SRY dans le milieu de culture des embryons aux jours 1, 3, 5 et 6 du développement embryonnaire. La croix indique une détection de la séquence dans le milieu.*

b) Le cas de la mucoviscidose

La mutation Delta F508 du gène *CFTR* correspond à une délétion de trois paires de bases responsable de la perte d'une phénylalanine en position 508 de la protéine. Cette mutation est impliquée dans la forme classique de la mucoviscidose, et a été recherchée dans le milieu de culture d'embryons issus d'un couple (BES VIR) porteur de cette délétion. Les milieux de culture en J1, J3, J5 et J6 sont en cours d'analyse pour cette altération génétique et les résultats seront comparés avec ceux d'un test génétique préimplantatoire effectué sur une biopsie de blastomère en J3 de chacun de ces embryons pour déterminer la présence ou l'absence de la mutation.

F. Conclusion, discussion et perspectives

L'ADN extracellulaire présent dans le milieu de culture des embryons constitue une approche prometteuse non invasive pour évaluer la qualité de l'embryon lors d'une FIV et détecter des anomalies génétiques chez des embryons à risque pour certaines maladies graves.

- **Quantification de l'ADNcf dans le milieu de culture des embryons**

Nous avons retrouvé de l'ADN extracellulaire nucléaire et mitochondrial dans les neuf milieux de culture frais testés, provenant de trois fournisseurs distincts. Nos données alertent sur la possibilité d'une contamination des milieux de culture embryonnaire utilisés durant le processus de fécondation *in vitro* par de l'ADNcf. En effet, la présence d'une contamination, même faible, pourrait influencer l'analyse et rendre l'interprétation des résultats difficile. Hammond *et al.* ont comparé le taux d'ADNcf nucléaire et mitochondrial dans les milieux de culture des embryons à des contrôles internes ainsi qu'à trois milieux commerciaux frais (Vitrolife et Irvine Scientific) par PCR digitale et Q-PCR *(271)*. Ils ont suggéré que l'ADN présent dans le milieu de culture ne peut pas être utilisé pour l'évaluation génétique de l'embryon à cause de son origine mixte exogène et embryonnaire. La contamination des milieux de culture par de l'ADN n'est pas surprenante. En effet, ces milieux contiennent de l'albumine sérique humaine, qui est importante pour le développement embryonnaire *ex-vivo (272)*, mais qui peut également interagir avec des molécules d'ADN *(270)*, ce qui expliquerait la présence d'ADN extracellulaire dans ces milieux.

Par ailleurs, des observations récentes ont montré le rôle de l'ADN extracellulaire dans le transfert horizontal de l'information et la communication intercellulaire. En effet, des fragments d'ADN *(108)* ou de l'ADN contenu dans des vésicules *(101)* sont capables de pénétrer dans les cellules et de modifier leur biologie en se localisant dans leur noyau. Ces observations soulèvent le problème d'une internalisation probable d'ADN exogène contaminant dans les cellules de l'embryon, et, par voie de conséquence d'effets potentiels sur sa croissance et sa qualité. Il est donc nécessaire de connaitre la composition exacte des

milieux de culture des embryons et leurs effets probables sur le développement embryonnaire *(273)*.

Nos résultats ont également montré que les embryons libèrent de l'ADN extracellulaire nucléaire et mitochondrial dans le milieu de culture pendant la FIV après normalisation vis-à-vis du taux d'ADN basal présent dans le milieu. En effet, la présence d'un contrôle interne constitué d'une goutte de milieu sans embryon incubée dans les mêmes conditions est indispensable et l'approche non-invasive nous semble fiable sous réserve de normaliser les quantités d'ADN détectées dans le milieu par rapport à la quantité basale initiale.

La relation entre le taux d'ADN extracellulaire présent dans le milieu et la qualité de l'embryon a été recherchée dans plusieurs études. Une fragmentation cellulaire est caractéristique d'un embryon de mauvaise qualité, et ce phénomène devrait normalement impliquer la libération d'ADN nucléaire et mitochondrial dans le milieu. Stigliani *et al.* ont identifié une corrélation significative entre la libération d'ADN et la fragmentation des embryons aux jours 2 et 3 du développement embryonnaire et ont montré que les milieux de culture d'embryons non fragmentés ou présentant un léger degré de fragmentation contenaient moins d'ADNcf nucléaire et mitochondrial que ceux d'embryons fragmentés *(265)*. Nos résultats concernant l'ADNcf nucléaire sont en accord avec ces observations. Une augmentation du taux d'ADNnu extracellulaire a été souvent observée entre le troisième et le cinquième jour du développement embryonnaire lorsque le grade de l'embryon diminue, ce qui pourrait s'expliquer par la libération de cet ADN suite à la fragmentation de l'embryon. Par contre, nos observations sur l'ADNcf mitochondrial contredisent celles déjà publiées. En effet, nous avons observé une diminution du taux de cet ADN dans le milieu lorsque le grade de l'embryon diminue, donc lorsque ce dernier est de moins bonne qualité. Cette observation pourrait s'expliquer par le fait que les embryons qui ne parviennent pas à s'implanter sont associés à une augmentation de leur contenu en nombre de copies d'ADNmt *(274)*. Diez-Juan *et al. (245)* ont montré des résultats similaires, et Fragouli *et al.* ont suggéré que quand le contenu en ADNmt dépasse un certain seuil, aucune grossesse n'a lieu *(275)*. Ces résultats ont également été validés par d'autres études *(246, 247)*. De plus, l'ADNmt est connu pour ne se répliquer dans les cellules de l'embryon qu'à partir du stade de blastocyste ou après implantation, et qu'au stade blastocyste, le déclenchement de la réplication est spécifique des cellules du trophectoderme *(276)*. Donc la quantité d'ADNmt semble rester stable tout au long des stades

de clivage. Ces observations, associées au fait qu'un embryon de mauvaise qualité contient plus d'ADNmt, pourraient être en faveur d'une libération plus faible d'ADN mitochondrial par un embryon de mauvaise qualité dans le milieu de culture.

Nos résultats doivent cependant être validés avec les milieux de culture des embryons testés provenant d'autres patientes.

- **L'analyse de l'ADNcf pour la détection des maladies génétiques**

La discrimination d'un embryon mâle et d'un embryon femelle est une information importante dans le cas des maladies liées au sexe *(120)*. Ces maladies, dues à la présence de gènes défectueux sur le chromosome X, sont en effet dominantes chez les mâles. Nous avons déterminé, de manière non invasive par la détection du gène *SRY* dans le milieu de culture, que 48% des embryons testés étaient de sexe masculin. La séquence du gène *SRY* a été détectée au premier, troisième, cinquième ou sixième jour du développement embryonnaire. Le sexe des embryons restants qui peuvent autant être femelles que mâles n'a pas pu être déterminé compte tenu de la quantité d'ADN insuffisante dans le milieu. Etant donné les contraintes éthiques, aucune validation par biopsie de cellule de l'embryon n'a été effectuée, ce qui n'a pas non plus permis de calculer le taux de prédiction du sexe. La détermination du sexe de l'embryon par analyse de l'ADN extracellulaire libéré dans son milieu de culture a été validée par plusieurs études *(120, 264)*, et Ho *et al.* ont montré une concordance de 81.3% et 70% en J3 et J5 respectivement avec une biopsie de l'embryon *(262)*.

L'ADN extracellulaire constitue donc une approche possible de test génétique préimplantatoire (PGT) non invasif. En plus de la détermination du sexe de l'embryon dans le cas des maladies liées à l'X, il a d'ailleurs été utilisé pour la caractérisation du statut de ploïdie de l'embryon *(257, 259)* et la détection d'altérations génétiques responsables de certaines maladies *(263)*. Afin d'évaluer l'utilité de l'ADNcf comme approche non invasive de PGT, notre but est de rechercher la mutation Delta F508 du gène *CFTR* impliquée dans la forme classique de la mucoviscidose dans le milieu de culture d'embryons issus d'un couple porteur de cette délétion.

La détection de séquences d'ADN spécifiques de l'embryon pourrait permettre de s'affranchir du problème de contamination, et de discriminer l'ADN provenant de l'embryon de celui résultant de la contamination du milieu. Toutefois, d'autres approches, semblables à celles utilisées pour l'analyse de l'ADN fœtal dans le sang maternel, pourraient également permettre de distinguer l'ADN embryonnaire de l'ADN contaminant, comme, par exemple, l'étude de la différence du profil de taille entre ADN fœtal circulant et ADN maternel *(136, 277)*.

- **Perspectives**

Nos observations sur la relation entre la quantité d'ADNcf nucléaire et mitochondrial dans le milieu de culture des embryons et le grade de l'embryon ainsi que sa qualité doivent être validées par l'analyse des résultats du reste des patientes.

Il a été montré que le rapport entre la concentration d'ADNcf mitochondrial et nucléaire est plus élevé dans le milieu de culture des embryons de bonne qualité que dans celui des embryons de mauvaise qualité *(266)*, et que, associé au classement morphologique, il pourrait permettre de mieux prédire la qualité de l'embryon *(267)*. Il serait alors intéressant de calculer notre MNR (Mitochondrial to Nuclear Ratio) pour nos divers embryons testés et d'étudier sa variation en fonction du jour du développement embryonnaire, de la qualité de l'embryon et de la réussite ou l'échec de l'implantation.

Enfin, nous prévoyons de poursuivre les analyses de détection de la mutation Delta F508 du gène *CFTR* à l'origine de la mucoviscidose dans le milieu de culture d'embryons à risque, et d'étudier la concordance de nos résultats avec ceux de la biopsie effectuée sur les mêmes embryons.

V. Conclusion générale

La découverte de l'ADN circulant provenant de la tumeur et du fœtus dans le plasma des patients cancéreux et des femmes enceintes a ouvert de nombreuses nouvelles possibilités pour la détection du cancer et le développement de tests génétiques préimplantatoires (PGT) non invasifs *(278)*.

Durant ma thèse, nous avons validé l'utilité des paramètres quantitatifs et structuraux de l'ADN circulant d'origine nucléaire et mitochondriale comme une approche potentielle pour la détection précoce du cancer. Nous avons également suggéré, avec une validation plus poussée, que l'ADN extracellulaire pouvait être utilisé pour évaluer la qualité de l'embryon lors de la fécondation *in vitro*, et potentiellement servir comme PGT non invasif.

Nous savons qu'une tumeur, comme un fœtus, libère de l'ADN extracellulaire nucléaire et mitochondrial, et que cet ADN présente de multiples similitudes avec l'ADN fœtal libéré *in vitro* ou dans la circulation de la mère. Tout en étant d'origine différente, ces ADN montrent, par exemple, des profils de fragmentation similaires, avec une taille inférieure à celle de l'ADN circulant provenant de la mère ou des cellules non tumorales *(278)*.

Les points communs entre le développement d'un embryon et celui d'une tumeur pourraient probablement expliquer cette ressemblance *(279)*. En effet, des études ont révélé que certaines voies sont activées ou inhibées de façon similaire au cours de l'embryogenèse et de la carcinogenèse. Ainsi, il apparait que le développement d'une tumeur se fait par la réactivation de voies de signalisation embryonnaires qui sont réprimées dans les cellules non tumorales, comme ceux impliquées dans la transition épithélio-mésenchymateuse *(280)*, qui constitue une étape essentielle dans ces deux processus. Les voies Wnt *(281)*, Hedgehog et Notch *(280)*, indispensables pour l'embryogenèse, et qui sont dérégulés dans le cas du cancer, en constituent un exemple. Le gène *TP53* a aussi été identifié comme marqueur qui est exprimé d'une façon similaire lors du développement embryonnaire précoce et de celui du cancer *(282)*.

Les cellules embryonnaires et les cellules tumorales, ont une capacité de prolifération, d'invasion et de migration fortement augmentée *(283, 284)*. De plus, les cellules souches pluripotentes ainsi que certains types de cellules cancéreuses sont connus pour obtenir

l'énergie principalement par glycolyse, réduisant souvent l'activité mitochondriale et la phosphorylation oxydative *(285)*. Il n'est donc pas aberrant de retrouver des profils d'ADN circulant très similaires entre celui provenant d'une tumeur et celui provenant d'un embryon.

Pour conclure, l'analyse de l'ADN circulant est prometteuse pour le développement d'une nouvelle génération de tests diagnostiques, et la connaissance des propriétés biologiques de cet ADN est essentielle pour son utilisation optimale dans différents contextes cliniques dont l'oncologie et le diagnostic préimplantatoire.

VI. Autres contributions

Durant ma thèse, j'ai également contribué à différents travaux au sein de l'équipe, dont certains ont fait l'objet des publications suivantes :

A. Nouvelles connaissances sur les caractéristiques structurelles et la détection optimale de l'ADN tumoral circulant déterminé par analyse de l'ADN simple brin

Titre : New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA analysis

Auteurs

Cynthia Sanchez, Matthew W. Snyder, **Rita Tanos**, Jay Shendure et Alain R. Thierry

Ce travail a été publié dans *Genomic Medicine* le 23 Novembre 2018.

Résumé

L'ADN circulant a suscité un intérêt croissant en oncologie dans le domaine du diagnostic, de la médecine personnalisée et de l'oncologie. Cependant, les caractéristiques structurales de l'ADNcir sont mal définies. Plus précisément, des divergences existent en ce qui concerne le profil de taille de cet ADN. Nous avons réalisé une étude en aveugle de la distribution de la taille des fragments d'ADNcir dans le plasma de patients cancéreux (n = 11), par différentes approches de PCR quantitative (Q-PCR) et d' « ultra-deep-sequencing ». Le séquençage du génome entier par préparation d'une banque d'ADN simple brin (SSP-S pour Whole-genome sequencing of single-stranded DNA library preparation) a révélé que presque la moitié des fragments d'ADNcir a une taille inférieure à 120 nucléotides. Ces fragments ne sont pas facilement détectables par les protocoles standard de préparation de banques d'ADN double brin (DSP pour double-stranded DNA library preparation). L'utilisation de méthodes basées sur la SSP-S ou sur la Q-PCR a montré une distribution de taille très similaire de l'ADNcir pour

un patient cancéreux, et a montré également que l'ADNcir de masse moléculaire élevée (plus de 350 pb) est un composant mineur (~2%) de l'ADNcir total. Ces petits fragments d'ADNcir détectés peuvent principalement résulter de coupures se produisant dans la circulation sanguine dans l'un ou les deux brins d'ADN. Ils sont ensuite révélés par l'étape de dénaturation lors des procédures de SSP et de Q-PCR. L'analyse détaillée des données suggère que la plupart des fragments d'ADNcir détectables dans le sang ont une empreinte nucléosomique (répétitions périodiques de ~10-pb). Le nucléosome est donc la structure la plus stabilisante de l'ADN dans la circulation. Les molécules d'ADNcir, qui sont initialement présentes sous forme de chromatine, sont libérées par les cellules et sont ensuite dégradées dynamiquement dans le sang, à l'intérieur et entre les nucléosomes ou les sous-complexes associés aux facteurs de transcription. Bien que cette étude apporte de nouvelles perspectives sur les profils de taille de l'ADNcir, harmonisant les résultats obtenus par séquençage et Q-PCR, nos données valident l'utilisation d'une méthode spécifique de Q-PCR et de SSP-S pour obtenir un signal optimal pour l'analyse qualitative ainsi que quantitative de ces ADN.

Contributions

J'ai contribué à la rédaction et à la mise en forme de l'article.

npj | Genomic Medicine

www.nature.com/npjgenmed

ARTICLE OPEN

New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA analysis

Cynthia Sanchez[1,2,3,4], Matthew W. Snyder[5], Rita Tanos[1,2,3,4], Jay Shendure[5] and Alain R. Thierry[1,2,3,4]

Circulating cell-free DNA (cfDNA) has received increasing interest as an apparent breakthrough approach in diagnostics, personalized medicine, and tumor biology. However, the structural features of cfDNA are poorly characterized. Specifically, the literature has discrepancies with regards to cfDNA size profile. We performed a blinded study of the distribution of cfDNA fragment sizes in cancer patient plasma ($n = 11$), by various ultra-deep-sequencing approaches and quantitative PCR (Q-PCR). Whole-genome sequencing of single-stranded DNA library preparation (SSP-S) revealed that nearly half of the total cfDNA fragment number are below 120 nucleotides, which are not readily detectable by standard double-stranded DNA library preparation (DSP) protocols. Fractional size distribution of cancer patient circulating DNA was very similar using both SSP-S-based or Q-PCR-based methods also revealing that high molecular weight (over 350 bp) cfDNA is a minor component (~2%). These extra small detected cfDNA fragments may mostly result from nicks occurring in blood circulation in one or both DNA strands, which are subsequently revealed through the denaturation step of the SSP and Q-PCR procedures. Detailed analysis of the data suggested that most of the detectable cfDNA in blood has a nucleosome footprint (~10-bp periodicity repeats). The nucleosome is thus the most stabilizing structure of DNA in the circulation. cfDNA molecules, which are initially packed in chromatin, are released from cells and are then dynamically degraded in blood both within and between nucleosomes or transcription factor-associated subcomplexes. While this study provides new insights into cfDNA size profiles harmonizing sequencing and Q-PCR findings, our data validate the use of a specific Q-PCR method and SSP-S for obtaining an optimal qualitative and quantitative analytical signal.

npj Genomic Medicine (2018)3:31 ; doi:10.1038/s41525-018-0069-0

INTRODUCTION

The discovery of small amount of DNA circulating freely in human blood[1] led to growing scrutiny with regards to its potential use in various clinical fields.[1–5] Circulating cell-free DNA (cfDNA) analysis is currently applied in prenatal diagnosis[6] and shows potential for clinical use in other fields including organ transplant, autoimmune diseases, trauma, myocardial infarction, and sepsis.[1–3] CfDNA concentration is significantly increased in cancer subjects[7,8] and a significant proportion of cfDNA may derive from the tumor, providing diagnostic and prognostic potential information. Therefore, the characterization of genetic and epigenetic alterations of tumor cells may be obtained in a non-invasive way by analyzing cfDNA from the plasma or serum of cancer subjects.[3,8] Thus, detection of mutations leading to resistance to targeted therapies, personalized therapeutic monitoring, and non-invasive follow-up of the disease may be possible in the course of cancer management care.

While cfDNA analysis appears clinically useful for several pathologies and specific physiological conditions, a precise understanding of its origins and nature, including the size distribution of cfDNA fragments, has not been established.[1] Nevertheless, the link to histones has been well established by various reports [9–13]. That said, in the oncology field it is imperative to have detailed information about the size distribution of cfDNA fragments, to examine the maximum cfDNA concentration and to obtain a high level of sensitivity and specificity, especially when detecting rare genetic alterations. Excluding mass spectrometry, all current methods of cfDNA analysis, including sequencing or PCR-based methods, require cfDNA size definition. It has to be noted that for non-invasive prenatal testing several methods have been described for fetal fraction determination that do not need size distribution, for example, SeqFF and Sanefalcon.[14]

CfDNA structure and size depend on the mechanisms of cfDNA release from cells. While we do not know their respective proportions with regards to cfDNA amount shed into the bloodstream, several have been proposed, including necrosis, apoptosis, phagocytosis, active release, and exosome/microparticle release.[1]

In addition to the diversity of cfDNA release mechanisms in the literature, discrepancies in cfDNA fragment size distribution in healthy individuals or cancer patients are apparent. The most reported size distribution is dominated by mononucleosomes and oligonucleosomes,[9] and has become the basic premise concerning the structure of cfDNA for many years.[1,10] In particular, conventional next-generation sequencing (NGS) methods, as well as earlier works based on electrophoretic mobility, such as PAGE

[1]IRCM – Institut de Recherche en Cancérologie de Montpellier, Montpellier 34298, France; [2]INSERM, U1194, Montpellier 34298, France; [3]Université de Montpellier, Montpellier 34090, France; [4]Institut régional du Cancer de Montpellier, Montpellier 34298, France and [5]Department of Genome Sciences, University of Washington, 3720 15th Avenue NE, Seattle, WA, USA
Correspondence: Alain R. Thierry (alain.thierry@inserm.fr)

Received: 23 April 2018 Accepted: 19 October 2018
Published online: 23 November 2018

Published in partnership with CEGMR, King Abdulaziz University

npj nature partner journals

or the Agilent platform, have revealed clearly a high proportion of cfDNA fragments ranging from 170 to 200 bp in both healthy or cancer individuals.[11,12] Apoptotic DNA cleavage produces a characteristic pattern ladder of 180–200 bp, or multiples thereof (oligonucleosomes). DNA wrapped around the histone octamer is 147 bp in length and the linker DNA is 20–90 bp (mainly, 20 bp). Association of these fragments with nucleosomes presumably assures structural integrity by protecting DNA from enzymatic degradation in the circulatory system.[13,15,16] Contrarily, other studies have shown the presence of large-sized fragments of many kilobases (kbp), which may indicate a necrotic release mechanism.[17] However, these observations should be taken with caution because of the uncertainty of pre-analytical factors, especially with regard to the contamination of DNA derived from blood-cell degradation.

In contrast, we have shown by alternative methods including a Q-PCR-based method and atomic force microscopy (AFM), the majority of cfDNA in cancer patients is <145 bp.[8,18] By amplifying DNA sequences of increasing size within the same DNA region, Chan et al.[19] and Diehl et al.[4] demonstrated the presence of a significant fraction of short cfDNA fragments below 180 bp in healthy individual and cancer patient cfDNA, respectively.[18] We later showed, for the first time, that the shortest is the amplicon (down to 60 bp), while the highest is the quantification in either healthy or cancer subjects,[8,18,20] and that mutant cfDNA fragments are shorter than wild-type fragments.[20] Moreover, this observation suggested that the detection of amplicons <100 bp is more relevant for optimally quantifying cfDNA.[8] This has been confirmed and now most of the Q-PCR-based methods involve the amplification of DNA sequences <100 bp, while targeting ~150 bp length sequences is known to be optimal for quantifying non-fragmented genomic DNA.[21–23]

CfDNA size distribution obtained by Q-PCR has shown high discordance with other methods, in particular conventional NGS, which precludes drawing any general conclusions. This issue therefore appears critical in respect to diagnostic performance (especially in oncology when testing for genetic alterations, which may be of rare frequencies among cfDNA fragments) as it relies on the number of examined copies. Our two groups pooled our respective expertise in deep-sequencing and Q-PCR methods, which are poorly associated in the literature, to determine in a blinded study the cfDNA size profile in plasma DNA extracts of cancer patients and evaluating performance of various methodologies of both approaches to optimally recover cfDNA copies.

RESULTS

Double-strand library preparation (DSP) is used typically for cell-free DNA analysis for several reasons: it is an easy protocol to carry out in the laboratory as it takes only a few hours compared to single-strand library preparation protocols, which typically take much longer; it is cheap on a per-sample basis; and a lot of optimization has been done on DSP so that it is more efficient and has less adapter ligation bias.[24] Single-strand library preparation (SSP) was recently designed to bypass the limits of the conventional DNA library in order to recover damaged and short double-strand DNA fragments, especially in the paleontology field.[25,26] High-throughput sequencing of cfDNA from SSP (SSP-S) has been used only by Snyder et al.[27] and Burnham et al.[28] In this study, we compared the cfDNA size profile obtained, on one hand, by sequencing from DSP and SSP and, on the other hand, in a blinded study by comparing SSP-S with a Q-PCR-based method (SI-1).

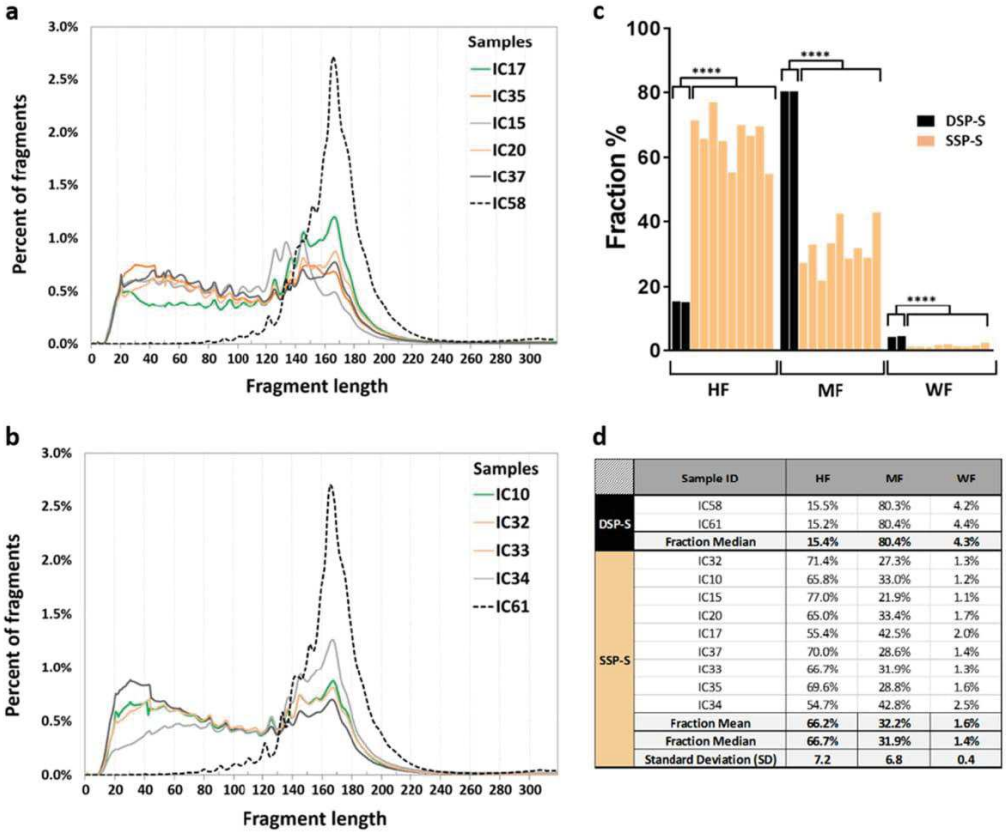
Although SSP-S and Q-PCR are based on the detection of double-stranded DNA (dsDNA) as the final analytical signal, cfDNA fragment size using these methods is determined from the size of single-strand DNA (ssDNA) fragments resulting from the denaturation step of both SSP and Q-PCR processes. Consequently, the number of nucleotides (nt) is used as the fragment size distribution length unit when using SSP-S or Q-PCR.

Size profile obtained from DSP-S of plasma from two cancer patients

DSP library was sequenced to 96-fold and 105-fold coverage (1.5 and 1.6 billion fragments). DSP-S analysis of two lung cancer patient plasma samples (IC58 and IC61) resulted in very similar size profiles (Fig. 1a, b, respectively). Both showed a Gaussian cfDNA size distribution ranging from 100 to 240 bp, peaking at 166 bp with a half peak width between 150 and 200 bp. Series of smaller peaks at a periodicity of ~10 bp could be observed between 81 and 166 bp in both samples at exactly the same peaking sizes (Fig. 1a, b and Table 1a). The proportion of fragments below 145 bp and over 180 bp are 18.5% and 24%, respectively. DSP-S did not reveal a significant amount of fragments lower than 80 bp.

Size profile obtained from SSP-S of plasma from nine cancer patients

SSP library were sequenced to 30-fold coverage (779M fragments). Nine samples showed similar cfDNA size profile using SSP-S. The size distribution was non-Gaussian with two apparent populations: the first ranging between 30 and 120 nt and the other between 120 and 220nt (Fig. 1a, b). The proportion of fragments below 145 bp and over 180 bp are 66.2% and 7.6%, respectively. Whereas size profile by SSP-S peaked at 166–168nt correspondingly to the peak size observed by DSP-S (166 bp), the shorter-length cfDNA fragment population was unique to SSP-S as compared to DSP-S (Fig. 1a, b). Series of smaller peaks at the periodicity of ~10 nt is detected from 41 to 167 nt when using SSP-S, while the same periodicity could be observed only from 81 to 166 bp when using DSP-S (Table 1a, b). The values of this small periodic peak size are strikingly equivalent between the nt and bp number as obtained from SSP-S and DSP-S, respectively (Table 1b). While maximum peak size is identical not only among all tested samples by either SSP-S or DSP-S, the ~10 nt subpeaks revealed by SSP-S corresponded to lengths being 2 to 4 nt shorter than that obtained by DSP-S (Table 1a). Note, size profile pattern among those nine samples as determined by SSP-S showed very high consistency as it exists likewise among the two previously tested samples by DSP-S as well as those previously assessed.[11–13,27] Note, the fifth and the sixth subpeaks showed exactly the same corresponding size following either SSP-S (83 and 94nt, respectively) or DSP-S (81 and 92 bp), and second, the periodicity between both subpeaks is identical following SSP-S and DSP-S analysis (11nt and 11 bp, respectively). No peak is detectable at lengths superior to 260 nt by SSP-S, while a weak DNA fragment sub-population (<3% of total reads) peaking at 307 and 308 bp (IC58 and IC61, respectively), which may correspond to the size of DNA contained in a dinucleosome, is detectable by DSP-S (SI-2). It should be noted that DSP-S was performed on plasma from lung cancer, while SSP-S was performed on four lung (IC10, IC32, IC15, and IC20), two colorectal (IC33 and IC37), two breast (IC34 and IC35), and one liver (IC17) cancers. Use of SSP generates higher recovery of fragments below 130 bp and the observation of fragments below 80nt as compared to DSP irrespective of the various types of cancer. In addition, CfDNA size profile obtained by SSP-S and DSP-S showed clear discrepancies, especially when comparing fragment size distribution by three size ranges we set based on the previous cfDNA sizing biological paradigm: MF, corresponding to the length of the DNA sequence compacted in a mononucleosome (145–249nt); WF, corresponding to weakly fragmented cfDNA (>249nt); and HF, corresponding to highly fragmented cfDNA (<145nt) (Fig. 1c, d). SSP-S showed, as compared to DSP-S, much higher HF fraction proportion (66.7% median±7.2SD and 15.4% median, respectively; $p = 5{,}41.10\text{-}6$), and



d

	Sample ID	HF	MF	WF
DSP-S	IC58	15.5%	80.3%	4.2%
	IC61	15.2%	80.4%	4.4%
	Fraction Median	**15.4%**	**80.4%**	**4.3%**
SSP-S	IC32	71.4%	27.3%	1.3%
	IC10	65.8%	33.0%	1.2%
	IC15	77.0%	21.9%	1.1%
	IC20	65.0%	33.4%	1.7%
	IC17	55.4%	42.5%	2.0%
	IC37	70.0%	28.6%	1.4%
	IC33	66.7%	31.9%	1.3%
	IC35	69.6%	28.8%	1.6%
	IC34	54.7%	42.8%	2.5%
	Fraction Mean	**66.2%**	**32.2%**	**1.6%**
	Fraction Median	**66.7%**	**31.9%**	**1.4%**
	Standard Deviation (SD)	**7.2**	**6.8**	**0.4**

Fig. 1 CfDNA fragments shorter than 100 nt are accessible for sequencing following single-strand library preparation (SSP). Comparison of the size profiles of cancer patient cfDNA obtained from DSP-S (dotted lines) as compared to SSP-S (full lines) DSP-S was performed on plasma from lung cancer (IC58 and IC61), while SSP-S was performed on four lung (IC10, IC32, IC15, and IC20), two colorectal (IC33 and IC37), two breast (IC34 and IC35), and one liver (IC17) cancers. In order to differentiate SSP-S derived size profile curves, two sets of data are presented (IC15, IC17, IC20, IC35, IC37, **a**; and IC10, IC32, IC33, IC34, **b**) each containing one of the two cfDNA size profile as determined by DSP-S. The nine cfDNA extracts from cancer patients analyzed by SSP-S were examined in the blinded study comparing SSP-S with Q-PCR size profile analysis. **c** presents the proportion of three size ranges we set based on the previous cfDNA sizing biological paradigm: MF, corresponding to the length of the DNA sequence compacted in a mononucleosome (145–249nt); WF, corresponding to weakly fragmented cfDNA (249nt); and HF, corresponding to highly fragmented cfDNA (<145nt). **d** presents the percent value of the three size ranges obtained from each patients either by SSP-S or DSP-S. ****, p < 0.0001

lower MF (31.9% median ± 6.8 SD and 80.4% median, respectively; $p = 5{,}33.10\text{-}6$) and WF (1.4% median ± 0.4 SD and 4.3% median, respectively; $p = 1{,}74.10\text{-}5$) proportion. Note, HF, MF, and WF proportion are rather homogeneous for either SSP-S (0.9095; 0.9098; 0.883 R^2, respectively; Fig. 1c, d) or DSP-S (Fig. 1c, d).

Blinded study comparing SSP-based sequencing and Q-PCR

To shed light on the seemingly divergent findings in the literature, SSP-S and Q-PCR analysis of the same cfDNA extract were compared in a set of nine cancer patient samples (IC10, 15, 17, 20, 32, 33, 34, 35, and 37). The determination of the fractional size distribution using the Q-PCR method is illustrated in Fig. 2 for the IC17 patient as indicated in SI-1 (Fig. 2a–d). The fractional size distribution of all cancer patients is represented in SI-3 and Table 2a. HF, MF, and WF size range proportion values were determined either by SSP-S (SI-4a, SI-5, Fig. 1d) or Q-PCR analysis as illustrated in Fig. 2b–d in case of the IC17 patient. Data obtained from the IC17, IC34, and IC35 patients showed that HF was the main fraction ranging from 55.4% to 69.6% and from 68.5% to 78.8% determined by SSP-S and Q-PCR, respectively; MF was the second main fraction ranging from 28.8% to 42.8% and from 17.7% to 28.9% determined by SSP-S and Q-PCR, respectively; WF was by far the lowest fraction ranging from 1.6% to 2.5% and from 2.6% to 4.3% determined by SSP-S and Q-PCR, respectively (SI-6a, SI-7a). Consistent with the previous observation for IC17, IC34, and IC35, HF and MF + WF values determined by Q-PCR methods in all nine patient plasma tested in blind (IC10, 15, 17, 20, 32, 33, 34, 35, and 37) were similar (HF: 69.8 ± 5.1%) (SI-6b, SI-7b). Fractional size distribution according to HF, MF + WF, or WF determined by SSP-S and Q-PCR correlated significantly ($R^2 = 0.83$; $p < 0.0001$) (Fig. 2e, f), SI-6, Table 2 and SI-8). The percentage decrease from HF to MF and to WF was similar for both SSP-S and Q-PCR. The fraction decreased by four-fold to two-fold from HF to MF, and by 15-fold to 80-fold from HF to WF. Because of the paucity of the extracted DNA for carrying out Q-PCR analysis, only the HF and MF + WF fractions could be compared for some samples. Fractional size distribution by SSP-S was also determined according to the size of the amplicon length detected by Q-PCR (Table 2b).The cfDNA proportion shorter than 145 nt made up a substantial proportion of cfDNA determined by SSP-S and Q-PCR (66.2 ± 7.2 % and 69.8 ±

Table 1. Detailed characterization of the ~10 nt periodicity subpeaks observed from size distribution of cfDNA of cancer patients as determined by DSP-S and SSP-S

a

	Subpeak corresponding size											
	SSP-S (nt)									DSP-S (bp)		Mean SSP-S minus mean DSP-S
Peak	IC17	IC35	IC15	IC10	IC20	IC32	IC33	IC34	IC37	IC 58	IC61	
1	42	43	41	41	42	42	43	43	42	–	–	
2	53	52	52	53	52	52	52	53	52	–	–	
3	62	63	61	62	61	63	60	63	60	–	–	
4	73	72	73	74	73	71	73	73	73	–	–	
5	83	83	83	83	83	83	83	83	83	81	81	2.0
6	94	94	94	94	94	94	94	94	94	92	92	2.0
7	104	104	104	103	104	104	103	103	103	102	102	1.6
8	114	114	114	114	114	114	115	114	114	111	111	3.1
9	125	126	126	125	125	125	125	126	125	122	122	3.3
10	137	136	133	137	136	137	137	137	136	134	134	2.2
11	145	146	144	145	145	145	145	145	145	145	142	1.5
12	156	152	–	156	156	157	156	156	–	152	152	3.6
13	166	165	166	167	167	166	167	166	166	166	166	0.2

b

Periodicity	Subpeak periodicity										
	SSP-S (nt)									DSP-S (bp)	
	IC17	IC35	IC15	IC10	IC20	IC32	IC33	IC34	IC37	IC 58	IC61
(1–2)	11	9	8	12	7	10	9	10	7	–	–
(2–3)	9	11	9	9	9	11	8	10	8	–	–
(3–4)	11	9	12	12	12	8	13	10	13	–	–
(4–5)	10	11	10	9	10	12	10	10	10	–	–
(5–6)	11	11	11	11	11	11	11	11	11	11	11
(6–7)	10	10	10	9	10	10	9	9	9	10	10
(7–8)	10	10	10	11	10	10	12	11	11	9	9
(8–9)	11	12	12	11	11	11	10	12	11	12	12
(9–10)	12	10	7	12	11	12	12	11	11	12	12
(10–11)	8	10	11	8	9	8	8	8	9	11	8
(11–12)	11	6	–	11	11	12	11	11	21	7	10
(12–13)	10	13	–	11	11	9	11	10		14	14

a, DNA size corresponding to the periodic subpeaks for each patient. Peak numbers are ranked from the smallest to the largest size as observed by SSP-S and DSP-S expressed as nt and bp, respectively. b, Characterization of the fragment periodicity as determined by the lengths between two consecutive subpeaks, which are observed in the size profile of each cfDNA plasma extracts. DNA lengths obtained by SSP-S and DSP-S are expressed as nt and bp, respectively. (–), not detected

5.1%, respectively) (SI-7b). Altogether, the data showed the higher proportion of the fraction below 145 nt (Fig. 2e) and the statistical correlation between value obtained by both methods (Fig. 2f).

Post hoc cfDNA size profile examination of cancer patient plasma by Q-PCR analysis

In order to consolidate our observations on cfDNA size distribution by Q-PCR, we enlarged the number of cancer patients by including an additional seven plasma samples (SI-9 and SI-10). All seven plasma DNA exhibited similar size profiles, which were equivalent to the nine samples tested in the blinded study (SI-4b). A higher proportion of cfDNA lower than 145 nt was observed in these additional plasma samples. The 16 samples of the blinded study ($n = 9$) and the post hoc study ($n = 7$) showed size profile homogeneity with HF and MF + WF size fractions of 71.6 ± 5% and 28.4 ± 5%, respectively (SI-4b and SI-8a).

Estimation of average DNA molecule length

Estimation of average DNA molecule length was analyzed by the method of Deagle et al.[29] The mean fragment size for IC17, IC35, IC34, IC15, IC37, IC20, IC32, IC33, IC10, and IC104 was found as 63, 83, 83, 77, 91, 63, 77, 91, 71, and 62 nt, respectively (SI-11a). A correlation of the average fragment size determined by the Deagle et al.[29] method and the proportion of cfDNA MF + WF fraction was found (SI-11b).

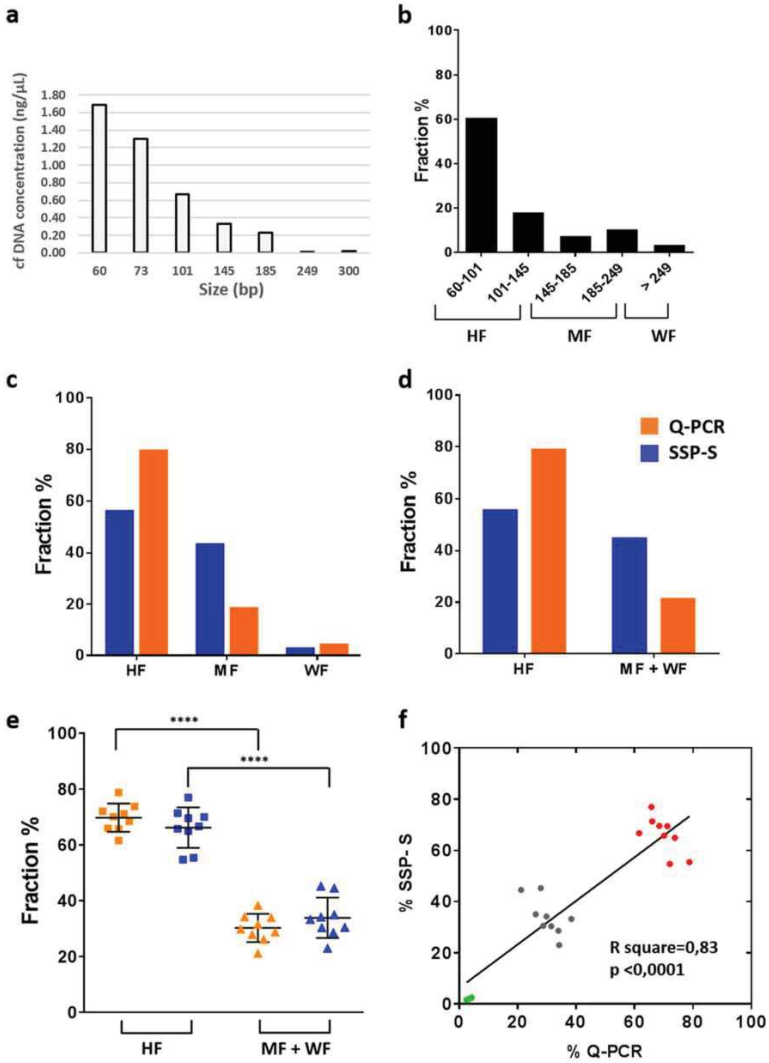


Fig. 2 Determination of cfDNA fragment size distribution by Q-PCR from plasma of a liver cancer patient (IC17). **a** Fragment length distribution is obtained by detecting variable amplicon lengths (bp) within the same DNA region. As cfDNA is randomly fragmented in plasma, the fragment amount, as measured by PCR, decreases monotonically with amplicon length, with a gradient that is a function of the underlying fragment length distribution. **b** Fractional size distribution, as determined under Methods Online, corresponds to the proportion of cfDNA per size range as a percentage of the total observed cfDNA by Q-PCR. **c**, **d**, blinded comparison of the fractional fragment size profile obtained by SSP-S (blue) and Q-PCR (orange). HF, highly fragmented DNA (<145nt); MF, mononucleosome fragmented DNA (145–249nt); WF, weakly fragmented DNA (>249nt); MF + WF (>145nt). **e** Combined data from the nine cancer patient plasma examined in the Q-PCR vs. SSP-S comparative study. cfDNA fractional fragment size distribution from cancer patients ($n = 9$) by Q-PCR and SSP-S methods is displayed upon two fractions: HF, highly fragmented DNA (<145nt) and MF + WF (>145 nt). For both analytical approaches, HF fraction is statistically different to the MF + WF fraction when using either Q-PCR or SSP-S (****$p = 1.65 \times 10^{-11}$ and, ****$p = 5.81 \times 10^{-8}$, respectively). **f** Correlation of the cfDNA fractional fragment size distribution determined by SSP analysis with that obtained from Q-PCR from the same samples of the blinded study ($n = 9$). Green, gray, and red dots correspond to the WF, MF + WF, and HF size fraction, respectively. Due to the low DNA concentrations in some patient plasma samples, cfDNA fractional distribution was not analyzed by all target sequence sizes among the examined samples. Data are expressed as a percentage of the cfDNA fractional fragment size

Study of the efficacy of targeting short sequences with respect to fragment size

By analyzing PCR with agarose gel electrophoresis, we showed that targeting a DNA sequence of the same size or longer than the input DNA fragment produced a similar PCR yield (SI-12). This demonstrates that efficacy of PCR targeting short DNA sequence is not restricted when the starting material contains short fragments equal or close to the amplicon size, and that does not lead to bias by favoring the amplification of longer fragments.

DISCUSSION

When directly comparing the cfDNA size profiles of cancer patient cfDNA extract obtained by SSP-S and DSP-S, a high discrepancy

Table 2. Fractional fragment size distribution of cfDNA of cancer patients by Q-PCR analysis (a) and SSP sequencing (b)

a

Sample ID	Patient gender	Cancer type	Stage	cfDNA yield (ng/ml)	Fraction						
					60–73	73–101	101–145	145–185	185–249	249–300	>300
IC17	M	Liver	IV	39	23.0%	37.6%	18.1%	7.5%	10.3%	2.7%	0.8%

Sample ID	Patient gender	Cancer type	Stage	cfDNA yield (ng/ml)	Fraction				
					60–101	101–145	145–185	185–249	>249
IC35	F	Breast	IV	16.2	50.1%	18.4%	8.2%	20.7%	2.6%
IC34	F			33.6	66.3%	5.8%	8.2%	15.5%	4.3%

Sample ID	Patient gender	Cancer type	Stage	cfDNA yield (ng/ml)	Fraction			
					60–101	101–145	145–185	>185
IC15	N	Lung	IV	22.5	51.2%	14.6%	15.6%	18.6%
IC37	F	Colorectal		15.9	61.4%	9.8%	4.9%	23.9%

Sample ID	Patient gender	Cancer type	Stage	cfDNA yield (ng/ml)	Fraction		
					60–101	101–145	>145
IC20	M	Lung	IV	21.9	54.7%	19.2%	26.2%
IC32	F			9.6	50.4%	15.7%	34.0%
IC33	M	Colorectal		13.8	61.4%	0.2%	38.4%

Sample ID	Patient gender	Cancer type	Stage	cfDNA yield (ng/ml)	Fraction	
					<145	>145
IC10	F	Lung	IV	11.4	70.1%	29.9%

b

Origin	Sample ID	Patient gender	Cancer type	Stage	cfDNA yield (ng/ml)	Fraction					
						SSP sequencing					
						30–59	60–100	101–145	146–180	181–249	250–1000
Blind study	IC32	F	Lung	IV	9.6	25.5%	24.0%	21.9%	22.0%	5.3%	1.3%
	IC10	F			11.4	20.5%	22.6%	22.7%	26.5%	6.4%	1.2%
	IC15	M			22.5	19.2%	23.2%	34.6%	18.6%	3.3%	1.1%
	IC20	M			21.9	18.9%	21.4%	24.7%	26.8%	6.5%	1.7%
	IC17	M	Liver		39	12.4%	16.4%	26.7%	35.3%	7.3%	2.0%
	IC37	F	Colorectal		15.9	21.2%	24.4%	24.4%	23.5%	5.1%	1.4%
	IC33	M			13.8	20.3%	23.0%	23.4%	25.2%	6.7%	1.3%
	IC35	F	Breast		16.2	22.7%	22.9%	24.0%	23.7%	5.1%	1.6%
	IC34	F			33.6	12%	18.9%	23.8%	34.5%	8.2%	2.5%
Fraction mean						19.2%	21.9%	25.1%	26.3%	6,00%	1.6%
Standard deviation (SD)						4.4	2.6	3.7	5.5	1.4	0.4

The same DNA extract was analyzed by both methodological approaches under blinded conditions. Data are expressed as percentage of the cfDNA size fraction

was found. The population corresponding to the DNA size wrapped around a mononucleosome (120–220-bp range, peaking up to 167 bp) was observed using both methods. Such a profile was analogous to that of several reports analyzing cfDNA[11,12,27] and the plasma of pregnant women or organ transplant recipients,[6] suggesting DNA fragmentation during apoptosis.[9,17] However, SSP-S revealed a substantial cfDNA fragment population ranging from 30 to 130 nt, especially from 30 to 80 nt, which was not detectable using the DSP library. CfDNA appeared more accessible for sequencing following SSP-S, as previously reported by our group[27] and then by Burnham et al.[28] It should be noted that the degree of depletion of short DNA molecules using double-stranded library preparation can be different across different studies, which can be due to adaptors clean-up steps.

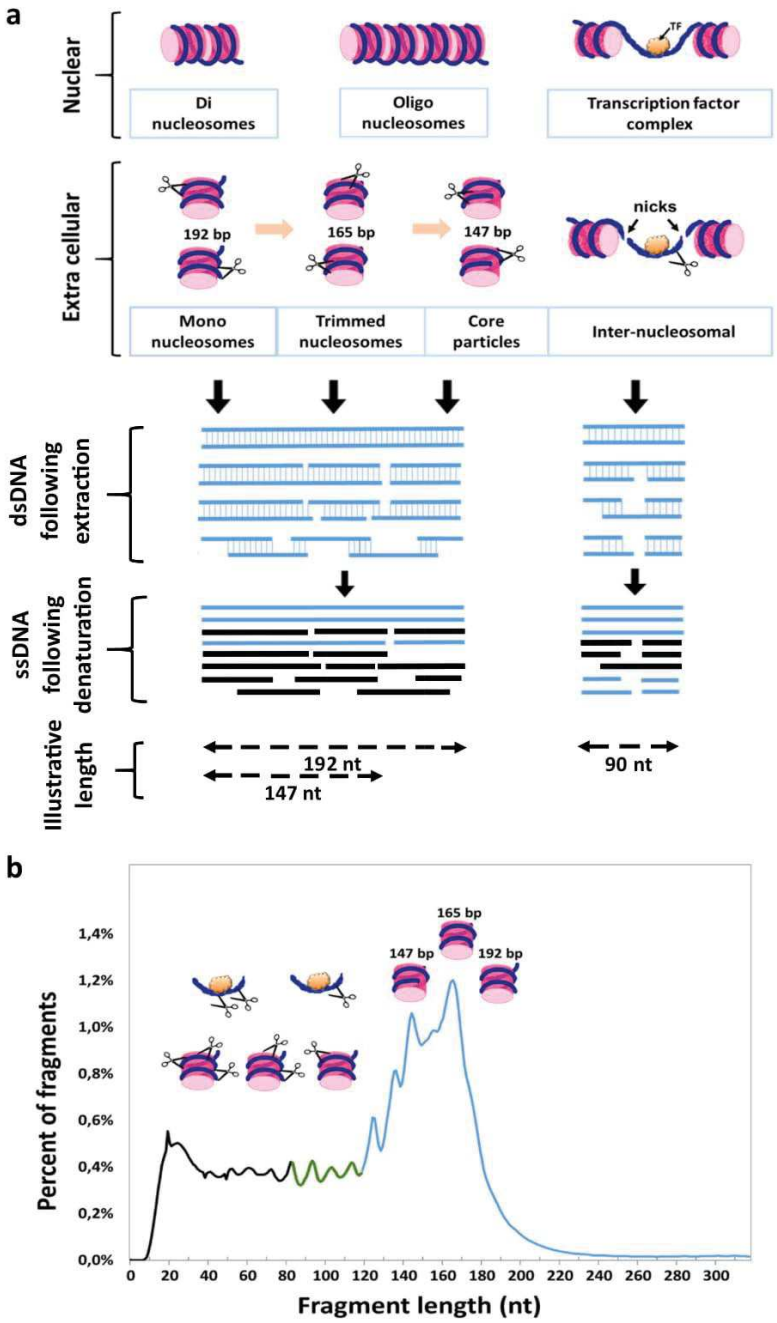
Fragments shorter than 100 nt are in abundance in cancer patient-derived plasma, but conventional DSP-S methods appeared insensitive to ultra-short cfDNA, emphasizing the need to use SSP-S for optimally examining cfDNA profiles. The SSP library has been recently described and used to generate high-resolution genomes when examining paleontological ancient DNA.[25,26] This method uses a single-strand DNA ligase and a 5′-phosphorylated and biotinylated adapter oligonucleotide to capture and bind single-strand DNA molecules to beads without prior end repair.[25] dsDNA is generated through use of primers from this ligation, and subsequently receives a second adaptor via blunt-end ligation. Completion of the adaptor sequence through an amplification reaction is then carried out from finished single strands obtained by heating the previously obtained molecules.

SSP-S offers various advantages over DSP-S with regard to the detection of cfDNA. The two most important reasons seem to be the different ligases used by each protocol and the denaturing step in the single-strand preparation. First, circligase, used in the SSP, is more efficient on small fragments than on longer fragments, and is almost certainly more efficient than T4 DNA ligase for short cfDNA fragments. Second, fragments that are damaged, for example, with nicks or abasic sites, are likely to be lost during DSP, but are retained during SSP, so if a sample has DNA damage (e.g., cfDNA), SSP is likely to capture the damaged molecules. DNA molecules with single-strand breaks on one or both strands may be present in cfDNA (Fig. 3). Whereas such molecules are completely lost under DSP procedure, SSP results to DNA break down into several fragments during heat denaturation and each fragment has an independent chance of being recovered in the library.[26] Third, through the initial biotinylation of cfDNA, all SSP reaction steps are performed while the DNA is tightly bound to the streptavidin-coated beads.[26] Loss of molecules in the DNA purification steps using silica spin columns or carboxylated beads, which are integral parts of DSP methods, are avoided. Fourth, DSP requires multiple bead-based, size-selective steps eliminating unwanted adapter-dimer products, whereas SSP does not require size-selective steps that eliminate shorter fragments.[28] Consequently, SSP libraries may contain a larger fraction of shorter molecules than those produced by the double-strand method as demonstrated by Bennett et al.[25] They observed that SSP improved the recovery of a higher proportion of mapped reads at almost every bin size, which could decrease for increasing fragment lengths although this is still controversial in the literature.[25] Hence, we cannot totally rule out that SSP-S enriched short over longer fragments and might generate a bias in the representation of the natural distribution.

Our blinded study using a Q-PCR method[18,20] on the same DNA extracts used for SSP-S showed strikingly similar fractional size distribution as those obtained using SSP-S. We previously showed using the Q-PCR method[30] that size distribution of cfDNA fragments is significantly lower than the conventional paradigm wherein the lowest size is that of the DNA sequence wrapped around a single histone octamer (147–200 bp, 180 bp mean).[18,20] Here, direct comparison of SSP-S and Q-PCR analysis showed a higher distribution of cfDNA fragments lower than 145 nt (145 bp corresponds to DNA wrapped around a nucleosomal core unit (167 bp) minus a linker fragment DNA of ~20 bp). Q-PCR is a robust and validated technique, but a few reports have scrutinized its efficiency and variation in quantifying short DNA fragments[25,31] (SI-12). We demonstrated in this study that our fractional size profile determination by Q-PCR assay does not show any bias in artificially enriching short vs. longer fragments.

Naked DNA is very rapidly degraded in the blood circulation as the half-life of intact DNA without a double-strand break has been estimated to be less than a minute.[32] Consequently, only cfDNA protected by stable structures can be detected in the blood-stream. Nuclear cfDNA fragmentation results from mapping locations of the chromatin organization along the genome, which protect/packed DNA with mononucleosomes as the lower unit. At least two key DNA/protein complexes enabling DNA protection from blood nucleases may be considered: DNA wrapped around a histone octamer or DNA bound to transcription factor (TF). Since linker DNA between nucleosomes is vulnerable to digestion, lengths corresponding to one nucleosomal subunit appear to be the most prevalent and conserved size with di- and tri-nucleosomal lengths showing much lower proportions.[16] Stable nucleosome associated structures may be 192 bp (mononucleosome plus linker), 165 bp (trimmed mononucleosome), or 147 bp (core particle: nucleosome excluding the DNA connected to the peripheral histone H1, which adds ~20 bp; Fig. 3). Trimmed mononucleosome cfDNA-associated structures (165 bp) appear to be preferentially protected as shown by its prevalence in the cfDNA size profile. Note, as already observed by Chan et al.[19] our data indicate that the fraction cfDNA fragments over the size of DNA wrapped in a mononucleosome as determined by Q-PCR as well as by SSP-S or DSP-S is very minor ($< 1.8\%$ and $< 6\%$, respectively). Whereas size distribution analysis by Q-PCR is not limited to DNA size over ~40 bp, size profile analysis through DSP- or SSP-S, as performed here, is limited to cfDNA of fragments under ~1000 bp and thus precludes examination of cfDNA of higher molecular weight. Nevertheless, cfDNA quantitations were similar when using Q-PCR and SSP-S, suggesting that high molecular weight (over 350 bp) cfDNA is a minor component (~2%) in terms of genome equivalent copy number in cancer patients (Fig. 3). Altogether, this suggest that presence of DNA circulating within di- or oligonucleosomes is minor and that high molecular weight DNA poorly circulate in cancer patients' blood. Note, this is observed when stringent protocol for the pre-analytical conditions are used. We may postulate that the significant presence of high molecular weight in a cancer patient cfDNA extract could indicate a possible contamination of genomic DNA from lysed blood cells and may be a pre-analytical parameter to assess as quality control of the cfDNA extract.

The presence of cfDNA fragments lower than 100 bp may be explained by various hypotheses. First, degradation at both linker extremities of pieces of DNA protected from TFs previously bound to linker DNA between two nucleosomes (with length ranging between 20 and 90 bp and varying among different species, or tissues; Fig. 3) may release protected short double-stranded fragments into the blood circulation. Second, and more likely, DNA double-stranded or single-stranded breaks may occur in bloodstream inside or outside both types of DNA/protein complexes, inside or outside cells. Following DNA denaturation during PCR or SSP, the resulting single strands may be of varying size. Since we clearly observed the detection of polymerized short double-stranded DNA and sequencing of short sequences from SSP, it is reasonable to assume that the possible sources of detected short ssDNA fragments include both short double-stranded cfDNA (<145 bp) or nicked double-standed cfDNA of higher size. The ~10nt periodicity, within the 41–166nt range, observed with using SSP-S demonstrated the presence of nucleosome-derived degradation since this pattern has been attributed to the internal nucleosome cleavage of accessible nucleotides that lie further from the surface of the histone core at each helical turn as DNA wraps around the core.[15,33] Observation of periodicity lower than 145 bp down to 81 bp by DSP-S might reveal the presence in blood of short double-stranded DNA associated to nucleosomes. Calculation of the number of reads cannot provide a true estimation of the percentage of nicked intranucleosomal cfDNA. However, since both SSP-S and DSP-S show the same peak at 166 bp we could consider that, at this size, a fraction of cfDNA molecule fragments, at least in one strand, are free of nicks, as illustrated in Fig. 3. DSP-S analysis showed that cfDNAs are principally associated with histones and that the lowest dsDNA fragment length is approximately 80 bp. SSP-S analysis, on the other hand, showed that the detected single-



strand cfDNA fragments below the size covered in a mono-nucleosomal core (<145 nt) were initially mostly associated with histones. This suggests that there is at most only a minor fraction of short histone-free fragments. Since histone association implies dsDNA secondary structure, our data suggest that there is negligible single-stranded DNA circulating in blood.

Note, our previous observations by AFM analysis support the existence of short ds cfDNA structures as a significant proportion of cfDNA from cancer patients was ranging between 100 and 145 bp.[8] In addition, previous studies of DNase I cleavage patterns identified two dominant classes of fragments: longer fragments associated with cleavage between nucleosomes, and shorter fragments associated with cleavage adjacent to TF-binding sites.[34] By generating maps of genome-wide in vivo nucleosome occupancy, we previously found that short cfDNA fragments (35–80 bp) harbor footprints of TFs.[27] Although additional observations from the literature are needed to estimate the proportion or significance of TF-associated cfDNA, higher scrutiny of those short cfDNA fragments might provide a new diagnostic potential based on TF presence. Note, the use of gel

Fig. 3 Schematic diagrams of circulating DNA fragmentation from nucleosomes. **a** Two hypotheses are presented: DNA wrapping around a histone octamer or bound to transcription factor (TF). Different types of dsDNA fragments are schematically represented and may exhibit nicks. DNA double-strand or single-strand breaks may occur inside or outside both types of DNA/protein complexes, inside or outside cells. Following DNA denaturation (such as under PCR or SSP preparation), the resulting single strands may be of varying size from several oligonucleotides to few hundreds of nucleotides. The lengths given are indicative. They are based on nucleosome consisting of an octamer of core histone proteins wrapped ~1.65 times by 147 bp of DNA and on the presence of linker DNA describing the non-nucleosomal DNA connecting two or more nucleosomes in an array with length ranging between 20 and 90 bp and varying among different species, or tissues. SsDNA fragments produced by SSP or Q-PCR are subsequently replicated, and sequenced or quantified by SSP-S or Q-PCR, respectively. **b** cfDNA structures and fragmentation with regards to size profile as determined by SSP-S of cfDNA extracted as illustrated from the IC17 patient. Three fractions of the size profile could approximately be distinguished in light of our observations and other works:[12,13,16,27] Blue curve, DNA fragments originating from cfDNA packed within mononucleosome without any intranucleosomal nicks revealed by both SSP-S and DSP-S; green curve, DNA fragments originating from cfDNA packed within mononucleosome exhibiting nuclease nicks, or within TFs without any nicks, observed by both SSP-S and DSP-S; and black curve, DNA fragments originating from cfDNA packed within mononucleosome with more nicks or within TFs with nicks, which are only observed by SSP-S

electrophoresis assay for cfDNA sizing, which is only based on detecting dsDNA never showed cfDNA fragment length peaking below 180 bp.[1]

All those various reasons concur to the significant discrepancy found in size profile, and that cfDNA structures are of high diversity spanning from tightly packed long dsDNA, mononucleosomes or oligonucleosomes, heminucleosome formation, short-sized TF-binding dsDNA, long-sized and short-sized DNA-associated microparticles, short-sized lipoproteonucleic complexes, and cell or cell-part association. These structures would be all subject to endonuclease and exonuclease degradation as soon as they are released from cells into the blood circulation. Our data showed that nucleosomal structures are one of the least degradable cfDNA structures with, to a lesser extent, TF-associated cfDNA. Apoptosis might appear as the main source of cfDNA; however, short-sized nucleosomal structures could also be the results of the progressive nuclease degradation of higher-sized cfDNA originating from necrosis, phagocytosis, microparticle-containing DNA, or active release from lymphocytes.[1] CfDNA structure diversity therefore results from different biological phenomena: various cellular mechanisms of release, dynamic nucleic or proteic degradation in the circulation, and potential association with blood constituents.

We recently demonstrated that deep-sequencing cfDNA for mapping genome-wide in vivo nucleosome occupancy may reveal its tissues of origin.[27] The many structures and mechanical origins have shown that cfDNA is a complex entity. Sizing following cfDNA extraction cannot fully account for characterizing their structures. Nevertheless, we may hypothesize that the level of fragmentation vary upon cfDNA origins (mitochondrial, nuclear, tumor or healthy cells, lymphocytes, tumor microenvironment cells, metastatic cells, etc.) and information on sizing may be key in accurately detecting and quantifying cfDNA. In the light of this assumption, optimal detection and discrimination of cell-free DNA collected from other body constituents may rely on sizes specific to their origin.

This study shows that much higher cfDNA copies may be readily recovered by selecting/targeting short single-strand fragments, consequently providing higher sensitivity when detecting genetic or epigenetic alterations when testing cancer patient plasma. Based on our initial observation on cfDNA size distribution[18] and the necessity of targeting short DNA sequences (50–80 bp) for optimal detection by Q-PCR, this strategy was first taken into consideration to an allele specific with blocker Q-PCR method (IntPlex), which demonstrated very high sensitivity,[35,36] and afterwards when accordingly setting other PCR-based methods, such as single locus Q-PCR,[22] Beaming,[21] or dPCR,[23] or by sequencing and selective amplification.[37,38] Since cfDNA fragments, and especially mutant cfDNA in cancer patient, may be poorly represented in blood, optimal recovery of cfDNA is required for its analysis. Several reports have showed that SSP-S appears better suited than conventional DSP-S for obtaining an optimal analytical signal. Alternatively, Moser et al.[39] did not observe a preferential enrichment of circulating DNA.[39] Thus, it is still debatable as to whether or not SSP will definitively improve the quantification performance, and whether a shift towards SSP-S for analyzing cfDNA in a clinical practice is warranted.

As well as improving cfDNA recovery for optimal detection, knowledge on sizing may also enable subject stratification. We reported that total cfDNA[18] as well as mutant cfDNA[20] of cancer patients is more fragmented than that of healthy individuals or of wild-type cfDNA, respectively, by using a Q-PCR-based method. The presence of more fragmented DNA molecules in cancer patients was further elucidated in another study with the use of sequencing technology based on double-stranded library preparation.[20]

The limitations of this study are detailed in the supplementary notes. Briefly, these concern not examining the presence of mitochondria-derived cfDNA, the cancer stages other than stage IV, and potential bias with the extraction procedure. Moreover, the main potential theoretical factors that might contribute to the difference between % SSP-S and % Q-PCR reside in the analytical size window of the methods used here: our ultra-deep-sequencing method spans from ~30 to ~1000 bp and our Q-PCR method over 60 bp. Consequently, our comparative study should be limited to the 60 to ~1000 bp fragment size range. In addition, the study could not determine whether fragment size profiles in cfDNA are associated with tissue types and cancer types as previously reported.[27,40–42] Furthermore, this study only focused on cancer patient plasma and all resulting observations should not automatically be applied to healthy individual plasma. Although ultra-deep-sequencing analysis showed in previous reports a roughly similar size distribution pattern in healthy and cancer plasma, previous works reported various distinguishing characteristics.[11,18,20,41,42] We cannot rule out that in-depth scrutiny of size profile may reveal discriminating clear-cut assessment between healthy and cancer patient.

In conclusion, this study confirms the crucial importance of examining the structural features of any analytes circulating in blood, in particular with regards to their association with hetero-compounds. We compared DSP, SSP, and Q-PCR analysis in a blinded study to update and assimilate previous knowledge of cfDNA size profiles. The fragment length distribution of cfDNA, extracted from plasma of cancer patients, was very similar with the SSP-S and Q-PCR methods, which both rely on the analysis of single-stranded DNA as the initial matrix. Both approaches were clearly effective in optimally measuring cfDNA copy number, because a substantial fraction of cfDNA found by these methods consisted of short fragments that are not readily detectable by standard DSP protocols. We also observed that most of the detectable cfDNA in blood, as well as most of the shortest cfDNA fragments (down to ~40 nt), have footprint of a nucleosome, which appears the most stabilizing structure for DNA in the circulation. We conclude that cellular DNAs, initially packaged in

Table 3. Patient clinical characteristics

	Sample ID	Center	Clinical diagnosis	Stage	Gender	Age	cfDNA yield (ng/ml)
DSP-S	IC58	Plasma Lab	Lung cancer (squamous cell carcinoma)	IV	M	63	8.1
	IC61		Lung cancer (small cell)		M	73	7.2
Blind study (SSP-S VS Q-PCR)	IC10		Lung cancer (adenocarcinoma)		F	65	11.4
	IC32		Lung cancer (small cell)		F	69	9.6
	IC33		Colorectal cancer (adenocarcinoma)		M	65	13.8
	IC34		Breast cancer (invasive/infiltrating lobular carcinoma)		F	62	33.6
	IC15		Lung cancer (small cell)		M	70	22.5
	IC17		Liver cancer (hepatocellular carcinoma)		M	62	39
	IC20		Lung cancer (squamous cell carcinoma)		M	60	21.9
	IC35		Breast cancer (ductal carcinoma in situ)		F	76	16.2
	IC37		Colorectal cancer (adenocarcinoma)		F	58	15.9
Post hoc (Q-PCR)	IC101	ICM	Colorectal cancer (adenocarcinoma)	IV	M	72	53.9
	IC102		Colorectal cancer (adenocarcinoma)		F	71	14.2
	IC103		Colorectal cancer (adenocarcinoma)		M	–	11.7
	IC104		Colorectal cancer (adenocarcinoma)	III	M	72	29.2
	IC105		Colorectal cancer (adenocarcinoma)	IV	–	–	22.2
	IC108		Colorectal cancer (adenocarcinoma)		M	52	24
	IC109		Colorectal cancer (adenocarcinoma)		M	71	21

Two patient plasma were subjected to DSP-S; nine patient plasma were examined in blind by Q-PCR and SSP-S; seven patient plasma were subjected to Q-PCR in a post hoc analysis. F female, M male, – not available

chromatin, are released by different biological phenomena in the extracellular compartment in various structures undergoing degradation down to nucleosomes or to a lesser extent TF-associated subcomplexes, resulting from continuous dynamic internucleosomal and intranucleosomal nuclease activity. Thus, detectable cfDNA are mostly composed of a complex mixture of highly degraded DNA as regards to their primary, secondary, or tertiary structures. As sensitivity is clearly a limitation of cfDNA applications, delineating the structural features of cfDNAs may help adapt optimal analytical approaches to study cancer progression or tumor biology.

METHODS

This research was conducted in accordance with all relevant guidelines and procedures, approved by the Institute of Research in Cancerology, and the INSERM.

Clinical samples

Blood samples were collected from individuals with stage III and IV cancer (IC $n = 18$) (Table 3): colorectal cancer ($n = 9$), lung cancer ($n = 6$), breast cancer ($n = 2$), and liver cancer ($n = 1$). Samples were obtained from Conversant Bio ($n = 11$, Huntsville, AL, USA) and for the post hoc study from the Cancer Institute of Montpellier (ICM; $n = 7$, Val d'Aurelle, Montpellier, France). All patients signed an informed consent and CRC samples from the ICM were obtained from the study EUDRACT 2016-001490-33. Samples were handled accordingly with a pre-analytical guideline previously established by our group.[43] The study followed the REMARK reporting guidelines. Written informed consent was obtained from all participants.

Plasma isolation and cfDNA extraction

All blood samples were collected in 4-ml EDTA tubes and plasma DNA were extracted as described in detail in Supplementary Information Materials and methods.

Preparation of sequencing libraries

Between 0.5 and 10.0 ng of cfDNA were used as inputs for all libraries. Library amplification for all samples was monitored by real-time PCR to avoid over-amplification, and was typically terminated after 4–6 cycles.

Preparation of the double-stranded sequencing library. Conventional, double-stranded sequencing libraries were prepared with the ThruPLEX DNA-seq 48D or ThruPLEX Plasma-seq Kits (Rubicon Genomics), comprising a proprietary series of end-repair, ligation, and amplification reactions.

Preparation of the single-stranded sequencing library. Single-stranded sequencing libraries were prepared according to a protocol adapted from Gansauge et al.[26] with using a double-stranded adapter (SI-13) as described in detail in Supplementary Information Materials and methods.

Size profile analysis by deep sequencing

All libraries were sequenced on HISeq 2000 or NextSeq 500 instruments (Illumina) as described in detail in Supplementary Information Materials and methods.

Size profile analysis by Q-PCR

The oligonucleotide primers target DNA sequences of increasing size in human *KRAS* region intron 2 (SI-14). The size of the amplicons was 60, 73, 101, 145, 185, 249, and 300 bp. The reverse primer used was the same for all sizes. Our Q-PCR experiments followed the MIQE guideline.[44] Q-PCR amplifications and analysis were performed as described in detail in Supplementary Information Materials and methods

Estimation of average DNA molecule length

The average DNA molecule length was estimated according to the method set by Deagle et al.[29] for quantifying damage in DNA recovered from highly degraded samples[31] as described in detail in Supplementary Information Materials and methods.

Analysis of the amplification of short mutant synthetic DNA fragments by Q-PCR

In order to confirm that targeting short sequences amplify the expected fragment size and that no bias exists in the preferential amplification of larger fragment, we amplified two fragments of mutant synthetic DNA of 61 and 103 bp (SI-12) and performed Q-PCR and agarose gel electrophoresis analysis as described in detail in Supplementary Information Materials and methods.

Statistical analysis

Statistical analysis was performed using the GraphPad Prism V6.01 software. The Student's *t* test was used to compare means. A probability of <0.05 was considered to be statistically significant: $*p=0.05$, $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$. The Pearson's test was used for correlation analysis.

DATA AVAILABILITY

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

ACKNOWLEDGEMENTS

We thank Marc Ychou, Brice Pastor, Amaëlle Otandault, Zahra Al Amir Dache, and Safia El Messaoudi. We also thank Philippe Blache, Thierry Grange, Eva-Maria Geigl, and Andrew Bennett for helpful discussion.

AUTHOR CONTRIBUTIONS

C.S., M.W.S, J.S., and A.R.T. analyzed and interpreted the patient data regarding the sequencing and PCR analysis. C.S., M.W.S, R.T., J.S., and A.R.T. contributed in writing the manuscript. All authors read and approved the final manuscript. AR.T. is the guarantor and is supported by INSERM. This study was partly supported by the SIRIC Montpellier Grant « INCa-DGOS-Inserm 6045», France.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Genomic Medicine* website (https://doi.org/10.1038/s41525-018-0069-0).

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Thierry, A. R., El Messaoudi, S., Gahan, P. B., Anker, P. & Stroun, M. Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastas-. Rev.* **35**, 347–376 (2016).
2. Stroun, M. et al. Neoplastic characteristics of the DNA found in the plasma of cancer patients. *Oncology* **46**, 318–322 (1989).
3. Wan, J. C. M. et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).
4. Diehl, F. et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci. USA* **102**, 16368–16373 (2005).
5. Mandel, P. & Metais, P. Les acides nucléiques du plasma sanguin chez l'homme. *C. R. Seances Soc. Biol. Fil.* **142**, 241–243 (1948).
6. Lo, Y. M. D. et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2**, 61ra91 (2010).
7. Leon, S. A., Shapiro, B., Sklaroff, D. M. & Yaros, M. J. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res.* **37**, 646–650 (1977).
8. Mouliere, F., El Messaoudi, S., Pang, D., Dritschilo, A. & Thierry, A. R. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol. Oncol.* **8**, 927–941 (2014).
9. Holdenrieder, S. et al. Circulating nucleosomes in serum. *Ann. N.Y. Acad. Sci.* **945**, 93–102 (2001).
10. Aucamp, J., Bronkhorst, A. J., Badenhorst, C. P. S. & Pretorius, P. J. A historical and evolutionary perspective on the biological significance of circulating DNA and extracellular vesicles. *Cell. Mol. Life Sci.* **73**, 4355–4381 (2016).
11. Jiang, P. et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl. Acad. Sci.* **112**, E1317–E1325 (2015).
12. Underhill, H. R. et al. Fragment length of circulating tumor DNA. *PLoS Genet.* **12**, e1006162 (2016).
13. Ivanov, M., Baranova, A., Butler, T., Spellman, P. & Mileyko, V. Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. *BMC Genom.* **16**, S1 (2015).
14. Kim, S. K. et al. Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. *Prenat. Diagn.* **35**, 810–815 (2015).
15. Szerlong, H. J. & Hansen, J. C. Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochem. Cell. Biol.* **89**, 24–34 (2011).
16. Chandrananda, D., Thorne, N. P. & Bahlo, M. High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. *BMC Med. Genom.* **8**, 29 (2015).
17. Jahr, S. et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **61**, 1659–1665 (2001).
18. Mouliere, F. et al. High fragmentation characterizes tumour-derived circulating DNA. *PLoS ONE* **6**, e23418 (2011).
19. Chan, K. C. A. et al. Size distributions of maternal and fetal DNA in maternal plasma. *Clin. Chem.* **50**, 88–92 (2004).
20. Mouliere, F. et al. Circulating cell-free DNA from colorectal cancer patients may reveal high KRAS or BRAF mutation load. *Transl. Oncol.* **6**, 319 (2013).
21. Heitzer, E. et al. Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer. *Int. J. Cancer* **133**, 346–356 (2013).
22. Andersen, R. F., Spindler, K.-L. G., Brandslund, I., Jakobsen, A. & Pallisgaard, N. Improved sensitivity of circulating tumor DNA measurement using short PCR amplicons. *Clin. Chim. Acta* **439**, 97–101 (2015).
23. Garlan, F. et al. Circulating tumor DNA measurement by picoliter droplet-based digital PCR and vemurafenib plasma concentrations in patients with advanced BRAF-mutated melanoma. *Target. Oncol.* **12**, 365–371 (2017).
24. Ng, S. B. et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
25. Bennett, E. A. et al. Library construction for ancient genomics: single strand or double strand? *Biotechniques* **56**, 289–290 (2014). 292–6, 298, passim.
26. Gansauge, M.-T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).
27. Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* **164**, 57–68 (2016).
28. Burnham, P. et al. Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. *Sci. Rep.* **6**, 27859 (2016).
29. Deagle, B. E., Eveson, J. P. & Jarman, S. N. Quantification of damage in DNA recovered from highly degraded samples—a case study on DNA in faeces. *Front. Zool.* **3**, 11 (2006).
30. Thierry, A. R. et al. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Res.* **38**, 6159–6175 (2010).
31. Guimaraes, S. et al. Owl pellets: a wise DNA source for small mammal genetics. *J. Zool.* **298**, 64–74 (2016).
32. Thierry, A. et al. Characterization of liposome-mediated gene delivery: expression, stability and pharmacokinetics of plasmid DNA. *Gene Ther.* **4**, 226–237 (1997).
33. Langley, S. A., Karpen, G. H. & Langley, C. H. Nucleosomes shape DNA polymorphism and divergence. *PLoS Genet.* **10**, e1004457 (2014).
34. Vierstra, J., Wang, H., John, S., Sandstrom, R. & Stamatoyannopoulos, J. A. Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. *Nat. Methods* **11**, 66–72 (2014).
35. Thierry, A. R. et al. Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. *Nat. Med.* **20**, 430–435 (2014).
36. Thierry, A. R. et al. Clinical utility of circulating DNA analysis for rapid detection of actionable mutations to select metastatic colorectal patients for anti-EGFR treatment. *Ann. Oncol.* **28**, 2149–2159 (2017).
37. Reckamp, K. L. et al. A highly sensitive and quantitative test platform for detection of NSCLC EGFR mutations in urine and plasma. *J. Thorac. Oncol.* **11**, 1690–1700 (2016).
38. Cohen, J. D. et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).
39. Moser, T. et al. Single-stranded DNA library preparation does not preferentially enrich circulating tumor DNA. *Clin. Chem.* **63**, 1656–1659 (2017).
40. Lehmann-Werman, R. et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc. Natl. Acad. Sci. USA* **113**, E1826–E1834 (2016).
41. Guo, S. et al. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat. Genet.* **49**, 635–642 (2017).

42. Widschwendter, M. et al. The potential of circulating tumor DNA methylation analysis for the early detection and management of ovarian cancer. *Genome Med.* **9**, 116 (2017).
43. El Messaoudi, S., Rolet, F., Mouliere, F. & Thierry, A. R. Circulating cell free DNA: preanalytical considerations. *Clin. Chim. Acta* **424**, 222–230 (2013).
44. Bustin, S. A. et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).

B. Progrès récents dans les applications cliniques des acides nucléiques circulants en oncologie

Titre : Recent advances in circulating nucleic acids in oncology

Auteurs

Amaelle Otandault, Philippe Anker, Zahra Al Amir Dache, Vanessa Guillaumon, Romain Meddeb, Brice Pastor, Ekaterina Pisareva, Cynthia Sanchez, **Rita Tanos**, Geoffroy Tousch, Heidi Schwarzenbach et Alain R. Thierry

Cette revue a été publiée dans *Annals of Oncology* le 7 Fevrier 2019.

Résumé

L'ADNcir est l'un des domaines en oncologie, qui a connu une croissance rapide au cours de ces dernières années. Ses utilisations potentielles en clinique couvrent maintenant chaque phase de la prise en charge des patients atteints de cancer (l'information prédictive, la détection de la maladie minimale résiduelle, la détection précoce de la résistance et la surveillance du traitement, la surveillance de la récurrence ainsi que la détection précoce et le dépistage du cancer). Cette revue décrit les progrès récents dans l'application de l'ADN ou de l'ARN circulant en oncologie en s'appuyant sur des résultats et des travaux initiaux ou encore non-publiés présentés lors du 10ème symposium international sur les acides nucléiques circulants dans le plasma et le sérum (CNAPS pour Circulating Nucleic Acids in Plasma and Serum) qui s'est tenu à Montpellier du 20 au 22 septembre 2017. Durant ce congrès, les congressistes ont présenté leurs récentes observations et résultats portant notamment sur (i) la biologie, les structures des ADNcir et leurs implications pour une détection optimale ; (ii) le rôle des ADNcir dans les processus métastatiques ou immunologiques ; (iii) l'évaluation des panels d'ARNm pour le suivi des patients cancéreux ; (iv) la détection de la maladie minimale résiduelle ; (v) l'évaluation du dépistage du cancer par l'analyse de l'ADNcir ; et (vi) des recommandations préanalytiques pour cette analyse. Cette revue liste non seulement les avancées récentes dans le domaine, présentées lors du congrès, mais aussi des résultats récemment publiés dans la littérature. Elle propose une vue

d'ensemble de la recherche fondamentale et de son potentiel, ainsi que de la mise en œuvre et des défis actuels dans l'utilisation des acides nucléiques circulants en oncologie.

Contributions

J'ai participé à l'organisation du 10ème symposium international sur les acides nucléiques circulants dans le plasma et le sérum (CNAPS) qui a eu lieu du 20 au 22 Septembre 2017 à Montpellier, et à la rédaction des parties de la revue portant sur les dernières études et travaux effectués sur l'ADN circulant dans le domaine du dépistage du cancer ainsi que la maladie minimale résiduelle et la récurrence.

ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

Annals of Oncology 30: 374–384, 2019
doi:10.1093/annonc/mdz031
Published online 7 February 2019

REVIEW

Recent advances in circulating nucleic acids in oncology

A. Otandault[1,2,3,4], P. Anker[1,2,3,4], Z. Al Amir Dache[1,2,3,4], V. Guillaumon[4,5], R. Meddeb[1,2,3,4], B. Pastor[1,2,3,4], E. Pisareva[1,2,3,4], C. Sanchez[1,2,3,4], R. Tanos[1,2,3,4], G. Tousch[1,2,3,4], H. Schwarzenbach[6] & A. R. Thierry[1,2,3,4]*

[1]IRCM, Institute of Research in Oncology of Montpellier, Montpellier; [2]INSERM, U1194, Montpellier; [3]Department of Oncology, Montpellier University, Montpellier; [4]Regional Institute of Cancer of Montpellier, Montpellier; [5]SIRIC, Integrated Cancer Research Site, Montpellier, France; [6]Department of Tumor Biology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Correspondence to: Dr Alain R. Thierry, INSERM, U1194, IRCM, 208 rue des Apothicaires, 34298 Montpellier cedex 5, France. Tel: +33-6-63-82-19-94; E-mail: alain.thierry@inserm.fr

Circulating cell-free DNA (cfDNA) is one of the fastest growing and most exciting areas in oncology in recent years. Its potential clinical uses cover now each phase of cancer patient management care (predictive information, detection of the minimal residual disease, early detection of resistance, treatment monitoring, recurrence surveillance, and cancer early detection/screening). This review relates the recent advances in the application of circulating DNA or RNA in oncology building on unpublished or initial findings/work presented at the 10th international symposium on circulating nucleic acids in plasma and serum held in Montpellier from the 20th to the 22nd of September 2017. This year, presenters revealed their latest data and crucial observations notably in relation to (i) the circulating cell-free (cfDNA) structure and implications regarding their optimal detection; (ii) their role in the metastatic or immunological processes; (iii) evaluation of miRNA panels for cancer patient follow up; (iv) the detection of the minimal residual disease; (v) the evaluation of a screening tests for cancer using cfDNA analysis; and (vi) elements of preanalytical guidelines. This work reviews the recent progresses in the field brought to light in the meeting, as well as in the most important reports from the literature, past and present. It proposes a broader picture of the basic research and its potential, and of the implementation and current challenges in the use of circulating nucleic acids in oncology.

Key words: cancer, circulating DNA, diagnostic, treatment, resistance, screening

Introduction

Use of circulating DNA (Figure 1) or RNA (Figure 2) received high scrutiny (Figure 3), especially during the last decade in various oncology fields (Figure 4). Circulating nucleic acids in plasma and serum (CNAPS) meetings (supplementary data S1, available at *Annals of Oncology* online) are unique in that they focus on and link together basic and applied research in various pathologies for nearly 20 years. Held once every 2 years, they are typically the occasion for the early presentation of new observations, and gave us the opportunity to relate these to previous milestone reports and recent literature, and to review the potential of these new avenues of research, as well as the challenges, which exist in developing the use of circulating nucleic acids in oncology.

Structures and tissue origins

Circulating cell-free DNA (cfDNA) can be present in the form of protein-associated DNA fragments or extracellular vesicles (EV), in the physiological circulating fluids of healthy and diseased individuals as recently reviewed [1, 2]. cfDNA is derived not only from genomic DNA but also from extrachromosomal mitochondrial DNA. While various clinical applications of cfDNA are currently progressing [3, 4], the identification of structural characteristics remains under investigation.

Confirming pioneering work on size profiling [5, 6], a higher fragmentation of cfDNA in metastatic colorectal and non-small-cell lung cancer (NSCLC) patients compared with healthy donors, and a significant difference in size profile between mutated and wild type DNA in cancer patients were

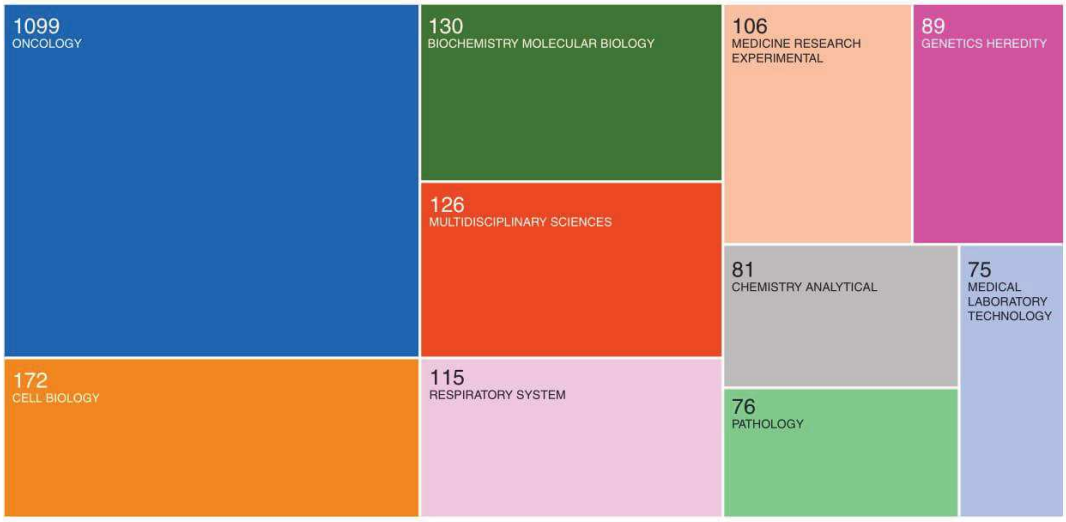


Figure 1. Publication number reporting use of circulating DNA in cancer upon research areas in 2017 and 2018. Web of science citation reports showing 2113 records for TOPIC: (cancer circulating DNA) in 2017 and 2018. This citation level corresponds to 36% to the total number of records up to 2018 (5800 records). A 51% of the publications concerned use in clinical oncology.

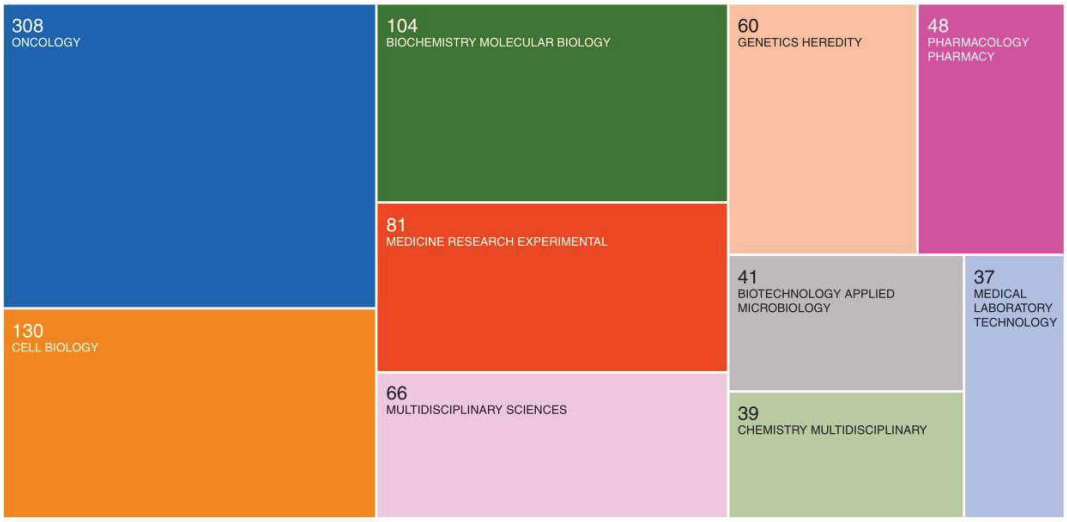


Figure 2. Publication number reporting use of circulating RNA in cancer upon research areas in 2017 and 2018. Web of Science citation reports showing 912 records for TOPIC: (cancer circulating RNA). This citation level corresponds to 24% to the total number up to 2018 (3751 records). A 34% of publications concerned use in clinical oncology. Number of records during that period of time was more than twofold less than records reporting circulating DNA in cancer (2113). The number of reports on the use of microRNA was nearly threefold higher than those on mRNA.

demonstrated, during the CNAPS, by Garlan et al., using the BIABooster system [7]. Similarly, Piskorz et al. presented their previous work which showed that in an ovarian cancer mouse, circulating tumor DNA fragments are shorter than nontumor cell-free DNA in plasma, and that the selection of fragments between 90 and 150 bp, using targeted and whole-genome sequencing (WGS) approaches, could enrich the tumor DNA up to 11-fold [8].

During the meeting, a blinded study by Sanchez et al. [9] demonstrated that circulating tumor DNA from cancer patients showed a similar size profile using either whole genome deep sequencing from single strand library preparation or using a q-PCR based method, revealing a higher proportion of shorter cfDNA fragments (below 80 bp) not readily detectable by standard double-stranded DNA library preparation protocols (as previously shown by Underhill et al.). This showed the higher capacity of single strand DNA template analysis to detect cfDNA fragments, and revealed that cfDNA molecules, which are initially packed in chromatin, are released from cells and then dynamically degraded in blood, both within and between nucleosomes or transcription factor-associated subcomplexes. In addition, this study revealed that cfDNA size distribution obtained from the

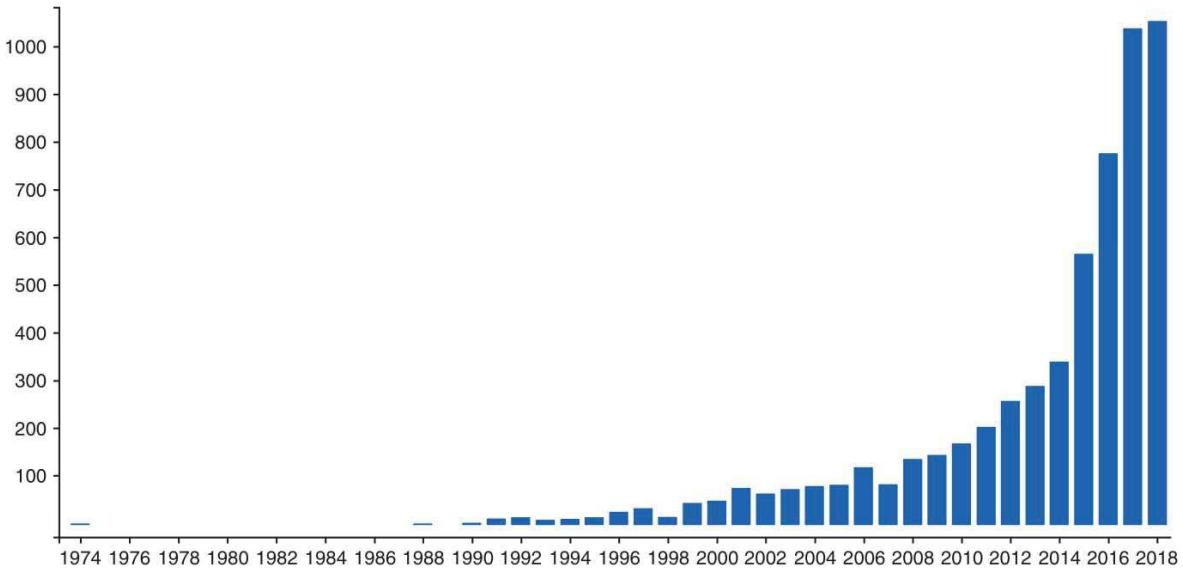


Figure 3. Publication record. Web of science publication records per year reached up to 1050 in 2018 for TOPIC: (cancer circulating DNA). The sum of publication records up to now is 5800. The sum of times cited per year was in exponential phase in the last decade up to 28 000 in 2018.

conventional WGS from a double-stranded DNA library should be distinguished from that obtained with q-PCR or WGS from a single-stranded DNA library which use single-strand DNA as a first template. Using standard WGS, Markus et al. showed that, in contrast to plasma where cfDNA fragments of mononucleosome length are predominant, the length of the most abundant fragments in urine is short (80–81 bp), which is consistent with DNA wrapped in a subnucleosomal structure and undergoing DNA digestion.

Bronkhorst et al. had previously demonstrated that 143B cancer cell line is able to actively release 2000–3000 bp sized segments of heterochromatin [10]. In their new work, they found that these fragments are of a repetitive nature, including alpha satellite, mini satellite and active transposons originating from the centromeric and the pericentromeric regions of chromosomes 1, 9 and 16. They suggest that this secretion into the extracellular environment can induce a wide range of detrimental biological effects.

Following initial observations from Lehmann-Werman et al. [11], Lo et al. [12] and Snyder et al. [13], the tissue origin of cfDNA has become of great interest especially in determining a particular biological state leading to a higher diagnostic value. These studies, among others, revealed the high proportion of cfDNA released by hematopoietic cells. Moss et al. had previously proposed a novel strategy for the identification of tissue origin of cfDNA, based on the discovery of the DNA methylation profile by Illumic Epic array [11]. Using this approach, the team was recently able to detect and monitor breast cancer. Moreover, in the CNAPS meeting they confirmed that this strategy is useful for the simultaneous identification of multiple tissue contributors of cfDNA in healthy individuals, transplant recipients and cancer patients [14]. This report followed the previous demonstration of the noninvasive monitoring of infection and rejection after lung transplantation [15] using the same approach. Lam et al. presented their published data demonstrating that erythroid DNA represents a median of 30.1% of plasma DNA in healthy subjects [16]. Simon et al. had earlier found that cfDNA is predominantly released as pro-inflammatory chromatin from cells of the hematopoietic lineage in response to acute exercise [17].

Specific cfDNA structure, as observed by size profile analysis [9, 18, 19], methylation [14, 20] or nucleosome positioning [11, 13] may benefit to higher capacity in diagnostics or cancer screening.

Biology

A number of previous studies had supported the concept that cfDNA can penetrate host cells and modify the biology of those cells. In their prior studies, Mittra et al. considered that DNA which exists in the extracellular compartment, including the circulation, had local and systemic biological functions that could form the basis of some human diseases. His team had demonstrated previously that DNA molecules which originate from dying cells due to chemotherapy and radiotherapy have the ability to integrate into genomes of bystander healthy cells in order to induce DNA damage and trigger apoptotic and inflammatory responses. They have shown that this patho-physiological effect can be prevented by using DNA-neutralizing and degrading agents [21]. Their preliminary data showed that circulating DNA released from hypoxic cells is directly involved in the activation of paracrine factors that modulate cells of the tumor microenvironment. In addition, Glebova et al. suggested that oxidized DNA can act as genetic elements involved in horizontal gene transfer, as shown in experiments with MCF-7 cells able to uptake and transcript previously oxidized cfDNA. In addition, Raghuram et al. have already confirmed that this phenomenon of horizontal gene transfer with cfDNA is a function of cfDNA size [22]. Their study showed that high molecular weight DNA is not able to enter cells, while sonicated DNA from different sources with sizes ranging from 300 to 3000 bp is able to do so, regardless of species or kingdom boundaries. They also proved that sonicated DNA has a

biological effect, such as the activation of the proinflammatory transcription factor. Here, Neuberger et al. have demonstrated the immunological effects of cfDNA. A higher gene expression of *IL1B*, *TNF* and *IL8* in THP1 cell line was observed after their stimulation with nucleosomal DNA released by netosis, compared with stimulation with purified genomic DNA. Netosis is a very recently discovered pathogen-killing mechanism in which neutrophils produce and release neutrophil extracellular traps that are composed of DNA and granular proteins [23]. Furthermore, using whole-transcriptome sequencing, Semenov et al. have recently reported that the incubation of A549 cell line with exosomes purified from human blood plasma induced the upregulation of innate immunity genes and the activation of toll-like receptors (TLR) and p53 signaling pathway.

Additionally, Rykova et al. discussed data published in the literature that encourages the development of immunomodulatory DNA-based therapeutics. The consequence of nucleic acids and their metabolites accumulation in the bloodstream due to obstruction of clearance functions, to specifically activate immune receptors is significant. For example, TLR9 is activated by unmethylated CpG motif [24], and TBK1 receptor is activated by cytoplasmic double stranded DNA [25]. Rykova also showed results proving that synthetic immunomodulatory CpG ODNs (oligodeoxynucleotides) can be used as vaccine adjuvants against different types of cancer and many other diseases [26, 27]. Taking into consideration all these supporting results, further studies should be carried out to develop potential specifically targeted nontoxic immunomodulatory DNA-based therapeutics.

It has previously been shown that telomeric sequences in the pool of cfDNA are able to block TLR9 and inhibit the above-mentioned immune response [28]. Korabecna et al. studied the stimulatory effects on monocytic THP1 cells of telomeric cfDNA present in plasma and serum, before and after treatment with DNases [29]. TNF-α messenger RNA (mRNA) expression in stimulated THP1 cells increased significantly after DNase treatment. Thus, they confirmed the involvement of cell-free telomeric sequences in the regulation of immune functions.

Mittra hypothesized that cfDNA could integrate into genomes of healthy cells to trigger inflammatory effects. Finally, because the circulatory system is crucial for cancer propagation, many researchers are now working on genomestasis theory, which considers nucleic acids in the circulation as a driver of metastasis [30, 31].

Additional work should be done to elucidate more completely the biological properties of cfDNA, in order to extend our knowledge of the subject with a mind toward clinical implementation of cfDNA analysis. New data combining structural and biological features of cfDNA have promising possibilities for the diagnosis and treatment of cancer and other diseases.

Theragnostics/companion test

The utility of circulating tumoral DNA (mutant cfDNA) as a predictive biomarker for the therapeutic adaptation/response of patients suffering from various types of cancers, as foreseen by Stroun et al. [32, 33], Leon et al. [34], and first clinically evaluated by Diehl et al. [35], is now widely described in the literature [3, 36]. Recently, Thierry et al. demonstrated the clinical utility of using cfDNA analysis by rapidly detecting *RAS* and *BRAF* mutations in metastatic colorectal cancer (mCRC) patient plasma showing a median data turnaround-time of 2 and 16 days for mutant cfDNA and tumor-tissue analysis, respectively [37]. In addition, the authors found a higher presence of mutations (24% of WT scored patients for both genes, instead of 45% by tumor-tissue analysis). It should be noted that the mutation profile, as determined from cfDNA analysis using various detection thresholds, highlights the importance of the test sensitivity. Kopetz et al. and Laurent-Puig et al. showed results of a clinical study named AGEO RASANC, already presented by Bachet at ASCO 2017 and published after the CNAPS meeting. This focused on the concordance between mutant cfDNA- and tissue-analysis in mCRC using next-generation sequencing (NGS) [38]. For instance, Laurent-Puig et al. obtained 92% of sensitivity, 94% of specificity and 93% of concordance between plasma and tissue analysis.

Lyskjær et al. presented unpublished results of a proof-of-concept phase II study about the prediction of Irinotecan efficacy in a cohort of 18 mCRC patients using mutant cfDNA assessed by droplet digital PCR (ddPCR). The aim of this study was to test if early alterations in plasma mutant cfDNA can be used as a marker for therapeutic effectiveness in mCRC patients treated with first line Irinotecan. The authors showed that mutant cfDNA levels decreased during the first 40 days of treatment, in accordance with the fact that none of the 18 patients had relapsed by the time of the first scan. Patients with a high baseline level of ctDNA (>median) were associated with an earlier progression. In addition, authors found an approximatively fivefold increase in the relative risk of progression in patients with <10-fold reduction in mutant cfDNA levels at day 40 post-treatment start. In sum, mutant cfDNA is a promising marker for early prediction of Irinotecan response in mCRC patients.

Keller et al. also presented results which were published after the CNAPS meeting which focused on the correlation between the results of medical imaging and the variation of cfDNA assessed using ddPCR [39] in a cohort of melanoma patients treated by immunotherapy, and found statically significant results (Fisher's exact test: $P = 0.034$). They also conducted an analysis of 13 patients with blood samples taken at baseline (BL) and during the early course of treatment (cycle C2 or C3). The results showed significantly lower progression-free survival (PFS) in patients with increased mutant cfDNA levels between BL and C3, compared with patients with decreased mutant cfDNA (21 versus 145 days, respectively). From these results, they suggested early mutant cfDNA variation as a predictor of tumor response to immunotherapy in melanoma patients.

The presentation of Rosenfeld et al. pointed out the recent approval for use by the European Medicines Agency (EMA) and the Food and Drug Administration (FDA) of cobas *EGFR* Mutation real-time PCR Test v2, using plasma specimens as a companion test for the detection of exon 19 deletions or exon 21 substitution (*L858R*) mutations in the epidermal growth factor receptor (*EGFR*) gene, necessary for identifying patients with metastatic NSCLC eligible for treatment with Tarceva® (erlotinib). Rosenfeld et al. also presented the published results of Remon about the Tagrisso® (Osimertinib) benefit in *EGFR*-mutant NSCLC patients with *T790M*-mutation detected by mutant cfDNA [40]. In this study, the authors showed that among assessable patients Osimertinib gave a partial response rate of 62.5%

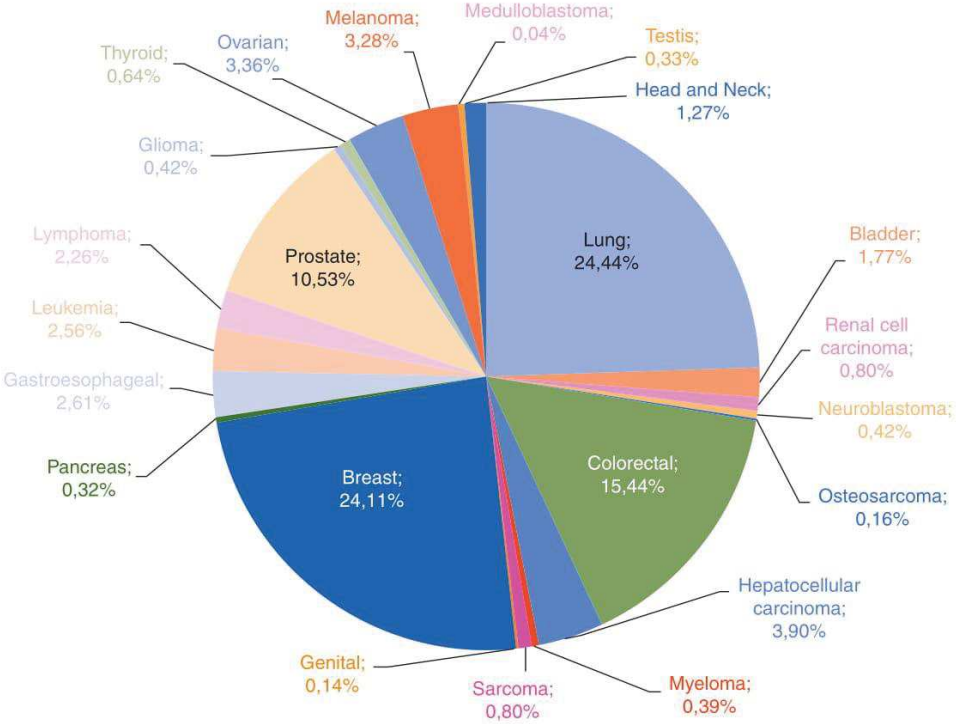


Figure 4. Proportion of the publication number reporting use of circulating DNA in cancer upon cancer types. Web of science citation reports showing 5800 records for TOPIC: (cancer circulating DNA) in up to end 2018.

and a stable disease rate of 37.5%, and they concluded that mutant cfDNA can be used as a substitute marker for *T790M* detection. In sum, cfDNA qualitative (mutation testing) and/or quantitative analysis shows considerable promise in the prediction of therapeutic tumor response in various types of cancer.

Data as regards the theragnostic/companion test field is also very encouraging. The recently approved use of cobas *EGFR* Mutation real-time PCR Test v2 using plasma samples in metastatic NSCLC should open the door for the approval of other tests for the prediction of therapeutic response. However, prospective randomized clinical trials to test the clinical utility of cfDNA as a stand-alone diagnostic test are still needed.

Minimal residual disease and recurrence

CfDNA analysis presents the opportunity to detect minimal residual disease and monitor recurrence, which might help improve patients' management and outcomes. However, the sensitivity of the analysis is often limited by the small amount of mutant DNA. This topic was widely discussed in the meeting and was the subject of numerous publications briefly before or after the conference. Tie et al. reported that cfDNA analysis may help in detecting the minimal residual disease, suggesting the capacity to select cancer patients for adjuvant therapy based on the detection of mutant cfDNA [41]. Rosenfeld suggested that massively multiplexed assays are needed to detect low levels of mutant cfDNA, hence the TAilored PAnel Sequencing (TAPAS) method that analyses 100s–1000s of mutant loci. Wan et al. demonstrated that this method's sensitivity allows the detection of residual levels of tumor DNA in the plasma and urine samples of melanoma patients receiving therapy [3]. Mutant cfDNA's ability to track tumor progression and provide evidence of minimal residual disease was investigated by Openshaw et al. [42], by tracking *TP53* mutation frequency in tumor tissues and corresponding plasma samples from patients with gastroesophageal adenocarcinoma. A different study by Phallen et al. used massively parallel sequencing to identify somatic mutations to be used as mutant cfDNA markers in CRC patients for the detection of minimal residual disease and to monitor changes in tumor burden during a 3-year follow-up period. Mutant cfDNA was detected in all relapsing patients but not in nonrelapsing patients [43]. Lin et al. evaluated the utility of a sensitive 54-cancer gene digital NGS assay, which covers all major melanoma driver genes, in melanoma patients with regional metastasis following curative surgical resection [44]. They showed that cfDNA mutation levels correlated with tumor burden ($P = 0.019$) and that monitoring these enabled earlier detection of recurrence than by monitoring serum LDH levels ($P = 0.01$). In the plasma of gastric cancer (GC) patients, Ogino et al. showed that the *HER2* ratio determined by ddPCR is a repeatable noninvasive approach for the detection of recurrence and for real-time evaluations of the *HER2* status to monitor the effects of treatments for patients with *HER2*-positive GC [45]. A copy-number aberration score was calculated by Silva et al. for melanoma patients with active disease and patients with recently excised disease [46]. cfDNA was higher in the first group and was

considered a good predictor of the presence of active disease with an AUC of 0.89.

Another group, Ehlert et al. developed a specific q-PCR approach, which was presented in the meeting, employing a first round blocking primer, *KRAS* mutation-specific ARMS primers and short LNA-probes to quantify minute amounts of mutant cfDNA. Martinson examined variant frequencies by using single-strand library preparation technology, an approach developed by Abbosh et al. [47], to study extremely fragmented DNA in baseline and postsurgical NSCLC samples. He showed in the conference that the fraction of variant-positive cases decreased in postsurgical samples. In another study, Kothari et al. evaluated the feasibility of detecting cfDNA and using it to noninvasively characterize genetic profiles of high-risk relapsed neuroblastoma.

In sum, the ability of mutant cfDNA detection to provide evidence of residual disease and identify patients at very high risk of relapse was validated in different studies, and work on the implementation of this analysis in clinical practice is ongoing.

Treatment and resistance monitoring

Drug resistance is clearly a challenge in oncology. Since previous pioneering works from L. Diaz and Bardelli's group [4, 48–50] demonstrating that resistant clones emerge in the tumor causing resistance to targeted therapy, the use of cfDNA analysis has emerged as a potential tool in detecting early resistance in the course of treatment. Works presented at the CNAPS 2017 showed advances in the identification of *de novo* or acquired methylations and point mutations leading to drug resistance [48]. In this context, a previous study by Thierry et al. showed the need for an ultrasensitive method of detecting *RAS* and *BRAF* point mutations in cfDNA before treatments, or when tracking secondary resistance in mCRC patients treated with chemotherapy and cetuximab [51]. Results revealed that 64% and 21% of point mutations detected in cfDNA would be missed if the limit of detection was below 1% or 0.1%, respectively [51]. Bardelli's team also showed in 2016 that, in mCRC patients, EGFR extracellular domain (ECD) point mutations can emerge during anti-EGFR therapy, leading to secondary resistance [49]. EGFR ECD point mutations induce a change in the EGFR structure, leading to nonrecognition by anti-*EGFR* monoclonal antibodies (cetuximab or panitumumab). To override this problem, Bardelli prepared a murine avatar from a G465E *EGFR*-mutant tumor derived from an mCRC patient. The resulting 2D cell culture was treated with a combination of anti-*EGFR* antibodies (MM-151), showing a dramatic reduction in cell viability compared with cetuximab or panitumumab alone [52]. Kopetz's team showed in the CNAPS meeting one of the first examples of cfDNA genomic landscape, in 1397 patients with CRC. Firstly, data showed that the mutational prevalence of the 20 most commonly mutated genes in cfDNA was strongly associated with the mutational prevalence in tumor tissue ($R^2 = 0.95$; $P < 0.0001$). Secondly, data revealed that 91% of patients with *EGFR* ECD mutations also harboring multiple alterations including *RAS* and *BRAF* point mutations, leading to resistance [53]. The emergence of point mutations is not the only way to monitor treatment. The modulation of mutant cfDNA concentration could also predict the efficacy of various treatments. Laurent-Puig's team suggested that in their published data [54] early decreases in mutant cfDNA concentration (*KRAS* point mutation or *WIF1* and *NPY* hypermethylation) was a marker of therapeutic efficacy in mCRC patients. Moreover, Mouliere et al. designed a patient-TAPAS approach to track mutation levels and fragmentation patterns in cfDNA. They observed changes in mutant cfDNA levels in the first 2 days following treatment initiation, including the emergence of subclonal mutations. In the same way, Reinert et al. showed earlier in CRC patients after local or systemic interventions that changes in mutant cfDNA level showed a good agreement with changes in tumor volume ($\kappa = 0.41$, Spearman's $\rho = 0.4$) [55]. Heitzer et al. developed the mFAST SeqS, an untargeted assessment of tumor DNA fraction, to monitor treatment response in metastatic breast cancer patients. They showed before the meeting that changing in levels of mutant cfDNA (*z*-score FAST-Seq) correlate with treatment response [56].

As the development of cfDNA analysis continues, more and more work is now focused on its implementation in nonmetastatic stage cancer [57, 58], and in investigating predictive response to chemotherapy [59]. It should be noted that, in addition to mutation analysis, methylation was found to be a great surrogate marker for following up metastatic cancer patients [60]. cfDNA analysis in oncology is a promising tool in the identification of early drug sensitivity or resistance, and in doing so guide the clinician for adapting therapies to tumor evolution. The next step in its development will be to use interventional clinical studies to investigate whether cfDNA analysis can increase patient survival (PFS and OS) by adapting therapies in real time.

Circulating RNA in oncology

In the session on cfRNA in oncology, the speakers presented their data on cfRNA, which is also a highly important research area in the field of liquid biopsy (Figure 2). In the bloodstream, cfRNA consists of diverse RNA subpopulations that are protected from degradation by association with lipids/proteins or by incorporation into EV such as exosomes.

Tewari et al. presented his data on RNA-seq, which are increasingly being used for profiling of cfRNA present in plasma and other biofluids. Many different methods for performing RNA-seq have been described to date. Tewari et al. improved the protocol- and sequence-specific bias of RNA-seq by using adapters for ligation with randomized end-nucleotides and computational correction factors. They detected that EV populations exhibit a common mRNA signature, carry tumor-specific alterations and can be interrogated as a source of cancer-derived cargo [61]. Garcia-Olmo showed the clinical relevance of perioperative detection of mRNA encoding *hTERT* and *GAPDH* in plasma from laryngeal and hypopharyngeal cancer patients by real-time PCR. Detection of *hTERT* mRNA before surgery had diagnostic value and was related to subsequent recurrence. Mean levels of plasma *GAPDH* mRNA in untreated patients were significantly higher than in healthy subjects [62]. Epigenetic biomarkers are increasingly recognized as promising diagnostic and prognostic tools in CRC. Using an ELISA-based assay, Gezer demonstrated that the levels of trimethylation of histone H4 lysine 20 (H4K20me3), a repressive histone methylation mark usually enriched in pericentromeric heterochromatin, decrease in circulating nucleosomes in blood of CRC patients [20].

Several studies in recent years have focused on circulating miRNAs, short single-stranded noncoding RNA that inhibits mRNA expression. In this area, Laktionov presented his screening of miRNAs in urine from prostate cancer patients. Expression of four pairs of miR-15a/miR-26b, miR-19b/miR-26b, miR-22-3p/miR-26b, miR-26b/miR-378a differentiated 80% of these patients with absolute specificity [63]. Using TaqMan real-time PCR and microarrays, Schwarzenbach's laboratory revealed that a network of deregulated cell-free and exosomal miRNAs involved in the regulation of cancer-associated signaling pathways, such as the phosphatidylinositol 3-kinase/AKT way, are associated with a particular biology of breast and ovarian carcinomas. Furthermore, they found an excessive secretion of exosomes in these patients, something associated with the prevalence of miRNAs in exosomes, when compared with cell-free miRNAs [64, 65].

Overall, the recent advances in cfRNA highlight the significant potential of cfRNAs as future tumor markers and their therapeutic value.

Cancer screening

Early cancer detection can improve survival and decrease mortality by increasing the chances of successful treatment. Papadopoulos' team managed to increase the sensitivity of cancer detection by combining mutation detection in more than one bodily fluid (saliva and blood) from the same individual, or by associating mutation detection with other markers or other clinical information. In a study called CancerSeek, published in January 2018 and including more than 1000 patients with 8 different cancer types (pancreas, colon, breast, liver, lung, ovarian, stomach and esophagus) and healthy individuals, they also showed an increased sensitivity in cancer detection while preserving specificity by combining mutation detection in the plasma of 16 genes commonly mutated in cancer and 41 protein markers [66]. While the CancerSeek test showed unrivaled performance, it nonetheless presents certain limitations in the context of universal cancer screening; for instance, it is principally used to confirm a previous diagnosis, with a low median sensitivity of only 43% for stage I cancers, whereas universal cancer screening should be carried out on an ostensibly healthy population [67].

Dennis Lo's group developed a test for nasopharyngeal cancer (NPC) by detecting the Epstein–Barr virus (EBV) involved in most NPC cases. Among the 20 174 male subjects recruited and screened, 5.5% showed a positive first test for EBV DNA, 1.5% had 2 consecutive positive plasma EBV DNA tests and 34 cases of NPC were identified [68]. Volition, a multi-national life sciences company, developed an age-adjusted algorithm of four normalized assays, each of which captures intact cell-free nucleosomes and labels a specific structural feature by double antibody ELISA. This algorithm improved precancerous polyp/adenoma detection and early stage CRC detection [69].

A study conducted by Rykova et al. [70] assessed the diagnostic potential of combinations of circulating aberrant methylated DNA and microRNAs (miRNAs) markers for lung cancer. Adleff et al. [43] proposed an approach called Targeted Error Correction Sequencing (TEC-Seq) that was able to detect mutant cfDNA in over half of early stage patients (I and II) and over three quarters of late stage patients (III and IV) with breast, lung, colorectal and ovarian cancers ($n = 200$), while no mutations were identified in the control population ($n = 44$). The authors suggest that this technique could be applicable for noninvasive early cancer detection, screening and cancer patient management. Tanos et al. developed a multiparametric cfDNA screening test using a specific q-PCR-based method and targeting nuclear and mitochondrial sequences in the plasma of healthy individuals and CRC patients. This study showed a significant difference between these two groups, and gave a strong indication of the test's capacity to discriminate healthy and cancer patients [71].

Cancer screening and early detection by using cfDNA analysis start to be evaluated. Moore et al., who used targeted sequencing for *TP53* on plasma samples of 130 women with high-grade serous ovarian cancer (HGSOC). The technique was able to detect mutant cfDNA in patients with stage I HGSOC, as well as 50% of all cancer cases, although optimization is necessary to improve sensitivity and increase detection. In the LUCID study, Heider et al. developed a technique for early stage mutant cfDNA detection based on patient-specific panels, using a parallel analysis of large numbers of mutations to assess mutant cfDNA levels at diagnosis. Several other studies were based on the detection of methylated markers, such as that of Mijnes et al., who examined the tumor-specific hypermethylation of *SPAG6*, *NKX2* and *PER1* promoters, which appears to be a promising tool for blood-based breast cancer detection. Indeed, it showed a 70% sensitivity and a 80% specificity for ductal carcinoma in situ detection. Fleischhacker et al. presented preliminary data suggesting the ability of a panel combining the methylation markers *mSHOX2* and *mSEPT9* to discriminate between patients with hepatocellular carcinoma from patients with alcohol-induced liver cirrhosis.

All of these studies encourage the opinion that cfDNA analysis presents a promising avenue for cancer screening and early detection. That said, further investigation of these emerging biomarkers is vital, since most studies underline the necessity of combining different parameters and markers in order to increase sensitivity and specificity for a better discrimination between early stage cancer patients and healthy individuals.

Preanalytics

Dana Tsui (New York, USA) and Michael Fleischhacker (Halle/Saale, Germany) organized a workshop dedicated to preanalytical procedures for the analysis of circulating nucleic acids. This session consisted of two parts. The first involved the oral presentations of Markus Havell and Ørntoft Mai-britt Worm on the evaluation of preanalytical factors affecting plasma DNA analysis, and the establishment of a standard operating procedure for cell-free DNA purification. The second part was an interactive workshop involving the contribution of all participants.

The lack of standardization of the preanalytical conditions hinders the use of cfDNA as a clinically robust biomarker in oncology [72, 73]. The main long-term objective of this workshop was to reach an agreement on the optimal handling of blood samples for cfDNA analysis, by presenting the importance of preanalytical factors, by exchanging on current practices in different laboratories, and efforts to standardize these procedures.

The discussion focused on time collection, collection tubes, sample processing delay, DNA extraction methods, methods of

Figure 5. Comparison of the number of publications reporting use of NGS- and PCR-based methods upon research areas in 2017 and 2018. (A) Web of science citation reports showing 146 records for TOPIC: (cancer circulating DNA NGS) in 2017 and 2018. This citation level corresponds to 67% to the total number up to 2018 (219 records). (B) Web of science citation reports showing 278 records for TOPIC: (cancer circulating DNA PCR-based) in 2017 and 2018. This citation level corresponds to 28% to the total number up to 2018 (988). PCR-based methods are still twice as much used when compared with NGS-based methods. However, increase of using NGS-based methods in the last 2 years is about twofold higher than the increase of using of PCR-based methods. A 62% of publication reporting NGS-based methods concern use in clinical oncology, while 30% of publication reporting PCR-based methods concerned use in clinical oncology. The numbers of publications reporting use of NGS- and PCR-based methods in clinical oncology were similar in the 2017–2018 period.

DNA quantification, and the storage of plasma and DNA samples. Preanalytical procedures still appear to vary from one laboratory to another. In addition, NGS- or PCR-based methods are among others the most used methods for analyzing cfDNA (Figure 5A and B). Nevertheless, with regard to sample processing, there is a general consensus in the scientific community on double centrifugation and a maximum delay in sample processing of 4 h in order to avoid genomic DNA contamination. The use of different extraction kits and quantification techniques in different laboratories were seen to influence the quality and quantity of extracted cfDNA.

Other comments on the origin of circulating DNA have emphasized the need not to quantify the mitochondrial DNA as nuclear DNA, but to calculate the copy number in DNA samples.

Method reproducibility and cost of the analysis are additional important factors to be taken into account in considering the use of circulating nucleic acids in clinical practice.

Despite routine analysis in clinical areas such as prenatal testing or oncology, and active investigation toward applying cfDNA analysis, the significance of preanalytics and demographic variables is not widely known. Consequently, the lack of standard operating procedures impedes the maturing of this new and promising diagnostic tool. Recent efforts from Meddeb et al. [74] to characterize and optimize cfDNA quantification in humans will help to address this.

Nomenclature

Historically, since the discovery of circulating DNA, no homogeneous nomenclature has been given to total circulating DNA in the bloodstream, unlike circulating DNA from tumor cells named ctDNA. For this purpose, during the CNAPS symposium a survey was taken to determine a common consensus on the matter. The three terms most widely used in the literature (cfDNA for cell-free DNA, ccfDNA for circulating cell-free DNA and cfDNA for circulating DNA) were proposed to the scientific community, who were also allowed to suggest other denominations. Sixty-three people participated in this vote. A 76% of the participants chose cfDNA, 16% ccfDNA, 5% cirDNA and 3% suggested another denomination. The results of this survey confirm 'cfDNA' as a term that can be used to commonly designate all extracellular circulating DNA in the blood.

In the oncology field, most reports use the term ctDNA (circulating tumor DNA), as illustrated by this small survey. One can discriminate among cfDNA origins three constituents in the blood of a cancer patient: malignant tumor cell-derived DNA, nonmalignant tumor cell-derived DNA and normal cells germline DNA [67]. However, ctDNA is employed in the literature to mean cfDNA bearing genetic or epigenetic alterations. Alternatively, ctDNA would word-by-word mean circulating DNA deriving from the tumor which is constituted of malignant and nonmalignant cells (stromal, endothelial and hematopoietic cells). Since the concentration of wild type cfDNA deriving from the microenvironment cells increases with the progression of the tumor, part of cfDNA is specific to the tumor growth. Readers could also be confused by the fact that ctDNA concentration values are biased in the literature, since in most cases ctDNA is determined from mutations of a few or a panel of genes, which inevitably leads to a bias in considering the total amount of cfDNA deriving from all the malignant cells. Because of these considerations, instead of the term ctDNA we used the term mutant cfDNA in this review.

This survey also addressed a second issue. The term 'biopsy' was first introduced by Besnier in 1879 [75] to designate the technique used to collect an organ or tissue fragment (cell samples) in order to establish an accurate diagnosis. According to the definition established by the National Cancer Institute (NCI Dictionary of Cancer Terms; available at https://www.kucancercenter.org/cancer-information/cancer-terms-dictionary), a liquid biopsy is a test done on a blood sample to look for cancer cells or pieces of tumor DNA found in the bloodstream. Given that cfDNA has nothing in common with DNA released from circulating tumor cells, and that cfDNA can also derive from nontumor cells, the question might be raised as to whether or not circulating DNA and RNA analysis ought to be considered as a 'liquid biopsy' properly speaking? A 80% of the voters supported this opinion, 10% disapproved and 10% had no opinion regarding the issue. F. Diehl (Hamburg, Germany) who organized the workshop session acknowledged the vote showing the need from the researchers from the field to unify the nomenclature despite inconstancies.

In conclusion, cfDNA with PD-1 and cfDNA with PD-L1 checkpoint pathway immunotherapies are the two fastest growing and most exciting areas in oncology in recent years. Following their first clinical evaluation in predictive medicine, all six potentials of cfDNA analysis in oncology (predictive information, detection of the minimal residual disease, early detection of resistance, treatment monitoring, recurrence surveillance and cancer early detection/screening) are now well covered by a comprehensive literature on each phase of cancer patient management care. A new and exciting therapeutic challenge now faces the scientific/clinical community: the use of cfDNA analysis to guide immunotherapy. Meanwhile, broadening our knowledge of the structural and biological features of cfDNA, as well as standardizing the preanalytical conditions, remain key steps to be taken toward the clinical implementation of cfDNA analysis, and may provide new avenues of research in tumor biology.

Acknowledgements

A special thanks and tribute to Maurice Stroun who past away just before the CNAPS Xth meeting. Maurice Stroun was remembered as one of the 'founding parents' of the circulating nucleic acids community. In particular, he founded the CNAPS conference series together with Anker [76]. Maurice was able to see his lifelong pursuit of circulating nucleic acids become an important part of scientific research and medical diagnostics. The field will continue to grow, and while those working in it will miss Maurice's physical presence, his indomitable pioneering spirit and his remarkable contributions will remain with the CNAPS community even as it reaches new frontiers.

The authors would like to thank all of the CNAPS scientific committee: P. Anker (Geneva, Switzerland), F. Diehl (Baltimore, USA), L. Diaz (New York, USA), M. Fleischhacker (Halle/Saale, Germany), P. Gahan (Montpellier, France), D. Lo (Hong Kong, China), N. Rosenfeld (Cambridge, UK), H. Schwarzenbach (Hamburg, Germany), M. Stroun (Geneva, Switzerland), and A. R. Thierry (Montpellier, France). We thank to Safia El Messaoudi who co-organized with A. R. Thierry the Xth CNAPS meeting with the critical contribution of the French Society of Nucleic Acids (SFAC) and Sabine Dejasse, SFAC President. Authors thank Marc Ychou, Karine Saget and Aurore Marquis from the SIRIC Montpellier. The authors would also like to thank Stefan Holdenrieder, Barbara Ottolini, Jean-Yves Cance, Serge Thierry and Philippe Grandjean.

Funding

This work was supported by the SFAC (Société Française des Acides Nucléiques Circulants/French Society of Circulating Nucleic Acids), the University of Montpellier (France), the GSO canceropole (France), the Languedoc-Roussillon region (France), Lilly (France) and the SIRIC Montpellier Grant

(INCa-DGOS Inserm 6045), France. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 755333 (LIMA). ART is supported by the INSERM (Institut National de la Santé et de la Recherche Médicale, France).

Disclosure

ART is cofounder, shareholder of DiaDx SAS and co-inventor of patent applications on cfDNA analytical methods. PA, CS and HS are co-inventors of patent applications on cfDNA analytical methods. All remaining authors have declared no conflicts of interest.

References

1. Thierry AR, El Messaoudi S, Gahan PB et al. Origins, structures, and functions of circulating DNA in oncology. Cancer Metastasis Rev 2016; 35(3): 347–376.
2. Pös O, Biró O, Szemes T, Nagy B. Circulating cell-free nucleic acids: characteristics and applications. Eur J Hum Genet 2018; 26(7): 937–945.
3. Wan JCM, Massie C, Garcia-Corbacho J et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer 2017; 17(4): 223–238.
4. Diaz LA, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol 2014; 32(6): 579–586.
5. Mouliere F, Robert B, Peyrotte EA et al. High fragmentation characterizes tumour-derived circulating DNA. PLoS One 2011; 6(9): e23418.
6. Mouliere F, Messaoudi SE, Gongora C et al. Circulating cell-free DNA from colorectal cancer patients may reveal high *KRAS* or *BRAF* mutation load. Transl Oncol 2013; 6(3): 319.
7. Andriamanampisoa C-L, Bancaud A, Boutonnet-Rodat A et al. BIABooster: online DNA concentration and size profiling with a limit of detection of 10 fg/μL and application to high-sensitivity characterization of circulating cell-free DNA. Anal Chem 2018; 90(6): 3766–3774.
8. Mouliere F, Piskorz AM, Chandrananda D et al. Selecting short DNA fragments in plasma improves detection of circulating tumour DNA. Sci Transl Med 2018; 10(466). doi: https://doi.org/10.1101/134437.
9. Sanchez C, Snyder MW, Tanos R et al. New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA analysis. NPJ Genomic Med 2018; 3(1): 31.
10. Bronkhorst AJ, Wentzel JF, Aucamp J et al. Characterization of the cell-free DNA released by cultured cancer cells. Biochim Biophys Acta 2016; 1863(1): 157–165.
11. Lehmann-Werman R, Neiman D, Zemmour H et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci USA 2016; 113(13): E1826–E1834.
12. Sun K, Jiang P, Chan KCA et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc Natl Acad Sci USA 2015; 112(40): E5503–E5512.
13. Snyder MW, Kircher M, Hill AJ et al. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell 2016; 164(1–2): 57.
14. Moss J, Magenheim J, Neiman D et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 2018; 448142.
15. De Vlaminck I, Martin L, Kertesz M et al. Noninvasive monitoring of infection and rejection after lung transplantation. Proc Natl Acad Sci USA 2015; 112(43): 13336–13341.
16. Lam WKJ, Gai W, Sun K et al. DNA of erythroid origin is present in human plasma and informs the types of anemia. Clin Chem 2017; 63(10): 1614–1623.
17. Tug S, Helmig S, Deichmann ER et al. Exercise-induced increases in cell free DNA in human plasma originate predominantly from cells of the haematopoietic lineage. Exerc Immunol Rev 2015; 21: 164–173.
18. Chandrananda D, Thorne NP, Bahlo M. High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. BMC Med Genomics 2015; 8: 29.
19. Underhill HR, Kitzman JO, Hellwig S et al. Fragment length of circulating tumor DNA. PLoS Genet 2016; 12(7): e1006162.
20. Gezer U, Yörüker EE, Keskin M et al. Histone methylation marks on circulating nucleosomes as novel blood-based biomarker in colorectal cancer. IJMS 2015; 16(12): 29654–29662.
21. Mittra I, Pal K, Pancholi N et al. Prevention of chemotherapy toxicity by agents that neutralize or degrade cell-free chromatin. Ann Oncol 2017; 28(9): 2119–2127.
22. Raghuram GV, Gupta D, Subramaniam S et al. Physical shearing imparts biological activity to DNA and ability to transmit itself horizontally across species and kingdom boundaries. BMC Mol Biol 2017; 18(1): 21.
23. Bonaventura A, Liberale L, Carbone F et al. The pathophysiological role of neutrophil extracellular traps in inflammatory diseases. Thromb Haemost 2018; 118(1): 6–27.
24. Roers A, Hiller B, Hornung V. Recognition of endogenous nucleic acids by the innate immune system. Immunity 2016; 44(4): 739–754.
25. Corrales L, McWhirter SM, Dubensky TW, Gajewski TF. The host STING pathway at the interface of cancer and immunity. J Clin Invest 2016; 126(7): 2404–2411.
26. Scheiermann J, Klinman DM. Clinical evaluation of CpG oligonucleotides as adjuvants for vaccines targeting infectious diseases and cancer. Vaccine 2014; 32(48): 6377–6389.
27. Wittig B, Schmidt M, Scheithauer W, Schmoll H-J. MGN1703, an immunomodulator and toll-like receptor 9 (TLR-9) agonist: from bench to bedside. Crit Rev Oncol Hematol 2015; 94(1): 31–44.
28. Gursel I, Gursel M, Yamada H et al. Repetitive elements in mammalian telomeres suppress bacterial DNA-induced immune activation. J. Immunol 2003; 171(3): 1393–1400.
29. Zinkova A, Brynychova I, Svacina A et al. Cell-free DNA from human plasma and serum differs in content of telomeric sequences and its ability to promote immune response. Sci Rep 2017; 7(1): 2591.
30. Chaudhary S, Raghuram GV, Mittra I. Is inflammation a direct response to dsDNA breaks? Mutat Res 2018; 808: 48–52.
31. Olmedillas-López S, García-Olmo DC, García-Arranz M et al. Liquid biopsy by NGS: differential presence of exons (DPE) in cell-free DNA reveals different patterns in metastatic and nonmetastatic colorectal cancer. Cancer Med 2018; 7(5): 1706–1716.
32. Stroun M, Anker P, Maurice P et al. Neoplastic characteristics of the DNA found in the plasma of cancer patients. Oncology 1989; 46(5): 318–322.
33. Stroun M, Anker P, Maurice P, Gahan PB. Circulating nucleic acids in higher organisms. Int Rev Cytol 1977; 51: 1–48.
34. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res 1977; 37(3): 646–650.
35. Diehl F, Li M, Dressman D et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci USA 2005; 102(45): 16368–16373.
36. Moati E, Taly V, Didelot A et al. Role of circulating tumor DNA in the management of patients with colorectal cancer. Clin Res Hepatol Gastroenterol 2018; 42(5): 396–402.
37. Thierry AR, El Messaoudi S, Mollevi C et al. Clinical utility of circulating DNA analysis for rapid detection of actionable mutations to select metastatic colorectal patients for anti-EGFR treatment. Ann Oncol 2017; 28(9): 2149–2159.
38. Bachet JB, Bouché O, Taieb J et al. *RAS* mutation analysis in circulating tumor DNA from patients with metastatic colorectal cancer: the AGEO RASANC prospective multicenter study. Ann Oncol 2018; 29(5): 1211–1219.
39. Keller L, Guibert N, Casanova A et al. Early circulating tumour DNA variations predict tumour response in melanoma patients treated with immunotherapy. Acta Derm Venerol 2019; 99(2): 206–210.
40. Remon J, Caramella C, Jovelet C et al. Osimertinib benefit in EGFR-mutant NSCLC patients with T790M-mutation detected by circulating tumour DNA. Ann Oncol 2017; 28(4): 784–790.

41. Tie J, Wang Y, Tomasetti C et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Sci Transl Med 2016; 8(346): 346ra92.
42. Openshaw MR, Richards CJ, Guttery DS et al. The genetics of gastroesophageal adenocarcinoma and the use of circulating cell free DNA for disease detection and monitoring. Expert Rev Mol Diagn 2017; 17(5): 459–470.
43. Phallen J, Sausen M, Adleff V et al. Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med 2017; 9(403). doi: 10.1126/scitranslmed.aan2415
44. Lin SY, Huang SK, Huynh KT et al. Multiplex gene profiling of cell-free DNA in patients with metastatic melanoma for monitoring disease. JCO Precision Oncol 2018; (2): 1–30.
45. Ogino S, Konishi H, Ichikawa D et al. Detection of fusion gene in cell-free DNA of a gastric synovial sarcoma. World J Gastroenterol 2018; 24(8): 949–956.
46. Silva S, Danson S, Teare D et al. Genome-wide analysis of circulating cell-free DNA copy number detects active melanoma and predicts survival. Clin Chem 2018; 64(9): 1338–1346.
47. Abbosh C, Birkbak NJ, Wilson GA et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature 2017; 545(7655): 446–451.
48. Misale S, Yaeger R, Hobor S et al. Emergence of *KRAS* mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature 2012; 486(7404): 532–536.
49. Van Emburgh BO, Arena S, Siravegna G et al. Acquired *RAS* or *EGFR* mutations and duration of response to EGFR blockade in colorectal cancer. Nat Commun 2016; 7: 13665.
50. Diaz LA, Williams R, Wu J et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature 2012; 486(7404): 537–540.
51. Thierry AR, Pastor B, Jiang Z-Q et al. Circulating DNA demonstrates convergent evolution and common resistance mechanisms during treatment of colorectal cancer. Clin Cancer Res 2017; 23(16): 4578–4591.
52. Arena S, Siravegna G, Mussolin B et al. MM-151 overcomes acquired resistance to cetuximab and panitumumab in colorectal cancers harboring EGFR extracellular domain mutations. Sci Transl Med 2016; 8(324): 324ra14.
53. Strickler JH, Loree JM, Ahronian LG et al. Genomic landscape of cell-free DNA in patients with colorectal cancer. Cancer Discov 2018; 8(2): 164–173.
54. Garlan F, Laurent-Puig P, Sefrioui D et al. Early evaluation of circulating tumor DNA as marker of therapeutic efficacy in metastatic colorectal cancer patients (PLACOL study). Clin Cancer Res 2017; 23(18): 5416–5425.
55. Schøler LV, Reinert T, Ørntoft M-BW et al. Clinical implications of monitoring circulating tumor DNA in patients with colorectal cancer. Clin Cancer Res 2017; 23(18): 5437–5445.
56. Belic J, Koch M, Ulz P et al. mFast-SeqS as a monitoring and prescreening tool for tumor-specific aneuploidy in plasma DNA. Adv Exp Med Biol 2016; 924: 147–155.
57. Tie J, Cohen JD, Wang Y et al. Serial circulating tumour DNA analysis during multimodality treatment of locally advanced rectal cancer: a prospective biomarker study. Gut 2018; doi:10.1136/gutjnl-2017-315852.
58. Schou JV, Larsen FO, Sørensen BS et al. Circulating cell-free DNA as predictor of treatment failure after neoadjuvant chemo-radiotherapy before surgery in patients with locally advanced rectal cancer. Ann Oncol 2018; 29(3): 610–615.
59. Yao J, Zang W, Ge Y et al. *RAS/BRAF* circulating tumor DNA mutations as a predictor of response to first-line chemotherapy in metastatic colorectal cancer patients. Can J Gastroenterol Hepatol 2018; 2018: 4248971.
60. Boeckx N, Op de Beeck K, Beyens M et al. Mutation and methylation analysis of circulating tumor DNA can be used for follow-up of metastatic colorectal cancer patients. Clin Colorectal Cancer 2018; 17(2): e369–e379.
61. Conley A, Minciacchi VR, Lee DH et al. High-throughput sequencing of two populations of extracellular vesicles provides an mRNA signature that can be detected in the circulation of breast cancer patients. RNA Biol 2017; 14(3): 305–316.
62. García–Olmo DC, Contreras JD, Picazo MG et al. Potential clinical significance of perioperative levels of mRNA in plasma from patients with cancer of the larynx or hypopharynx. Head Neck 2017; 39(4): 647–655.
63. Bryzgunova OE, Konoshenko MY, Laktionov PP. MicroRNA-guided gene expression in prostate cancer: literature and database overview. J Gene Med 2018; 20(5): e3016.
64. Meng X, Müller V, Milde-Langosch K et al. Diagnostic and prognostic relevance of circulating exosomal miR-373, miR-200a, miR-200b and miR-200c in patients with epithelial ovarian cancer. Oncotarget 2016; 7(13): 16923–16935.
65. Eichelser C, Stückrath I, Müller V et al. Increased serum levels of circulating exosomal microRNA-373 in receptor-negative breast cancer patients. Oncotarget 2014; 5(20): 9650–9663.
66. Cohen JD, Li L, Wang Y et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science 2018; 359(6378): 926–930.
67. Thierry AR. A step closer to cancer screening by blood test. Clin Chem 2018; 64(10):1420–1422.
68. Lam WKJ, Jiang P, Chan KCA et al. Sequencing-based counting and size profiling of plasma Epstein–Barr virus DNA enhance population screening of nasopharyngeal carcinoma. Proc Natl Acad Sci USA 2018; 115(22): E5115–E5124.
69. Rahier J-F, Druez A, Faugeras L et al. Circulating nucleosomes as new blood-based biomarkers for detection of colorectal cancer. Clin Epigenetics 2017; 9: 53.
70. Zaporozhchenko IA, Ponomaryova AA, Rykova EY, Laktionov PP. The potential of circulating cell-free RNA as a cancer biomarker: challenges and opportunities. Expert Rev Mol Diagn 2018; 18(2): 133–145.
71. Tanos R, Thierry AR. Clinical relevance of liquid biopsy for cancer screening. Transl Cancer Res 2018; 7(S2): S105–S129.
72. Schmidt B, Reinicke D, Reindl I et al. Liquid biopsy - performance of the PAXgene® blood ccfDNA tubes for the isolation and characterization of cell-free plasma DNA from tumor patients. Clin Chim Acta 2017; 469: 94–98.
73. Schmidt B, Fleischhacker M. Is liquid biopsy ready for the litmus test and what has been achieved so far to deal with pre-analytical issues? Transl Cancer Res 2018; 7(S2): S130–S139.
74. Meddeb R, Al Amir Dache Z, Thezenas S et al. Quantifying circulating cell-free DNA in humans. Sci Rep 2018.
75. Nezelof C, Guinebretière J-M. [1879, Ernest Besnier inventor of the word "biopsy"]. Rev Prat 2006; 56(18): 2081–2085.
76. Anker P, Stroun M (eds). Circulating Nucleic Acids in Plasma or Serum, Vol. 906. New York, NY: New York Academy of Sciences 2000; 1–188.

C. Quantification de l'ADN circulant chez l'homme

Titre : Quantifying cell-free DNA in humans

Auteurs

Romain Meddeb, Zahra Al Amir Dache, Simon Thezenas, Amaëlle Otandault, **Rita Tanos**, Brice Pastor, Cynthia Sanchez, Joelle Azzi, Geoffroy Tousch, Simon Azan, Caroline Mollevi, Antoine Adenis, Safia El Messaoudi, Philippe Blache, et Alain R. Thierry

Ce travail a été publié dans *Scientific Reports* en Mars 2019.

Résumé

Ce travail est la première étude exhaustive portant sur l'influence de plusieurs paramètres pré-analytiques et démographiques pouvant être une source de variabilité dans la quantification de l'ADN circulant nucléaire et mitochondrial (ADNcirN et ADNcirM). Les données ont été obtenues sur un total de 222 sujets, dont 104 individus sains et 118 patients atteints d'un cancer colorectal métastatique (mCRC). Environ 50,000 et 3,000 fois plus de copies du génome mitochondrial que de copies du génome nucléaire ont été détectées dans le plasma de sujets en bonne santé et de patients atteints de mCRC respectivement. Chez les individus sains, la concentration d'ADNcirN est significativement influencée par l'âge ($p = 0,009$) et le sexe ($p = 0,048$). L'analyse multivariée avec régression logistique a montré qu'un âge supérieur à 47 ans est un facteur prédictif d'une concentration plus élevée d'ADNcirN (OR = 2,41 ; $p = 0,033$). La concentration de l'ADNcirM est indépendante de l'âge et du sexe chez les sujets en bonne santé. Chez les patients atteints de mCRC, les taux d'ADNcirN et d'ADNcirM étaient indépendants de l'âge, du sexe, du délai entre la prise de nourriture par rapport au temps du prélèvement sanguin et de l'aspect plasmatique, que ce soit en analyse univariée ou multivariée. Néanmoins, une étude *ad-hoc* suggère que la ménopause et le temps écoulé entre le prélèvement sanguin et son analyse auraient tendance à influencer la quantification de l'ADN circulaire. De plus, des différences significativement élevées ont été observées entre les patients atteints de mCRC et les individus sains pour l'ADNcirN ($p < 0,0001$), l'ADNcirM ($p < 0,0001$) et le rapport ADNcirM / ADNcirN ($p < 0,0001$). Finalement, les niveaux d'ADNcirN et

d'ADNcirM ne varient pas de la même manière dans les cas de cancers *vs* état physiologique normal, et également pour ce qui concerne les facteurs pré-analytiques et démographiques.

Contributions

J'ai participé à l'analyse des échantillons de plasma et à la quantification de l'ADNcir nucléaire et mitochondriale en particulier pour la cohorte des individus sains. J'ai également participé à la discussion des résultats.

www.nature.com/scientificreports

OPEN

Quantifying circulating cell-free DNA in humans

Romain Meddeb[1,2,3,4], Zahra Al Amir Dache[1,2,3,4], Simon Thezenas[1,2,3,4,5], Amaëlle Otandault[1,2,3,4], Rita Tanos[1,2,3,4], Brice Pastor[1,2,3,4], Cynthia Sanchez[1,2,3,4], Joelle Azzi[1,2,3,4], Geoffroy Tousch[1,2,3,4], Simon Azan[1,2,3,4], Caroline Mollevi[1,2,3,4,5], Antoine Adenis[1,2,3,4,6], Safia El Messaoudi[1,2,3,4], Philippe Blache[1,2,3,4] & Alain R. Thierry[1,2,3,4]

Received: 6 September 2018
Accepted: 9 January 2019
Published online: 26 March 2019

To our knowledge, this is the first comprehensive study on the influence of several pre-analytical and demographic parameters that could be a source of variability in the quantification of nuclear and mitochondrial circulating DNA (NcirDNA and McirDNA). We report data from a total of 222 subjects, 104 healthy individuals and 118 metastatic colorectal cancer (mCRC) patients. Approximately 50,000 and 3,000-fold more mitochondrial than nuclear genome copies were found in the plasma of healthy individuals and mCRC patients, respectively. In healthy individuals, NcirDNA concentration was statistically influenced by age ($p = 0.009$) and gender ($p = 0.048$). Multivariate analysis with logistic regression specified that age over 47 years-old was predictive to have higher NcirDNA concentration (OR = 2.41; $p = 0.033$). McirDNA concentration was independent of age and gender in healthy individuals. In mCRC patients, NcirDNA and McirDNA levels were independent of age, gender, delay between food intake and blood collection, and plasma aspect, either with univariate or multivariate analysis. Nonetheless, ad hoc study suggested that menopause and blood collection time might have tendency to influence cirDNA quantification. In addition, high significant statistical differences were found between mCRC patients and healthy individuals for NcirDNA ($p < 0.0001$), McirDNA ($p < 0.0001$) and McirDNA/NcirDNA ratio ($p < 0.0001$). NcirDNA and McirDNA levels do not vary in the same way with regards to cancer vs healthy status, pre-analytical and demographic factors.

Since Mandel and Metais discovered the presence of nucleic acids in serum in the 1940s[1], different studies have reported elevated levels of circulating DNA (cirDNA) in the blood of patients suffering from various diseases[2–6], especially cancer[7–10]. Despite scant early consideration, interest in the feasibility of cirDNA analysis has increased exponentially, over the last decade, among researchers working on a large range of disorders. CirDNA was first clinically implemented in prenatal diagnosis of sex-determination and pregnancy-associated disorders by assaying fetal DNA in maternal plasma[11–13]. The main sources of cirDNA are cell death, either by necrosis or apoptosis, and active release by viable cells, including exocytosis and NETosis[14,15]. Note, cirDNA may derive from either nuclear (NcirDNA) or mitochondrial DNA (McirDNA). To date, research and development of cirDNA analysis has focused on the qualitative rather than the quantitative information provided. For example, cirDNA analysis is now clinically validated for detecting specific sequences or mutations to guide the oncologist toward the most appropriate treatment. CirDNA analysis is also performed for prenatal and embryo-culture genetic testing. There have been several years of intensive studies validating cirDNA quantitation in different clinical scenarios, including sepsis, transplant recipients and immune disorders. CirDNA quantification is also now taken into consideration in oncology as, it was recently shown that, the level of mutant cirDNA is useful for following-up cancer patients to detect minimal residual disease and to monitor response to therapy and disease recurrence. Although total cirDNA levels were first examined in the early phase of cirDNA research and development, it is now not considered as a single biomarker because of its lack of specificity. Nevertheless, all cirDNA analysis relies on the optimal quantification of total amount of cirDNA. Biological biomarkers should be highly dynamic, and their diagnostic performance may vary depending on internal and external changes. The clinical efficacy of cirDNA will require the identification and the control of various patient-related confounders that may

[1]IRCM, Institute of Research in Oncology of Montpellier, Montpellier, France. [2]INSERM, U1194, Montpellier, France. [3]University of Montpellier, Montpellier, France. [4]Regional Institute of Cancer of Montpellier, Montpellier, France. [5]Biometry Unit, Regional Institute of Cancer of Montpellier, Montpellier, France. [6]Digestive Oncology Department, Regional Institute of Cancer of Montpellier, Montpellier, France. Correspondence and requests for materials should be addressed to A.R.T. (email: alain.thierry@inserm.fr)

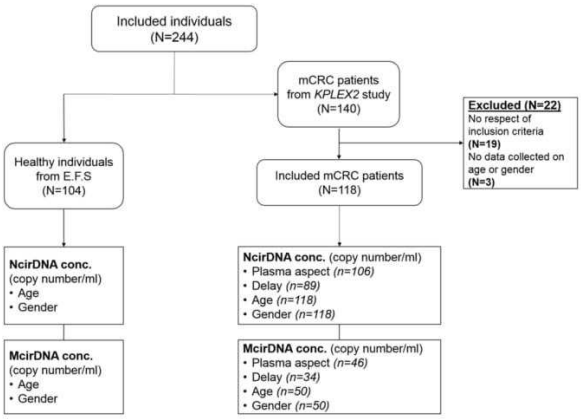


Figure 1. Flow chart of the study.

affect its measurement[16]. Human cellular aging is usually marked by senescence and cell death. Similarly, it often features a phenomenon, originally called "Inflamm-aging", that induces a chronic or low-grade inflammatory state[17–20]. Indeed, it was demonstrated that tissue damage and a pro-inflammatory environment increase release of cirDNA into the blood stream[14]. Additionally, we may speculate that sex-based differences, such as genetic dissimilarity and steroid hormones levels, could cause differences in cirDNA concentration between men and women. Likewise, blood component concentrations may vary with the circadian clock and upon food intake and potentially influencing cirDNA concentration[21]. Despite outstanding research in the field of cirDNA, relatively few studies have examined the effect of pre-analytical and demographic parameters as sources of intra- and inter-individual variability in cirDNA levels. There is currently no single operating procedure and there are relatively few clinical guidelines in the literature[22–25]. This study aims first at defining a framework for cirDNA analysis, to harmonize NcirDNA and McirDNA quantification and to explore potential sources of variability that could cause interpretation errors. Previous studies already documented various pre-analytical limitations and specified conditions for cirDNA analysis, including specific collection tubes, and specific plasma isolation and extraction protocols, including storage condition variables and limits on the number of freeze-thaw cycles[22–24,26]. Here, we study the influence of various pre-analytical (plasma aspect, delay between blood collection and last food intake) and demographic variables such age and gender in NcirDNA and McirDNA concentration determination, from plasma of 104 healthy donors and 118 mCRC patients (Fig. 1). In addition, we compared McirDNA and NcirDNA levels in healthy individuals and mCRC patients. Some others pre-analytical conditions such as blood collection tubes, blood collection time and blood stability were examined *ad hoc* study on healthy volunteers.

Results

Simultaneous quantification of NcirDNA and McirDNA. We used serial dilutions of genomic DNA and of mitochondrial plasmid DNA to validate a reproducible, linear and sensitive assay to quantify, both nuclear and mitochondrial, circulating DNA (NcirDNA and McirDNA). The NcirDNA assay is based upon the qPCR detection of a sequence of the *KRAS* gene and it can detect one copy of nuclear *KRAS* gene per 6 microliters of plasma (1 nuclear Genome Equivalent per 12 µL). The McirDNA assay can detect down to one copy of the mitochondrial *MT-CO3* gene per 1.7 microliter of plasma (1 mitochondrial Genome Equivalent per 1.7 µL). Note, the targeted MT-CO3 gene sequence was selected as being not mutated in the mitochondrial genome of cancer patients. Copy number calculation is performed by using a specific equation that eliminates bias in the nuclear or mitochondrial DNA calibration curves to allow simultaneous calculation of their real relative proportions. For example, in the cohort of 104 healthy individuals, we found a median NcirDNA and McirDNA plasma concentration of 1.64×10^3 and 8.32×10^7 copies/mL, respectively, corresponding to 5.43 and 1.36 ng/mL of plasma.

Comparison between NcirDNA and McirDNA levels. We compared NcirDNA and McirDNA concentration, in healthy individuals and mCRC patients groups, expressed either in copy number/ml and ng/ml of plasma. A supplementary table summarizes all quantification data and comparative tests in detail (Supplementary Table S1). There was a highly statistical difference between NcirDNA and McirDNA concentration expressed in copy number/ml, in healthy individuals (Fig. 2A; *Mann-Whitney U test, P value* < 0.0001) and equally in mCRC patients (Fig. 2B; *Mann-Whitney U test, P value* < 0.0001). We next compared McirDNA and NcirDNA concentration expressed in ng/ml, in healthy individuals and mCRC patients groups. Here, also, we observed a considerable statistical difference between NcirDNA and McirDNA concentrations in healthy individuals (Fig. 2C; *Mann-Whitney U test, P value* < 0.0001) and mCRC patients (Fig. 2D; *Mann-Whitney U test, P value* < 0.0001). McirDNA concentration is significantly higher than NcirDNA concentration when measured in units of copy number/ml, and conversely, the NcirDNA concentration was significantly higher than the McirDNA concentration when measured in ng/ml, either in healthy individuals and mCRC patients.

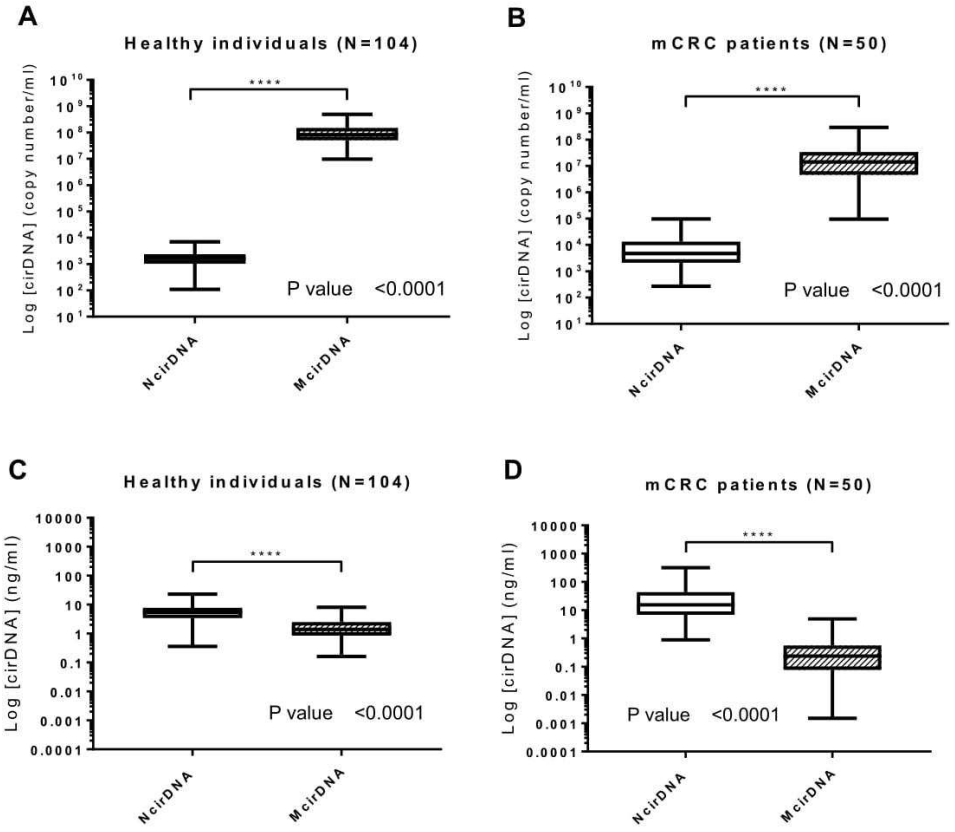


Figure 2. Respective values of NcirDNA and McirDNA plasma concentration. Boxplot analysis of cirDNA concentrations from healthy individuals (N = 104) (**A**,**C**) and mCRC patients (N = 50) (**B**,**D**). Values are expressed either as copy number/ml (**A**,**B**) or as ng/ml (**C**,**D**). CirDNA concentration was determined as described in Materials and Methods. Boxplot represent median with min to max of values and Mann-Whitney U test was performed for comparison. A probability of ≤ 0.05 was considered to be statistically significant; *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001, ****p ≤ 0.0001.

Effect of age and gender on cirDNA concentrations in healthy individuals. *NcirDNA.* **Age (n = 104):** We dichotomized the healthy individuals cohort in two groups around the median age, which was 47 (Table 1). The median NcirDNA copy number in the <47 years-old group (n = 52) and in the ≥47 years-old group (n = 52) were 1.36×10^3 and 1.73×10^3 copies/ml, respectively. A statistical difference was found between young and older healthy individuals groups (Fig. 3A; *Mann-Whitney U test, P value = 0.009).*

Gender (n = 104): The median NcirDNA copy number in the healthy male group (n = 62) and in the group of healthy females (n = 42) were 1.69×10^3 and 1.48×10^3 copies/ml, respectively. This difference in the NcirDNA copy number between healthy males and females was statistically significant (Fig. 3B; *Mann-Whitney U test, P value = 0.048*).

Multivariate analysis: Logistic regression analysis including age and gender specified that age over 47 years-old was predictive to have higher NcirDNA concentration (Fig. 3C; OR = 2.41, *P = 0.033*). Results of multivariate analysis and Odds ratios (OR) with 95% confidence intervals (CIs) are summarized in Supplementary Table S2A.

McirDNA. **Age (n = 104):** The median McirDNA copy number in the <47 (n = 52) and in the ≥47 year-old group (n = 52) were 7.77×10^7 and 8.40×10^7 copies/ml, respectively. There was no statistical difference in the McirDNA copy number between these groups (Fig. 3D; *Mann-Whitney U test, P value = 0.489*).

Gender (n = 104): The median McirDNA copy number in the healthy male group (n = 62) and in the group of healthy women (n = 42) were 8.03×10^7 and 9.39×10^7 copies/ml, respectively. No statistical difference in McirDNA copy number was observed between healthy males and females (Fig. 3E; *Mann-Whitney U test, P value = 0.485*).

Multivariate analysis: Logistic regression analysis including age and gender confirmed no statistically significant difference between studied groups (Fig. 3F). Results of multivariate analysis and Odds ratios (OR) with 95% confidence intervals (CIs) are summarized in Supplementary Table S2B.

Patient's characteristics					
Healthy individuals (N = 104)			**mCRC patients (N = 118)**		
Age (years)					
Mean	45		Mean	65	
Median	47		Median	65	
(min-max)	(18–69)		(min-max)	(22–91)	
Gender					
Males	62	59,6%	Males	68	57,6%
Females	42	40,4%	Females	50	42,4%
TOTAL	99		TOTAL	118	
Males (N)					
Mean age	45		Mean age	65	
Median age	47		Median age	65	
(min-max)	(19–69)		(min-max)	(34–88)	
Females (N)					
Mean age	44		Mean age	65	
Median age	45		Median age	67	
(min-max)	(18–63)		(min-max)	(22–91)	

Table 1. Characteristics of healthy individuals (N = 104) and mCRC patients (N = 118).

Effect of various parameters on cirDNA concentrations in mCRC patients. *NcirDNA*. **Plasma aspect (n = 106):** We first compared two groups, 27 abnormal plasmas (icteric and/or opaque plasmas) and 79 normal plasmas. There was no statistical difference in NcirDNA levels between abnormal plasmas and normal plasmas groups (Fig. 4A; *Mann-Whitney U test; P value = 0.266*). However, the median NcirDNA amount determined in abnormal plasmas group was slightly lower than in normal plasmas (3.62×10^3 vs 5.22×10^3 copies/ml).

Delay between blood collection time and last food intake (n = 89): We then compared two groups, "1 h < delay < 5 h" (n = 69) and "other delays" (n = 20). There was no statistical difference in NcirDNA levels between "1 h < delay < 5 h" and "other delays" (Fig. 4B; *Mann-Whitney U test; P value = 0.220*). Nonetheless, the "1 h < delay < 5 h" group showed lower median NcirDNA amount than the group "other delays" (3.68×10^3 vs 5.72×10^3 copies/ml).

Age (n = 118): We also dichotomized the mCRC cohort into two groups around the median age of 65 (Table 1). The median NcirDNA copy number in the <65 year-old group (n = 52) and in the ≥65 year-old group (n = 66) were 4.73×10^3 and 4.94×10^3 copies/ml, respectively. No statistical difference was found (Fig. 4C; *Mann-Whitney U test, P value = 0.757*). A comparative study using the same cut-off for both healthy and mCRC cohorts as the median age of all individuals tested here (N = 222, median age = 56 years) confirmed the statistical difference between young (N = 79) and older (N = 25) healthy individuals groups (*Mann-Whitney U test, P value = 0.0026)*, and also the no statistical difference between young (N = 25) and older (N = 93) mCRC groups (*Mann-Whitney U test, P value = 0.913*) (Supplementary Fig. S1A,B).

Gender (n = 118): The median NcirDNA copy number in the mCRC males (n = 68) and in and females (n = 50) mCRC patients were 4.65×10^3 and 5.40×10^3 copies/ml, respectively. No statistical difference was found (Fig. 4D; *Mann-Whitney U test, P value = 0.971*).

Multivariate analysis: Logistic regression analysis including all the parameters (plasma aspect, delay, age and gender) confirmed no statistical significant results (Fig. 4E). Results of multivariate analysis and Odds ratios (OR) with 95% confidence intervals (CIs) are summarized in Supplementary Table S2C.

McirDNA. **Plasma aspect (n = 46):** We next compared the abnormal plasmas (n = 11) and normal plasmas (n = 35) groups. There was no statistical difference in NcirDNA levels between abnormal and normal plasmas groups (Fig. 5A; *Mann-Whitney U test; P value = 0.263*). However, the median McirDNA amount determined in abnormal plasmas was slightly lower than in normal plasmas (6.64×10^6 vs 1.19×10^7 copies/ml).

Delay between day-time of blood draw and the last food intake (n = 34): There was no statistical difference in McirDNA copy number between the "1 h < delay < 5 h" group (n = 25) and the "other delay" group (n = 9) (Fig. 5B; *Mann-Whitney U test; P value = 0.509*). The "1 h < delay < 5 h" group showed higher median NcirDNA amount than the group "other delay" (1.68×10^7 vs 1.19×10^7 copies/ml).

Age (n = 50): The median McirDNA copy number in the < 65 year-old (n = 23) and ≥ 65 year-old (n = 27) groups were 1.12×10^7 and 1.84×10^7 copies/ml, respectively. There was no statistical difference in the McirDNA copy number between the groups (Fig. 5C; *Mann-Whitney U test, P value = 0.240*).

Gender (n = 50): The median McirDNA copy number in the mCRC males group (n = 25) and in the group of mCRC women (n = 25) were 1.19×10^7 and 1.68×10^7 copies/ml, respectively. We did not found any statistical difference in the McirDNA copy number between mCRC males and females (Fig. 5D; *Mann-Whitney U test, P value = 0.261*).

Multivariate analysis: Logistic regression analysis including all the parameters (plasma aspect, delay, age and gender) confirmed no statistically significant results (Fig. 5E). Results of multivariate analysis and Odds ratios (OR) with 95% confidence intervals (CIs) are summarized in Supplementary Table S2D.

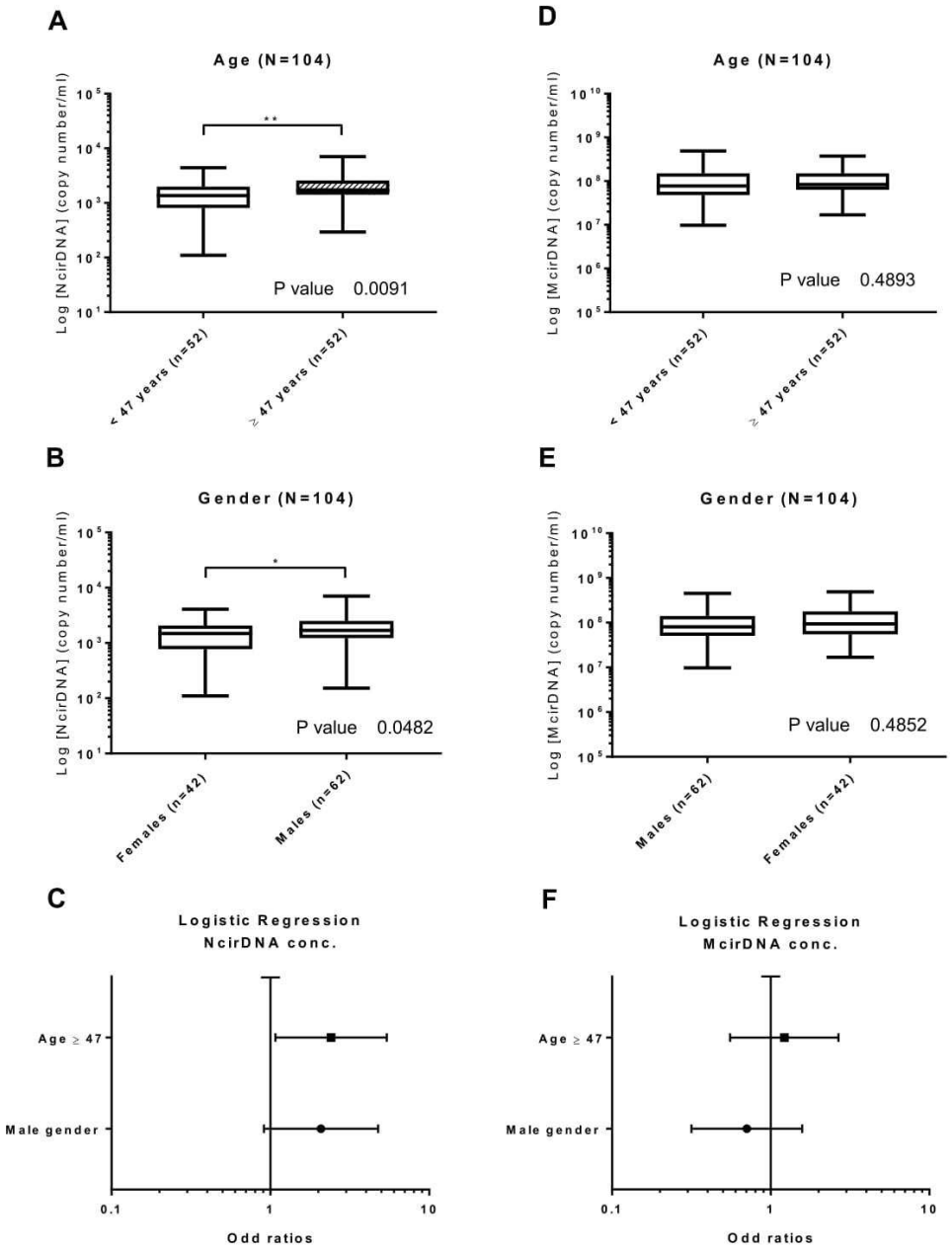


Figure 3. Influence of age and gender on cirDNA concentration in healthy individuals. Boxplot analysis of cirDNA concentration extracted from healthy individuals (N = 104), with regards to age (**A,D**) and gender (**B,E**). (**C,F**) Multivariate analysis representations. NcirDNA (**A–C**) and McirDNA (**D–F**) concentrations are expressed in copy number/ml of plasma. Boxplot represent median with min to max of values. Mann-Whitney U test was performed for univariate analysis and logistic regression was performed for multivariate analysis. Odds ratio (OR) with 95% confidence intervals (CIs) are represented. A probability of ≤ 0.05 was considered to be statistically significant; *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$.

Comparing cirDNA levels between mCRC patients and healthy individuals. *NcirDNA.* We compared the median NcirDNA amount between healthy individuals (n = 104) and mCRC patients (n = 118). The median NcirDNA concentration in healthy individuals and mCRC patients was 1.64×10^3 and 4.73×10^3 copies/ml, respectively, revealing a significant difference between healthy individuals and mCRC patients (Fig. 6A; *Mann-Whitney U test, P value < 0.0001*).

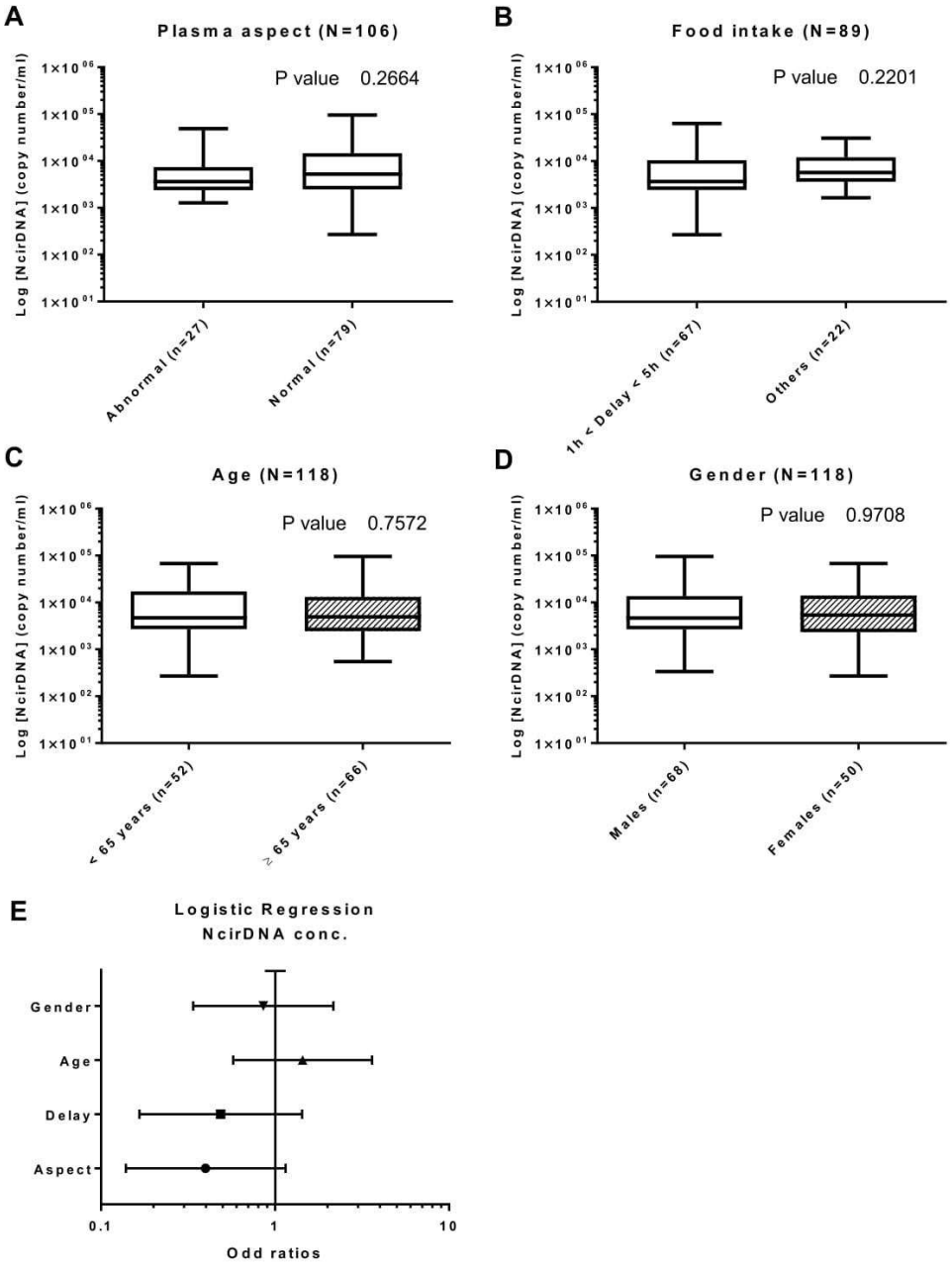


Figure 4. Influence of various factors on NcirDNA concentration in mCRC patients. Boxplot analysis of NcirDNA concentration extracted from mCRC patients (N = 118), with regards to plasma aspect (**A**); delay between blood collection and last food intake (**B**); age (**C**) and gender (**D**). (**E**) Multivariate analysis representation. Boxplot represent median with min to max of values. Mann-Whitney U test was performed for univariate analysis and logistic regression was performed for multivariate analysis. Odds ratio (OR) with 95% confidence intervals (CIs) are presented. A probability of ≤ 0.05 was considered to be statistically significant; *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001, ****p ≤ 0.0001.

McirDNA. Next we compared the median McirDNA amount between healthy individuals (n = 104) and mCRC group (n = 50). The median McirDNA concentration in healthy individuals and mCRC patients was statistically different at 8.32×10^7 and 1.44×10^7 copies/ml, respectively (Fig. 6B; *Mann-Whitney U test, P value < 0.0001*).

McirDNA/NcirDNA ratio. Finally we compared the median McirDNA/NcirDNA ratio between healthy individuals (n = 104) and mCRC patients (n = 50). The median McirDNA/NcirDNA ratio in healthy individuals

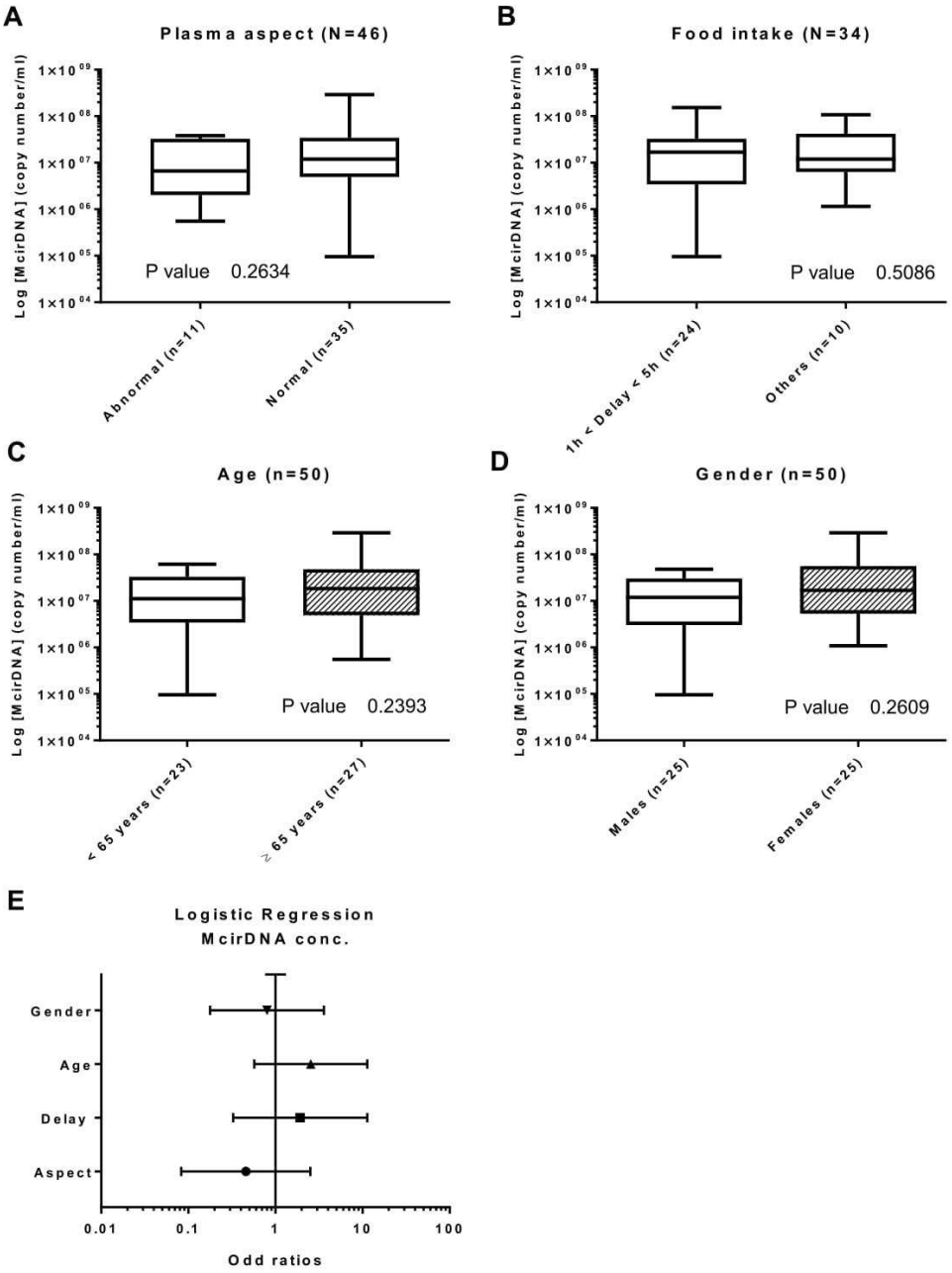


Figure 5. Influence of various factors on McirDNA concentration in mCRC patients. Boxplot analysis of McirDNA concentration extracted from mCRC patients (N = 50), with regards to plasma aspect (**A**); delay between blood collection and last food intake (**B**); age (**C**) and gender (**D**). (**E**) Multivariate analysis representation. Boxplot represent median with min to max of values. Mann-Whitney U test was performed for univariate analysis and logistic regression was performed for multivariate analysis. Odds ratio (OR) with 95% confidence intervals (CIs) are presented. A probability of ≤ 0.05 was considered to be statistically significant; *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001, ****p ≤ 0.0001.

and mCRC patients were statistically different at 5.41×10^4 and 2.70×10^3 copies/ml, respectively (Fig. 6C; *Mann-Whitney U test, P value < 0.0001*). The median McirDNA/NcirDNA ratio was 20-fold higher in the healthy individuals group than in the mCRC patients group. We did the same comparison using McirDNA and NcirDNA concentrations expressed in ng/ml and we obtained the same results (Supplementary Fig. S2, *Mann-Whitney U test, P value < 0.0001)*.

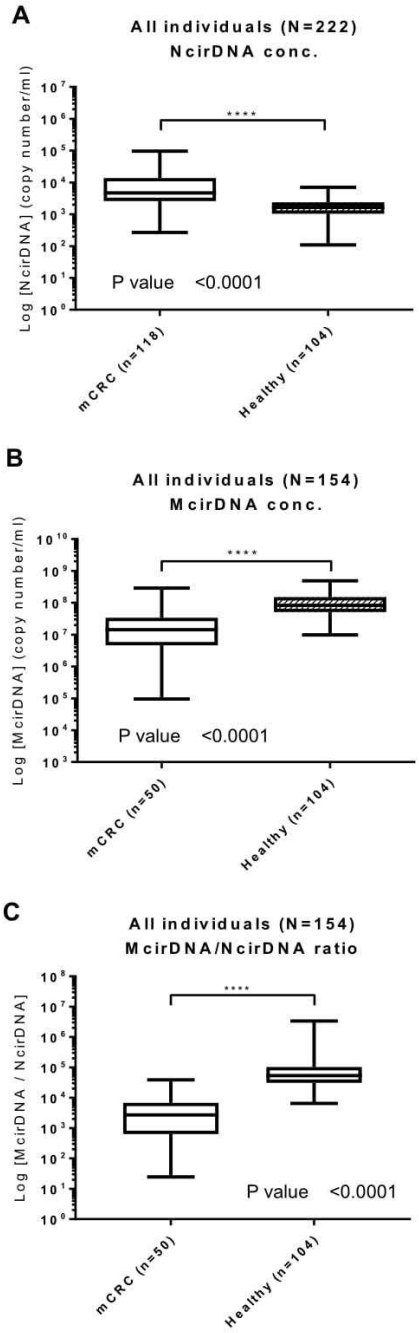


Figure 6. Biomarker capacity of respective NcirDNA and McirDNA concentration for discriminating healthy individuals and mCRC patients. The cohort was dichotomized in two populations (mCRC patients and healthy individuals). Boxplot analysis of the amount of NcirDNA (**A**) and McirDNA (**B**) extracted from plasma of all individuals (N = 222 and N = 154, respectively). (**C**) Boxplot analysis of the McirDNA/NcirDNA ratio of all individuals (N = 154). The boxplots represent medians with min to max of values and Mann-Whitney U test was performed for comparison. A probability of ≤ 0.05 was considered to be statistically significant; *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$.

CirDNA stability in whole blood samples of healthy volunteers. NcirDNA concentration in blood collected in EDTA tubes increased by more than two-fold and three-fold respectively, at days 2 and 5 after blood collection (Supplementary Fig. S3A). In contrast, there was no difference between days 0, 2, 5 and 7 following

blood collection in Cell-Free DNA BCT® STRECK (BCT) tubes (Supplementary Fig. S3B). Note, median values at day 0 are similar in plasma from blood collected in EDTA or BCT tubes. In addition, The McirDNA concentration in blood collected in EDTA tubes approximately increased more than 12-fold and 27-fold respectively, at days 2 and 5 after blood collection (Supplementary Fig. S3C).

Effect of the blood collection time on cirDNA levels in healthy volunteers. *NcirDNA.* In EDTA tubes at 9.00 AM (fasted state), 12.00 PM (2 hours after breakfast), 3.00 PM (2 hours after lunch) and 6.00 PM, the median NcirDNA copy numbers were 1500, 480, 715 and 709 copies/ml respectively. An additional figure shows this in more detail (Supplementary Fig. S4A,E). Taking the four healthy individuals together, the median NcirDNA concentration strongly changed from the fasted state to three hours after breakfast. Median NcirDNA concentration was slightly higher at the earliest collection time in BCT tubes, over the same time-course of the day, at 874, 562, 728 and 313 copies/ml, respectively. An additional figure shows this in more detail (Supplementary Fig. S4B,F). Altogether our data from both collection methods showed that NcirDNA content in healthy individuals declined from 9.00 AM to 12.00 PM then plateaued up to the 6.00 PM collection time. An additional figure shows this in more detail (Supplementary Fig. S4C,D,G,H).

McirDNA. In EDTA tubes from healthy individuals, the median NcirDNA copy numbers were unchanged over the time course at 1.05, 1.25, 1.17 and 1.26 million copies/ml. An additional figure shows this in more detail (Supplementary Fig. S5A,B). No clear difference was found when comparing McirDNA amounts collected at the various collection time points. Note, plasma appearance become opaque at 6.00 PM. An additional picture illustrates this observation (Supplementary Fig. S6).

Discussion

Here, we present a comprehensive study on the quantification of nuclear and mitochondrial cirDNA in the plasma of healthy individuals and a homogenous cohort of mCRC patients. To our knowledge, this work is the first to address altogether the influence of various preanalytical, analytical and demographical factors on NcirDNA and McirDNA levels in a large set of individuals ($n > 200$). Solid tumor mass is composed of a variety of cells, mostly consisting of malignant/cancer, stromal, endothelial and immunological cells. In order to avoid any confusion in nomenclature, we define as 'tumor cells' all the cells composing the tumor mass. The concentration values presented in this study correspond to total cell-free DNA in plasma, either of nuclear or mitochondrial origin. Our observations are summarized in Table 2 and Supplementary Table S3.

Total cirDNA levels. This study design relies on the analytical performance of the assay. We used here a qPCR-based method with an unmatched combination, of reproducibility, sensitivity and specificity for quantifying plasma cirDNA (see Methods section). Furthermore, our assay benefits from a clinically-validated optimal pre-analytic process that we previously set up for plasma preparation[22]; and adapted from Chiu *et al.*, DNA extraction and sample handling[27]. The accuracy of the cirDNA concentration measurement in this study is supported by two assessments: (i) total cirDNA concentration by targeting a *KRAS* sequence was routinely controlled by quantifying a *BRAF* sequence. In addition, this control quality enable to detect and exclude sample with loss of heterozygosity (LOH) or gene amplification which have been reported in CRC patients[28]. Moreover, since *KRAS* amplification is an infrequent event in CRC (0.67%)[28], this level will not modify our observations or the values described in our manuscript; and (ii) the study of cirDNA measurement under Poisson law distribution revealed single copy detection of nuclear cirDNA.

Respective NcirDNA and McirDNA proportions inverted when the plasma cirDNA concentrations were calculated according to copy number or mass (Fig. 2A,C). For example, data obtained from 104 healthy individuals revealed that median McirDNA plasma concentration was approximately 50,000-fold higher than median NcirDNA plasma concentration in terms of copy number/mL (99.998% McirDNA and 0.002% NcirDNA), corresponding to approximately 4-fold lower in terms of ng/ml (25.0% McirDNA and 75.0% NcirDNA). This highlights the much lower size of the mitochondrial genome (16569 bp) than the nuclear genome (about 3×10^9 bp), and also the high number of mitochondrial genome copies within a cell. Each human cell, depending on its type, contains a number of mitochondria ranging from 500 to 2,000 and each mitochondria holds between 2 and 10 mitochondrial DNA molecules. Therefore, each cell may contain approximatively 1,000 to 20,000 mitochondrial DNA molecules and only one nuclear genome equivalent, which is consistent with our results. Our method of calculation, described here, appears as the most rigorous means to simultaneously quantify mitochondrial and nuclear DNA concentrations. A higher coefficient of variation was routinely observed for mitochondrial than for nuclear DNA, despite the higher analytical signal for McirDNA copy number. The median concentration levels we found here appear to be similar to the average values observed in several studies[27,29]. In addition, performance of our assay support the accuracy of the values presented throughout this study such as 5.43 and 1.36 ng/mL in healthy subject plasma of N- and McirDNA, respectively. Considering the significant percentage of McirDNA (i.e.: 25% in healthy subjects) among total cirDNA mass existing in plasma, we propose to always specify the origin of cirDNA when examining total circulating WT DNA.

Influence of age and gender on NcirDNA levels in healthy individual plasma. Most studies found no relationship between NcirDNA levels and any demographic parameters, such as age and gender, in healthy individuals[30–32]. A few studies showed opposite observations[33–35]. Here, we observed an influence of age on NcirDNA plasma concentration in the healthy individual cohort as a whole ($p = 0.009$). Our data also revealed a statistical difference in NcirDNA amount between healthy males and females ($p = 0.048$). We performed logistic regression for multivariate analysis including age and gender and showed that age ≥ 47 was predictive of a high NcirDNA concentration (OR = 2.41, $p = 0.033$). This observation is in accordance with a study by Jylhävä *et al.*

Demographical considerations	Group	N	In accordance with previous works
NcirDNA concentration			
A statistical difference between males and females	Healthy	104	Catarino R, *et al.*[63]
No statistical difference between males and females	mCRC	118	Hao TB, *et al.*[31]
A statistical difference with regard to age	Healthy	104	Jylhävä J, *et al.*[35]
No statistical difference with regard to age	mCRC	118	Van der Drift MA, *et al.*[30]
Statistical increase in mCRC patients as compared to healthy individuals	Healthy mCRC	104 118	Bedin C, *et al.*[64]
McirDNA concentration			
No statistical difference between males and females	Healthy mCRC	104 50	New observation
No statistical difference with regard to age	Healthy mCRC	104 50	Jylhävä J, *et al.*[35] New observation
Statistical decrease in mCRC patients as compared to healthy individuals	Healthy mCRC	104 50	New observation
McirDNA vs NcirDNA concentration			
Necessity of independently quantifying NcirDNA and McirDNA			
About 50,000-fold more McirDNA copy number as compared to NcirDNA in healthy individuals	Healthy	104	New observation
About 3,000-fold more McirDNA copy number as compared to NcirDNA in cancer patients	mCRC	50	New observation
About 20-fold more McirDNA/NcirDNA ratio in healthy individuals as compared to mCRC patients	Healthy mCRC	104 50	New observation

Table 2. Summary of the observations made on the influence of demographical factors.

that consisted of 12 nonagenarian women (age > 90 years) and 11 healthy control female (22 < age < 37 years) that showed a higher concentration of NcirDNA in nonagenarians than in control women[33]. The authors explained this increase in the amount of cirDNA with age, as the accentuation of a senescence phenomenon and cell death, caused by an inflammation associated with age and even by decreased clearance and phagocytic capacity. Zhong *et al.* showed a significant increase in total plasma cirDNA concentration in women over 60, compared to younger women, which appears to be consistent with Jylhävä's study. We may speculate that menopause could be an explanation of the statistically higher NcirDNA concentration in healthy males as compared to healthy females ($p = 0.048$), whereas no difference between mCRC males and females was observed ($p = 0.971$). This speculation is based on two observations: (i) Median menopausal age in the European population is 51 years-old; 10–15% of women go into menopause before the age of 45, and globally 90–95% reach menopause by 55 years of age. By applying these categories of age to our women cohorts, we observed no statistical difference and no tendency between mCRC women <45, 45–54 and ≥55 year-old (*Kruskal-Wallis rank test, P value = 0.592*) while we showed a statistical difference between healthy women <45, 45–54 and ≥55 year-old, with a concentration gradient: 1275, 1440 and 2610 median copy number, respectively (*Kruskal-Wallis rank test, P value = 0.026*); and (ii) CRC females are at a high risk of chemotherapy-induced menopause or menstrual disorders like amenorrhea or a decrease of steroid hormone levels[36,37], and metachronous mCRC patients may attain a menopausal state earlier. An additional figure shows these results in more detail (Supplementary Fig. S7).

Influence of various parameters on NcirDNA levels in mCRC patients plasma. Despite the large number of studies that aimed at determining if age and gender might influence NcirDNA levels, no clear results have been demonstrated[30–32,38]. There are discrepancies in the literature with regards to the influence of age and gender in patient populations suffering from various types of cancers. These discrepancies could result from use of serum, or pre-analytical or analytical factors. Note, a study by Hohaus *et al.* showed that patients with Hodgkin and non-Hodgkin's lymphoma over 60 year-old (n = 142) had higher levels of cirDNA in plasma than younger patients ($p = 0.018$)[38]. Conversely, our data showed no statistical influence of age ($p = 0.757$) and gender ($p = 0.971$) on NcirDNA concentration in mCRC patients and these results was confirmed with a multivariate analysis using logistic regression. We also demonstrated no influence of pre-analytical factors like plasma aspect ($p = 0.266$) or delay between last food intake and blood collection ($p = 0.220$). On the other hand, our results also confirm those previously published by our laboratory[25] and by many teams[39,40], namely that NcirDNA concentration is significantly higher cancer patients than in healthy individuals ($p < 0.0001$), whether male or female. For more than a decade it was suggested that total NcirDNA could be a cancer biomarker[25,41,42]. However, previous attempts to apply total cirDNA quantity as a screening test for cancer lacked a strong statistical demonstration[43]. High standards, with regard to pre-analytical factors and quantification, could lead to its use as one marker, among other, for tumor burden.

McirDNA levels in plasma. Relatively few reports have quantitatively analyzed McirDNA and there are discrepancies among them. In one study there was no significant difference between young and aged healthy subjects[35]. For Pinti *et al.* however, McirDNA concentration would increase with age[18]. In this study, McirDNA content was analyzed in 831 plasma samples from subjects with different healthy status, aged from 1 to 104 years; McirDNA content significantly increased after fifty years-of-age and it peaked in nonagenarians. Elevated McirDNA levels might

help maintain the low-grade chronic inflammation that is common in elderly individuals. With regards to McirDNA level in cancer patients, Mengel-From *et al.* measured McirDNA copy number in blood cells from 1,067 subjects aged 18 to 93 and conversely, observed a tendency for lower mitochondrial DNA copy number with advanced age[44]. These findings are consistent with other studies, performed on different types of tissue, such as skeletal muscle and pancreatic islets[45,46]. This age-related tissue-specific depletion of cellular mitochondrial DNA could lead to a proportional reduction of McirDNA copy-number in plasma. Inversely, there was no association between serum McirDNA levels and demographic parameters (age/gender) in urological malignancies[47,48]. Likewise, in breast cancer, there was no significant difference in McirDNA content in blood samples of stage I patients with respect to their age[49]. Our data revealed no influence of age and gender on McirDNA concentration, either in plasma of healthy individuals or mCRC patients. These results were confirmed in multivariate analysis. We also reported no significant influence of plasma aspect ($p = 0.263$) and delay between last food intake and blood collection ($p = 0.509$). Nonetheless, our data revealed a significant higher McirDNA than NcirDNA concentration in plasma, whether for mCRC patients, healthy individuals, male or female, and regardless of the age of the subject. Note, while median NcirDNA concentration is much higher in mCRC patients than in healthy individuals, median McirDNA concentration is conversely lower in mCRC patients, revealing a proportionally lower McirDNA release from cancer cells. This might be explained by the fact that cancer cells, in comparison to healthy cells, may have, among other differences, fewer mitochondria per cell and less DNA within their mitochondria[34,50]. However, this is still controversial and explanation of our striking observation is under active investigation in our team. Nevertheless, we may speculate that the McirDNA/NcirDNA ratio might have some power in discriminating healthy individuals from cancer patients. McirDNA/NcirDNA ratio is undergoing clinical validation as potential biomarker for tumor burden or diagnosis in a large study involving broader scope of cancer patients with various malignancies and stages. In light of the high copy number of McirDNA and its tendency to be mutated in cancer[51], our observations confirm the gradual acceptance of McirDNA as a new potent diagnostic and prognostic biomarker for many solid tumors[52].

Blood stability for cirDNA plasma assessment. As previously reported[23,53], NcirDNA concentration determined from blood collected in EDTA tubes increased with time highlighting release of genomic DNA resulting from blood cell lysis when stored at room temperature or +4 °C and consequently to contamination of cell-derived DNA. Note, whole blood stored in EDTA tubes at +4 C° showed no change in cirDNA concentration for up to one day suggesting their potential use within this time period (data not shown)[23,53]. We propose routine clinical analyses use plasma stored in EDTA tubes for up to 6 hours, given the uncertainty of maintaining the temperature of samples in the course of blood processing, as we earlier described[54]. Conversely, BCT tubes appeared to conserve blood cell integrity, since no DNA concentration increase was observed up to 7 days following blood collection[23,55]. Thus, BCT tubes maintain the true cirDNA concentration and are good tools to conserve/stabilize blood for optimal quantification of NcirDNA for up to 7 days following collection. Cell-preserving tubes greatly allows postal shipment of whole-blood within this time period and it allows interventional analysis as well as enabling clinical centers that lack lab facilities to immediately prepare plasma. While being cheaper by themselves, use of EDTA tubes necessitates plasma preparation within a short time frame and immediate subsequent storage under frozen conditions until analysis requiring costly shipment when plasma originate from a single patient. We first reported that McirDNA concentration determined from blood collected in EDTA tubes strongly increased with time. We may assume that this results as well blood cell lysis and blood cell-derived mitochondrial DNA contamination.

Effect of blood collection time on cirDNA plasma concentration. There is currently no indication in the literature on the optimal time for blood collection when analyzing cirDNA. Our data seem to indicate that NcirDNA median levels are 2- to 3-fold higher at 9.00 AM, which is the earliest time-point examined, compared to later blood-collection time-points (12.00, 3.00 and 6.00 PM) when the NcirDNA level stabilizes. Decrease from 9.00 AM to 12.00 PM might be explained by the postprandial effect of the breakfast being taken at 10.00 AM. This hypothesis is supported by several observations. First, we reported that NcirDNA plasma levels in blood collected between one and five hours after food intake were lower than in blood collected on patients under fasting conditions. It was previously showed that plasma triglyceride increased one hour after food intake, peaked ≈3 hours after intake of a test meal and baseline values were restored back to initial values after 5 hours[56]. Second, NcirDNA concentration was lower in opaque than in non-opaque plasmas. Multivariate analysis including age, gender, plasma aspect and delay between food intake and blood collection revealed no statistical influence but abnormal plasma aspect showed a clear tendency to have lower NcirDNA concentration ($OR = 0.399$; $p = 0.089$). These observations are all consistent with postprandial effects. Food intake with high lipid content may result in hyperlipidemia which can be characterized by opaque plasma and high triglyceride concentrations. However, despite the large examined cohort data, we cannot state that postprandial is the explanation since no statistical difference was found. This may be due to various factors: (i) blood triglyceride levels largely depend on fat distribution and body weight, lifestyle choices, and also genetic factors[57]; (ii) there were considerable within- and between-subject variations in non-fasting plasma triglycerides[58]; and (iii) the subjects had a chronic illness, mCRC. We speculate that the postprandial effect could occur because the presence of lipids or proteins may interfere with DNA extraction yield from plasma. We cannot exclude the possibility that NcirDNA levels depend on circadian clocks and metabolism, resulting in more elevated concentration in the morning. Moreover, we cannot exclude the possibility that other metabolic changes during fasting/feeding alter cirDNA yield. Nevertheless, our data suggest that fasting blood samples should be included when studying or clinically examining cirDNA to improve its diagnostic performance, especially when low mutation frequency in cancer patients or prenatal testing is considered. Note, in addition to opaque plasma, we remarked on various occasions that, icteric plasma had aberrant cirDNA concentration when qPCR was the analytical method. Therefore, we propose observation of icteric, hemolytic and opaque plasma as criteria of blood sample exclusion.

Limitations of the study. While the study was carried out with statistically sufficient number of subjects to support the observations, the *ad hoc* study is limited by the low number of tested individuals since *ad hoc* study experiments are cumbersome and time-consuming. Thus, blood stability and blood collection time was only carried out on a few number of healthy volunteers (n = 5) and not on mCRC patients. This not allowed us to provide statistical analysis while the results showed clear tendencies. A specific study should be performed to statistically confirm these results on a larger cohort of healthy volunteers as well as mCRC patients. Although we routinely experienced that abnormal plasmas resulted in lowering cirDNA concentration values, we cannot fully discriminate the implication of postprandial effects like triglycerides serum level, to the possible involvement of biological changes dues to the circadian rhythm. In order to definitively address this issue, it would be interesting to compare NcirDNA levels at 9.00 AM and 12.00 AM, with and without breakfast, in order to determine the impact of food intake and circadian rhythm, respectively. In addition, the influence of the menopause on cirDNA concentration with regard to our observations of gender, age and pathological status is only speculative. A specific study on the difference of NcirDNA concentration between postmenopausal and premenopausal healthy females, as well as the difference between premenstrual and postmenstrual young women, should be performed to definitively address this issue. Conclusions drawn here in respect to cancer plasma samples should be restricted to mCRC patients and extension to other malignancies or even to localized disease is speculative.

In conclusion, the levels of mitochondrial and nuclear circulating DNA differently vary with regards to pre-analytical and demographic factors. Those variables should be taken into consideration when evaluating cirDNA analysis in clinical setting and perhaps in the future clinical practice when cirDNA quantification is directly or indirectly used as a biomarker. In addition, our study highlights the potential for combining the analysis of NcirDNA and McirDNA since examining their respective levels may have diagnostic value.

Methods

Patients. Blood samples collected from 104 healthy donors were provided by the Etablissement Français du Sang (E.F.S), the blood transfusion center of Montpellier (Convention EFS-PM N° 21PLER2015-0013). Blood samples collected by the E.F.S were analyzed (virology, serology, immunology, blood numeration). If an anomaly is detected, the sample is ruled out and the donor is warned then by mail. NcirDNA concentration and McirDNA copy number were determined for all healthy individuals. Data on age and gender were collected for all healthy individuals. mCRC patients data are taken from a study comparing the detection of *KRAS* exon 2 and *BRAF V600E* mutations by circulating DNA (cirDNA) analysis to conventional detection by tumor tissue analysis[59]. This study (KPLEX2) was performed and presented under the STARD criteria. 140 patients have been included in 11 clinical centers in France, over a period of 12 months. Eligible patients were male or female, aged ≥18 years, with a proven histologically mCRC, a measurable disease as defined by response evaluation criteria in solid tumors (RECIST v1.1) and untreated by radiotherapy or chemotherapy in the last 4 weeks before inclusion. There is no possibility that cirDNA from mCRC patients can be affected by therapy since eligible patients were untreated by radiotherapy or chemotherapy in the last 4 weeks before blood collection (inclusion criteria). Written consent was obtained from the part of all patients. Inclusion criteria were described previously[59]. 19 were excluded from analysis for no respect of the inclusion criteria (due to various inclusion criteria) and 3 were excluded due to lack of data collected on age or gender. NcirDNA analysis was performed on 118 patients and McirDNA analysis on 50/118 mCRC patients. While study on age and gender effect were carried out on all the included mCRC patients for NcirDNA (N = 118), cohort patient number varied when studying plasma aspect and delay for NcirDNA, or age and gender for McirDNA because of two main reasons: non-reported information for delay and plasma aspect for to NcirDNA and insufficient plasma volume needed to carry out supplementary analysis for McirDNA. All data on age and gender were collected. Plasma aspect (normal, abnormal: opaque or/and icteric) was noted for 106/118 mCRC patients and delay between time of blood draw and the last food intake was informed for 89/118 mCRC patients.

Ethics approval and consent to participate. Blood samples collected from 104 healthy donors were provided by the Etablissement Français du Sang (E.F.S), the blood transfusion center of Montpellier (Convention EFS-PM N° 21PLER2015-0013). Plasma samples from mCRC patients were obtained from the Kplex2 study registration number EUDRACT 2016-001490-33 with ethic committee approval ("Comité de Protection des Personnes", Nimes, France). All methods were performed in accordance with the relevant guidelines and regulations. The study obtained informed consent from all participants for the study.

Blood stability. 7 samples for each healthy donor were collected at 9.00 a.m. on Day 0 (fasted state): 3 with EDTA K2 tubes and 4 with Cell-Free DNA BCT® STRECK tubes. Each tube was processed as we early described, at day 0, 2 and 5 for EDTA tubes and day 0, 2, 5 and 7 for BCT tubes. EDTA tubes were stored at +4 °C and BCT tubes were conserved at room temperature before isolation. Healthy individual N°1 (HI1) and N°2 (HI2) are 57 year-old and 29 year-old men, respectively, both with no known disease. Healthy individual N°3 (HI3), N°4 (HI4) and N°5 (HI5) are 25 year-old, 24 year-old and 28 year-old women, respectively, both with no known diseases. NcirDNA analysis for HI1 was not performed due to clotting in BCT tubes.

Blood collection time. 4 blood collection times were defined 9.00 a.m. (fasted state), 12.00 p.m. (2 hours after breakfast), 3.00 p.m. (2 hours after lunch) and 6.00 p.m. (4 hours after lunch). One EDTA K2 and one Cell-Free DNA BCT® STRECK tube per day-time were collected for each donor. Healthy individual N°6 (HI6) and N°7 (HI7) are 29 year-old and 27 year-old men, respectively, both with no known diseases. Healthy individual N°8 (HI8) and N°9 (HI9) are 28 year-old and 30 year-old women respectively, both with no known diseases. Each donors took the same meal during breakfast: one butter croissant, one chocolate croissant and one coffee with sugar; and for the lunch: a dish of tomato rice with sausage, bread and one coffee with sugar.

Sample characteristics and preparation. Samples were collected and treated in accordance with a pre-analytical guideline previously established by our group[22]. In summary, blood was collected in EDTA K3 tubes and plasma was isolated within 2 hours. The isolation technique consist of a double centrifugation. Initially, tubes were centrifuged for 10 minutes at 4 °C and 1,200 g in a Heraeus Multifuge LR centrifuge. The supernatant was collected while carefully avoiding the buffy-coat. The second centrifugation was conducted for 10 minutes at 4 °C and 16,000 g. The supernatant was transferred to 1.5 ml tubes before performing the extraction of cirDNA or being stored at −20 °C. CirDNA extraction was performed with the Qiagen Blood Mini kit, following all steps of the protocol. In all, 0.2 to 1 ml of plasma was extracted in several successive passes on a column. The final elution volume was 80 to 130 µl and eluates were frozen at −20 °C prior to analysis by qPCR. Freeze-thaw cycles should be avoided to reduce the phenomenon of cirDNA fragmentation and the extracts are not kept longer than 3 months at −20 °C.

Q-PCR analysis. CirDNA analysis was performed by a qPCR technique developed in our laboratory, and clinically validated previously[60]. The method is based on an innovative design of short amplicons (60–100 bp ± 10 bp) targeting a wild-type sequence of the gene, (here the *KRAS* nuclear gene and the mitochondrial Cytochrome oxidase III gene, *MT-CO3)*. Quantification of this amplicon gives an estimation of the total NcirDNA and McirDNA concentration, respectively. For the quantification of NcirDNA, we amplified of a 67 bp-length sequence of the KRAS gene with the following primers: forward (5′ CCTTGGGTTTCAAGTTATATG 3′) and reverse (5′ CCCTGACATACTCCCAAGGA 3′). For McirDNA, we amplified a 67 bp-length sequence of the cytochrome oxidase sub-unit 3 mitochondrial gene with the following primers: forward (5′ GACCCACCAATCACATGC 3′) and reverse (5′ TGAGAGGGCCCCTGTTAG 3′). These primers were designed using the Primer 3 software according to the following requirements: (i) T_m ranging from 50 to 64 °C; (ii) GC-content between 40 and 60%; (iii) size from 18 to 23 NT; (iv) amplicon size ranging from 60 to 100 bp. We performed local-alignment analyses with the BLAST program to confirm the specificity of the designed primers. All sequences were checked for self- or inter-molecular annealing with nucleic-acid-folding software (Mfold and oligoAnalyzer 1.2). Oligonucleotides were synthesized and HPLC-purified by Eurofins (Ebersberg, Germany) and quality control of the oligonucleotides was performed by MALDI-TOF. For all analyses, negative controls and standard curves were used. All tests are performed in triplicate with 5 µl of DNA extract in a 25 µl reaction volume, on a CFX96 instrument using CFX manager software (Bio-Rad). This method (qPCR, primer design, program) and technical validation have been described previously[25]. The mCRC patient's blood samples were excluded if the total cirDNA concentration, due to a problem of pre-analytic treatment or even for unknown reasons, was below a quality threshold.

The DNA concentration quality threshold was 3 ng/mL (about 900 copies/mL) for cancer patients. Note, this value corresponds to about the half of the median concentration found for healthy individuals (N = 104, median NcirDNA concentration = 5.43 ng/mL of plasma corresponding to 1645 copies/mL). In addition to quantifying cirDNA by targeting two different sequences on two different chromosomes, an experiment based on Poisson law distribution showed the accuracy of our cirDNA amount measurement (Supplementary Fig. S8). It should be noted that our Q-PCR systems enable the detection of a single genome copy (Supplementary Table S4), and that we have previously shown that targeting a DNA sequence of the same size or longer than the input DNA fragment produced a similar PCR yield[61]. The measurement of the total cirDNA concentration by targeting a *KRAS* sequence was routinely controlled by quantifying a *BRAF* sequence. Control quality is acceptable when the *KRAS*-based value is 1.3 to 1.8-fold higher than that of the *BRAF*-based value when using the reported Q-PCR primer systems; otherwise, a second analysis is performed. Note, *BRAF* analysis data from 33 healthy individuals were not available. In addition, we excluded 26 patient plasmas (17 mCRC and 9 healthy) in which the *KRAS*/*BRAF* ratio was over 3 or below 0.5. Data revealed that the *KRAS*-based concentration value was positively correlated with the *BRAF*-based concentration values in the 62 healthy individuals (Spearman analysis; r = 0.762, *P value* < *0.0001*) and in the 101 mCRC patients (Spearman analysis; r = 0.882, *P value* < *0.0001*) (Supplementary Fig. S9). We already addressed this issue in our previous report (Spearman analysis; r = 0.966, *P value* < *0.001*)[25].The *KRAS* and *BRAF* genes are monogenic and poorly amplified in both healthy and cancer individuals[62]. Supplementary section figures present NcirDNA concentration determined from targeting *BRAF* with using a primer set of similar size. Data revealed that the same observations could be made: the NcirDNA amount, as determined using *BRAF* sequence targeting, is statistically different in healthy (N = 62) and mCRC individuals (N = 101) (Supplementary Fig. S10; Mann-Whitney U test, *P value* < *0.0001*), and fully correlates with our observation based on *KRAS* sequence targeting. We may therefore indicate that amplification of the *KRAS* gene will not have any influence on the observations and conclusions made in this study. Valtorta *et al.* detected *KRAS* amplification in 7/1,039 (0.67%) evaluable CRC specimens, demonstrating that *KRAS* amplification is an infrequent event in CRC[28]. Thus this level will not modify significantly the observations or values described in our manuscript.

CirDNA calibration assay. *NcirDNA.* A genomic DNA extract from human wild-type *KRAS* colorectal cells was used for the NcirDNA calibration assay. Initial genomic DNA solution concentration and purity were determined by measuring optic density at λ = 260 nm, 230 nm and 280 nm, with an Eppendorf BioPhotometer® D30. Starting genomic DNA concentration was adjusted to 1800 pg/µl for the first dilution point, according to optic density measurement at λ = 260 nm. A qPCR standard curve was obtained by 6 successive dilutions of the vector solution (1800, 180, 45, 20, 10 and 5 pg/µl). The standard curve was used to determine the NcirDNA concentration of the mCRC patients and healthy individuals and calculate the NcirDNA copy number per milliliter of plasma.

McirDNA. A 3382-pb human ORF vector with a 786-pb *MT-CO3* insert was obtained from ABM good® (accession no.YP_003024032) and used for the McirDNA calibration assay. Initial vector solution concentration

and purity were determined by measuring optic density at $\lambda = 260$ nm, 230 nm and 280 nm, with an Eppendorf BioPhotometer® D30. Starting vector concentration was adjusted at 1800 pg/µl for the first dilution point, according to optic density measurement at $\lambda = 260$ nm. A qPCR standard curve was obtained by 6 successive dilutions of the vector solution (1800, 180, 45, 20, 10 and 5 pg/µl). The standard curve was used to determine the McirDNA concentration of the mCRC patients and healthy individuals and calculate the McirDNA copy number per milliliter of plasma.

CirDNA copy number calculation. *NcirDNA.* NcirDNA copy number per milliliter of plasma, in all analyses, was determined with the following calculation:

$$Qnuclear = \left(\frac{c}{3,\ 3}\right) * \left(\frac{Vélution}{Vplasma}\right)$$

$Q_{nuclear}$ is the NcirDNA copy number per milliliter, c is the NcirDNA concentration (pg/µl) determined by qPCR targeting the nuclear *KRAS* gene sequence and 3.3 pg is the human haploid genome mass. $V_{élution}$ is the volume of cirDNA extract (µl) and V_{plasma} is the volume of plasma used for the extraction (ml).

McirDNA. McirDNA copy number per milliliter of plasma, in all analyses, was determined with the following calculation:

$$Qmito = \left(\frac{c * Na}{2 * MW * Lvector}\right) * \left(\frac{Vélution}{Vplasma}\right)$$

$Q_{mitochondrial}$ is the McirDNA copy number per milliliter, 'c' is the McirDNA mass concentration (g/µl) determined by a qPCR targeting the mitochondrial *MT-CO3* gene. N_A is Avogadro's number ($6.02 * 10^{23}$ molecules per mole), L_{vector} is the plasmid length (nucleotides) and MW is the molecular weight of one nucleotide (g/mol). $V_{elution}$ is the elution volume of cirDNA extract (µl) and V_{plasma} is the volume of plasma used for the extraction (ml).

Statistical analysis. For continuous variables, median and range were computed. To investigate their associations with the biologic parameters, univariate statistical analyses were performed using Mann-Whitney U test or Kruskal-Wallis rank test for continuous variables. Moreover, multivariate analyses were carried out using logistic regressions, with a stepwise selection procedure, to investigate known predictive. Odds ratio (OR) with 95% confidence intervals (CIs) are presented. The power of analysis was reduced due to all patients did not have measurements for all variables. All P values reported are two sided. A probability of ≤ 0.05 was considered to be statistically significant; *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$. Statistical analysis was performed using the STATA 13.1 software (Stata Corporation, College Station, TX).

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Mandel, P. & Metais, P. [Not Available]. *C. R. Seances Soc. Biol. Fil.* **142**, 241–243 (1948).
2. Malik, A. N. *et al.* Altered circulating mitochondrial DNA and increased inflammation in patients with diabetic retinopathy. *Diabetes Res. Clin. Pract.* **110**, 257–265 (2015).
3. Dhondup, Y. *et al.* Low Circulating Levels of Mitochondrial and High Levels of Nuclear DNA Predict Mortality in Chronic Heart Failure. *J. Card. Fail.* **22**, 823–828 (2016).
4. Timmermans, K., Kox, M., Scheffer, G. J. & Pickkers, P. Plasma Nuclear and Mitochondrial DNA Levels, and Markers of Inflammation, Shock, and Organ Damage in Patients with Septic Shock. *Shock Augusta Ga* **45**, 607–612 (2016).
5. Zhang, S. *et al.* Elevated plasma cfDNA may be associated with active lupus nephritis and partially attributed to abnormal regulation of neutrophil extracellular traps (NETs) in patients with systemic lupus erythematosus. *Intern. Med. Tokyo Jpn.* **53**, 2763–2771 (2014).
6. Lam, N. Y. L., Rainer, T. H., Chan, L. Y. S., Joynt, G. M. & Lo, Y. M. D. Time course of early and late changes in plasma DNA in trauma patients. *Clin. Chem.* **49**, 1286–1291 (2003).
7. Yaros, M. J. S. A. L. Free DNA in the Serum of Cancer Patients and the Effect of Therapy. *Cancer Research* **37**, 647 (1977).
8. Zachariah, R. R. *et al.* Levels of circulating cell-free nuclear and mitochondrial DNA in benign and malignant ovarian tumors. *Obstet. Gynecol.* **112**, 843–850 (2008).
9. Mahmoud, E. H., Fawzy, A., Ahmad, O. K. & Ali, A. M. Plasma Circulating Cell-free Nuclear and Mitochondrial DNA as Potential Biomarkers in the Peripheral Blood of Breast Cancer Patients. *Asian Pac. J. Cancer Prev. APJCP* **16**, 8299–8305 (2015).
10. Gautschi, O. *et al.* Circulating deoxyribonucleic Acid as prognostic marker in non-small-cell lung cancer patients undergoing chemotherapy. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **22**, 4157–4164 (2004).
11. Lo, Y. M. *et al.* Presence of fetal DNA in maternal plasma and serum. *Lancet Lond. Engl.* **350**, 485–487 (1997).
12. Lo, Y. M. *et al.* Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. *Am. J. Hum. Genet.* **62**, 768–775 (1998).
13. Lo, Y. M. Fetal DNA in maternal plasma: biology and diagnostic applications. *Clin. Chem.* **46**, 1903–1906 (2000).
14. Thierry, A. R., Messaoudi, S. E., Gahan, P. B., Anker, P. & Stroun, M. Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev.* **35**, 347–376 (2016).
15. Stroun, M., Lyautey, J., Lederrey, C., Olson-Sand, A. & Anker, P. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. *Clin. Chim. Acta Int. J. Clin. Chem.* **313**, 139–142 (2001).
16. Aucamp, J. *et al.* Kinetic analysis, size profiling, and bioenergetic association of DNA released by selected cell lines in vitro. *Cell. Mol. Life Sci.* **74**, 2689–2707 (2017).
17. Franceschi, C. Inflammaging as a Major Characteristic of Old People: Can It Be Prevented or Cured? *Nutr. Rev.* **65**, S173–S176 (2007).

18. Pinti, M. *et al.* Circulating mitochondrial DNA increases with age and is a familiar trait: Implications for 'inflamm-aging'. *Eur. J. Immunol.* **44**, 1552–1562 (2014).
19. Hsu, F.-C. *et al.* Association between inflammatory components and physical function in the health, aging, and body composition study: a principal component analysis approach. *J. Gerontol. A. Biol. Sci. Med. Sci.* **64**, 581–589 (2009).
20. Franceschi, C. & Campisi, J. Chronic Inflammation (Inflammaging) and Its Potential Contribution to Age-Associated. *Diseases. J. Gerontol. Ser. A* **69**, S4–S9 (2014).
21. Tóth, K. *et al.* Circadian Rhythm of Methylated Septin 9, Cell-Free DNA Amount and Tumor Markers in Colorectal Cancer Patients. *Pathol. Oncol. Res.* **23**, 699–706 (2017).
22. El Messaoudi, S., Rolet, F., Mouliere, F. & Thierry, A. R. Circulating cell free DNA: Preanalytical considerations. *Clin. Chim. Acta Int. J. Clin. Chem.* **424**, 222–230 (2013).
23. Parpart-Li, S. *et al.* The Effect of Preservative and Temperature on the Analysis of Circulating Tumor DNA. *Clin. Cancer Res.* **23**, 2471–2477 (2017).
24. Schmidt, B. & Fleischhacker, M. Is liquid biopsy ready for the litmus test and what has been achieved so far to deal with pre-analytical issues? *Transl. Cancer Res.* **7**, S130–S139 (2017).
25. Mouliere, F., El Messaoudi, S., Pang, D., Dritschilo, A. & Thierry, A. R. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol. Oncol.* **8**, 927–941 (2014).
26. Kloten, V. *et al.* Liquid biopsy in colon cancer: comparison of different circulating DNA extraction systems following absolute quantification of KRAS mutations using Intplex allele-specific PCR. *Oncotarget* **8**, 86253–86263 (2017).
27. Chiu, R. W. K. *et al.* Quantitative analysis of circulating mitochondrial DNA in plasma. *Clin. Chem.* **49**, 719–726 (2003).
28. Valtorta, E. *et al.* KRAS gene amplification in colorectal cancer and impact on response to EGFR-targeted therapy. *Int. J. Cancer* **133**, 1259–1265 (2013).
29. Ye, W. *et al.* Accurate quantitation of circulating cell-free mitochondrial DNA in plasma by droplet digital PCR. *Anal. Bioanal. Chem.* **409**, 2727–2735 (2017).
30. van der Drift, M. A. *et al.* Circulating DNA is a non-invasive prognostic factor for survival in non-small cell lung cancer. *Lung Cancer Amst. Neth.* **68**, 283–287 (2010).
31. Hao, T. B. *et al.* Circulating cell-free DNA in serum as a biomarker for diagnosis and prognostic prediction of colorectal cancer. *Br. J. Cancer* **111**, 1482–1489 (2014).
32. Kim, K. *et al.* Circulating cell-free DNA as a promising biomarker in patients with gastric cancer: diagnostic validity and significant reduction of cfDNA after surgical resection. *Ann. Surg. Treat. Res.* **86**, 136–142 (2014).
33. Jylhävä, J. *et al.* Aging is associated with quantitative and qualitative changes in circulating cell-free DNA: the Vitality 90+ study. *Mech. Ageing Dev.* **132**, 20–26 (2011).
34. Zhong, X. Y., Hahn, S., Kiefer, V. & Holzgreve, W. Is the quantity of circulatory cell-free DNA in human plasma and serum samples associated with gender, age and frequency of blood donations? *Ann. Hematol.* **86**, 139–143 (2007).
35. Jylhävä, J. *et al.* Characterization of the role of distinct plasma cell-free DNA species in age-associated inflammation and frailty. *Aging Cell* **12**, 388–397 (2013).
36. Cercek, A., Siegel, C. L., Capanu, M., Reidy-Lagunes, D. & Saltz, L. B. Incidence of Chemotherapy-Induced Amenorrhea in Premenopausal Women Treated With Adjuvant FOLFOX for Colorectal Cancer. *Clin. Colorectal Cancer* **12**, 163–167 (2013).
37. Wan, J., Gai, Y., Li, G., Tao, Z. & Zhang, Z. Incidence of Chemotherapy- and Chemoradiotherapy-Induced Amenorrhea in Premenopausal Women With Stage II/III Colorectal Cancer. *Clin. Colorectal Cancer* **14**, 31–34 (2015).
38. Hohaus, S. *et al.* Cell-free circulating DNA in Hodgkin's and non-Hodgkin's lymphomas. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. ESMO* **20**, 1408–1413 (2009).
39. Kohler, C. *et al.* Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors. *Mol. Cancer* **8**, 105 (2009).
40. Stroun, M., Anker, P., Lyautey, J., Lederrey, C. & Maurice, P. A. Isolation and characterization of DNA from the plasma of cancer patients. *Eur. J. Cancer Clin. Oncol.* **23**, 707–712 (1987).
41. Aucamp, J., Bronkhorst, A. J., Badenhorst, C. P. S. & Pretorius, P. J. A historical and evolutionary perspective on the biological significance of circulating DNA and extracellular vesicles. *Cell. Mol. Life Sci. CMLS* **73**, 4355–4381 (2016).
42. Rago, C. *et al.* Serial Assessment of Human Tumor Burdens in Mice by the Analysis of Circulating DNA. *Cancer Res.* **67**, 9364–9370 (2007).
43. Tanos, R. & Thierry, A. R. Clinical relevance of liquid biopsy for cancer screening. *Transl. Cancer Res.* **7**, S105–S129 (2018).
44. Mengel-From, J. *et al.* Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. *Hum. Genet.* **133**, 1149–1159 (2014).
45. Nile, D. L. *et al.* Age-Related Mitochondrial DNA Depletion and the Impact on Pancreatic Beta Cell Function. *Plos One* **9** (2014).
46. Short, K. R. *et al.* Decline in skeletal muscle mitochondrial function with aging in humans. *Proc. Natl. Acad. Sci. USA* **102**, 5618–5623 (2005).
47. Ellinger, J., Müller, S. C., Wernert, N., von Ruecker, A. & Bastian, P. J. Mitochondrial DNA in serum of patients with prostate cancer: a predictor of biochemical recurrence after prostatectomy. *BJU Int.* **102**, 628–632 (2008).
48. Ellinger, J. *et al.* Circulating mitochondrial DNA in serum: a universal diagnostic biomarker for patients with urological malignancies. *Urol. Oncol.* **30**, 509–515 (2012).
49. Xia, P. *et al.* Decreased mitochondrial DNA content in blood samples of patients with stage I breast cancer. *BMC Cancer* **9**, 454 (2009).
50. Zhong, X. Y. *et al.* Elevated level of cell-free plasma DNA is associated with breast cancer. *Arch. Gynecol. Obstet.* **276**, 327–331 (2007).
51. Chatterjee, A., Mambo, E. & Sidransky, D. Mitochondrial DNA mutations in human cancer. *Oncogene* **25**, 4663–4674 (2006).
52. Yu, M. Generation, function and diagnostic value of mitochondrial DNA copy number alterations in human cancers. *Life Sci.* **89**, 65–71 (2011).
53. Kang, Q. *et al.* Comparative analysis of circulating tumor DNA stability In K3EDTA, Streck, and CellSave blood collection tubes. *Clin. Biochem.* **49**, 1354–1360 (2016).
54. El Messaoudi, S. *et al.* Circulating DNA as a Strong Multimarker Prognostic Tool for Metastatic Colorectal Cancer Patient Management Care. *Clin. Cancer Res.* **22**, 3067–3077 (2016).
55. Medina Diaz, I. *et al.* Performance of Streck cfDNA Blood Collection Tubes for Liquid Biopsy Testing. *PloS One* **11**, e0166354 (2016).
56. Boquist, S. *et al.* Alimentary Lipemia, Postprandial Triglyceride-Rich Lipoproteins, and Common Carotid Intima-Media Thickness in Healthy, Middle-Aged Men. *Circulation* **100**, 723–728 (1999).
57. Truong, V. *et al.* Blood triglyceride levels are associated with DNA methylation at the serine metabolism gene PHGDH. *Sci. Rep.* **7**, 11207 (2017).
58. Larsen, L. F., Bladbjerg, E.-M., Jespersen, J. & Marckmann, P. Effects of Dietary Fat Quality and Quantity on Postprandial Activation of Blood Coagulation Factor VII. *Arterioscler. Thromb. Vasc. Biol.* **17**, 2904–2909 (1997).
59. Thierry, A. R. *et al.* Clinical utility of circulating DNA analysis for rapid detection of actionable mutations to select metastatic colorectal patients for anti-EGFR treatment. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **28**, 2149–2159 (2017).
60. Thierry, A. R. *et al.* Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. *Nat. Med.* **20**, 430–435 (2014).

61. Sanchez, C., Snyder, M. W., Tanos, R., Shendure, J. & Thierry, A. R. New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA analysis. *Npj Genomic Med.* **3**, 31 (2018).
62. Corcoran, R. B. *et al.* BRAF Gene Amplification Can Promote Acquired Resistance to MEK Inhibitors in Cancer Cells Harboring the BRAF V600E Mutation. *Sci Signal* **3**, ra84–ra84 (2010).
63. Catarino, R. *et al.* Circulating DNA: diagnostic tool and predictive marker for overall survival of NSCLC patients. *PloS One* **7**, e38559 (2012).
64. Bedin, C. *et al.* Diagnostic and prognostic role of cell-free DNA testing for colorectal cancer patients. *Int. J. Cancer* **140**, 1888–1898 (2017).

Acknowledgements

We are grateful to A. Bauer and B. Ottolini for their help. We would like to thank the clinical investigators from the Kplex2 study: J.L. Raoul, R. Guimbaud, D. Pezet, P. Artru, E. Assenat, C. Borg, M. Mathonnet, C. De La Fouchardière, O. Bouché, and C. Gavoille for collecting the blood samples. The authors would like to thank Kevin Billings and Streck for providing the Cell-Free DNA BCT CE tubes. This work was supported by the INSERM (Institut National de la Santé et de la Recherche Médicale), Lilly (France), and the SIRIC Montpellier Grant (INCa-DGOS Inserm 6045), France.

Author Contributions

R.M. and A.R.T. designed the study, developed the methodology, analyzed the data and prepared the manuscript. R.M., Z.A.A.D., A.O., R.T., B.P., C.S., J.A., G.T., S.A. and S.E.M. realized the experiments. R.M., S.T. and C.M. performed the statistical analysis. All of the authors (R.M., Z.A.A.D., S.T., A.O., R.T., B.P., C.S., J.A., G.T., S.A., C.M., A.A., S.E.M., P.B. and A.R.T.) discussed the results and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-41593-4.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

D. Le sang contient des mitochondries libres circulantes

Titre : Blood contains circulating cell free respiratory competent mitochondria

Auteurs

Zahra Al Amir Dache, Amaëlle Otandault, **Rita Tanos**, Brice Pastor, Romain Meddeb, Cynthia Sanchez, Giuseppe Arena, Laurence Lasorsa, E. Andrew Bennett, Thierry Grange, Safia El Messaoudi, Thibault Mazard, Corinne Prevostel, et Alain R. Thierry

Ce travail est sous press dans *The Faceb Journal (Septembre 2019)*.

Résumé

Les mitochondries sont considérées comme les unités génératrices d'énergie de la cellule en raison de leur rôle clé dans le métabolisme énergétique et la signalisation cellulaire. Cependant, des composants mitochondriaux peuvent se trouver dans l'espace extracellulaire, sous forme de fragments ou encapsulés dans des vésicules. En outre, cet organite intact a récemment été signalé comme étant libéré par les plaquettes exclusivement dans des conditions spécifiques. Ici, nous démontrons, pour la première fois, que la préparation du sang avec des plaquettes non-activées, contient des mitochondries entières et fonctionnelles dans un état physiologique normal. De même, nous montrons que les cellules normales et les cellules tumorales en culture sont capables de sécréter leurs mitochondries. En utilisant la centrifugation en série ou la filtration suivie par des méthodes basées sur la PCR et le séquençage du génome entier, nous détectons de l'ADN mitochondrial extracellulaire intègre dans des particules de plus de 0,2 µm contenant des protéines spécifiques des membranes mitochondriales. Nous identifions ces particules comme étant des mitochondries extracellulaires intactes en utilisant l'analyse FACS, la microscopie à fluorescence et la microscopie électronique à transmission. L'analyse de la consommation d'O2 a révélé que ces mitochondries sont métaboliquement fonctionnelles. Compte tenu du fait que les mitochondries peuvent jouer un rôle dans le transfert intercellulaire, cette découverte pourrait grandement élargir la portée de la biologie de la communication intercellulaire. D'autres analyses doivent être élaborées pour étudier le rôle potentiel des mitochondries en tant qu'organite de signalisation à l'extérieur de la cellule et pour déterminer si ces unités

circulantes pourraient être pertinentes pour la détection précoce et le pronostic de diverses maladies.

Contributions

J'ai participé dans ce travail à la réalisation de certaines expériences, ainsi qu'à la discussion des résultats avec les différents auteurs du papier.

Blood contains circulating cell free respiratory competent mitochondria

Running Title: Blood contains extracellular mitochondria

Zahra Al Amir Dache[1], Amaëlle Otandault[1], Rita Tanos[1], Brice Pastor[1], Romain Meddeb[1], Cynthia Sanchez[1], Giuseppe Arena[2], Laurence Lasorsa[1], E. Andrew Bennett[3], Thierry Grange[3], Safia El Messaoudi[1], Thibault Mazard[1], Corinne Prevostel[1], and Alain R. Thierry[1*]

[1] IRCM, Institut de Recherche en Cancérologie de Montpellier, INSERM U1194, Université de Montpellier, Institut régional du Cancer de Montpellier, Montpellier, F-34298, France

[2]Department of Developmental and Stem Cell Biology, Institut Pasteur, CNRS, Paris, France.

[3] Institut Jacques Monod, Université Paris Diderot, Paris, France.

* Corresponding author and lead contact to : Dr Alain R. Thierry, INSERM, U1194, IRCM, 208 rue des Apothicaires, Montpellier Cedex 5, 34298 , France, Tel: +33-6-63-82-19-94; E-mail: alain.thierry@inserm.fr

Non-standard Abbreviations:

ROS: Reactive oxygen species

DNA: Deoxyribonucleic acid

cfDNA: Cell-free Deoxyribonucleic acid

McfDNA: Mitochondrial cell-free Deoxyribonucleic acid

NcfDNA: Nuclear cell-free Deoxyribonucleic acid

FACS: Fluorescence-activated cell sorting (FACS)

PCR: Polymerase chain reaction

Q-PCR: Quantitative polymerase chain reaction

bp: Base pair

Kbp: Kilo base pair

DII: DNA integrity index

TOM 22: Translocase of outer membrane 22

TIM 23: Translocase of inner membrane 23

OCR: Oxygen consumption rate

EM: Electron microscopy

TFAM: Transcription factor A, mitochondrial

DAMPs: Damage associated molecular patterns

NLRP3: NOD-like receptor family, pyrin domain containing 3

cGAS-STING: Cyclic GMP-AMP Synthase- Stimulator of Interferon Genes

WGS: Whole genome sequencing

LS: Low speed

HS: High speed

PPAP: Protocol precluding activation of platelets

F: Filtrated

NF: Non filtrated

LR: Long range

16gP: 16,000g pellet

IM: Isolated mitochondria

CCCM: Centrifuged cell culture media

TMPD: N, N, N′, N′-tetramethyl-p-phenylenediamine

EDTA: Ethylenediaminetetraacetic acid

CTAD: Citrate–theophylline–adenosine–dipyridamole

ACD-A: Anticoagulant Citrate Dextrose Solution A

MIB: Mitochondrial isolation buffer

FSC: Forward scatter

SSC: Side scatter

PBS: Phosphate-buffered saline

FCCP: Carbonyl cyanide-4-phenylhydrazone

HEPES: 2-[4-(2-hydroxyethyl) piperazin-1-yl]ethanesulfonic acid

BSA: Bovine serum albumin

MAS: Mitochondrial assay solution

Abstract:

Mitochondria are considered as the power-generating units of the cell due to their key role in energy metabolism and cell signaling. However, mitochondrial components could be found in the extracellular space, as fragments or encapsulated in vesicles. In addition, this intact organelle has been recently reported to be released by platelets exclusively in specific conditions. Here, we demonstrate for the first time, that blood preparation with resting platelets, contains whole functional mitochondria in normal physiological state. Likewise, we show, that normal and tumor cultured cells are able to secrete their mitochondria. Using serial centrifugation or filtration followed by PCR-based methods, Whole Genome Sequencing, we detect extracellular full-length mitochondrial DNA in particles over 0.2µm holding specific mitochondrial membrane proteins. We identify these particles as intact cell-free mitochondria using FACS analysis, fluorescence microscopy and transmission electron microscopy. O_2 consumption analysis revealed that these mitochondria are respiratory-competent. In view of previously described mitochondrial potential in intercellular transfer, this discovery could greatly widen the scope of cell-cell communication biology. Further steps should be developed to investigate the potential role of mitochondria as a signaling organelle outside the cell and to determine whether these circulating units could be relevant for early detection and prognosis of various diseases.

Keywords:

Blood, mitochondria, mitochondrial genome, circulating DNA, respiratory-competent.

Introduction:

The presence of mitochondria in unicellular to mammalian organisms originates from an ancient symbiosis between primitive eukaryotic cells and free-living aerobic prokaryotes (1, 2). Mitochondria are crucial organelles for central cell functions (3), and they are the principal nutrient-up-taking and energy-producing cell organelle; they also take part in calcium signaling, ROS production, cell death and diverse cell signaling events (4–7). Mitochondria have retained many of their ancestral bacterial features including length, proteome, double membrane and circular genome (8).

Cell-derived mitochondrial components, including mitochondrial DNA, have been found in the extracellular space (9, 10). Those DNA fragments were found in the physiological circulating fluid of healthy subjects and patients with various diseases (11). Lately, mitochondrial cell-free DNA (McfDNA) has emerged as an attractive circulating biomarker due to its potential role in diagnostic applications in multiple diseases (e.g., Diabetes, acute myocardial infarction, cancer…) (12–14), and in physio-pathological conditions (e.g., trauma) (15). Despite the promising future of McfDNA in clinical applications, knowledge regarding its origin, composition and function is still lacking. In addition, the structure of McfDNA is currently unknown. In contrast, the structure of circulating DNA of nuclear origin is being characterized (16) and mono and di- nucleosomes and, to a lesser extent, transcription factors are found as stabilized cfDNA-associated structures in the blood-stream (17, 18). It is expected there are considerable configuration differences between nuclear and mitochondrial circulating DNA, because mitochondrial DNA is a small circular genome, without protective histones, and thus is more sensitive to degradation in the circulation. However, by revealing recently that there are approximately 50,000-fold more copies of the mitochondrial genome than the nuclear genome in the plasma of healthy individuals (19), we confirmed that McfDNA is sufficiently stable to be detected and quantified (12, 20), implying the presence of stable structures protecting these DNA molecules.

The present study aims at identifying the structures containing mitochondrial DNA in peripheral blood. By examining McfDNA integrity, and associated structure size and density, we revealed the presence of stable particles with full-length mitochondrial genomes. We characterized the structures by fluorescence and electron

microscopy, flow cytometry and we identified the presence of intact mitochondria in the circulation. Oxygen consumption assays suggested the functional viability of at least some of these extracellular mitochondria. Our work demonstrates for the first time the presence in blood of circulating cell-free respiratory competent mitochondria. Overall, in view of the potential roles of mitochondria in cell to cell communication, immune response and inflammation, our discovery has broad implications in homeostasis and disease, and paves the way for new paths towards the treatment and prevention of diseases.

Materials and Methods:

Plasma isolation:

Blood samples from healthy volunteers were provided by the "Etablissement Français du Sang (E.F.S)", the blood transfusion center of Montpellier, France (Convention EFS-PM N° 21PLER2015-0013). Blood samples from 50 mCRC patients were provided by a clinical study comparing the detection of KRAS exon 2 and BRAF V600E mutations by cfDNA analysis to conventional detection by tumor tissue analysis. All blood samples were processed within 4 h after collection. Plasma was extracted by various protocols, depending on the experiments.

<u>For mCRC patient:</u>

Blood was collected in EDTA tubes, and plasma was isolated by a single centrifugation, performed at 1,200g for 10 min at 4°C.

<u>For healthy individuals:</u>

- Plasma isolation using Ficoll:

Fresh blood was collected in EDTA tubes. Plasma was isolated by a Ficoll density-gradient centrifugation, performed at 400g for 30 min at 18°C (Ficoll® paque plus GE Healthcare, Fisher Scientific, Illkirch, France).

- Plasma isolation by centrifugation:

Fresh blood was collected in EDTA tubes. Plasma was isolated by a single centrifugation, performed at 1 200g for 10 min at 4°C (21).

- Plasma isolation without platelet activation:

Fresh blood was collected in a BD Vacutainer™ CTAD tubes (Ozyme, Montigny-le-Bretonneux, France). Plasma was isolated via incremental centrifugations, all performed for 10 min at room temperature without a break: two successive centrifugations were first performed at 200g and were followed by a third centrifugation at 300g. Preheated (37°C) ACD-A buffer (0.1M Trisodium citrate, 0.11M Glucose and 0.08 citric acid) and Prostaglandin E1 (1µM) (Sigma-Aldrich, St. Quentin Fallavier Cedex) were then added to the plasma, which was further centrifuged at 1 100g and finally at 2 500g.

Cell lines:

Human colon cancer cell lines (DLD-1/SW620) were obtained from the American type culture collection (ATCC) and a normal immortalized cell line (CCD-18Co) was obtained from Andrei Turtoi's laboratory (IRCM, Montpellier, France). SW620 and CCD-18Co cells were grown in RPMI 1640 and DLD-1 cells in DMEM (Gibco, Fisher Scientific, Illkirch, France), both supplemented with 10% fetal bovine serum (Eurobio, les Ulis, France) and 1X streptomycin/penicillin (Gibco, Fisher Scientific, Illkirch, France). Cell culture, for all cell lines, was performed at 37°C in 5% CO_2.

1.5 million cells were seeded in a T-75 flask with 10 ml of appropriate supplemented medium. Cells were incubated for either 24h or 60h. Culture media were replaced with fresh medium 24h before experiments. Collected media from cultured cells were centrifuged at 600g for 10min at 4°C, to precipitate both floating cells and cells debris, and were further processed by various protocols, depending on the experiments.

Quantification of NcfDNA and McfDNA in healhy individuals and cancer patients:

NcfDNA and McfDNA in healthy individuals (Figure 1A-C), and in mCRC patients (Figure S 1A-B) were quantified from a plasma supernatant obtained following the first centrifugation step mentioned before at 1 200g for 10 min at 4°C and a second centrifugation step at 16 000g for 10 min at 4°C. Following this centrifugation, total cfDNA was extracted from supernatant and analysed with q-PCR.

Kinetic study of cfDNA stability:

First, SW620 culture medium was removed, centrifuged at 1 200 g and then at 16 000g for 10 min, both at 4°C, and further incubated at 37°C in 5% CO_2 for 4 days. An aliquot of cell culture media was withdrawn every day for DNA extraction. NcfDNA and McfDNA were quantified by q-PCR using specific primers (MIT MT-CO3 F and MIT MT-CO3 R2) (Table S 2).

cfDNA extraction:

CfDNA was extracted with a Qiagen Blood Mini Kit (Qiagen, Courtaboeuf, France) according to the manufacturer's recommendations, except that extraction was performed with 1mL of plasma sequentially loaded on a single column and that the cfDNA was eluted with 130µL of elution buffer. The cfDNA was stored at −20°C for further analysis. Freeze-thawing was avoided to reduce cfDNA fragmentation.

Q-PCR analysis:

CfDNA was quantified by q-PCR according to an innovative design of short (60–100 bp ± 10bp) and long (300bp ± 10bp) amplicons targeting the wild-type sequences of specific genes: the KRAS nuclear gene and the mitochondrial Cytochrome oxidase III gene, MT-CO3 (Table S 2). Quantification of the short and long amplicons provides an estimation of the concentrations of the total nuclear circulating free DNA (NcfDNA) and the mitochondrial circulating free DNA (McfDNA) respectively. Q-PCR amplifications were performed in triplicate in a 25µL final reaction volume controlled by the CFX manager software of a CFX96 touch™ Real-Time PCR detection system (Bio-Rad). Each PCR reaction mixture was composed of 12.5µL of SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad, Marnes-la-Coquette, France), 2.5µL of free water (Qiagen), 2.5µL of forward and 2.5µL reverse primers (3 pmol/µL) and 5µL of template. Thermal cycling conditions were as follows: 95°C (3:00) + [95°C (0:10) + 60°C (0:30)] × 40 cycles. Melting curves were investigated by increasing the temperature from 60°C to 90°C, reading the plate every 0.2°C. Each q-PCR run was performed with 1.8ng/µL of genomic

DNA extracted from the DiFi cell line (ATCC) for the standard curve and without DNA as the control condition. Q-PCR amplification was validated by melt curve differentiation.

CfDNA calibration assay and copy number calculation:

NcfDNA quantification: A genomic DNA extract from human wild-type KRAS colorectal cells was used for the NcfDNA calibration assay. Initial genomic DNA solution concentration and purity were determined by measuring optic density at λ=260 nm, 230 nm and 280 nm, with an Eppendorf BioPhotometer® D30. Starting genomic DNA concentration was adjusted to 1800pg/µl for the first dilution point, according to optical density measurement at λ=260 nm. A q-PCR standard curve was obtained by 6 successive dilutions of the vector solution (to 1800, 180, 45, 20, 10 and 5pg/µl). The standard curve was used to determine the NcfDNA concentration per milliliter of plasma and cell media supernatant. The NcfDNA copy number was calculated as follows:

$$Qnuclear = \left(\frac{c}{3{,}3}\right) * \left(\frac{Vélution}{Vplasma}\right)$$

$Q_{nuclear}$ is the NcfDNA copy number per milliliter; c is the NcfDNA mass concentration (pg/µl) determined by a q-PCR targeting the nuclear *KRAS* gene sequence and 3.3pg is the human haploid genome mass. $V_{elution}$ is the volume of cfDNA extract (µl) and V_{plasma} is the starting volume of plasma used for the extraction (ml).

McfDNA quantification: A 3382-pb human ORF vector with a 786-pb MT-CO3 insert was obtained from ABM (accession no.YP_003024032) and used for the McfDNA calibration assay.

Initial vector solution testing, starting concentration adjustments and q-PCR standard curves were performed as for NcfDNA above.

The standard curve was used to determine the McfDNA concentration per milliliter of plasma and of cell media supernatant. The McfDNA copy number was calculated as follows:

$$Qmito = \left(\frac{c * Na}{2 * MW * Lvector}\right) * \left(\frac{Vélution}{Vplasma}\right)$$

$Q_{mitochondrial}$ is the McfDNA copy number per milliliter, 'c' is the McfDNA mass concentration (g/µl) determined by q-PCR targeting the mitochondrial *MT-CO3* gene sequence. N_A is Avogadro's number ($6.02*10^{23}$ molecules

per mole), L_{vector} is the plasmid length (nucleotides) and MW is the molecular weight of one nucleotide (g/mol); $V_{elution}$ is the elution volume of cfDNA extract (µl) and V_{plasma} is the starting volume of plasma used for the extraction (ml).

DNA Integrity Index calculation:

The degree of cfDNA fragmentation was assessed simultaneously by targeting *KRAS* and MT-CO3 sequences from each plasma DNA sample by calculating the DII (DNA integrity index). The DII was determined by calculating the ratio of the concentration determined using the primer set amplifying a large target (KRAS F/R2 and MT-CO3 F/R2) to the concentration determined using the primer set amplifying a short target (KRAS F/R1 and MT-CO3 F/R1)(Table S 2).

Examination of the percentage of McfDNA short and long fragments:

CfDNA was extracted from a pool of 80 healthy individual's plasma, with the Maxwell RSC ccfDNA plasma kit (Promega, Charbonnières-les-Bains, France). To obtain highly concentrated cfDNA, extracts were subjected to a second extraction with the same method followed by a Qiamp DNA blood Mini Kit extraction resulting in a final volume of 30µL. The samples were then electrophoresed on a 2% agarose gel and DNA fractions of short (<500bp) and long (>500bp) fragments were extracted from the gel with a QIAquick Gel extraction Kit (Qiagen). The quantity of mitochondrial DNA fragments was then assessed with q-PCR by using mitochondrial specific primer (MT-CO3 F/R1) (Table S 2).

Library preparation for whole genome sequencing:

Dual-indexed single-stranded libraries were prepared from 1-11 ng human DNA using TL137 as a linker oligo (22). To allow sequencing with single-stranded libraries, custom double-stranded adapters were generated, identical to those used in the single-stranded method, by annealing a short oligo SLP4 (23). Ligations were performed with an NEBNext Quick Ligation Module kit (New England Biolabs) with 0.04µM of each annealed

adapter in a 50µl total volume, incubated for 30min at 20°C. Adapters were elongated by adding 50µl OneTaq 1x Master Mix (New England Biolabs) to the ligation mixture and incubating for 20min at 60°C. All oligos were purchased from Eurogentec (Kaneka Eurogentec, Seraing, Belgium). Single-stranded libraries were quantified via qPCR, with a diluted fraction, in a LightCycler 2 (Roche Applied Science, Mannheim, Germany). Samples were then amplified using Taq polymerase and an optimal number of cycles for each sample to avoid plateau phase, as calculated from the Ct of the diluted qPCR. Amplified samples were then purified by two rounds of Macherey-Nagel NGS beads at 1.3x volume, to retain shorter inserts, and eluted in 30µL EBT. Purified libraries were then visualized on a Bioanalyzer 2100 (Agilent, Santa Clara, CA) and quantified via qPCR, Qubit 2 Fluorometer (Life Technologies, Grand Island, NY), and Bioanalyzer 2100. Libraries were then pooled in equimolar amounts and sequenced on an Illumina MiSeq platform using two MiSeq v3 150 (2x75) kits, with sequencing primer CL72 replacing the first read sequencing primer.

Sequence analysis:

Adapter sequences were removed with cutadapt 1.3 (24) and reads were aligned to the human genome reference (hg19) with BWA aln (25) using the default parameters and filtered for mapping quality 20 with SAMtools 1.5 (26). Duplicate removal and histogram generation was performed with MarkDuplicates and CollectInsertSizeMetrics, respectively, from Picard tools 1.88.

Differential centrifugation and filtration of plasma and cell media:

Schematic view and details are presented in Figure S 5.

Isolation of intracellular mitochondria:

Mitochondria from cultured cells were isolated as a positive control. Cells were scraped from culture dishes, washed with PBS and centrifuged at 1,300 rpm for 5 min at room temperature. Pelleted cells were resuspended in 1 mL of MIB (0.2M sucrose, 0.01M tris, 0.001M EGTA, 1X protease inhibitor) and gently lysed using an IKA

T18 Basic Dispersers Homogenizer (Ultraturrax) at speed 2 for 10 s until obtaining 80 to 90% intact nuclei. Lysed cells were then centrifuged at 600g at 4°C for 10 min to remove cell debris and nuclei, to recover supernatant containing both cytoplasm and mitochondria. Pelleted intact nuclei were washed twice in 1ml MIB and centrifuged again at 600*g* for 10min at 4°C, to be used as a negative control in experiments with flow cytometry, while the resulting supernatants were pooled together with the previous supernatants to further isolate mitochondria. For this, the 3ml of supernatant were centrifuged at 600*g* at 4°C for 10 min to eliminate contaminating nuclei and cell debris. The resulting supernatant was then further centrifuged at 8,000*g* for 10 min at 4°C to pellet mitochondria. The supernatant was collected again and centrifuged again at 8,000*g* for 10 min at 4°C. These pellets were pooled together and gently resuspended in 500uL MIB, transferred to 1.5mL tubes and centrifuged again at 8,000*g* for 10 min at 4°C. The final supernatant was carefully discarded to collect the pelleted mitochondria.

Isolation of extracellular mitochondria:

Mitochondria in the plasma extracted by Ficoll gradient, were isolated by sequential centrifugations at 600g, then 1,200g and finally 2,000g for 10 min at 4°C to remove any contaminating blood cells or platelets. The resulting supernatant was centrifuged at 16,000g to collect the extracellular mitochondria pellet.

Mitochondria in the plasma obtained without platelet activation were pelleted by a one-step centrifugation at 16,000g for 10 min at 4°C.

Mitochondria in the cell media supernatant were isolated by sequential centrifugations at 600 g, then 1,200 g for 10 min at 4°C to remove any contaminating cells. The pellet of extracellular mitochondria was collected from a subsequent centrifugation at 16 000g.

Note that an extracellular mitochondrial pellet was also recovered from a centrifugation at 8,000g instead of 16,000g to protect the mitochondrial membrane.

Amplification of the mitochondrial genome:

The DNA extracted from cell media pellets (40 flask T-75/cell line), and from two different plasma pools (without platelet activation) was selectively amplified with the repli g mitochondrial DNA kit (Qiagen), according to the manufacturer's recommendations. The amplified mitochondrial genome was then amplified by long range PCR performed in a 50µL total volume in a Mastercycle® nexus eco thermal cycler (Eppendorf). Each PCR reaction mixture was composed of 30.5µL of free water (Qiagen), 5µL of 10× LA PCR Buffer II (Mg^{2+} plus), 8µL of dNTPs (2.5mM each), 5µL of mixed forward and reverse primers (10µM each), 1µL of template (DNA mass between 10 and 100ng, calculated with a Qubit broad range kit) and 0.5µl of TAKARA LA Taq (5 U/µl) (Ozyme, Saint Quentin Yvelines, France). Thermal cycling conditions were as follows: 95 °C (2:00) + [95 °C (0:15) + 68 °C (10:00)] × 30 cycles + 68 °C (20:00) + 4 °C (∞).PCR amplifications were performed with five pairs of overlapping primers (Mito1/Mito2/Mito3/hmt1/hmt2) (Integrated DNA Technologies, Leuven, Belgium). All PCR amplified products were loaded on a 0.8% agarose gel to check the size of amplicons with reference to the 1Kb plus DNA ladder (Thermo Fischer).

Mitochondrion staining:

The 16,000g pellets from both, the plasma without platelet activation (pool from 5 healthy individuals) and from the cell media (2 flask T-75/ cell line), were stained with 200nM MitoTracker® Green FM (Thermofisher) for 45 min, washed twice with PBS, and resuspended in 10µl of PBS to be visualized by ApoTome microscopy (ZEISS Axio Imager 2). Images were edited with ZEN Software. In parallel, cell culture media pellets (8 T-75 flask/cell line), Ficoll-isolated plasma pellet and positive and negative controls were resuspended in 150 µl of PBS and analyzed by Gallios flow cytometry (Beckman Coulter). Mitochondria isolated from cultured cells were used as a positive control for the MitoTracker® specificity and as a standard to delineate both the appropriate gate and voltages for the flow cytometry. All sample plots were collected according to size (Forward scatter FSC) and granularity (side scatter SSC) in logarithmic mode. Voltages were adjusted for submicron particles and the flow rate was set at 'low'. Doublets were excluded by plotting FSC-Area versus FSC-H, both in logarithmic mode.

Collected events in the newly created gate were gated as singlet. Following confirmation that all events recorded in this gate were MitoTracker Green positive, the gate was applied to all samples. Plot histograms were created for both unstained and stained samples. Results were analyzed with Kaluza analysis 1.5 software.

Western Blot:

Pellets obtained at 8,000g from both the plasma (pool of 5 healthy individuals) and the cell media supernatant (100 flask t-75/cell line) were resuspended in 50µl of 1X Laemli buffer, loaded on a 12% Acrylamide/Bis-Acrylamide gel and transferred onto nitrocellulose membranes (GE Healthcare). Membranes were blocked in 10% milk in 1X PBS-tween for 1h. Blotting was performed with either 1/500 diluted primary mouse anti-TOM22 (Sigma , St. Quentin Fallavier Cedex) or with 1/1,000 diluted primary mouse anti-TIM23 (BD Transduction Laboratorie, France) antibodies over-night at 4°C and further with 1/10,000 diluted HRP-conjugated secondary Rabbit anti-mouse antibodies (Merck, Île-de-France, France) for 1h at room temperature. Immunoblots were visualized by ECL (Perkin Elmer, Villebon sur Yvette, France).

Electron microscopy:

8 000g pellets from either cell media (30 flask T-75/cell line) or plasma prepared without platelet activation (pool from 3 healthy individuals) were immersed in a solution of 2.5% glutaraldehyde in PHEM buffer (1X, pH 7.4) overnight at 4°C, washed in PHEM and post-fixed in 0.5% osmic acid for 2h in the dark at room temperature. Samples were then washed twice in PHEM buffer, dehydrated in a graded series of 30 to 100% ethanol solutions and finally embedded in EmBed 812 using an Automated Microwave Tissue Processor for Electronic Microscopy (Leica EM AMW). 70nm sections (Leica-Reichert Ultracut E), collected at different levels of each block, were counterstained, with 1.5% uranyl acetate in 70% Ethanol and lead citrate, and observed with a Tecnai F20 transmission electron microscope at 200KV.

Metabolic activity:

Mitochondrial bioenergetics function was determined by oxygen consumption rate OCR, with a Seahorse XF-96 extracellular flux analyzer (Agilent). We performed the electron flow assay that allows the functional assessment of selected mitochondrial complexes together in the same period. After isolation, 64 µg of freshly isolated DLD-1 mitochondria (positive control), and 128 µg of DLD1 cell media pellet and plasma pool pellet prepared without platelets activation, were plated in each well in a volume of 25 µL containing MAS 1X (70mM sucrose, 220mM mannitol, 10mM KH_2PO_4, 5mM $MgCl_2$, 2mM HEPES, 1mM EGTA and 0.2% fatty acid-free BSA, pH 7.2) supplemented with 10mM pyruvate, 2mM malate and 4µM FCCP. Note, pyruvate and malate drive respiration via complex I, and FCCP uncouples the mitochondrial function. The XF plate was then centrifuged for 20 min at 2 000g at 4°C. After centrifugation, 155 µL of electron flow substrate-containing 1xMAS was added to each well, and the plate was warmed up in a 37°C non-CO_2 incubator for 5-10 min. 10-fold concentrated compounds were loaded in the ports of the cartridge, using the "loading helper" plate: port A, 50µL of 20µM rotenone, a complex I inhibitor (2µM final): port B, 55µL of 100mM succinate, a complex II activator (10mM final); port C, 60µL of 40µM Antimycin A, a complex III inhibitor (4µM final); port D, 65µL of 100mM ascorbate plus 1mM TMPD, complex IV activator (10mM and 100µM final, respectively). After calibration of the cartridge by the XF machine, the XF plate was introduced and the assay continued using Agilent's protocol for isolated mitochondria.

Statistical analysis:

Statistical analysis was performed with GraphPad Prism software (version 7.01). A probability of ≤ 0.05 was considered to be statistically significant by Student's t-test: * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

Results:

Unlike nuclear cell free DNA, mitochondrial cell free DNA is stable in plasma and cell culture media.

McfDNA is sufficiently stable to be detected and quantified in plasma (Fig 1A; Fig S 1A). However, mitochondrial DNA (16 kilo base pairs (kbp)) was thought to be more susceptible to degradation and less stable than nuclear DNA because of its lack of histones. In our initial investigation, we found, to our surprise, that mitochondrial cell-free DNA is actually more stable than nuclear cell-free DNA, in fetal bovine serum supplemented culture medium (Fig 1B). To establish mitochondrial genome cell-free DNA integrity, we determined the DII (DNA integrity index) we previously used for the study of NcfDNA fragmentation (27), and data revealed that all or most McfDNA fragments are over 300 bp (Fig 1C; Fig S 1B). We next examined a plasma DNA extract pool, from eighty healthy humans, by agarose gel electrophoresis and q-PCR. Approximately 10% only, of the McfDNA mass in healthy samples was in fragments below 500 base pairs, demonstrating that this DNA is not as fragmented as NcfDNA (Fig 1D).

Whole Genome Sequencing of healthy plasma revealed a wide-ranging population of short mitochondrial DNA fragments mainly decreasing in abundance from 30 to 300 nucleotides. In contrast, NcfDNA had a more homogenous distribution, conventionally peaking at 167 base pairs, corresponding to mono-nucleosomes (Fig 1E) (17, 28). Paired-end sequencing enables high-resolution measurement of DNA size but has a practical upper limit of read-out at ~1,000bp. Since no McfDNA was detected between 500bp and 1,000bp, Whole Genome Sequencing only detected the minor fraction formerly revealed by gel electrophoresis (Fig 1D). Overall, these data suggest that a large proportion of the McfDNA fragments are over 1,000bp, and that the minor fraction of McfDNA fragments below 300bp is highly degraded. We therefore set out to characterize the structural elements that stabilize mitochondrial DNA in the circulation.

Plasma and cell culture media contain McfDNA in dense structures larger than 0.22 µm.

We next used differential centrifugation of plasma to separate the different structures associated with cell-free nuclear and mitochondrial DNAs. First, we isolated the plasma from whole blood at 400g on a ficoll gradient.

This was followed by incremental centrifugations at 16, 40 and 200 thousand g to remove the large organelles and remaining membrane debris, the micro-vesicles and finally the exosomes, respectively (Fig S 5). Q-PCR of the resulting supernatant revealed, surprisingly, that McfDNA is packed in denser structures than NcfDNA (Fig 2A,C).

We therefore sought to further confirm our surprising result, that 16,000g centrifugation precipitated large amounts of mitochondrial DNA but very little nuclear DNA, by performing a refined plasma isolation protocol that precludes platelet activation, based on sequential centrifugations at increasing strengths: 2 x 200g, 300g, 1 100g, 2 500g and 16 000g, each for 10min at 20°C with no break in-between. Note, this protocol excludes blood cells from the final plasma preparation without activating or degrading them in the process (29, 30) (Fig 2A) and this sequence (series) of centrifugations was used in all the following work described here. Notably, in the final plasma preparation, the nuclear DNA passed through a 0.2µm filter whereas the mitochondrial DNA did not (Fig 2A,D,E; Fig S 2A-C). No statistically significant variation in NcfDNA content occurs, in contrast to McfDNA contents showing up to 7-fold difference when using these various procedures (Fig 2D-E); this demonstrated that mitochondrial DNA detected in plasma are included in cell-free structures and that plasma preparation is devoid of nucleated blood cells.

Next, we sought to completely eliminate the possibility that mitochondria or free mitochondrial DNA is actively released into the plasma by platelets, during handling of the blood samples. Therefore, we repeated all the above analysis on culture media supernatants from various cancer and normal cell lines, DLD-1, SW620 and CCD-18CO, and again here the McfDNA sedimented at 16,000g (Fig 2B; Fig S 3A-C), and was retained on 0.22 filters (Fig S 2D-E). 73% ± 11% and 78% ± 15% of the McfDNA precipitated into the pellet from the cell culture supernatant and plasma, respectively (Fig 2C; Fig S 3C; Table S 1). It has to be noted that mitochondria are between 0.4 and 1.5 microns in diameter and the conventional speed used for their sedimentation is between 7,000 and 20,000g. Mitochondrial DNA in the 16,000g pellet of cell culture media was not degraded by DNase I treatment, in contrast to nuclear DNA (Fig S 3D). Overall, these experiments demonstrate that plasma and cell culture media contain McfDNA in dense and stable structures above 0.22 µm, excluding platelets as its source.

Both plasma and cell culture media contain structures with full-length mitochondrial genomes.

We next investigated the integrity of the mitochondrial DNA in the 16,000g pellets of two different plasma pools and in DLD-1 and SW620 culture media. After Highly Uniform Whole Mitochondrial Genome Amplification of the DNA extracts, we used long-range PCR with five mitochondria-specific primer sets (Fig 3A) (31). While no signal was detected in the negative controls, agarose gel electrophoresis sizing clearly revealed the expected mitochondrial amplicons from the cell media and plasma 16,000g pellets, thus demonstrating the presence of intact full-length mitochondrial DNA in cell culture media and in healthy individuals plasma (Fig 3B), consistent with the high integrity index of plasma mitochondrial DNA in Fig 1C.

Both plasma and cell culture media contain structurally intact mitochondria:

To test for the presence of structurally intact mitochondria or material derived from mitochondria in plasma, we used, the specific live mitochondrial marker, Mitotracker green, on the 16,000g pellet derived from cell cultures media and on the plasma of five healthy individuals. Positively labeled mitochondria were specifically observed by Fluorescence microscopy in all biological sources (Fig 5A-C). In addition, Flow cytometry analysis was performed on the Mitotracker green-stained preparations. Nuclei (Fig 4B, Fig S 4B) and mitochondria (Fig 4A, Fig S 4A), purified from the DLD-1 and SW620 cells, were included as negative and positive control respectively. 16,000g pellet from fresh cell media was also used as a negative control (Fig 4C, Fig S 4C). Similarly to the positive control, fluorescence signals were detected in the 16,000g pellet of both cell culture media (Fig 4D-E; Fig S 4D-E) and plasma (Fig 4F; Fig S 4F-I), indicating the presence of mitochondria. In addition, immunoblotting revealed the presence of the outer and inner membrane mitochondrial transport proteins TOM 22 and TIM 23 in the 16,000g pellets from the DLD-1, SW620 and CCD-18Co cell culture media and the plasma supernatants (Fig 4G-H).

To validate the presence of free intact mitochondria, we examined the pellet obtained from the cell culture media of DLD-1 and SW620, and from isolated plasma without platelet activation, by electron microscopy. Structures

with a double membrane and a similar size to mitochondria were detected in all the samples (Fig 5D-F). The 16,000g pellets contained structures with all the morphological characteristics of mitochondria, including the inner and outer membranes and the corresponding intermembrane space and matrix. It has to be noted that not all these structures match perfectly the conventional morphology of mitochondria inside a cell. This is expected due to the various morphological alterations that mitochondria may undergo especially in different pathological conditions (32).

Both plasma and cell culture media contain respiratory-competent mitochondria

To determine whether these mitochondria are functional, we evaluated the metabolic activity by measuring oxygen consumption rate (OCR) with the Seahorse XF extracellular flux analyzer technology (Agilent). The electron flow assay clearly indicated that the pellets isolated from DLD-1 culture media, as well as from plasma pool of healthy individuals, consume oxygen and are sensitive to complex I inhibition by rotenone, suggesting that the extracellular mitochondria remain competent for respiration (Fig 6A-B).

Discussion:

To sum up, we report here that blood contains intact cell-free full-length mitochondrial DNA in dense and biologically stable structures over 0.22µm in diameter and that these structures have specific mitochondrial proteins, double membranes and a morphology resembling that of mitochondria. We further demonstrate that these structurally intact cell-free mitochondria in the blood circulation are respiratory competent. While several studies reported extracellular mitochondria in specific conditions resulting in platelet activation (33), or encapsulated within microvesicles (34), it is remarkable how the presence of intact cell-free mitochondria was unnoticed in normal physiological state. This could be explained by the high dilution of these organelles in the plasma and cell culture media. Note, our finding first corroborate previous collateral observations of plasma filtrates suggesting existence of particles containing mitochondrial DNA in the circulation (10). Intact full mitochondrial genomes were also observed in the cell free DNA fraction of both healthy individuals and patients

with mitochondrial disease. However it was assumed that the circular nature of mitochondrial DNA delays its degradation by circulating nucleases, so the presence of intact mitochondria was not suspected and the structural characteristics associated with McfDNA were not investigated (35).

We estimate that there are between 200,000 and 3.7 million cell-free intact mitochondria per mL of plasma, based on the McfDNA copy number that was either pelleted at 16,000g or filtered (Table S 3). We believe that circulating cell-free intact mitochondria have crucial biological and physiological roles because mitochondria are already known as systemic messengers in cell-cell communication by transferring hereditary and non-hereditary constituents. Mitochondria were recently discovered to translocate from one cell to the other (36). Indeed intercellular mitochondria trafficking was observed *in vitro* and *in vivo*, both in physiological and pathophysiological conditions including tissue injury and cancer (37). In particular, the mitochondria transfer between differentiated and mesenchymal stem cells was observed in cardiomyocytes and endothelial cells as a means to rescue injured tissues (38). Similarly, mitochondria are able to be internalized by different types of cells, by macro pinocytosis which is an endocytic pathway specific for large vesicles (39). Clinical transplantation of mitochondria between cells is an active area of research but the specific mechanisms and critical factors involved in natural mitochondria transfer between the donor cells and recipient cells are remain to be fully characterized (40).

All cell-free mitochondria in human plasma or cell culture media supernatant as observed by EM were not surrounded by a bi- or multi-layer phospholipid membrane. A few studies suggested that mitochondrial DNA could be encapsulated in microvesicles, and act as efficient messengers in many biological systems (34). Cells may communicate with each other by secreting extracellular vesicles, such as exosomes that are secreted by the endosomal Traffic System, micro-vesicles that are released directly from the plasma membrane, or apoptotic bodies. Extracellular vesicles may play an important role in promoting inflammation and neutrophil infiltration in a variety of diseases (41). One can speculate that the previously described biological effect of cell-free

mitochondrial DNA enriched micro-particles could, as well, be performed by cell-free intact mitochondria as their presence in blood was not known before the study of Boudreau et al on activated platelet preparation (33).

Like extracellular vesicles then, cell-free intact mitochondria could contribute to genetic material exchange between cells, either directly or via body fluids. This supports the notion that intercellular exchange of genetic information is biologically significant in mammals (42). Mitochondria structures could thus be a stable potential delivery vehicle for intercellular exchange of bioactive molecules. We therefore think it is plausible that mitochondrial clearance and degradation occurs principally by phagocytosis. Since this study is the first showing free structurally intact mitochondria in plasma, it is not yet known how they are degraded there, in contrast to intracellular mitophagy, which is well described (43). Presumably, mitochondria are degraded in plasma, and their contents are released in the blood stream.

We hypothesize that the minor fraction of short cell-free mitochondrial DNA fragments, we observed here, are packaged in TFAM-complex (Mitochondrial transcription factor A) nucleoids (44). TFAM-complex nucleoids may originate from mitochondria or extra-cellular vesicles degradation in blood and might have a biological role. An increasing number of studies have revealed that elevated plasma mitochondrial DNA concentration is associated with damage associated molecular patterns (DAMPs) across traumatic, infectious, and neurodegenerative diseases like Alzheimer's, and in intensive care unit patients (45, 46). Cell-free mitochondrial DNA can stimulate immune cells and induce inflammatory activation through Toll-like receptors, NLRP3 and the cGAS-STING signal pathway (47, 48). Mitochondrial DNA is involved in host immune responses and inflammatory activation and cell-free mitochondria may be involved in communication networks within and between cells (36–40, 42). Finally, our observation should be placed in the context of the endosymbiotic origin of mitochondria (49) and the on-going natural transfer of mitochondrial DNA into the nuclear genome of eukaryotic cells (50).

Our work demonstrates for the first time the presence of a new constituent of the largest and visceral organ in the body, the blood. This component of the blood that delivers our cellular and organismal energy needs is none other than the quintessential ancient symbiotic energy-providing component of all eukaryotic cells, the mitochondrion. This new discovery has broad implications in homeostasis and disease; it could open up a considerable field of investigation to better understand the mechanisms governing the newly demonstrated intercellular transfer of biological material, including mitochondria between various types of donor and host cells, in physiological and pathological processes. Our data thus represent a quantum leap since cell-free intact mitochondria might have a wider role as a novel systemic messenger. In addition, circulating mitochondria could be considered as a non-invasive source of molecular biomarkers for early detection, diagnosis, prognosis and monitoring of various diseases.

Acknowledgments:

The authors would like to thank P. Blache, T. Salehzada and C. Jay-Allemand for helpful comments. We are grateful to J. Venables (Science Sense) for help in editing the manuscript. We thank Beatrice Chabi (INRA, UMR 0866 DMEM Dynamique Musculaire et Métabolisme, Centre de recherche de Montpellier, Montpellier, France) for her contribution to the oxygraph experiment. We are grateful to the MRI platform (Montpellier Bio campus) for access to imaging facilities, and to METAMONTP platform for access to seahorse technology. We thank A. Turtoi for providing us CCD-18Co cell line and reagents. ART is supported by the INSERM (Institut National de la Santé et de la Recherche Médicale).

Funding: The work is supported by the SIRIC Montpellier Grant (INCa-DGOS-Inserm 6045), France.

Author contributions: ZAAD, CP and ART designed the study, developed the methodology, analyzed the data and prepared the manuscript. ZAAD, AO, RT, BP, RM, CS, GA, LL, AB, TG and SEM realized the experiments.

All of the authors (ZAAD, AO, RT, BP, RM, CS, GA, LL, AB, TG, SEM, TM, CP and ART) discussed the results and approved the manuscript.

Conflicts of interest: The authors declare no competing financial interests.

Materials and correspondence: The published article includes all data generated or analyzed during this study. Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact: alain.thierry@inserm.fr.

References:

1. Roberts, R. G. (2017) Mitochondria—A billion years of cohabitation. *PLOS Biology* **15**, e2002338
2. Lane, N. and Martin, W. (2010) The energetics of genome complexity. *Nature* **467**, 929–934
3. Nunnari, J. and Suomalainen, A. (2012) Mitochondria: In Sickness and in Health. *Cell* **148**, 1145–1159
4. Rizzuto, R. Mitochondria as sensors and regulators of calcium signalling. 13
5. Wang, C. and Youle, R. J. (2009) The Role of Mitochondria in Apoptosis. *Annual Review of Genetics* **43**, 95–118
6. Tait, S. W. G. and Green, D. R. (2012) Mitochondria and cell signalling. *Journal of Cell Science* **125**, 807–815
7. Murphy, E., Ardehali, H., Balaban, R. S., DiLisa, F., Dorn, G. W., Kitsis, R. N., Otsu, K., Ping, P., Rizzuto, R., Sack, M. N., Wallace, D., and Youle, R. J. (2016) Mitochondrial Function, Biology, and Role in Disease: A Scientific Statement From the American Heart Association. *Circulation Research* **118**, 1960–1991
8. Friedman, J. R. and Nunnari, J. (2014) Mitochondrial form and function. 21
9. Thierry, A. R., El Messaoudi, S., Gahan, P. B., Anker, P., and Stroun, M. (2016) Origins, structures, and functions of circulating DNA in oncology. *Cancer and Metastasis Reviews* **35**, 347–376
10. Chiu, R. W. K., Chan, L. Y. S., Lam, N. Y. L., Tsui, N. B. Y., Ng, E. K. O., Rainer, T. H., and Lo, Y. M. D. (2003) Quantitative Analysis of Circulating Mitochondrial DNA in Plasma. *Clinical Chemistry* **49**, 719–726
11. Yu, M. (2012) Circulating cell-free mitochondrial DNA as a novel cancer biomarker: opportunities and challenges. *Mitochondrial DNA* **23**, 329–332
12. Kohler, C., Radpour, R., Barekati, Z., Asadollahi, R., Bitzer, J., Wight, E., Bürki, N., Diesch, C., Holzgreve, W., and Zhong, X. (2009) Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors. *Molecular Cancer* **8**, 105
13. Malik, A. N., Parsade, C. K., Ajaz, S., Crosby-Nwaobi, R., Gnudi, L., Czajka, A., and Sivaprasad, S. (2015) Altered circulating mitochondrial DNA and increased inflammation in patients with diabetic retinopathy. *Diabetes Research and Clinical Practice* **110**, 257–265
14. Sudakov, N. P., Popkova, T. P., Katyshev, A. I., Goldberg, O. A., Nikiforov, S. B., Pushkarev, B. G., Klimenkov, I. V., Lepekhova, S. A., Apartsin, K. A., Nevinsky, G. A., and Konstantinov, Yu. M. (2015) Level of blood cell-free circulating mitochondrial DNA as a novel biomarker of acute myocardial ischemia. *Biochemistry (Moscow)* **80**, 1387–1392

15. Zhang, Q., Itagaki, K., and Hauser, C. J. (2010) Mitochondrial DNA is released by shock and activates neutrophils via p38 MAP kinase. *Shock* **34**, 55
16. Otandault, A., Anker, P., Al Amir Dache, Z., Guillaumon, V., Meddeb, R., Pastor, B., Pisareva, E., Sanchez, C., Tanos, R., Tousch, G., Schwarzenbach, H., and Thierry, A. R. (2019) Recent advances in circulating nucleic acids in oncology. *Annals of Oncology*
17. Sanchez, C., Snyder, M. W., Tanos, R., Shendure, J., and Thierry, A. R. (2018) New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA analysis. *npj Genomic Medicine* **3**
18. Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M., and Shendure, J. (2016) Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**, 57–68
19. Meddeb, R., Al Amir Dache, Z., Thezenas, S., Otandault, A., Tanos, R., Pastor, B., Sanchez, C., Azzi, J., Tousch, G., Azan, S., Mollevi, C., Adenis, A., Messaoudi, S. E., Blache, P., and Thierry, A. R. (2019) Quantifying circulating cell-free DNA in humans. *Scientific Reports* **9**, 5220
20. Fliss, M. S. (2000) Facile Detection of Mitochondrial DNA Mutations in Tumors and Bodily Fluids. *Science* **287**, 2017–2019
21. Meddeb, R., Pisareva, E., and Thierry, A. R. (2019) Guidelines for the Preanalytical Conditions for Analyzing Circulating Cell-Free DNA. *Clinical Chemistry* clinchem.2018.298323
22. Gansauge, M.-T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., Riehl, L. M., Schmidt, A., and Meyer, M. (2017) Single-stranded DNA library preparation from highly degraded DNA using *T4* DNA ligase. *Nucleic Acids Research* gkx033
23. Bennett, E. A., Massilani, D., Lizzo, G., Daligault, J., Geigl, E.-M., and Grange, T. (2014) Library construction for ancient genomics: Single strand or double strand? *BioTechniques* **56**
24. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12
25. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760
26. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
27. Mouliere, F., El Messaoudi, S., Pang, D., Dritschilo, A., and Thierry, A. R. (2014) Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Molecular Oncology* **8**, 927–941
28. Zhang, R., Nakahira, K., Guo, X., Choi, A. M. K., and Gu, Z. (2016) Very Short Mitochondrial DNA Fragments and Heteroplasmy in Human Plasma. *Scientific Reports* **6**
29. Rigothier, C., Daculsi, R., Lepreux, S., Auguste, P., Villeneuve, J., Dewitte, A., Doudnikoff, E., Saleem, M., Bourget, C., Combe, C., and Ripoche, J. (2016) CD154 Induces Matrix Metalloproteinase-9 Secretion in Human Podocytes. *Journal of Cellular Biochemistry* **117**, 2737–2747
30. Coumans, F. A. W., Brisson, A. R., Buzas, E. I., Dignat-George, F., Drees, E. E. E., El-Andaloussi, S., Emanueli, C., Gasecka, A., Hendrix, A., Hill, A. F., Lacroix, R., Lee, Y., van Leeuwen, T. G., Mackman, N., Mäger, I., Nolan, J. P., van der Pol, E., Pegtel, D. M., Sahoo, S., Siljander, P. R. M., Sturk, G., de Wever, O., and Nieuwland, R. (2017) Methodological Guidelines to Study Extracellular Vesicles. *Circulation Research* **120**, 1632–1648
31. Dames, S., Eilbeck, K., and Mao, R. (2015) A High-Throughput Next-Generation Sequencing Assay for the Mitochondrial Genome. In *Mitochondrial Medicine* (Weissig, V. and Edeas, M., eds) vols. I, Probing Mitochondrial Function, pp. 77–88, New York, NY
32. Wiemerslage, L., Ismael, S., and Lee, D. (2016) Early alterations of mitochondrial morphology in dopaminergic neurons from Parkinson's disease-like pathology and time-dependent neuroprotection with D2 receptor activation. *Mitochondrion* **30**, 138–147
33. Boudreau, L. H., Duchez, A.-C., Cloutier, N., Soulet, D., Martin, N., Bollinger, J., Pare, A., Rousseau, M., Naika, G. S., Levesque, T., Laflamme, C., Marcoux, G., Lambeau, G., Farndale, R. W., Pouliot, M.,

Hamzeh-Cognasse, H., Cognasse, F., Garraud, O., Nigrovic, P. A., Guderley, H., Lacroix, S., Thibault, L., Semple, J. W., Gelb, M. H., and Boilard, E. (2014) Platelets release mitochondria serving as substrate for bactericidal group IIA-secreted phospholipase A2 to promote inflammation. *Blood* **124**, 2173–2183

34. Sansone, P., Savini, C., Kurelac, I., Chang, Q., Amato, L. B., Strillacci, A., Stepanova, A., Iommarini, L., Mastroleo, C., Daly, L., Galkin, A., Thakur, B. K., Soplop, N., Uryu, K., Hoshino, A., Norton, L., Bonafé, M., Cricca, M., Gasparre, G., Lyden, D., and Bromberg, J. (2017) Packaging and transfer of mitochondrial DNA via exosomes regulate escape from dormancy in hormonal therapy-resistant breast cancer. *PNAS* **114**, E9066–E9075
35. Newell, C., Hume, S., Greenway, S. C., Podemski, L., Shearer, J., and Khan, A. (2018) Plasma-derived cell-free mitochondrial DNA: A novel non-invasive methodology to identify mitochondrial DNA haplogroups in humans. *Molecular Genetics and Metabolism*
36. Torralba, D., Baixauli, F., and Sánchez-Madrid, F. (2016) Mitochondria Know No Boundaries: Mechanisms and Functions of Intercellular Mitochondrial Transfer. *Frontiers in Cell and Developmental Biology* **4**
37. Rodriguez, A.-M., Nakhle, J., Griessinger, E., and Vignais, M.-L. (2018) Intercellular mitochondria trafficking highlighting the dual role of mesenchymal stem cells as both sensors and rescuers of tissue injury. *Cell Cycle* **17**, 712–721
38. Mahrouf-Yorgov, M., Augeul, L., Da Silva, C. C., Jourdan, M., Rigolet, M., Manin, S., Ferrera, R., Ovize, M., Henry, A., Guguin, A., Meningaud, J.-P., Dubois-Randé, J.-L., Motterlini, R., Foresti, R., and Rodriguez, A.-M. (2017) Mesenchymal stem cells sense mitochondria released from damaged cells as danger signals to activate their rescue properties. *Cell Death and Differentiation* **24**, 1224–1238
39. Patel, D., Rorbach, J., Downes, K., Szukszto, M. J., Pekalski, M. L., and Minczuk, M. (2017) Macropinocytic entry of isolated mitochondria in epidermal growth factor-activated human osteosarcoma cells. *Scientific Reports* **7**
40. Wang, J., Li, H., Yao, Y., Zhao, T., Chen, Y., Shen, Y., Wang, L., and Zhu, Y. (2018) Stem cell-derived mitochondria transplantation: a novel strategy and the challenges for the treatment of tissue injury. *Stem Cell Research & Therapy* **9**
41. Cai, Y., Xu, M.-J., Koritzinsky, E. H., Zhou, Z., Wang, W., Cao, H., Yuen, P. S. T., Ross, R. A., Star, R. A., Liangpunsakul, S., and Gao, B. (2017) Mitochondrial DNA–enriched microparticles promote acute-on-chronic alcoholic neutrophilia and hepatotoxicity. *JCI Insight* **2**
42. Mittelbrunn, M. and Sánchez-Madrid, F. (2012) Intercellular communication: diverse structures for exchange of genetic information. *Nature Reviews Molecular Cell Biology* **13**, 328–335
43. Pua, H. H., Guo, J., Komatsu, M., and He, Y.-W. (2009) Autophagy Is Essential for Mitochondrial Clearance in Mature T Lymphocytes. *The Journal of Immunology* **182**, 4046–4055
44. Kukat, C., Wurm, C. A., Spahr, H., Falkenberg, M., Larsson, N.-G., and Jakobs, S. (2011) Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *Proceedings of the National Academy of Sciences* **108**, 13534–13539
45. Collins, L. V., Hajizadeh, S., Holme, E., Jonsson, I.-M., and Tarkowski, A. (2004) Endogenously oxidized mitochondrial DNA induces in vivo and in vitro inflammatory responses. *Journal of Leukocyte Biology* **75**, 995–1000
46. Garcia-Martinez, I., Santoro, N., Chen, Y., Hoque, R., Ouyang, X., Caprio, S., Shlomchik, M. J., Coffman, R. L., Candia, A., and Mehal, W. Z. (2016) Hepatocyte mitochondrial DNA drives nonalcoholic steatohepatitis by activation of TLR9. *JCI*, March 1, 2016
47. Hu, Q., Ren, H., Ren, J., Liu, Q., Wu, J., Wu, X., Li, G., Wang, G., Gu, G., Guo, K., Hong, Z., Liu, S., and Li, J. (2018) Released Mitochondrial DNA Following Intestinal Ischemia Reperfusion Induces the Inflammatory Response and Gut Barrier Dysfunction. *Scientific Reports* **8**
48. Fang, C., Wei, X., and Wei, Y. (2016) Mitochondrial DNA in the regulation of innate immune responses. *Protein & Cell* **7**, 11–16
49. Martin, W. F., Garg, S., and Zimorski, V. (2015) Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20140330

50. Portugez, S., Martin, W. F., and Hazkani-Covo, E. (2018) Mosaic mitochondrial-plastid insertions into the nuclear genome show evidence of both non-homologous end joining and homologous recombination. *BMC Evolutionary Biology* **18**

Figure legends:

Figure 1: Both plasma of healthy individuals and cell culture media contain cell-free mitochondrial DNA that is well preserved and not fragmented, unlike nuclear cell-free DNA.

(A) NcfDNA and McfDNA quantification in plasma samples of 99 healthy individuals. McfDNA copy number is significantly higher than that of NcfDNA ($p<0.0001$). **(B)** Kinetic study of cfDNA stability in cell culture media. n=2. **(C)** NcfDNA and McfDNA integrity index calculated for plasma samples of 13 healthy individuals, demonstrating that McfDNA is less fragmented than NcfDNA ($p<0.0001$). **(D)** Size fractionation after agarose gel electrophoresis of a pool of cfDNA extracts obtained from 80 healthy individuals'plasma. McfDNA fragments were quantified by q-PCR from extracts of the excised agarose gel slices above (Long) and below (Short) 500bp. **(E)** DNA fragment size profile at single base resolution obtained by paired-end massively parallel WGS of cfDNA extracted from a healthy individual's plasma. (Blue for NcfDNA and red for McfDNA). The lines inside the boxes and the upper and lower limits of the boxes indicate the median, 75th and 25th percentiles, respectively. The upper and lower horizontal bars indicate the maximum and minimum values, respectively.

Figure 2: Plasma of healthy individuals and cell culture media contain structures enclosing mitochondrial cell-free DNA.

(A) cfDNA content following three different subsequent treatments of plasma: (i) by high-speed centrifugation, (ii) by filtration (0.22µM) in Ficoll isolated plasma and (iii) by high-speed centrifugation in plasma isolated by a protocol precluding activation of platelets (PPAP). LS and HS, low and high-speed centrifugation at 16 000g; NF and F, non-filtered and filtered plasma (Fig S 3); n=8 p=0.0630, n=7 p=0.0009 and n=4 p=0.0173, respectively.

(B) Decrease of the cfDNA content in the DLD-1/SW620/CCD-18Co cell culture media after an initial low-speed (LS, 600g) and a subsequent high-speed centrifugation (HS, 16 000g). (Methods and Fig S 4). n=28. p<0.0001. **(C)** Proportion of NcfDNA and McfDNA plasma contents after treatment of PPAP plasma (Table S 1). n=4. **(D)** NcfDNA copy number in healthy individuals' plasma prepared under PPAP and subsequently filtered and/or centrifuged at either 2 500 or 16 000g (n=5). **(E)** McfDNA copy number in healthy individuals' plasma prepared under PPAP and subsequently filtred and/or centrifuged at either 2,500 or 16,000g. n=5, p=0.0111, p=0.0114, p=0.0115, respectively.

Figure 3: Plasma of healthy individuals and cell culture media contain structures enclosing whole intact mitochondrial genome.

(A) Adapted schematic view of the Long-range (LR) PCR amplification strategy for testing mitochondrial genome integrity. Overlapping amplicons amplify the whole mitochondrial genome by using the two htm1 and htm2 sequences or the three mito1, mito2 and mito3 sequences. **(B)** Agarose gel analysis of LR-PCR amplified mitochondrial DNA demonstrating the presence of a preserved cell free mitochondrial genome in cell culture media and plasma pellet. Extracted DNA from isolated mitochondria from SW620 cells (lane b, positive control), the 16,000g pellet (16gP) from the SW620 cells culture medium (lane c), the 16gP from the DLD-1 cells culture medium (lane d), a mouse genomic DNA (lane e, negative control), the Repli-g mitochondrial DNA kit's Water (lane f, negative control), the PCR water (lane g, negative control), the 16gP from two different plasma pools (lane i, j), were amplified using the repli-g mitochondrial DNA kit and the 5 sets of primers for the Mito1 (3,968pb), Mito2 (5,513bp), Mito3 (7,814bp,) htm1 (9,289bp) and htm2 (7,626bp) amplicons. PCR-products analysis on a 0,8% agarose gel reveals the expected amplicons in the 16gPs from cell culture media and from plasma, demonstrating the presence of a whole mitochondrial genome. No signal was detected in negative control. (Lanes a. Lane h, 1kb DNA ladder).

Figure 4: Both plasma of healthy individuals and cell culture media contain cell free mitochondrial material. Flow cytometry analysis of preparations stained with mitochondrial specific dye MitoTracker Green of **(A)**, Isolated mitochondria (IM) from DLD-1 cells (positive control); **(B and C)**, two negative controls: isolated nucleus and 16gP obtained from fresh cell media and; **(D)**, the 16gP isolated from DLD-1 centrifuged cell culture media (CCCM); **(E)**, the four previous samples; and **(F)**, the 16gP isolated from the plasma of an healthy donor. Results were analyzed with Kaluza analysis 1.5 software. Immunoblotting of membranes by labelling the inner (Tim23) and outer (Tom22) mitochondrial membrane proteins **(G)**: Tim23 (lane a: 16gP from PPAP plasma; lane b: 16gP from CCD-18Co CCCM; lane c: 16gP from SW620 CCCM; lane d: 16gP from DLD-1 CCCM; lane e: IM from CCD-18Co cells; lane f: IM from SW620 cells; lane g: IM from DLD-1 cells; lane h: IM from DLD-1 cells). **(H)**: Tom 22 (lane a: IM from DLD-1 cells; lane: IM from CCD-18Co cells; lane c: IM from SW620 cells; lane d: IM from SW620 cells; lane e: 16gP from DLD-1 CCCM; lane f: 16gP from CCD-18Co CCCM; lane g: 16gP from SW620 CCCM; lane h: 16gP from PPAP plasma).

Figure 5: Both plasma of healthy individuals and cell culture media contain structurally intact mitochondria. Fluorescence microscopy images of the MitoTracker Green stained 16gP from PPAP plasma **(A)**, SW620 CCCM **(B)** and DLD1 CCCM **(C)**. Electron microscopy images of the 16gP from of a PPAP plasma **(D)**, SW620 CCCM **(E)** and DLD1 CCCM **(F).**

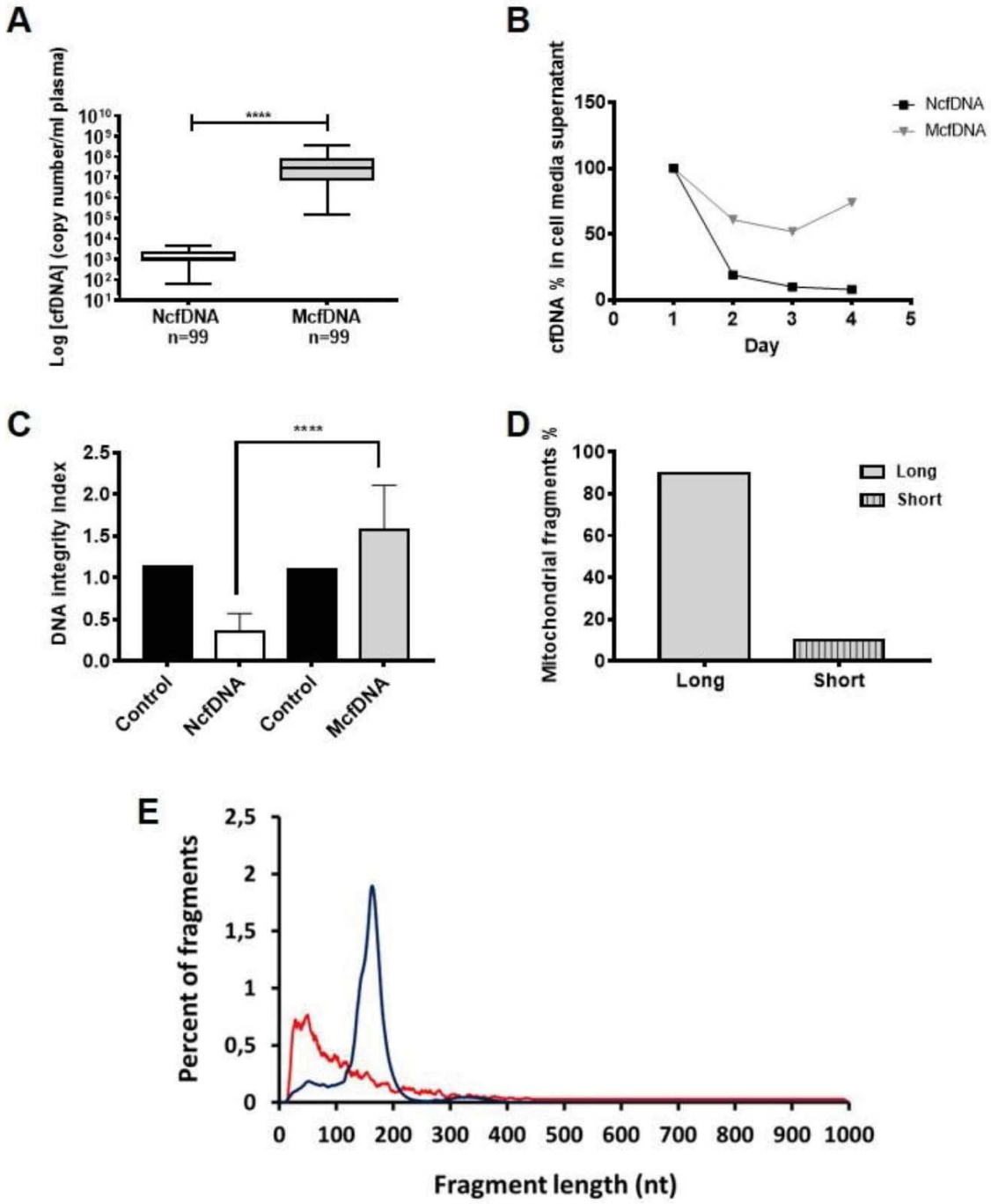
Figure 6: Both plasma of healthy individuals and cell culture media contain respiratory-competent mitochondria. The electron flow experiment performed using a seahorse XF96 instrument. Like the positive control (Isolated mitochondria from DLD-1), OCR was detectable in both 16gP from DLD-1 cell media and plasma pool, in the presence of pyruvate and malate that drive respiration via complex I **(A)**. There was a significant decrease in OCR when rotenone was injected, indicating an inhibition of complex I-driven respiration **(B)**. When complex II-driven respiration was initiated, by sequential addition of succinate, a slight increase in

OCR was observed, indicating that complex II may be functional. The inhibition of complex III (antimycin A) and the stimulation of complex IV (ascorbate + TMPD), led to a detectable activity of cytochrome C/complex IV in all the samples. Each value shown was from a minimum of five replicates. The data are presented in **(A)** as mean (symbol) ± SD (line). The four injections are shown as dashed lines. The data in **(B)** correspond to the mean ± SD, of the replicates in the last time point before adding Rotenone (Basal respiration), and the first time point after adding Rotenone (Rotenone). The statistical analysis was made between all the time points before (4) and after the addition of Rotenone (3).

29

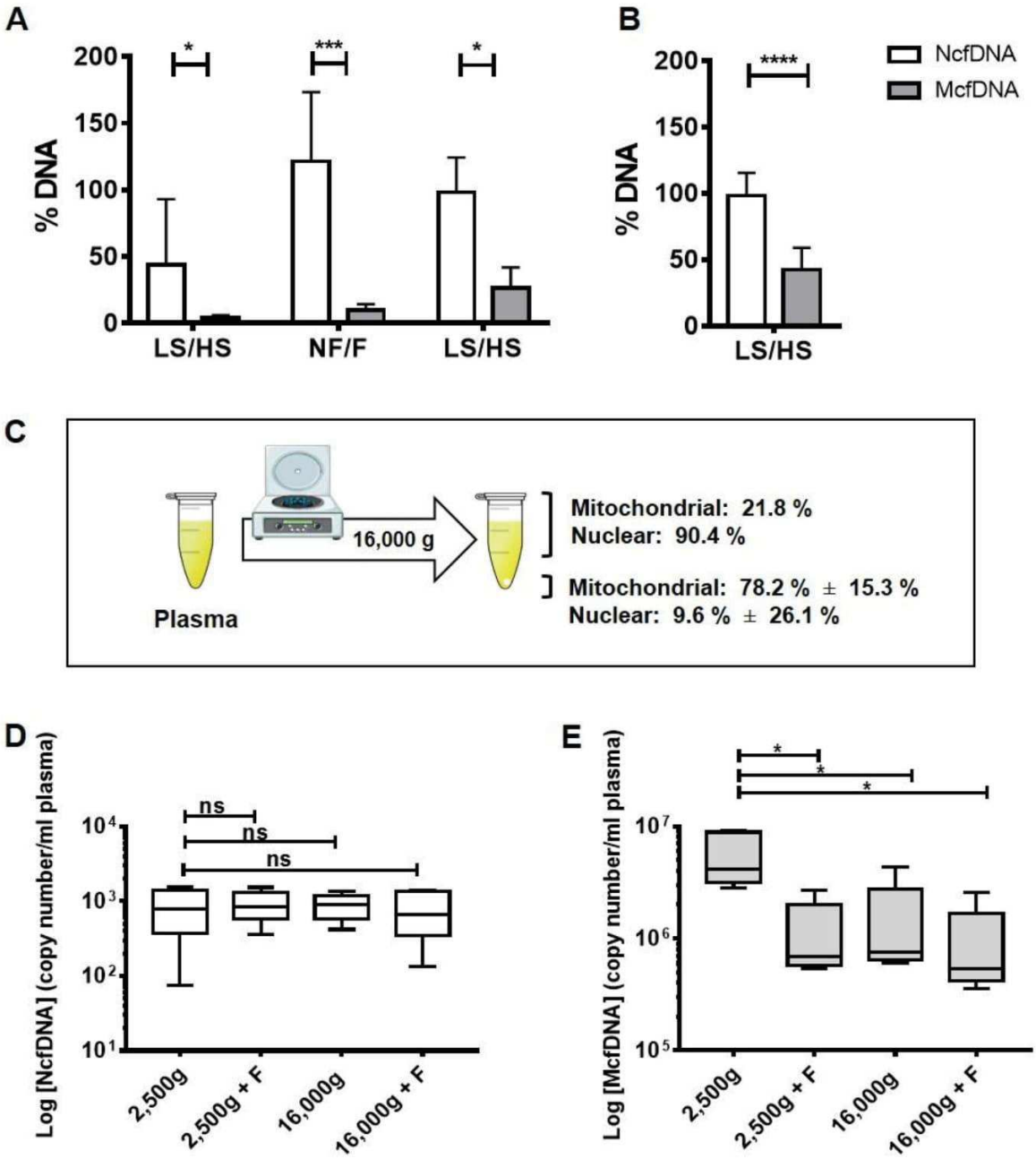
Figures:

Figure 1



30

Figure 2


A
% DNA
200
150
100
50
0
*
***
*
LS/HS
NF/F
LS/HS
B
% DNA
200
150
100
50
0
****
LS/HS
NcfDNA
McfDNA
C
16,000 g
Plasma
Mitochondrial: 21.8 %
Nuclear: 90.4 %
Mitochondrial: 78.2 % ± 15.3 %
Nuclear: 9.6 % ± 26.1 %
D
Log [NcfDNA] (copy number/ml plasma)
ns
ns
ns
2,500g
2,500g + F
16,000g
16,000g + F
E
Log [McfDNA] (copy number/ml plasma)
*
*
*
2,500g
2,500g + F
16,000g
16,000g + F

31

Figure 3

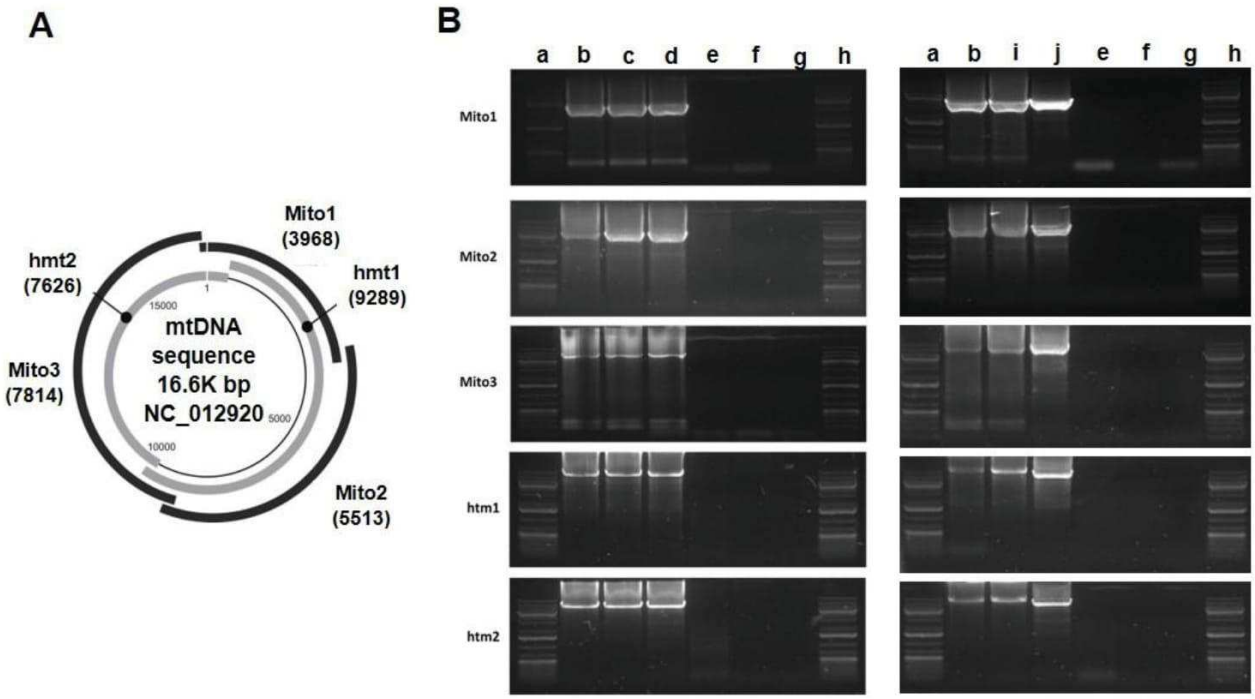
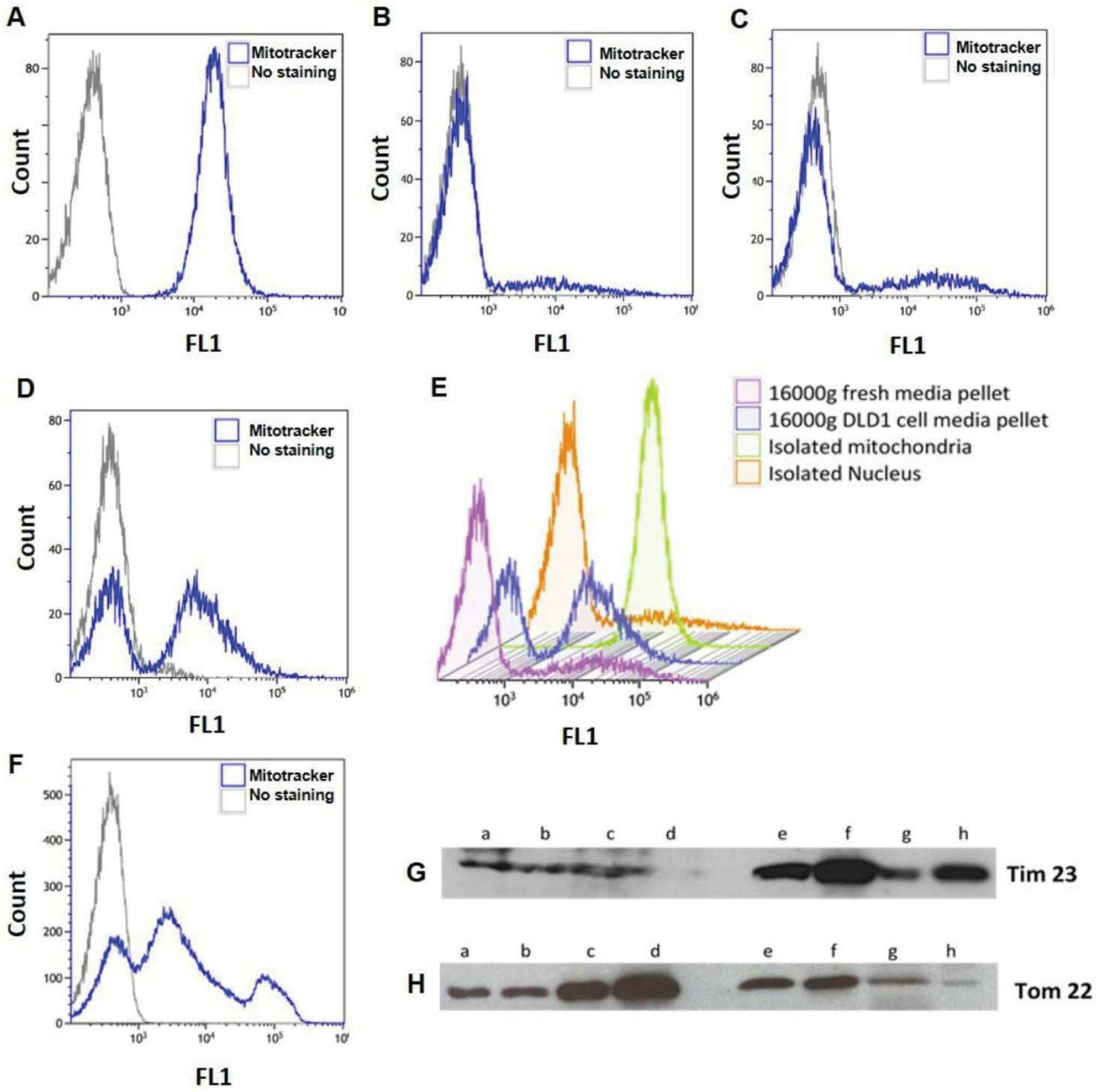

A
Mito1
(3968)
hmt1
(9289)
Mito2
(5513)
Mito3
(7814)
hmt2
(7626)
mtDNA
sequence
16.6K bp
NC_012920
1
15000
5000
10000
B
a b c d e f g h
a b i j e f g h
Mito1
Mito2
Mito3
htm1
htm2

Figure 4


A
Mitotracker
No staining
Count
FL1
B
Mitotracker
No staining
Count
FL1
C
Mitotracker
No staining
Count
FL1
D
Mitotracker
No staining
Count
FL1
E
16000g fresh media pellet
16000g DLD1 cell media pellet
Isolated mitochondria
Isolated Nucleus
FL1
F
Mitotracker
No staining
Count
FL1
G
a b c d e f g h
Tim 23
H
a b c d e f g h
Tom 22

33

Figure 5

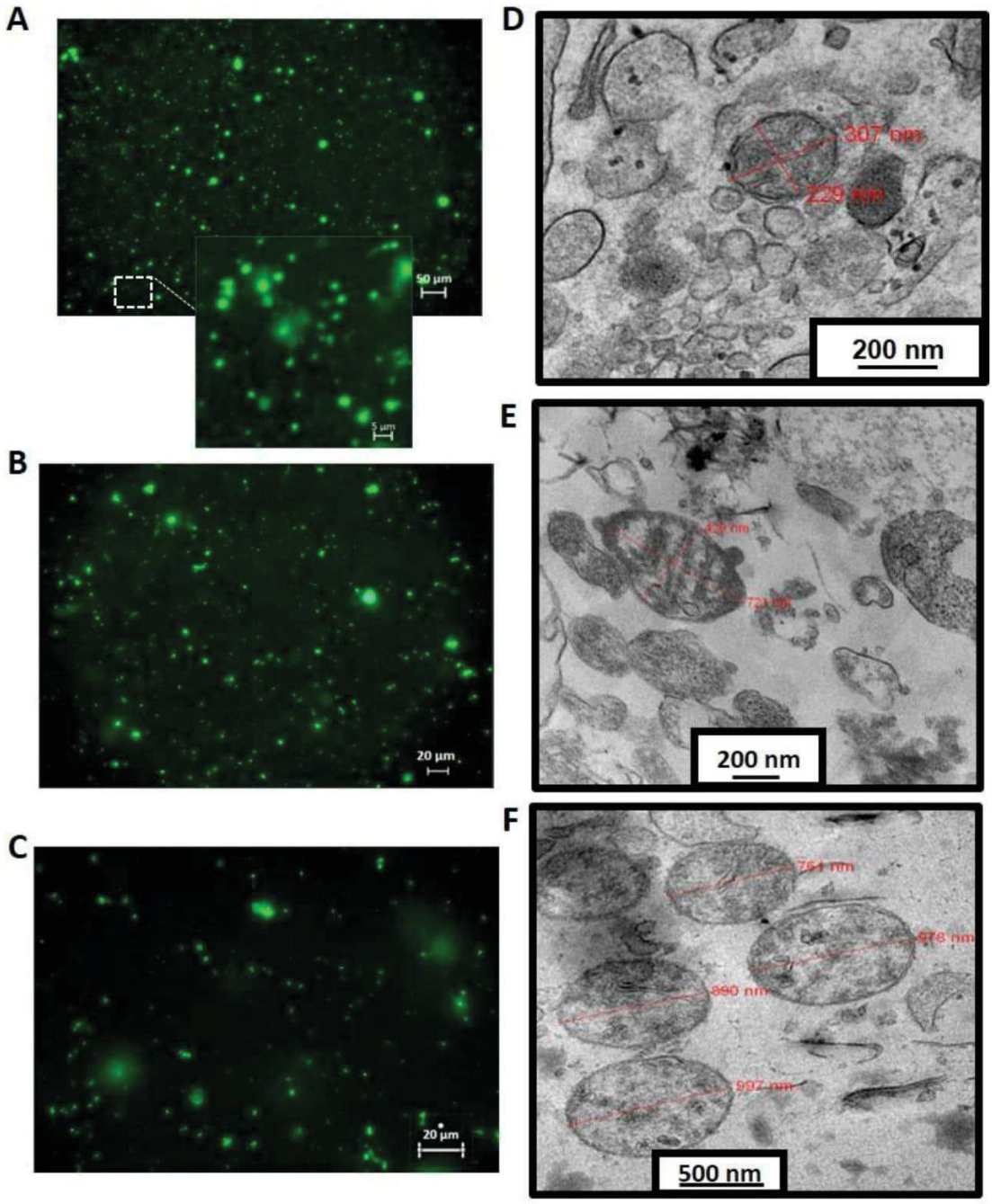
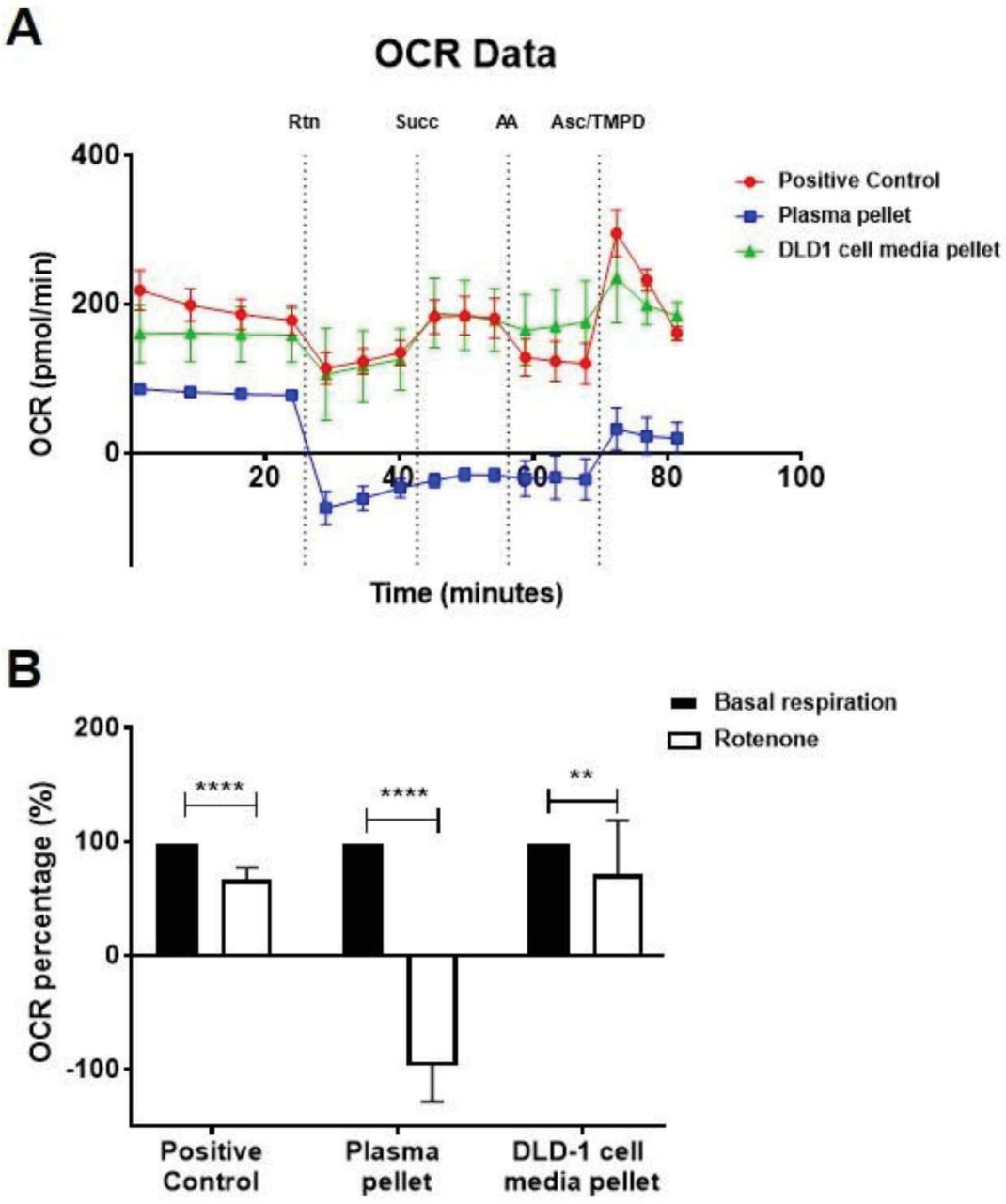

A
50 µm
5 µm
B
20 µm
C
20 µm
D
307 nm
229 nm
200 nm
E
200 nm
F
761 nm
876 nm
890 nm
997 nm
500 nm

Figure 6


A
OCR Data
Rtn
Succ
AA
Asc/TMPD
400
200
0
20
40
60
80
100
OCR (pmol/min)
Time (minutes)
Positive Control
Plasma pellet
DLD1 cell media pellet
B
Basal respiration
Rotenone
200
100
0
-100
OCR percentage (%)
****
****
**
Positive Control
Plasma pellet
DLD-1 cell media pellet

35

E. Taille et caractéristiques structurelles de l'ADN circulant

Titre : Sizing and structural features of circulating DNA

Auteurs

A.Bennett, Cynthia Sanchez, **Rita Tanos**, Thibault Mazard, Philippe Blache, E.M. Geigl, Thierry Grange et Alain R. Thierry

Ce travail est sous préparation pour être soumis à *Nucleic Acid Research*.

Résumé

Nous avons étudié la distribution de la taille de l'ADNcir dans le sang en utilisant la Q-PCR et le séquençage du génome entier à partir de la préparation de librairies d'ADN double et simple brin, pour étudier leur structure dans la circulation. Le profil de taille chez les individus sains est remarquablement homogène parmi les plasmas testés. Nous avons confirmé que le profil de taille de l'ADNcir est caractéristique de la fragmentation entre les nucléosomes. Dans l'ensemble, les données ont montré que la proportion d'ADNcir inséré dans les mono-nucléosomes, les di-nucléosomes ou de taille moléculaire supérieure (>1 000 pb) est de 87,5%, 1,1% et 11,4%. Ainsi, nos données suggèrent que la proportion d'ADNcir ayant une taille supérieure à celle existant dans les complexes de mono-nucléosomes ou de facteurs de transcription circulant dans le sang est mineure, ce qui suggère que l'ADNcir qui peut être détecté dans le plasma est principalement associé à ces deux structures. La distribution générale de la taille de l'ADNcir chez les patients atteints d'un cancer colorectal métastatique est semblable à celle des individus sains, cependant nos données ont révélé des différences subtiles mais fiables dans la fourchette de 30 à 220 pb, et au-delà de 1 000 pb. En plus de ces différences dans le profil de taille, nous avons observé dans l'ADNcir provenant d'un sujet atteint de cancer, contrairement à l'ADNcir d'un sujet sain, que la fragmentation variait également avec la fréquence des allèles mutés.

Contributions

Dans ce travail, j'ai réalisé certaines exprériences, et j'ai également discuté les résultats du papier avec les différents auteurs.

REFERENCES

1. P. Mandel, P. Metais, Les acides nucléiques du plasma sanguin chez l'homme., *C. R. Seances Soc. Biol. Fil.* **142**, 241–43 (1948).

2. M. Stroun, C. Mathon, J. Stroun, Modifications transmitted to the offspring provoked by heterograft in Solanum melongena, *Arch Sci* **16**, 225–245 (1963).

3. P. Anker, M. Stroun, Circulating nucleic acids and evolution, *Expert Opin. Biol. Ther.* **12**, S113–S117 (2012).

4. C. R. Darwin, The variation of animals and plants under domestication., *London John Murray. First Ed. first issue* **Volume 2** (1868).

5. E. M. Tan, P. H. Schur, R. I. Carr, H. G. Kunkel, Deoxyribonucleic acid (DNA) and antibodies to DNA in the serum of patients with systemic lupus erythematosus., *J. Clin. Invest.* **45**, 1732–40 (1966).

6. Y. M. D. Lo, N. Corbetta, P. F. Chamberlain, V. Rai, I. L. Sargent, C. W. G. Redman, J. S. Wainscoat, Presence of fetal DNA in maternal plasma and serum, *Lancet* **350**, 485–487 (1997).

7. S. A. Leon, B. Shapiro, D. M. Sklaroff, M. J. Yaros, Free DNA in the Serum of Cancer Patients and the Effect of Therapy, *Cancer Res.* **37**, 646–650 (1977).

8. M. Stroun, P. Anker, J. Lyautey, C. Lederrey, P. A. Maurice, Isolation and characterization of DNA from the plasma of cancer patients, *Eur. J. Cancer Clin. Oncol.* **23**, 707–12 (1987).

9. M. Stroun, P. Anker, P. Maurice, J. Lyautey, C. Lederrey, M. Beljanski, Neoplastic characteristics of the DNA found in the plasma of cancer patients, *Oncology* **46**, 318–322 (1989).

10. Liquid Biopsy: Using Tumor DNA in Blood to Aid Cancer Care*Natl. Cancer Inst.* (2017) (available at https://www.cancer.gov/news-events/cancer-currents-blog/2017/liquid-biopsy-detects-treats-cancer).

11. NCI Dictionary of Cancer Terms*Natl. Cancer Inst.* (available at https://www.cancer.gov/publications/dictionaries/cancer-terms).

12. A. R. Thierry, F. Mouliere, S. El Messaoudi, C. Mollevi, E. Lopez-Crapez, F. Rolet, B. Gillet, C. Gongora, P. Dechelotte, B. Robert, M. Del Rio, P.-J. Lamy, F. Bibeau, M. Nouaille, V. Loriot, A.-S. Jarrousse, F. Molina, M. Mathonnet, D. Pezet, M. Ychou, Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA, *Nat. Med.* **20**, 430–435 (2014).

13. A. R. Thierry, B. Pastor, Z.-Q. Jiang, A. D. Katsiampoura, C. Parseghian, J. M. Loree, M. J. Overman, C. Sanchez, S. El Messaoudi, M. Ychou, S. Kopetz, Circulating DNA Demonstrates Convergent Evolution and Common Resistance Mechanisms during Treatment of Colorectal Cancer, *Clin. Cancer Res.* **23**, 4578–4591 (2017).

14. G. Siravegna, B. Mussolin, M. Buscarino, G. Corti, A. Cassingena, G. Crisafulli, A. Ponzetti, C. Cremolini, A. Amatu, C. Lauricella, S. Lamba, S. Hobor, A. Avallone, E. Valtorta, G. Rospo, E. Medico, V. Motta, C. Antoniotti, F. Tatangelo, B. Bellosillo, S. Veronese, A. Budillon, C. Montagut, P. Racca, S. Marsoni, A. Falcone, R. B. Corcoran, F. Di Nicolantonio, F. Loupakis, S. Siena, A. Sartore-Bianchi, A. Bardelli, Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients, *Nat. Med.* **21**, 795–801 (2015).

15. S. Perakis, M. Auer, J. Belic, E. Heitzer, in *Advances in Clinical Chemistry*, (Elsevier, 2017), vol. 80, pp. 73–153.

16. M. Elazezy, S. A. Joosse, Techniques of using circulating tumor DNA as a liquid biopsy component

in cancer management., *Comput. Struct. Biotechnol. J.* **16**, 370–378 (2018).

17. Non-Invasive Prenatal Testing (NIPT) Market Value, Trends, Situation, Sales Area, Competitors, Industry Demand and Scope 2023 – Honest Version (available at https://honestversion.com/non-invasive-prenatal-testing-nipt-market-value-trends-situation-sales-area-competitors-industry-demand-and-scope-2023/).

18. Y. M. D. Lo, Fetal DNA in Maternal Plasma, *Ann. N. Y. Acad. Sci.* **906**, 141–147 (2006).

19. J. Zhang, J. Li, J. B. Saucier, Y. Feng, Y. Jiang, J. Sinson, A. K. McCombs, E. S. Schmitt, S. Peacock, S. Chen, H. Dai, X. Ge, G. Wang, C. A. Shaw, H. Mei, A. Breman, F. Xia, Y. Yang, A. Purgason, A. Pourpak, Z. Chen, X. Wang, Y. Wang, S. Kulkarni, K. W. Choy, R. J. Wapner, I. B. Van den Veyver, A. Beaudet, S. Parmar, L.-J. Wong, C. M. Eng, Non-invasive prenatal sequencing for multiple Mendelian monogenic disorders using circulating cell-free fetal DNA, *Nat. Med.* **25**, 439–447 (2019).

20. J. Camunas-Soler, H. Lee, L. Hudgins, S. R. Hintz, Y. J. Blumenfeld, Y. Y. El-Sayed, S. R. Quake, Noninvasive Prenatal Diagnosis of Single-Gene Disorders by Use of Droplet Digital PCR., *Clin. Chem.* **64**, 336–345 (2018).

21. L. J. Salomon, A. Sotiriadis, C. B. Wulff, A. Odibo, R. Akolekar, Risk of miscarriage following amniocentesis or chorionic villus sampling: systematic review of the literature and updated meta-analysis, *Ultrasound Obstet. Gynecol.* , uog.20353 (2019).

22. M. D. Pertile, M. Halks-Miller, N. Flowers, C. Barbacioru, S. L. Kinnings, D. Vavrek, W. K. Seltzer, D. W. Bianchi, Rare autosomal trisomies, revealed by maternal plasma DNA sequencing, suggest increased risk of feto-placental disease, *Sci. Transl. Med.* **9**, eaan1240 (2017).

23. A. R. Gregg, B. G. Skotko, J. L. Benkendorf, K. G. Monaghan, K. Bajaj, R. G. Best, S. Klugman, M. S. Watson, Noninvasive prenatal screening for fetal aneuploidy, 2016 update: a position statement of the American College of Medical Genetics and Genomics, *Genet. Med.* **18**, 1056–1065 (2016).

24. Y. M. Lo, P. Patel, C. N. Baigent, M. D. Gillmer, P. Chamberlain, M. Travi, M. Sampietro, J. S. Wainscoat, K. A. Fleming, Prenatal sex determination from maternal peripheral blood using the polymerase chain reaction., *Hum. Genet.* **90**, 483–8 (1993).

25. Y. M. D. Lo, N. M. Hjelm, C. Fidler, I. L. Sargent, M. F. Murphy, P. F. Chamberlain, P. M. K. Poon, C. W. G. Redman, J. S. Wainscoat, Prenatal Diagnosis of Fetal RhD Status by Molecular Analysis of Maternal Plasma, *N. Engl. J. Med.* **339**, 1734–1738 (1998).

26. V. Swarup, M. r. Rajeswari, Circulating (cell-free) nucleic acids – A promising, non-invasive tool for early detection of several human diseases, *FEBS Lett.* **581**, 795–799 (2007).

27. J. Atamaniuk, Y.-Y. Hsiao, M. Mustak, D. Bernhard, L. Erlacher, M. Fodinger, B. Tiran, K. M. Stuhlmeier, Analysing cell-free plasma DNA and SLE disease activity, *Eur. J. Clin. Invest.* **41**, 579–583 (2011).

28. Y. Xu, Y. Song, J. Chang, X. Zhou, Q. Qi, X. Tian, M. Li, X. Zeng, M. Xu, W. Zhang, D. S. Cram, J. Liu, High levels of circulating cell-free DNA are a biomarker of active SLE, *Eur. J. Clin. Invest.* **48**, e13015 (2018).

29. A. Truszewska, B. Foroncewicz, L. Pączek, The role and diagnostic value of cell-free DNA in systemic lupus erythematosus., *Clin. Exp. Rheumatol.* **35**, 330–336 (2016).

30. S. Tug, S. Helmig, J. Menke, D. Zahn, T. Kubiak, A. Schwarting, P. Simon, Correlation between cell free DNA levels and medical evaluation of disease progression in systemic lupus erythematosus patients., *Cell. Immunol.* **292**, 32–9 (2014).

31. S. R. Knight, A. Thorne, M. L. Lo Faro, Donor-specific Cell-free DNA as a Biomarker in Solid Organ

Transplantation. A Systematic Review., *Transplantation* **103**, 273–283 (2019).

32. T. M. Snyder, K. K. Khush, H. A. Valantine, S. R. Quake, Universal noninvasive detection of solid organ transplant rejection., *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6229–34 (2011).

33. Daly, K. P., Circulating donor-derived cell-free DNA: a true biomarker for cardiac allograft rejection?, *Ann. Transl. Med.* **3** (2015), doi:10.21037/5972.

34. S. K. Goh, H. Do, A. Testro, J. Pavlovic, A. Vago, J. Lokan, R. M. Jones, C. Christophi, A. Dobrovic, V. Muralidharan, The Measurement of Donor-Specific Cell-Free DNA Identifies Recipients With Biopsy-Proven Acute Rejection Requiring Treatment After Liver Transplantation, *Transplant. Direct* **5** (2019), doi:10.1097/TXD.0000000000000902.

35. E. Schütz, A. Fischer, J. Beck, M. Harden, M. Koch, T. Wuensch, M. Stockmann, B. Nashan, O. Kollmar, J. Matthaei, P. Kanzow, P. D. Walson, J. Brockmöller, M. Oellerich, A. Singal, Ed. Graft-derived cell-free DNA, a noninvasive early rejection and graft damage marker in liver transplantation: A prospective, observational, multicenter cohort study, *PLOS Med.* **14**, e1002286 (2017).

36. R. D. Bloom, J. S. Bromberg, E. D. Poggio, S. Bunnapradist, A. J. Langone, P. Sood, A. J. Matas, S. Mehta, R. B. Mannon, A. Sharfuddin, B. Fischbach, M. Narayanan, S. C. Jordan, D. Cohen, M. R. Weir, D. Hiller, P. Prasad, R. N. Woodward, M. Grskovic, J. J. Sninsky, J. P. Yee, D. C. Brennan, Circulating Donor-Derived Cell-Free DNA in Blood for Diagnosing Active Rejection in Kidney Transplant Recipients (DART) Study Investigators, Cell-Free DNA and Active Rejection in Kidney Allografts., *J. Am. Soc. Nephrol.* **28**, 2221–2232 (2017).

37. H. Zhang, L. Liu, C. Zheng, X. Li, Q. Fu, J. Li, Q. Su, H. Huang, M. Ye, C. Wang, The Role of Donor-Derived Cell-Free DNA in the Identification of Injury in Kidney Allografts with Antibody-Mediated Rejection or De Novo DSA, *Transplantation* **102**, S5–S6 (2018).

38. D. J. Dwivedi, L. J. Toltl, L. L. Swystun, J. Pogue, K.-L. Liaw, J. I. Weitz, D. J. Cook, A. E. Fox-Robichaud, P. C. Liaw, Prognostic utility and characterization of cell-free DNA in patients with severe sepsis, *Crit. Care* **16**, R151 (2012).

39. T. Huang, Z. Yang, S. Chen, J. Chen, [Predictive value of plasma cell-free DNA for prognosis of sepsis]., *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue* **30**, 925–928 (2018).

40. A. Shimony, D. Zahger, H. Gilutz, H. Goldstein, G. Orlov, M. Merkin, A. Shalev, R. Ilia, A. Douvdevani, Cell free DNA detected by a novel method in acute ST-elevation myocardial infarction patients, *Acute Card. Care* **12**, 109–111 (2010).

41. L. Wang, L. Xie, Q. Zhang, X. Cai, Y. Tang, L. Wang, T. Hang, J. Liu, J. Gong, Plasma nuclear and mitochondrial DNA levels in acute myocardial infarction patients., *Coron. Artery Dis.* **26**, 296–300 (2015).

42. K. V. Glebova, N. N. Veiko, A. A. Nikonov, L. N. Porokhovnik, S. V. Kostuyk, Cell-free DNA as a biomarker in stroke: Current status, problems and perspectives, *Crit. Rev. Clin. Lab. Sci.* **55**, 55–70 (2018).

43. G. C. O'Connell, A. B. Petrone, C. S. Tennant, N. Lucke-Wold, Y. Kabbani, A. R. Tarabishy, P. D. Chantler, T. L. Barr, Circulating extracellular DNA levels are acutely elevated in ischaemic stroke and associated with innate immune system activation, *Brain Inj.* **31**, 1369–1375 (2017).

44. A. R. Thierry, S. El Messaoudi, P. B. Gahan, P. Anker, M. Stroun, Origins, structures, and functions of circulating DNA in oncology, *Cancer Metastasis Rev.* **35**, 347–76 (2016).

45. J. Aucamp, A. J. Bronkhorst, C. P. S. Badenhorst, P. J. Pretorius, The diverse origins of circulating cell-free DNA in the human body: a critical re-evaluation of the literature, *Biol. Rev.* **93**, 1649–1683

(2018).

46. M. van der Vaart, P. J. Pretorius, The Origin of Circulating Free DNA, *Clin. Chem.* **53**, 2215 (2007).

47. M. Stroun, P. Maurice, V. Vasioukhin, J. Lyautey, C. Lederrey, F. Lefort, A. Rossier, X. Q. Chen, P. Anker, The Origin and Mechanism of Circulating DNA, *Ann. N. Y. Acad. Sci.* **906**, 161–168 (2000).

48. R. Aarthy, S. Mani, S. Velusami, S. Sundarsingh, T. Rajkumar, Role of Circulating Cell-Free DNA in Cancers, *Mol. Diagn. Ther.* **19**, 339–350 (2015).

49. S. Elmore, Apoptosis: a review of programmed cell death., *Toxicol. Pathol.* **35**, 495–516 (2007).

50. S. Nagata, Apoptotic DNA Fragmentation, *Exp. Cell Res.* **256**, 12–18 (2000).

51. S. Jahr, H. Hentze, S. Englisch, D. Hardt, F. O. Fackelmayer, R.-D. Hesch, R. Knippers, DNA Fragments in the Blood Plasma of Cancer Patients: Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells, *Cancer Res.* **61**, 1659–1665 (2001).

52. M. Stroun, J. Lyautey, C. Lederrey, A. Olson-Sand, P. Anker, About the possible origin and mechanism of circulating DNA: Apoptosis and active DNA release, *Clin. Chim. Acta* **313**, 139–142 (2001).

53. P. M. Rumore, C. R. Steinman, Endogenous circulating DNA in systemic lupus erythematosus. Occurrence as multimeric complexes bound to histone., *J. Clin. Invest.* **86**, 69–74 (1990).

54. V. V. Vlassov, P. P. Laktionov, E. Y. Rykova, Extracellular nucleic acids, *BioEssays* **29**, 654–667 (2007).

55. R. J. Youle, D. P. Narendra, Mechanisms of mitophagy, *Nat. Rev. Mol. Cell Biol.* **12**, 9–14 (2011).

56. K. Thurairajah, G. D. Briggs, Z. J. Balogh, The source of cell-free mitochondrial DNA in trauma and potential therapeutic strategies., *Eur. J. Trauma Emerg. Surg.* **44**, 325–334 (2018).

57. N. Jiang, C. F. Reich, D. S. Pisetsky, Role of macrophages in the generation of circulating blood nucleosomes from dead and dying cells, *Blood* **102**, 2243–2250 (2003).

58. J.-J. Choi, C. F. Reich, D. S. Pisetsky, The role of macrophages in the in vitro generation of extracellular DNA from apoptotic and necrotic cells, *Immunology* **115**, 55–62 (2005).

59. V. Brinkmann, U. Reichard, C. Goosmann, B. Fauler, Y. Uhlemann, D. S. Weiss, Y. Weinrauch, A. Zychlinsky, Neutrophil extracellular traps kill bacteria., *Science* **303**, 1532–5 (2004).

60. V. Brinkmann, A. Zychlinsky, Neutrophil extracellular traps: Is immunity the second function of chromatin?, *J. Cell Biol.* **198**, 773–783 (2012).

61. P. R. Cooper, L. J. Palmer, I. L. C. Chapple, Neutrophil extracellular traps as a new paradigm in innate immunity: friend or foe?, *Periodontol. 2000* **63**, 165–197 (2013).

62. J. Cools-Lartigue, J. Spicer, B. McDonald, S. Gowing, S. Chow, B. Giannias, F. Bourdeau, P. Kubes, L. Ferri, Neutrophil extracellular traps sequester circulating tumor cells and promote metastasis, *J. Clin. Invest.* **123**, 3446–3458 (2013).

63. T. Beiter, A. Fragasso, J. Hudemann, M. Schild, J. Steinacker, F. C. Mooren, A. M. Niess, Neutrophils release extracellular DNA traps in response to exercise., *J. Appl. Physiol.* **117**, 325–33 (2014).

64. M. Stroun, P. Anker, Nucleic acids spontaneously released by living frog auricles., *Biochem. J.* **128**, 100P-101P (1972).

65. P. Anker, M. Stroun, P. A. Maurice, Spontaneous release of DNA by human blood lymphocytes as shown in an in vitro system., *Cancer Res.* **35**, 2375–82 (1975).

66. P. B. Gahan, M. Stroun, The virtosome-a novel cytosolic informative entity and intercellular messenger., *Cell Biochem. Funct.* **28**, 529–38 (2010).

67. J. Aucamp, A. J. Bronkhorst, C. P. S. Badenhorst, P. J. Pretorius, A historical and evolutionary perspective on the biological significance of circulating DNA and extracellular vesicles, *Cell. Mol. Life Sci.* **73**, 4355–4381 (2016).

68. P. B. Gahan, P. Anker, M. Stroun, Metabolic DNA as the Origin of Spontaneously Released DNA?, *Ann. N. Y. Acad. Sci.* **1137**, 7–17 (2008).

69. M. ABOLHASSANI, J. TILLOTSON, J. CHIAO, Characterization of the release of DNA by a human leukimia-cell line HL-60, *Int. J. Oncol.* **4**, 417–421 (1994).

70. M. R. Fernando, C. Jiang, G. D. Krzyzanowski, W. L. Ryan, New evidence that a large proportion of human blood plasma cell-free DNA is localized in exosomes, *PLoS One* **12**, e0183915 (2017).

71. B. K. Thakur, H. Zhang, A. Becker, I. Matei, Y. Huang, B. Costa-Silva, Y. Zheng, A. Hoshino, H. Brazier, J. Xiang, C. Williams, R. Rodriguez-Barrueco, J. M. Silva, W. Zhang, S. Hearn, O. Elemento, N. Paknejad, K. Manova-Todorova, K. Welte, J. Bromberg, H. Peinado, D. Lyden, Double-stranded DNA in exosomes: a novel biomarker in cancer detection, *Cell Res.* **24**, 766–769 (2014).

72. M. Guescini, S. Genedani, V. Stocchi, L. F. Agnati, Astrocytes and Glioblastoma cells release exosomes carrying mtDNA, *J. Neural Transm.* **117**, 1–4 (2010).

73. G. van Niel, G. D'Angelo, G. Raposo, Shedding light on the cell biology of extracellular vesicles, *Nat. Rev. Mol. Cell Biol.* **19**, 213–228 (2018).

74. V. R. Minciacchi, M. R. Freeman, D. Di Vizio, Extracellular vesicles in cancer: exosomes, microvesicles and the emerging role of large oncosomes., *Semin. Cell Dev. Biol.* **40**, 41–51 (2015).

75. Y. Y. N. Lui, K.-W. Chik, R. W. K. Chiu, C.-Y. Ho, C. W. K. Lam, Y. M. D. Lo, Predominant Hematopoietic Origin of Cell-free DNA in Plasma and Serum after Sex-mismatched Bone Marrow Transplantation, *Clin. Chem.* **48** (2002).

76. K. Sun, P. Jiang, K. C. A. Chan, J. Wong, Y. K. Y. Cheng, R. H. S. Liang, W. Chan, E. S. K. Ma, S. L. Chan, S. H. Cheng, R. W. Y. Chan, Y. K. Tong, S. S. M. Ng, R. S. M. Wong, D. S. C. Hui, T. N. Leung, T. Y. Leung, P. B. S. Lai, R. W. K. Chiu, Y. M. D. Lo, Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments, *Proc. Natl. Acad. Sci.* **112**, E5503–E5512 (2015).

77. M. W. Snyder, M. Kircher, A. J. Hill, R. M. Daza, J. Shendure, Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin., *Cell* **164**, 57–68 (2016).

78. Y. M. Lo, L. Y. Chan, K. W. Lo, S. F. Leung, J. Zhang, A. T. Chan, J. C. Lee, N. M. Hjelm, P. J. Johnson, D. P. Huang, Quantitative analysis of cell-free Epstein-Barr virus DNA in plasma of patients with nasopharyngeal carcinoma., *Cancer Res.* **59**, 1188–91 (1999).

79. K. C. A. Chan, J. K. S. Woo, A. King, B. C. Y. Zee, W. K. J. Lam, S. L. Chan, S. W. I. Chu, C. Mak, I. O. L. Tse, S. Y. M. Leung, G. Chan, E. P. Hui, B. B. Y. Ma, R. W. K. Chiu, S.-F. Leung, A. C. van Hasselt, A. T. C. Chan, Y. M. D. Lo, Analysis of Plasma Epstein–Barr Virus DNA to Screen for Nasopharyngeal Cancer, *N. Engl. J. Med.* **377**, 513–22 (2017).

80. Z. Kang, S. Stevanović, C. S. Hinrichs, L. Cao, Circulating Cell-free DNA for Metastatic Cervical Cancer Detection, Genotyping, and Monitoring, *Clin. Cancer Res.* **23**, 6856–6862 (2017).

81. A. P. West, G. S. Shadel, Mitochondrial DNA in innate immune responses and inflammatory pathology, *Nat. Rev. Immunol.* **17**, 363–375 (2017).

82. M. J. Hotz, D. Qing, M. G. S. Shashaty, P. Zhang, H. Faust, N. Sondheimer, S. Rivella, G. S. Worthen, N. S. Mangalmurti, Red blood cells homeostatically bind mitochondrial DNA through TLR9 to maintain quiescence and to prevent lung injury, *Am. J. Respir. Crit. Care Med.* **197**, 470–480 (2018).

83. E. S. Morozkin, P. P. Laktionov, E. Y. Rykova, V. V Vlassov, Extracellular Nucleic Acids in Cultures of Long-Term Cultivated Eukaryotic Cells, *Ann. N. Y. Acad. Sci.* **1022**, 244–249 (2004).

84. S. Breitbach, S. Tug, P. Simon, Circulating cell-free DNA: an up-coming molecular marker in exercise physiology., *Sports Med.* **42**, 565–86 (2012).

85. S. Anderson, A. T. Bankier, B. G. Barrell, M. H. L. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden, I. G. Young, Sequence and organization of the human mitochondrial genome, *Nature* **290**, 457–465 (1981).

86. J.-W. Taanman, The mitochondrial genome: structure, transcription, translation and replication, *Biochim. Biophys. Acta - Bioenerg.* **1410**, 103–123 (1999).

87. M. Yu, Somatic mitochondrial DNA mutations in human cancers, *Adv. Clin. Chem.* **57**, 99–138 (2012).

88. S. Zhong, M. C. Y. Ng, Y. M. D. Lo, J. C. N. Chan, P. J. Johnson, Presence of mitochondrial tRNA(Leu(UUR)) A to G 3243 mutation in DNA extracted from serum and plasma of patients with type 2 diabetes mellitus, *J. Clin. Pathol.* **53**, 466–469 (2000).

89. J. Afrifa, T. Zhao, J. Yu, Circulating mitochondria DNA, a non-invasive cancer diagnostic biomarker candidate*Mitochondrion* **47**, 238–243 (2019).

90. M. Yu, Circulating cell-free mitochondrial DNA as a novel cancer biomarker: opportunities and challenges, *Mitochondrial DNA* **23**, 329–332 (2012).

91. Q. Hu, J. Ren, J. Wu, G. Li, X. Wu, S. Liu, G. Wang, G. Gu, J. Li, Elevated Levels of Plasma Mitochondrial DNA Are Associated with Clinical Outcome in Intra-Abdominal Infections Caused by Severe Trauma, *Surg. Infect. (Larchmt).* **18**, 610–618 (2017).

92. D. Lindqvist, O. M. Wolkowitz, M. Picard, L. Ohlsson, F. S. Bersani, J. Fernström, Å. Westrin, C. M. Hough, J. Lin, V. I. Reus, E. S. Epel, S. H. Mellon, Circulating cell-free mitochondrial DNA, but not leukocyte mitochondrial DNA copy number, is elevated in major depressive disorder, *Neuropsychopharmacology* , 1 (2018).

93. A. Pyle, R. Brennan, M. Kurzawa-Akanbi, A. Yarnall, A. Thouin, B. Mollenhauer, D. Burn, P. F. Chinnery, G. Hudson, Reduced cerebrospinal fluid mitochondrial DNA is a biomarker for early-stage Parkinson's disease, *Ann. Neurol.* **78**, 1000–1004 (2015).

94. P. Podlesniy, J. Figueiro-Silva, A. Llado, A. Antonell, R. Sanchez-Valle, D. Alcolea, A. Lleo, J. L. Molinuevo, N. Serra, R. Trullas, Low cerebrospinal fluid concentration of mitochondrial DNA in preclinical Alzheimer disease, *Ann. Neurol.* **74**, 655–668 (2013).

95. H. Lowes, A. Pyle, M. Duddy, G. Hudson, Cell-free mitochondrial DNA in progressive multiple sclerosis, *Mitochondrion* **46**, 307–312 (2019).

96. X. Gu, G. Wu, Y. Yao, J. Zeng, D. Shi, T. Lv, L. Luo, Y. Song, Intratracheal administration of mitochondrial DNA directly provokes lung inflammation through the TLR9-p38 MAPK pathway, *Free Radic. Biol. Med.* **83**, 149–158 (2015).

97. L. Xie, S. Liu, J. Cheng, L. Wang, J. Liu, J. Gong, Exogenous administration of mitochondrial DNA promotes ischemia reperfusion injury via TLR9-p38 MAPK pathway, *Regul. Toxicol. Pharmacol.* **89**, 148–154 (2017).

98. I. Mittra, N. K. Khare, G. V. Raghuram, R. Chaubal, F. Khambatti, D. Gupta, A. Gaikwad, P. Prasannan, A. Singh, A. Iyer, A. Singh, P. Upadhyay, N. K. Nair, P. K. Mishra, A. Dutt, Circulating nucleic acids damage DNA of healthy cells by integrating into their genomes., *J. Biosci.* **40**, 91–111 (2015).

99. B. P. Chelobanov, P. P. Laktionov, V. V Vlasov, Proteins involved in binding and cellular uptake of nucleic acids., *Biochemistry. (Mosc).* **71**, 583–96 (2006).

100. J. Meldolesi, Exosomes and Ectosomes in Intercellular Communication, *Curr. Biol.* **28**, R435–R444 (2018).

101. C. Tetta, E. Ghigo, L. Silengo, M. C. Deregibus, G. Camussi, Extracellular vesicles as an emerging mechanism of cell-to-cell communication, *Endocrine* **44**, 11–19 (2013).

102. A. Waldenström, N. Gennebäck, U. Hellman, G. Ronquist, Cardiomyocyte microvesicles contain DNA/RNA and convey biological messages to target cells, *PLoS One* **7** (2012), doi:10.1371/journal.pone.0034653.

103. A. V. Ermakov, M. S. Konkova, S. V. Kostyuk, N. A. Egolina, L. V. Efremova, N. N. Veiko, Oxidative stress as a significant factor for development of an adaptive response in irradiated and nonirradiated human lymphocytes after inducing the bystander effect by low-dose X-radiation, *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **669**, 155–161 (2009).

104. A. V. Ermakov, M. S. Konkova, S. V. Kostyuk, V. L. Izevskaya, A. Baranova, N. N. Veiko, Oxidized extracellular DNA as a stress signal in human cells*Oxid. Med. Cell. Longev.* (2013), doi:10.1155/2013/649747.

105. A. Picca, A. M. S. Lezza, C. Leeuwenburgh, V. Pesce, R. Calvani, M. Bossola, E. Manes-Gravina, F. Landi, R. Bernabei, E. Marzetti, Circulating Mitochondrial DNA at the Crossroads of Mitochondrial Dysfunction and Inflammation During Aging and Muscle Wasting Disorders, *Rejuvenation Res.* **21**, 350–359 (2018).

106. C. Poli, J. F. Augusto, J. Dauvé, C. Adam, L. Preisser, V. Larochette, P. Pignon, A. Savina, S. Blanchard, J. F. Subra, A. Chevailler, V. Procaccio, A. Croué, C. Créminon, A. Morel, Y. Delneste, H. Fickenscher, P. Jeannin, IL-26 Confers Proinflammatory Properties to Extracellular DNA., *J. Immunol.* **198**, 3650–3661 (2017).

107. A. Bendich, T. Wilczok, E. Borenfreund, Circulating DNA as a possible factor in oncogenesis, *Science (80-.).* **148**, 374–376 (1965).

108. G. V. Raghuram, D. Gupta, S. Subramaniam, A. Gaikwad, N. K. Khare, M. Nobre, N. K. Nair, I. Mittra, Physical shearing imparts biological activity to DNA and ability to transmit itself horizontally across species and kingdom boundaries, *BMC Mol. Biol.* **18**, 21 (2017).

109. D. C. García-Olmo, C. Domínguez, M. García-Arranz, P. Anker, M. Stroun, J. M. García-Verdugo, D. García-Olmo, Cell-free nucleic acids circulating in the plasma of colorectal cancer patients induce the oncogenic transformation of susceptible cultured cells., *Cancer Res.* **70**, 560–7 (2010).

110. C. Trejo-Becerril, E. Pérez-Cárdenas, L. Taja-Chayeb, P. Anker, R. Herrera-Goepfert, L. A. Medina-Velázquez, A. Hidalgo-Miranda, D. Pérez-Montiel, A. Chávez-Blanco, J. Cruz-Velázquez, J. Díaz-Chávez, M. Gaxiola, A. Dueñas-González, B. Lichty, Ed. Cancer Progression Mediated by Horizontal Gene Transfer in an In Vivo Model, *PLoS One* **7**, e52754 (2012).

111. D. García-Olmo, D. C. García-Olmo, J. Ontañón, E. Martinez, Horizontal transfer of DNA and the "genometastasis hypothesis", *Blood* **95**, 724–5 (2000).

112. J. Ehnfors, M. Kost-Alimova, N. L. Persson, A. Bergsmedh, J. Castro, T. Levchenko-Tegnebratt, L. Yang, T. Panaretakis, L. Holmgren, Horizontal transfer of tumor DNA to endothelial cells in vivo, *Cell*

Death Differ. **16**, 749–757 (2009).

113. A. Bergsmedh, A. Szeles, M. Henriksson, A. Bratt, M. J. Folkman, A. L. Spetz, L. Holmgren, Horizontal transfer of oncogenes by uptake of apoptotic bodies., *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6407–11 (2001).

114. R. Tanos, A. R. Thierry, Clinical relevance of liquid biopsy for cancer screening, *Transl. Cancer Res.* **7**, S105–S129 (2018).

115. A. R. Thierry, R. Tanos, La biopsie liquide une voie possible pour le dépistage du cancer, *médecine/sciences* **34**, 824–832 (2018).

116. J. D. Cohen, L. Li, Y. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A. A. Javed, F. Wong, A. Mattox, R. H. Hruban, C. L. Wolfgang, M. G. Goggins, M. Dal Molin, T.-L. Wang, R. Roden, A. P. Klein, J. Ptak, L. Dobbyn, J. Schaefer, N. Silliman, M. Popoli, J. T. Vogelstein, J. D. Browne, R. E. Schoen, R. E. Brand, J. Tie, P. Gibbs, H.-L. Wong, A. S. Mansfield, J. Jen, S. M. Hanash, M. Falconi, P. J. Allen, S. Zhou, C. Bettegowda, L. A. Diaz, C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. M. Lennon, N. Papadopoulos, Detection and localization of surgically resectable cancers with a multi-analyte blood test., *Science* **359**, 926–930 (2018).

117. S. Y. Shen, R. Singhania, G. Fehringer, A. Chakravarthy, M. H. A. Roehrl, D. Chadwick, P. C. Zuzarte, A. Borgida, T. T. Wang, T. Li, O. Kis, Z. Zhao, A. Spreafico, T. da S. Medina, Y. Wang, D. Roulois, I. Ettayebi, Z. Chen, S. Chow, T. Murphy, A. Arruda, G. M. O'Kane, J. Liu, M. Mansour, J. D. McPherson, C. O'Brien, N. Leighl, P. L. Bedard, N. Fleshner, G. Liu, M. D. Minden, S. Gallinger, A. Goldenberg, T. J. Pugh, M. M. Hoffman, S. V. Bratman, R. J. Hung, D. D. De Carvalho, Sensitive tumour detection and classification using plasma cell-free DNA methylomes, *Nature* **563**, 579–583 (2018).

118. H. N. Sallam, N. H. Sallam, S. H. Sallam, Non-invasive methods for embryo selection., *Facts, views Vis. ObGyn* **8**, 87–100 (2016).

119. M. I. Shamonki, H. Jin, Z. Haimowitz, L. Liu, Proof of concept: preimplantation genetic screening without embryo biopsy through analysis of cell-free DNA in spent embryo culture media, *Fertil. Steril.* **106**, 1312–1318 (2016).

120. S. Assou, O. Aït-Ahmed, S. El Messaoudi, A. R. Thierry, S. Hamamah, Non-invasive pre-implantation genetic diagnosis of X-linked disorders, *Med. Hypotheses* **83**, 506–508 (2014).

121. P. R. Brezina, D. S. Brezina, W. G. Kearns, Preimplantation genetic testing., *BMJ* **345**, e5908 (2012).

122. M. Fleischhacker, B. Schmidt, Circulating nucleic acids (CNAs) and cancer—A survey, *Biochim. Biophys. Acta - Rev. Cancer* **1775**, 181–232 (2007).

123. A. J. Bronkhorst, V. Ungerer, S. Holdenrieder, The emerging role of cell-free DNA as a molecular marker for cancer management., *Biomol. Detect. Quantif.* **17**, 100087 (2019).

124. H. Schwarzenbach, D. S. B. Hoon, K. Pantel, Cell-free nucleic acids as biomarkers in cancer patients, *Nat. Rev. Cancer* **11**, 426–437 (2011).

125. S. Holdenrieder, P. Stieber, H. Bodenmüller, M. Busch, J. Von Pawel, A. Schalhorn, D. Nagel, D. Seidel, Circulating nucleosomes in serum., *Ann. N. Y. Acad. Sci.* **945**, 93–102 (2001).

126. F. Mouliere, B. Robert, E. A. Peyrotte, M. Del Rio, M. Ychou, F. Molina, C. Gongora, A. R. Thierry, High Fragmentation Characterizes Tumour-Derived Circulating DNA, *PLoS One* **6**, e23418 (2011).

127. F. Mouliere, S. El Messaoudi, C. Gongora, A.-S. Guedj, B. Robert, M. Del Rio, F. Molina, P.-J. Lamy, E. Lopez-Crapez, M. Mathonnet, M. Ychou, D. Pezet, A. R. Thierry, Circulating Cell-Free DNA from Colorectal Cancer Patients May Reveal High KRAS or BRAF Mutation Load, *Transl. Oncol.* **6**, 319–328 (2013).

128. P. Jiang, C. W. M. Chan, K. C. A. Chan, S. H. Cheng, J. Wong, V. W.-S. Wong, G. L. H. Wong, S. L. Chan, T. S. K. Mok, H. L. Y. Chan, P. B. S. Lai, R. W. K. Chiu, Y. M. D. Lo, Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients, *Proc. Natl. Acad. Sci.* **112**, E1317–E1325 (2015).

129. A. R. Thierry, F. Mouliere, C. Gongora, J. Ollier, B. Robert, M. Ychou, M. Del Rio, F. Molina, Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts, *Nucleic Acids Res.* **38**, 6159–6175 (2010).

130. F. Mouliere, S. El Messaoudi, D. Pang, A. Dritschilo, A. R. Thierry, Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer, *Mol. Oncol.* **8**, 927–41 (2014).

131. H. R. Underhill, J. O. Kitzman, S. Hellwig, N. C. Welker, R. Daza, D. N. Baker, K. M. Gligorich, R. C. Rostomily, M. P. Bronner, J. Shendure, D. J. Kwiatkowski, Ed. Fragment Length of Circulating Tumor DNA, *PLOS Genet.* **12**, e1006162 (2016).

132. C. Sanchez, M. W. Snyder, R. Tanos, J. Shendure, A. R. Thierry, New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA analysis, *npj Genomic Med.* **3**, 31 (2018).

133. S. Hellwig, D. A. Nix, K. M. Gligorich, J. M. O'Shea, A. Thomas, C. L. Fuertes, P. J. Bhetariya, G. T. Marth, M. P. Bronner, H. R. Underhill, V. Adalsteinsson, Ed. Automated size selection for short cell-free DNA fragments enriches for circulating tumor DNA and improves error correction during next generation sequencing, *PLoS One* **13**, e0197333 (2018).

134. F. Mouliere, D. Chandrananda, A. M. Piskorz, E. K. Moore, J. Morris, L. B. Ahlborn, R. Mair, T. Goranova, F. Marass, K. Heider, J. C. M. Wan, A. Supernat, I. Hudecova, I. Gounaris, S. Ros, M. Jimenez-Linan, J. Garcia-Corbacho, K. Patel, O. Østrup, S. Murphy, M. D. Eldridge, D. Gale, G. D. Stewart, J. Burge, W. N. Cooper, M. S. van der Heijden, C. E. Massie, C. Watts, P. Corrie, S. Pacey, K. M. Brindle, R. D. Baird, M. Mau-Sørensen, C. A. Parkinson, C. G. Smith, J. D. Brenton, N. Rosenfeld, Enhanced detection of circulating tumor DNA by fragment size analysis., *Sci. Transl. Med.* **10**, eaat4921 (2018).

135. Y. M. D. Lo, K. C. A. Chan, H. Sun, E. Z. Chen, P. Jiang, F. M. F. Lun, Y. W. Zheng, T. Y. Leung, T. K. Lau, C. R. Cantor, R. W. K. Chiu, Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus., *Sci. Transl. Med.* **2**, 61ra91 (2010).

136. S. C. Y. Yu, K. C. A. Chan, Y. W. L. Zheng, P. Jiang, G. J. W. Liao, H. Sun, R. Akolekar, T. Y. Leung, A. T. J. I. Go, J. M. G. van Vugt, R. Minekawa, C. B. M. Oudejans, K. H. Nicolaides, R. W. K. Chiu, Y. M. D. Lo, Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing., *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8583–8 (2014).

137. Y. W. L. Zheng, K. C. A. Chan, H. Sun, P. Jiang, X. Su, E. Z. Chen, F. M. F. Lun, E. C. W. Hung, V. Lee, J. Wong, P. B. S. Lai, C.-K. Li, R. W. K. Chiu, Y. M. D. Lo, Nonhematopoietically derived DNA is shorter than hematopoietically derived DNA in plasma: a transplantation model., *Clin. Chem.* **58**, 549–58 (2012).

138. D. C. Wallace, Mitochondria and cancer, *Nat. Rev. Cancer* **12**, 685–698 (2012).

139. P. E. Porporato, N. Filigheddu, J. M. B.-S. Pedro, G. Kroemer, L. Galluzzi, Mitochondrial metabolism and cancer, *Cell Res.* **28**, 265–280 (2018).

140. M. Brandon, P. Baldi, D. C. Wallace, Mitochondrial mutations in cancer, *Oncogene* **25**, 4647–4662 (2006).

141. A. Chatterjee, E. Mambo, D. Sidransky, Mitochondrial DNA mutations in human cancer, *Oncogene* **25**, 4663–4674 (2006).

142. E. Reznik, M. L. Miller, Y. Şenbabaoğlu, N. Riaz, J. Sarungbam, S. K. Tickoo, H. A. Al-Ahmadie, W. Lee, V. E. Seshan, A. A. Hakimi, C. Sander, Mitochondrial DNA copy number variation across human cancers, *Elife* **5**, e10769 (2016).

143. J. Ellinger, D. C. Müller, S. C. Müller, S. Hauser, L. C. Heukamp, A. von Ruecker, P. J. Bastian, G. Walgenbach-Brunagel, Circulating mitochondrial DNA in serum: A universal diagnostic biomarker for patients with urological malignancies, *Urol. Oncol. Semin. Orig. Investig.* **30**, 509–515 (2012).

144. J. Ellinger, P. Albers, S. C. Müller, A. Von Ruecker, P. J. Bastian, Circulating mitochondrial DNA in the serum of patients with testicular germ cell cancer as a novel noninvasive diagnostic biomarker, *BJU Int.* **104**, 48–52 (2009).

145. X. Meng, H. Schwarzenbach, Y. Yang, V. Müller, N. Li, D. Tian, Y. Shen, Z. Gong, Circulating Mitochondrial DNA is Linked to Progression and Prognosis of Epithelial Ovarian Cancer, *Transl. Oncol.* **12**, 1213–1220 (2019).

146. J. S. Keserű, B. Soltész, J. Lukács, É. Márton, M. Szilágyi-Bónizs, A. Penyige, R. Póka, B. Nagy, Detection of cell-free, exosomal and whole blood mitochondrial DNA copy number in plasma or whole blood of patients with serous epithelial ovarian cancer, *J. Biotechnol.* **298**, 76–81 (2019).

147. M. Yu, Y.-F. Wan, Q.-H. Zou, Cell-free Circulating Mitochondrial DNA in the Serum: A Potential Non-invasive Biomarker for Ewing's Sarcoma, *Arch. Med. Res.* **43**, 389–394 (2012).

148. L. Li, H.-W. Hann, S. Wan, R. S. Hann, C. Wang, Y. Lai, X. Ye, A. Evans, R. E. Myers, Z. Ye, B. Li, J. Xing, H. Yang, Cell-free circulating mitochondrial DNA content and risk of hepatocellular carcinoma in patients with chronic HBV infection, *Sci. Rep.* **6**, 23992 (2016).

149. C. Kohler, R. Radpour, Z. Barekati, R. Asadollahi, J. Bitzer, E. Wight, N. Bürki, C. Diesch, W. Holzgreve, X. Y. Zhong, Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors, *Mol. Cancer* **8**, 105 (2009).

150. R. Meddeb, Z. A. A. Dache, S. Thezenas, A. Otandault, R. Tanos, B. Pastor, C. Sanchez, J. Azzi, G. Tousch, S. Azan, C. Mollevi, A. Adenis, S. El Messaoudi, P. Blache, A. R. Thierry, Quantifying circulating cell-free DNA in humans, *Sci. Reports 2019 91* **9**, 5220 (2019).

151. H. Lu, J. Busch, M. Jung, S. Rabenhorst, B. Ralla, E. Kilic, S. Mergemeier, N. Budach, A. Fendler, K. Jung, Diagnostic and prognostic potential of circulating cell-free genomic and mitochondrial DNA fragments in clear cell renal cell carcinoma patients, *Clin. Chim. Acta* **452**, 109–119 (2016).

152. J. C. M. Wan, C. Massie, J. Garcia-Corbacho, F. Mouliere, J. D. Brenton, C. Caldas, S. Pacey, R. Baird, N. Rosenfeld, Liquid biopsies come of age: towards implementation of circulating tumour DNA, *Nat. Rev. Cancer* **17**, 223–38 (2017).

153. C. A. Parkinson, D. Gale, A. M. Piskorz, H. Biggs, C. Hodgkin, H. Addley, S. Freeman, P. Moyle, E. Sala, K. Sayal, K. Hosking, I. Gounaris, M. Jimenez-Linan, H. M. Earl, W. Qian, N. Rosenfeld, J. D. Brenton, E. R. Mardis, Ed. Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study, *PLOS Med.* **13**, e1002198 (2016).

154. C. Abbosh, N. J. Birkbak, G. A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, J. Le Quesne, D. A. Moore, S. Veeriah, R. Rosenthal, T. Marafioti, E. Kirkizlar, T. B. K. Watkins, N. McGranahan, S. Ward, L. Martinson, J. Riley, F. Fraioli, M. Al Bakir, E. Grönroos, F. Zambrana, R. Endozo, W. L. Bi, F. M. Fennessy, N. Sponer, D. Johnson, J. Laycock, S. Shafi, J. Czyzewska-Khan, A. Rowan, T. Chambers, N. Matthews, S. Turajlic, C. Hiley, S. M. Lee, M. D. Forster, T. Ahmad, M. Falzon, E. Borg, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, D. Hafez, A. Naik, A. Ganguly, S. Kareht, R. Shah, L. Joseph, A. Marie Quinn, P. A. Crosbie, B. Naidu, G.

Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, D. Oukrif, A. U. Akarca, J. A. Hartley, H. L. Lowe, S. Lock, N. Iles, H. Bell, Y. Ngai, G. Elgar, Z. Szallasi, R. F. Schwarz, J. Herrero, A. Stewart, S. A. Quezada, K. S. Peggs, P. Van Loo, C. Dive, C. J. Lin, M. Rabinowitz, H. J. W. L. Aerts, A. Hackshaw, J. A. Shaw, B. G. Zimmermann, C. Swanton, C. Swanton, M. Jamal-Hanjani, C. Abbosh, S. Veeriah, S. Shafi, J. Czyzewska-Khan, D. Johnson, J. Laycock, L. Bosshard-Carter, G. Goh, R. Rosenthal, P. Gorman, N. Murugaesu, R. E. Hynds, G. A. Wilson, N. J. Birkbak, T. B. K. Watkins, N. McGranahan, S. Horswell, M. Al Bakir, E. Grönroos, R. Mitter, M. Escudero, A. Stewart, P. Van Loo, A. Rowan, H. Xu, S. Turajlic, C. Hiley, J. Goldman, R. K. Stone, T. Denner, N. Matthews, G. Elgar, S. Ward, J. Biggs, M. Costa, S. Begum, B. Phillimore, T. Chambers, E. Nye, S. Graca, K. Joshi, A. Furness, A. Ben Aissa, Y. N. S. Wong, A. Georgiou, S. A. Quezada, K. S. Peggs, J. A. Hartley, H. L. Lowe, J. Herrero, D. Lawrence, M. Hayward, N. Panagiotopoulos, S. Kolvekar, M. Falzon, E. Borg, T. Marafioti, C. Simeon, G. Hector, A. Smith, M. Aranda, M. Novelli, D. Oukrif, A. U. Akarca, S. M. Janes, R. Thakrar, M. D. Forster, T. Ahmad, S. M. Lee, D. Papadatos-Pastos, D. Carnell, R. Mendes, J. George, N. Navani, A. Ahmed, M. Taylor, J. Choudhary, Y. Summers, R. Califano, P. Taylor, R. Shah, P. Krysiak, K. Rammohan, E. Fontaine, R. Booton, M. Evison, P. A. Crosbie, S. Moss, F. Idries, L. Joseph, P. Bishop, A. Chaturvedi, A. M. Quinn, H. Doran, A. Leek, P. Harrison, K. Moore, R. Waddington, J. Novasio, F. Blackhall, J. Rogan, E. Smith, C. Dive, J. Tugwood, G. Brady, D. G. Rothwell, F. Chemi, J. Pierce, S. Gulati, B. Naidu, G. Langman, S. Trotter, M. Bellamy, H. Bancroft, A. Kerr, S. Kadiri, J. Webb, G. Middleton, M. Djearaman, D. A. Fennell, J. A. Shaw, J. Le Quesne, D. A. Moore, A. Thomas, H. Walter, J. Riley, L. Martinson, A. Nakas, S. Rathinam, W. Monteiro, H. Marshall, L. Nelson, J. Bennett, L. Primrose, G. Anand, S. Khan, A. Amadi, M. Nicolson, K. Kerr, S. Palmer, H. Remmen, J. Miller, K. Buchan, M. Chetty, L. Gomersall, J. F. Lester, A. Edwards, F. Morgan, H. Adams, H. Davies, M. Kornaszewska, R. Attanoos, S. Lock, A. Verjee, M. MacKenzie, M. Wilcox, H. Bell, N. Iles, A. Hackshaw, Y. Ngai, S. Smith, N. Gower, C. Ottensmeier, S. Chee, B. Johnson, A. Alzetani, E. Shaw, E. Lim, P. De Sousa, M. T. Barbosa, A. Bowman, S. Jordan, A. Rice, H. Raubenheimer, C. Proli, M. E. Cufari, J. C. Ronquillo, A. Kwayie, H. Bhayani, M. Hamilton, Y. Bakar, N. Mensah, L. Ambrose, A. Devaraj, S. Buderi, J. Finch, L. Azcarate, H. Chavan, S. Green, H. Mashinga, A. G. Nicholson, K. Lau, M. Sheaff, P. Schmid, J. Conibear, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, P. Russell, T. Light, T. Horey, S. Danson, J. Bury, J. Edwards, J. Hill, S. Matthews, Y. Kitsanta, K. Suvarna, P. Fisher, A. D. Keerio, M. Shackcloth, J. Gosney, P. Postmus, S. Feeney, J. Asante-Siaw, T. Constantin, R. Salari, N. Sponer, A. Naik, B. G. Zimmermann, M. Rabinowitz, H. J. W. L. Aerts, S. Dentro, C. Dessimoz, T. P. consortium, C. Swanton, Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution, *Nature* **545**, 446–451 (2017).

155. C. Bettegowda, M. Sausen, R. J. Leary, I. Kinde, Y. Wang, N. Agrawal, B. R. Bartlett, H. Wang, B. Luber, R. M. Alani, E. S. Antonarakis, N. S. Azad, A. Bardelli, H. Brem, J. L. Cameron, C. C. Lee, L. A. Fecher, G. L. Gallia, P. Gibbs, D. Le, R. L. Giuntoli, M. Goggins, M. D. Hogarty, M. Holdhoff, S.-M. Hong, Y. Jiao, H. H. Juhl, J. J. Kim, G. Siravegna, D. A. Laheru, C. Lauricella, M. Lim, E. J. Lipson, S. K. N. Marie, G. J. Netto, K. S. Oliner, A. Olivi, L. Olsson, G. J. Riggins, A. Sartore-Bianchi, K. Schmidt, L.-M. Shih, S. M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T. T. Harkins, S. Veronese, T.-L. Wang, J. D. Weingart, C. L. Wolfgang, L. D. Wood, D. Xing, R. H. Hruban, J. Wu, P. J. Allen, C. M. Schmidt, M. A. Choti, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, N. Papadopoulos, L. A. Diaz, Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies, *Sci. Transl. Med.* **6**, 224ra24 (2014).

156. A. M. Newman, S. V Bratman, J. To, J. F. Wynne, N. C. W. Eclov, L. A. Modlin, C. L. Liu, J. W. Neal, H. A. Wakelee, R. E. Merritt, J. B. Shrager, B. W. Loo Jr, A. A. Alizadeh, M. Diehn, An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage, *Nat. Med.* **20**, 548–554 (2014).

157. S. El Messaoudi, F. Mouliere, S. Du Manoir, C. Bascoul-Mollevi, B. Gillet, M. Nouaille, C. Fiess, E. Crapez, F. Bibeau, C. Theillet, T. Mazard, D. Pezet, M. Mathonnet, M. Ychou, A. R. Thierry, Circulating DNA as a Strong Multimarker Prognostic Tool for Metastatic Colorectal Cancer Patient Management

Care, *Clin. Cancer Res.* **22**, 3067–3077 (2016).

158. D. Pietrasz, N. Pécuchet, F. Garlan, A. Didelot, O. Dubreuil, S. Doat, F. Imbert-Bismut, M. Karoui, J.-C. Vaillant, V. Taly, P. Laurent-Puig, J.-B. Bachet, Plasma Circulating Tumor DNA in Pancreatic Cancer Patients Is a Prognostic Marker, *Clin. Cancer Res.* **23**, 116–123 (2017).

159. M.-H. Delfau-Larue, A. van der Gucht, J. Dupuis, J.-P. Jais, I. Nel, A. Beldi-Ferchiou, S. Hamdane, I. Benmaad, G. Laboure, B. Verret, C. Haioun, C. Copie-Bergman, A. Berriolo-Riedinger, P. Robert, R.-O. Casasnovas, E. Itti, Total metabolic tumor volume, circulating tumor cells, cell-free DNA: distinct prognostic value in follicular lymphoma., *Blood Adv.* **2**, 807–816 (2018).

160. S. Valpione, G. Gremel, P. Mundra, P. Middlehurst, E. Galvani, M. R. Girotti, R. J. Lee, G. Garner, N. Dhomen, P. C. Lorigan, R. Marais, Plasma total cell-free DNA (cfDNA) is a surrogate biomarker for tumour burden and a prognostic biomarker for survival in metastatic melanoma patients, *Eur. J. Cancer* **88**, 1–9 (2018).

161. A. Lièvre, J.-B. Bachet, D. Le Corre, V. Boige, B. Landi, J.-F. Emile, J.-F. Côté, G. Tomasic, C. Penna, M. Ducreux, P. Rougier, F. Penault-Llorca, P. Laurent-Puig, *KRAS* Mutation Status Is Predictive of Response to Cetuximab Therapy in Colorectal Cancer, *Cancer Res.* **66**, 3992–3995 (2006).

162. R. G. Amado, M. Wolf, M. Peeters, E. Van Cutsem, S. Siena, D. J. Freeman, T. Juan, R. Sikorski, S. Suggs, R. Radinsky, S. D. Patterson, D. D. Chang, Wild-Type *KRAS* Is Required for Panitumumab Efficacy in Patients With Metastatic Colorectal Cancer, *J. Clin. Oncol.* **26**, 1626–1634 (2008).

163. S. Chang-Hao Tsao, J. Weiss, C. Hudson, C. Christophi, J. Cebon, A. Behren, A. Dobrovic, Monitoring response to therapy in melanoma by quantifying circulating tumour DNA with droplet digital PCR for BRAF and NRAS mutations, *Sci. Rep.* **5**, 11198 (2015).

164. F. Diehl, K. Schmidt, M. A. Choti, K. Romans, S. Goodman, M. Li, K. Thornton, N. Agrawal, L. Sokoll, S. A. Szabo, K. W. Kinzler, B. Vogelstein, L. A. Diaz Jr, Circulating mutant DNA to assess tumor dynamics, *Nat. Med.* **14**, 985–990 (2008).

165. J. Tie, I. Kinde, Y. Wang, H. L. Wong, J. Roebert, M. Christie, M. Tacey, R. Wong, M. Singh, C. S. Karapetis, J. Desai, B. Tran, R. L. Strausberg, L. A. Diaz, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, P. Gibbs, Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer, *Ann. Oncol.* **26**, 1715–1722 (2015).

166. S.-J. Dawson, D. W. Y. Tsui, M. Murtaza, H. Biggs, O. M. Rueda, S.-F. Chin, M. J. Dunning, D. Gale, T. Forshew, B. Mahler-Araujo, S. Rajan, S. Humphray, J. Becq, D. Halsall, M. Wallis, D. Bentley, C. Caldas, N. Rosenfeld, Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer, *N. Engl. J. Med.* **368**, 1199–1209 (2013).

167. L. Cabel, F. Riva, V. Servois, A. Livartowski, C. Daniel, A. Rampanou, O. Lantz, E. Romano, M. Milder, B. Buecher, S. Piperno-Neumann, V. Bernard, S. Baulande, I. Bieche, J. Y. Pierga, C. Proudhon, F.-C. Bidard, Circulating tumor DNA changes for early monitoring of anti-PD1 immunotherapy: a proof-of-concept study, *Ann. Oncol.* **28**, 1996–2001 (2017).

168. R. Raja, M. Kuziora, P. Z. Brohawn, B. W. Higgs, A. Gupta, P. A. Dennis, K. Ranade, Early Reduction in ctDNA Predicts Survival in Patients with Lung and Bladder Cancer Treated with Durvalumab, *Clin. Cancer Res.* **24**, 6212–6222 (2018).

169. A. Marchetti, J. F. Palma, L. Felicioni, T. M. De Pas, R. Chiari, M. Del Grammastro, G. Filice, V. Ludovini, A. Brandes, A. Chella, F. Malorgio, F. Guglielmi, M. De Tursi, A. Santoro, L. Crinò, F. Buttitta, Early Prediction of Response to Tyrosine Kinase Inhibitors by Quantification of EGFR Mutations in Plasma of NSCLC Patients, *J. Thorac. Oncol.* **10**, 1437–1443 (2015).

170. N. Boeckx, K. Op de Beeck, M. Beyens, V. Deschoolmeester, C. Hermans, P. De Clercq, S. Garrigou,

C. Normand, E. Monsaert, K. Papadimitriou, P. Laurent-Puig, P. Pauwels, G. Van Camp, V. Taly, M. Peeters, Mutation and Methylation Analysis of Circulating Tumor DNA Can Be Used for Follow-up of Metastatic Colorectal Cancer Patients., *Clin. Colorectal Cancer* **17**, e369–e379 (2018).

171. F. Garlan, P. Laurent-Puig, D. Sefrioui, N. Siauve, A. Didelot, N. Sarafan-Vasseur, P. Michel, G. Perkins, C. Mulot, H. Blons, J. Taieb, F. Di Fiore, V. Taly, A. Zaanan, Early Evaluation of Circulating Tumor DNA as Marker of Therapeutic Efficacy in Metastatic Colorectal Cancer Patients (PLACOL Study), *Clin. Cancer Res.* **23**, 5416–5425 (2017).

172. L. A. Diaz, A. Bardelli, Liquid Biopsies: Genotyping Circulating Tumor DNA, *J. Clin. Oncol.* **32**, 579–86 (2014).

173. L. A. Diaz Jr, R. T. Williams, J. Wu, I. Kinde, J. R. Hecht, J. Berlin, B. Allen, I. Bozic, J. G. Reiter, M. A. Nowak, K. W. Kinzler, K. S. Oliner, B. Vogelstein, The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers, *Nature* **486**, 537–540 (2012).

174. S. Misale, R. Yaeger, S. Hobor, E. Scala, M. Janakiraman, D. Liska, E. Valtorta, R. Schiavo, M. Buscarino, G. Siravegna, K. Bencardino, A. Cercek, C.-T. Chen, S. Veronese, C. Zanon, A. Sartore-Bianchi, M. Gambacorta, M. Gallicchio, E. Vakiani, V. Boscaro, E. Medico, M. Weiser, S. Siena, F. Di Nicolantonio, D. Solit, A. Bardelli, Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer, *Nature* **486**, 532–536 (2012).

175. B. O. Van Emburgh, S. Arena, G. Siravegna, L. Lazzari, G. Crisafulli, G. Corti, B. Mussolin, F. Baldi, M. Buscarino, A. Bartolini, E. Valtorta, J. Vidal, B. Bellosillo, G. Germano, F. Pietrantonio, A. Ponzetti, J. Albanell, S. Siena, A. Sartore-Bianchi, F. Di Nicolantonio, C. Montagut, A. Bardelli, Acquired RAS or EGFR mutations and duration of response to EGFR blockade in colorectal cancer, *Nat. Commun.* **7**, 13665 (2016).

176. M. Murtaza, S.-J. Dawson, D. W. Y. Tsui, D. Gale, T. Forshew, A. M. Piskorz, C. Parkinson, S.-F. Chin, Z. Kingsbury, A. S. C. Wong, F. Marass, S. Humphray, J. Hadfield, D. Bentley, T. M. Chin, J. D. Brenton, C. Caldas, N. Rosenfeld, Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA, *Nature* **497**, 108–112 (2013).

177. K. Pantel, C. Alix-Panabières, Liquid biopsy and minimal residual disease — latest advances and implications for cure, *Nat. Rev. Clin. Oncol.* **16**, 409–424 (2019).

178. J. Tie, Y. Wang, C. Tomasetti, L. Li, S. Springer, I. Kinde, N. Silliman, M. Tacey, H.-L. Wong, M. Christie, S. Kosmider, I. Skinner, R. Wong, M. Steel, B. Tran, J. Desai, I. Jones, A. Haydon, T. Hayes, T. J. Price, R. L. Strausberg, L. A. Diaz, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, P. Gibbs, Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer., *Sci. Transl. Med.* **8**, 346ra92 (2016).

179. M. J. Overman, J.-N. Vauthey, T. A. Aloia, C. Conrad, Y. S. Chun, A. A. L. Pereira, Z. Jiang, S. Crosby, S. Wei, K. P. S. Raghav, V. K. Morris, M. Tan, A. Maslan, A. Talasaz, S. Mortimer, S. Kopetz, Circulating tumor DNA (ctDNA) utilizing a high-sensitivity panel to detect minimal residual disease post liver hepatectomy and predict disease recurrence., *J. Clin. Oncol.* **35**, 3522–3522 (2017).

180. Y.-H. Chen, B. A. Hancock, J. P. Solzak, D. Brinza, C. Scafe, K. D. Miller, M. Radovich, Next-generation sequencing of circulating tumor DNA to predict recurrence in triple-negative breast cancer patients with residual disease after neoadjuvant chemotherapy, *NPJ Breast Cancer* **3** (2017), doi:10.1038/s41523-017-0028-4.

181. S. Silva, S. Danson, D. Teare, F. Taylor, J. Bradford, A. J. G. McDonagh, A. Salawu, G. Wells, G. J. Burghel, I. Brock, D. Connley, H. Cramp, D. Hughes, N. Tiffin, A. Cox, Genome-Wide Analysis of Circulating Cell-Free DNA Copy Number Detects Active Melanoma and Predicts Survival., *Clin. Chem.* **64**, 1338–1346 (2018).

182. S. Y. Lin, S. K. Huang, K. T. Huynh, M. P. Salomon, S.-C. Chang, D. M. Marzese, R. B. Lanman, A. Talasaz, D. S. B. Hoon, Multiplex Gene Profiling of Cell-Free DNA in Patients With Metastatic Melanoma for Monitoring Disease, *JCO Precis. Oncol.* , 1–30 (2018).

183. G. Leszinski, J. Lehner, U. Gezer, S. Holdenrieder, Increased DNA Integrity in Colorectal Cancer, *In Vivo (Brooklyn).* **28**, 299–303 (2014).

184. A. M. Aravanis, M. Lee, R. D. Klausner, Next-Generation Sequencing of Circulating Tumor DNA for Early Cancer Detection, *Cell* **168**, 571–574 (2017).

185. J. Phallen, M. Sausen, V. Adleff, A. Leal, C. Hruban, J. White, V. Anagnostou, J. Fiksel, S. Cristiano, E. Papp, S. Speir, T. Reinert, M.-B. W. Orntoft, B. D. Woodward, D. Murphy, S. Parpart-Li, D. Riley, M. Nesselbush, N. Sengamalay, A. Georgiadis, Q. K. Li, M. R. Madsen, F. V. Mortensen, J. Huiskens, C. Punt, N. van Grieken, R. Fijneman, G. Meijer, H. Husain, R. B. Scharpf, L. A. Diaz, S. Jones, S. Angiuoli, T. Ørntoft, H. J. Nielsen, C. L. Andersen, V. E. Velculescu, Direct detection of early-stage cancers using circulating tumor DNA, *Sci. Transl. Med.* **9**, eaan2415 (2017).

186. H. Tang, E. T. Donnell, Application of a model-based recursive partitioning algorithm to predict crash frequency, *Accid. Anal. Prev.* **132**, 105274 (2019).

187. S. S. Panesar, R. N. D'Souza, F.-C. Yeh, J. C. Fernandez-Miranda, Machine Learning Versus Logistic Regression Methods for 2-Year Mortality Prognostication in a Small, Heterogeneous Glioma Database, *World Neurosurg. X* **2**, 100012 (2019).

188. G. Wichmann, C. Gaede, S. Melzer, J. Bocsi, S. Henger, C. Engel, K. Wirkner, J. R. Wenning, T. Wald, J. Freitag, M. Willner, M. Kolb, S. Wiegand, M. Löffler, A. Dietz, A. Tárnok, Discrimination of Head and Neck Squamous Cell Carcinoma Patients and Healthy Adults by 10-Color Flow Cytometry: Development of a Score Based on Leukocyte Subsets, *Cancers (Basel).* **11**, 814 (2019).

189. J. T. Jordan, D. E. McNiel, Characteristics of a suicide attempt predict who makes another attempt after hospital discharge: A decision-tree investigation, *Psychiatry Res.* **268**, 317–322 (2018).

190. Q. Q. Tiet, Y. E. Leyva, R. H. Moos, S. M. Frayne, L. Osterberg, B. Smith, Screen of Drug Use: Diagnostic Accuracy of a New Breif Tool for Primary Care, *JAMA Intern. Med.* **175**, 1371 (2015).

191. A. A. I. Sina, L. G. Carrascosa, Z. Liang, Y. S. Grewal, A. Wardiana, M. J. A. Shiddiky, R. A. Gardiner, H. Samaratunga, M. K. Gandhi, R. J. Scott, D. Korbie, M. Trau, Epigenetically reprogrammed methylation landscape drives the DNA self-assembly and serves as a universal cancer biomarker, *Nat. Commun.* **9**, 4915 (2018).

192. S. Tug, S. Helmig, E. R. Deichmann, A. Schmeier-Jürchott, E. Wagner, T. Zimmermann, M. Radsak, M. Giacca, P. Simon, Exercise-induced increases in cell free DNA in human plasma originate predominantly from cells of the haematopoietic lineage., *Exerc. Immunol. Rev.* **21**, 164–73 (2015).

193. M. M. Gil, M. S. Quezada, R. Revello, R. Akolekar, K. H. Nicolaides, Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis, *Ultrasound Obstet. Gynecol.* **45**, 249–266 (2015).

194. Y. M. D. Lo, M. S. C. Tein, T. K. Lau, C. J. Haines, T. N. Leung, P. M. K. Poon, J. S. Wainscoat, P. J. Johnson, A. M. Z. Chang, N. M. Hjelm, Quantitative Analysis of Fetal DNA in Maternal Plasma and Serum: Implications for Noninvasive Prenatal Diagnosis, *Am. J. Hum. Genet.* **62**, 768–775 (1998).

195. Y. M. Lo, J. Zhang, T. N. Leung, T. K. Lau, A. M. Chang, N. M. Hjelm, Rapid clearance of fetal DNA from maternal plasma., *Am. J. Hum. Genet.* **64**, 218–24 (1999).

196. V. C. Y. Lee, J. F. C. Chow, W. S. B. Yeung, Preimplantation genetic diagnosis for monogenic diseases, *Best Pract. Res. Clin. Obstet. Gynaecol.* **44**, 68–75 (2017).

197. S. C. Harris, N. L. Vora, Noninvasive Prenatal Testing: An Update on Prenatal Screening and Testing Options, *Neoreviews* **15**, e7–e16 (2014).

198. W. Tungwiwat, G. Fucharoen, T. Ratanasiri, K. Sanchaisuriya, S. Fucharoen, Non-invasive fetal sex determination using a conventional nested PCR analysis of fetal DNA in maternal plasma, *Clin. Chim. Acta* **334**, 173–177 (2003).

199. J.-M. Costa, A. Benachi, E. Gautier, New Strategy for Prenatal Diagnosis of X-Linked Disorders, *N. Engl. J. Med.* **346**, 1502–1502 (2002).

200. C. Chi, J. Hyett, K. Finning, C. Lee, R. Kadir, Non-invasive first trimester determination of fetal gender: a new approach for prenatal diagnosis of haemophilia, *BJOG An Int. J. Obstet. Gynaecol.* **113**, 239–242 (2006).

201. X. Y. Zhong, W. Holzgreve, S. Hahn, Detection of fetal Rhesus D and sex using fetal DNA from maternal plasma by multiplex polymerase chain reaction, *BJOG An Int. J. Obstet. Gynaecol.* **107**, 766–769 (2000).

202. R. W. K. Chiu, K. C. A. Chan, Y. Gao, V. Y. M. Lau, W. Zheng, T. Y. Leung, C. H. F. Foo, B. Xie, N. B. Y. Tsui, F. M. F. Lun, B. C. Y. Zee, T. K. Lau, C. R. Cantor, Y. M. D. Lo, Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma, *Proc. Natl. Acad. Sci.* **105**, 20458–20463 (2008).

203. R. W. K. Chiu, R. Akolekar, Y. W. L. Zheng, T. Y. Leung, H. Sun, K. C. A. Chan, F. M. F. Lun, A. T. J. I. Go, E. T. Lau, W. W. K. To, W. C. Leung, R. Y. K. Tang, S. K. C. Au-Yeung, H. Lam, Y. Y. Kung, X. Zhang, J. M. G. van Vugt, R. Minekawa, M. H. Y. Tang, J. Wang, C. B. M. Oudejans, T. K. Lau, K. H. Nicolaides, Y. M. D. Lo, Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study., *BMJ* **342**, c7401 (2011).

204. H. Zhang, Y. Gao, F. Jiang, M. Fu, Y. Yuan, Y. Guo, Z. Zhu, M. Lin, Q. Liu, Z. Tian, H. Zhang, F. Chen, T. K. Lau, L. Zhao, X. Yi, Y. Yin, W. Wang, Non-invasive prenatal testing for trisomies 21, 18 and 13: clinical experience from 146 958 pregnancies, *Ultrasound Obstet. Gynecol.* **45**, 530–538 (2015).

205. K. Gorzelnik, J. Bijok, J. G. Zimowski, G. Jakiel, T. Roszkowski, [Noninvasive prenatal diagnosis of trisomy 21, 18 and 13 using cell-free fetal DNA], *Ginekol. Pol.* **84**, 714–719 (2013).

206. E. Z. Chen, R. W. K. Chiu, H. Sun, R. Akolekar, K. C. A. Chan, T. Y. Leung, P. Jiang, Y. W. L. Zheng, F. M. F. Lun, L. Y. S. Chan, Y. Jin, A. T. J. I. Go, E. T. Lau, W. W. K. To, W. C. Leung, R. Y. K. Tang, S. K. C. Au-Yeung, H. Lam, Y. Y. Kung, X. Zhang, J. M. G. van Vugt, R. Minekawa, M. H. Y. Tang, J. Wang, C. B. M. Oudejans, T. K. Lau, K. H. Nicolaides, Y. M. D. Lo, Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma dna sequencing, *PLoS One* **6** (2011), doi:10.1371/journal.pone.0021791.

207. M. Allyse, M. A. Minear, E. Berson, S. Sridhar, M. Rote, A. Hung, S. Chandrasekharan, Non-invasive prenatal testing: A review of international implementation and challenges*Int. J. Womens. Health* **7**, 113–126 (2015).

208. Le diagnostic prénatal (DPN) - Vaincre les Maladies Lysosomales (available at https://www.vml-asso.org/le-diagnostic-prenatal-dpn).

209. M. Phylipsen, S. Yamsri, E. E. Treffers, D. T. S. L. Jansen, W. A. Kanhai, E. M. J. Boon, P. C. Giordano, S. Fucharoen, E. Bakker, C. L. Harteveld, Non-invasive prenatal diagnosis of beta-thalassemia and sickle-cell disease using pyrophosphorolysis-activated polymerization and melting curve analysis, *Prenat. Diagn.* **32**, 578–587 (2012).

210. L. Xiong, A. N. Barrett, R. Hua, T. Z. Tan, S. S. Y. Ho, J. K. Y. Chan, M. Zhong, M. Choolani, Non-invasive prenatal diagnostic testing for β-thalassaemia using cell-free fetal DNA and next generation

sequencing, *Prenat. Diagn.* **35**, 258–265 (2015).

211. M. Hill, P. Twiss, T. I. Verhoef, S. Drury, F. McKay, S. Mason, L. Jenkins, S. Morris, L. S. Chitty, Non-invasive prenatal diagnosis for cystic fibrosis: detection of paternal mutations, exploration of patient preferences and cost analysis, *Prenat. Diagn.* **35**, 950–958 (2015).

212. C. Guissart, C. Dubucs, C. Raynal, A. Girardet, F. Tran Mau Them, V. Debant, C. Rouzier, A. Boureau-Wirth, E. Haquet, J. Puechberty, E. Bieth, D. Dupin Deguine, P. Khau Van Kien, M. P. Brechard, V. Pritchard, M. Koenig, M. Claustres, M. C. Vincent, Non-invasive prenatal diagnosis (NIPD) of cystic fibrosis: an optimized protocol using MEMO fluorescent PCR to detect the p.Phe508del mutation, *J. Cyst. Fibros.* **16**, 198–206 (2017).

213. L. S. Chitty, Y. M. D. Lo, Noninvasive Prenatal Screening for Genetic Diseases Using Massively Parallel Sequencing of Maternal Plasma DNA, *Cold Spring Harb. Perspect. Med.* **5**, a023085 (2015).

214. F. C. K. Wong, Y. M. D. Lo, Prenatal Diagnosis Innovation: Genome Sequencing of Maternal Plasma, *Annu. Rev. Med.* **67**, 419–432 (2016).

215. Les étapes de la fécondation - Doctissimo (available at http://www.doctissimo.fr/html/sante/encyclopedie/sa_2221_fecondation_naturelle.htm).

216. fertility treatment BLOG – Sperm qualityAssessment and significance in fertility treatments (available at https://www.fertility-treatment-blog.com/sperm-qualityassessment-and-significance-in-fertility-treatments/).

217. The Stages of Embryo Development — Altora Health (available at https://altorahealth.com/the-stages-of-embryo-development/).

218. Quelle est la différence entre zygote, embryon et fœtus? (available at https://www.invitra.com/fr/differences-entre-zygote-embryon-et-foetus/).

219. Fécondation in vitro - Expert-Esthetique (available at https://expert-esthetique.com/fecondation-in-vitro/).

220. PMA : les techniques de procréation médicalement assistée | PARENTS.fr (available at https://www.parents.fr/envie-de-bebe/pma/pma-les-techniques-de-procreation-medicalement-assistee-77869).

221. S. X. Y. Wang, The past, present, and future of embryo selection in in vitro fertilization: Frontiers in Reproduction Conference., *Yale J. Biol. Med.* **84**, 487–90 (2011).

222. Fertility Specialist - Access Genomics (available at http://accessgenomics.ca/fertility-physician/).

223. Preimplantation genetic testing - UpToDate (available at https://www.uptodate.com/contents/preimplantation-genetic-testing).

224. A. Y. Wang, E. A. Sullivan, Z. Li, C. Farquhar, Day 5 versus day 3 embryo biopsy for preimplantation genetic testing for monogenic/single gene defects, *Cochrane Database Syst. Rev.* (2018), doi:10.1002/14651858.CD013233.

225. F. R. Parikh, A. S. Athalye, N. J. Naik, D. J. Naik, R. R. Sanap, P. F. Madon, Preimplantation Genetic Testing: Its Evolution, Where Are We Today?, *J. Hum. Reprod. Sci.* **11**, 306–314 (2018).

226. K. L. Scott, K. H. Hong, R. T. Scott, Selecting the optimal time to perform biopsy for preimplantation genetic testing, *Fertil. Steril.* **100**, 608–614 (2013).

227. D. K. Griffin, L. J. Wilton, A. H. Handyside, G. H. Atkinson, R. M. Winston, J. D. Delhanty, Diagnosis of sex in preimplantation embryos by fluorescent in situ hybridisation., *BMJ* **306**, 1382 (1993).

228. A. M. Gronowski, R. T. Scott, A. L. Caplan, L. J. Nelson, The ethical implications of preimplantation genetic diagnosis., *Clin. Chem.* **60**, 25–8 (2014).

229. S. Rechitsky, A. Kuliev, Preimplantation Genetic Testing (PGT) for non-traditional indications, *Reprod. Biomed. Online* **38**, e3 (2019).

230. Y. Verlinsky, S. Rechitsky, T. Sharapova, R. Morris, M. Taranissi, A. Kuliev, Preimplantation HLA Testing, *JAMA* **291**, 2079 (2004).

231. G. G. K. MUKHERJEE, *Practical guide in andrology and embryology.* (JAYPEE Brothers MEDICAL P, 2018; https://books.google.fr/books?hl=en&lr=&id=_3KSDwAAQBAJ&oi=fnd&pg=PA255&dq=embryo+viability+assessment+and+hla&ots=X-zJ9pIcnf&sig=YSz_fFN8LGEyLLe2qbHjkOYU9_0#v=onepage&q=embryo viability assessment and hla&f=false).

232. S. Mastenbroek, F. van der Veen, A. Aflatoonian, B. Shapiro, P. Bossuyt, S. Repping, Embryo selection in IVF, *Hum. Reprod.* **26**, 964–966 (2011).

233. M. OMIDI, A. FARAMARZI, A. AGHARAHIMI, M. A. KHALILI, Noninvasive imaging systems for gametes and embryo selection in IVF programs: a review, *J. Microsc.* **267**, 253–264 (2017).

234. P. Kovacs, Embryo selection: the role of time-lapse monitoring, *Reprod. Biol. Endocrinol.* **12** (2014), doi:10.1186/1477-7827-12-124.

235. P. Khosravi, E. Kazemi, Q. Zhan, J. E. Malmsten, M. Toschi, P. Zisimopoulos, A. Sigaras, S. Lavery, L. A. D. Cooper, C. Hickman, M. Meseguer, Z. Rosenwaks, O. Elemento, N. Zaninovic, I. Hajirasouliha, Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization, *npj Digit. Med.* **2**, 21 (2019).

236. R. L. Krisher, W. B. Schoolcraft, M. G. Katz-Jaffe, Omics as a window to view embryo viability*Fertil. Steril.* **103**, 333–341 (2015).

237. I. Noci, B. Fuzzi, R. Rizzo, L. Melchiorri, L. Criscuoli, S. Dabizzi, R. Biagiotti, S. Pellegrini, A. Menicucci, O. R. Baricordi, Embryonic soluble HLA-G as a marker of developmental potential in embryos, *Hum. Reprod.* **20**, 138–146 (2005).

238. G. SHER, L. KESKINTEPE, J. FISCH, B. ACACIO, P. AHLERING, J. BATZOFIN, M. GINSBURG, Soluble human leukocyte antigen G expression in phase I culture media at 46 hours after fertilization predicts pregnancy and implantation from day 3 embryo transfer, *Fertil. Steril.* **83**, 1410–1413 (2005).

239. Z. P. Nagy, D. Sakkas, B. Behr, Non-invasive assessment of embryo viability by metabolomic profiling of culture media ('metabolomics'), *Reprod. Biomed. Online* **17**, 502–507 (2008).

240. D. K. Gardner, M. Lane, J. Stevens, W. B. Schoolcraft, Noninvasive assessment of human embryo nutrient consumption as a measure of developmental potential., *Fertil. Steril.* **76**, 1175–80 (2001).

241. D. K. Gardner, P. L. Wale, R. Collins, M. Lane, Glucose consumption of single post-compaction human embryos is predictive of embryo sex and live birth outcome, *Hum. Reprod.* **26**, 1981–1986 (2011).

242. R. G. Sturmey, D. R. Brison, H. J. Leese, in *Reproductive BioMedicine Online*, (Reproductive Healthcare Ltd, 2008), vol. 17, pp. 486–496.

243. L. Scott, J. Berntsen, D. Davies, J. Gundersen, J. Hill, N. Ramsing, *Human oocyte respiration-rate measurement – potential to improve oocyte and embryo selection?* (2008).

244. A. Tejera, D. Castelló, J. M. de los Santos, A. Pellicer, J. Remohí, M. Meseguer, in *Fertility and*

Sterility, (Elsevier Inc., 2016), vol. 106, pp. 119-126.e2.

245. A. Diez-Juan, C. Rubio, C. Marin, S. Martinez, N. Al-Asmar, M. Riboldi, P. Díaz-Gimeno, D. Valbuena, C. Simón, Mitochondrial DNA content as a viability score in human euploid embryos: Less is better, *Fertil. Steril.* **104**, 534-541.e1 (2015).

246. K. Ravichandran, C. McCaffrey, J. Grifo, A. Morales, M. Perloe, S. Munne, D. Wells, E. Fragouli, Mitochondrial DNA quantification as a tool for embryo viability assessment: retrospective analysis of data from single euploid blastocyst transfers, *Hum. Reprod.* **32**, 1282–1292 (2017).

247. E. Fragouli, C. McCaffrey, K. Ravichandran, K. Spath, J. A. Grifo, S. Munné, D. Wells, Clinical implications of mitochondrial DNA quantification on pregnancy outcomes: A blinded prospective non-selection study, *Hum. Reprod.* **32**, 2340–2347 (2017).

248. M. J. de los Santos, A. Diez Juan, A. Mifsud, A. Mercader, M. Meseguer, C. Rubio, A. Pellicer, Variables associated with mitochondrial copy number in human blastocysts: what can we learn from trophectoderm biopsies?, *Fertil. Steril.* **109**, 110–117 (2018).

249. A. R. Victor, A. J. Brake, J. C. Tyndall, D. K. Griffin, C. G. Zouves, F. L. Barnes, M. Viotti, Accurate quantitation of mitochondrial DNA reveals uniform levels in human blastocysts irrespective of ploidy, age, or implantation potential., *Fertil. Steril.* **107**, 34-42.e3 (2017).

250. E. Scalici, S. Traver, N. Molinari, T. Mullet, M. Monforte, E. Vintejoux, S. Hamamah, Cell-free DNA in human follicular fluid as a biomarker of embryo quality, *Hum. Reprod.* **29**, 2661–2669 (2014).

251. S. Traver, E. Scalici, T. Mullet, N. Molinari, C. Vincens, T. Anahory, S. Hamamah, I. Polejaeva, Ed. Cell-free DNA in Human Follicular Microenvironment: New Prognostic Biomarker to Predict in vitro Fertilization Outcomes, *PLoS One* **10**, e0136172 (2015).

252. C. Farra, F. Choucair, J. Awwad, Non-invasive pre-implantation genetic testing of human embryos: an emerging concept, *Hum. Reprod.* **33**, 2162–2167 (2018).

253. S. Palini, L. Galluzzi, S. De Stefani, M. Bianchi, D. Wells, M. Magnani, C. Bulletti, Genomic DNA in human blastocoele fluid, *Reprod. Biomed. Online* **26**, 603–610 (2013).

254. M. C. Magli, A. Pomante, G. Cafueri, M. Valerio, A. Crippa, A. P. Ferraretti, L. Gianaroli, Preimplantation genetic testing: polar bodies, blastomeres, trophectoderm cells, or blastocoelic fluid?, *Fertil. Steril.* **105**, 676-683.e5 (2016).

255. E. R. Hammond, A. N. Shelling, L. M. Cree, Nuclear and mitochondrial DNA in blastocoele fluid and embryo culture medium: evidence and potential clinical use, *Hum. Reprod.* **31**, 1653–1661 (2016).

256. L. Galluzzi, S. Palini, S. De Stefani, F. Andreoni, M. Primiterra, A. Diotallevi, C. Bulletti, M. Magnani, Extracellular embryo genomic DNA and its potential for genotyping applications, *Futur. Sci. OA* **1** (2015), doi:10.4155/fso.15.62.

257. V. Kuznyetsov, S. Madjunkova, R. Antes, R. Abramov, G. Motamedi, Z. Ibarrientos, C. Librach, G. M. Kelly, Ed. Evaluation of a novel non-invasive preimplantation genetic screening approach, *PLoS One* **13**, e0197262 (2018).

258. L. Gianaroli, M. C. Magli, A. Pomante, A. M. Crivello, G. Cafueri, M. Valerio, A. P. Ferraretti, Blastocentesis: a source of DNA for preimplantation genetic testing. Results from a pilot study, *Fertil. Steril.* **102**, 1692-1699.e6 (2014).

259. J. Xu, R. Fang, L. Chen, D. Chen, J.-P. Xiao, W. Yang, H. Wang, X. Song, T. Ma, S. Bo, C. Shi, J. Ren, L. Huang, L.-Y. Cai, B. Yao, X. S. Xie, S. Lu, Noninvasive chromosome screening of human embryos by genome sequencing of embryo culture medium for in vitro fertilization., *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11907–11912 (2016).

260. R. Fang, W. Yang, X. Zhao, F. Xiong, C. Guo, J. Xiao, L. Chen, X. Song, H. Wang, J. Chen, X. Xiao, B. Yao, L.-Y. Cai, Chromosome screening using culture medium of embryos fertilised in vitro: a pilot clinical study, *J. Transl. Med.* **17**, 73 (2019).

261. M. Feichtinger, E. Vaccari, L. Carli, E. Wallner, U. Mädel, K. Figl, S. Palini, W. Feichtinger, Non-invasive preimplantation genetic screening using array comparative genomic hybridization on spent culture media: a proof-of-concept pilot study, *Reprod. Biomed. Online* **34**, 583–589 (2017).

262. J. R. Ho, N. Arrach, K. Rhodes-Long, A. Ahmady, S. Ingles, K. Chung, K. A. Bendikson, R. J. Paulson, L. K. McGinnis, Pushing the limits of detection: investigation of cell-free DNA for aneuploidy screening in embryos, *Fertil. Steril.* **110**, 467-475.e2 (2018).

263. H. Wu, C. Ding, X. Shen, J. Wang, R. Li, B. Cai, Y. Xu, Y. Zhong, C. Zhou, Medium-based noninvasive preimplantation genetic diagnosis for human α-thalassemias-SEA., *Medicine (Baltimore).* **94**, e669 (2015).

264. L. Yang, Q. Lv, W. Chen, J. Sun, Y. Wu, Y. Wang, X. Chen, X. Chen, Z. Zhang, Presence of embryonic DNA in culture medium, *Oncotarget* **8**, 67805–67809 (2017).

265. S. Stigliani, P. Anserini, P. L. Venturini, P. Scaruffi, Mitochondrial DNA content in embryo culture medium is significantly associated with human embryo fragmentation, *Hum. Reprod.* **28**, 2652–2660 (2013).

266. S. Stigliani, L. Persico, C. Lagazio, P. Anserini, P. L. Venturini, P. Scaruffi, Mitochondrial DNA in Day 3 embryo culture medium is a novel, non-invasive biomarker of blastocyst potential and implantation outcome, *MHR Basic Sci. Reprod. Med.* **20**, 1238–1246 (2014).

267. S. Stigliani, G. Orlando, C. Massarotti, I. Casciano, F. Bovis, P. Anserini, F. M. Ubaldi, V. Remorgida, L. Rienzi, P. Scaruffi, Non-invasive mitochondrial DNA quantification on day 3 predicts blastocyst development: a prospective, blinded, multi-centric study, *Mol. Hum. Reprod.* (2019), doi:10.1093/molehr/gaz032.

268. S. Hamamah, S. El Messaoudi, A. R. Thierry, S. Assou, Methods for determining the quality of an embryo (2014) (available at https://patents.google.com/patent/US20160138104A1/en).

269. S. El Messaoudi, F. Rolet, F. Mouliere, A. R. Thierry, Circulating cell free DNA: Preanalytical considerations, *Clin. Chim. Acta* **424**, 222–230 (2013).

270. H. Malonga, J. F. Neault, H. Arakawa, H. A. Tajmir-Riahi, DNA Interaction with Human Serum Albumin Studied by Affinity Capillary Electrophoresis and FTIR Spectroscopy, *DNA Cell Biol.* **25**, 63–68 (2006).

271. E. R. Hammond, B. C. McGillivray, S. M. Wicker, J. C. Peek, A. N. Shelling, P. Stone, L. W. Chamley, L. M. Cree, Characterizing nuclear and mitochondrial DNA in spent embryo culture media: genetic contamination identified., *Fertil. Steril.* **107**, 220-228.e5 (2017).

272. D. E. Morbeck, M. Paczkowski, J. R. Fredrickson, R. L. Krisher, H. S. Hoff, N. A. Baumann, T. Moyer, D. Matern, Composition of protein supplements used for human embryo culture, *J. Assist. Reprod. Genet.* **31**, 1703–1711 (2014).

273. A. Sunde, D. Brison, J. Dumoulin, J. Harper, K. Lundin, M. C. Magli, E. Van den Abbeel, A. Veiga, Time to take human embryo culture seriously, *Hum. Reprod.* **31**, 2174–2182 (2016).

274. J. Kim, E. Seli, Mitochondria as a Biomarker for IVF Outcome, *Reproduction* (2019), doi:10.1530/REP-18-0580.

275. E. Fragouli, K. Spath, S. Alfarawati, F. Kaper, A. Craig, C.-E. Michel, F. Kokocinski, J. Cohen, S. Munne, D. Wells, S. K. Kim, Ed. Altered Levels of Mitochondrial DNA Are Associated with Female Age,

Aneuploidy, and Provide an Independent Measure of Embryonic Implantation Potential, *PLOS Genet.* **11**, e1005241 (2015).

276. J. C. St. John, J. Facucho-Oliveira, Y. Jiang, R. Kelly, R. Salah, Mitochondrial DNA transmission, replication and inheritance: a journey from the gamete through the embryo and into offspring and embryonic stem cells, *Hum. Reprod. Update* **16**, 488–509 (2010).

277. K. C. A. Chan, J. Zhang, A. B. Y. Hui, N. Wong, T. K. Lau, T. N. Leung, K.-W. Lo, D. W. S. Huang, Y. M. D. Lo, Size distributions of maternal and fetal DNA in maternal plasma., *Clin. Chem.* **50**, 88–92 (2004).

278. P. Jiang, Y. M. D. Lo, The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics., *Trends Genet.* **32**, 360–371 (2016).

279. S. G. Holtan, D. J. Creedon, P. Haluska, S. N. Markovic, Cancer and Pregnancy: Parallels in Growth, Invasion, and Immune Modulation and Implications for Cancer Therapeutic Agents, *Mayo Clin. Proc.* **84**, 985–1000 (2009).

280. F. C. Kelleher, D. Fennelly, M. Rafferty, Common critical pathways in embryogenesis and cancer, *Acta Oncol. (Madr).* **45**, 375–388 (2006).

281. M.-A. Forget, S. Turcotte, D. Beauseigle, J. Godin-Ethier, S. Pelletier, J. Martin, S. Tanguay, R. Lapointe, The Wnt pathway regulator DKK1 is preferentially expressed in hormone-resistant breast tumours and in some common cancer types, *Br. J. Cancer* **96**, 646–653 (2007).

282. K. Shah, S. Patel, S. Mirza, R. M. Rawal, Unravelling the link between embryogenesis and cancer metastasis, *Gene* **642**, 447–452 (2018).

283. R. S. Stoika, R. R. Panchuk*, B. R. Stoika, Parallels and Antagonisms of Embryogenesis and Carcinogenesis, *Russ. J. Dev. Biol.* **35**, 57–62 (2004).

284. Y. Ma, P. Zhang, F. Wang, J. Yang, Z. Yang, H. Qin, The relationship between early embryo development and tumourigenesis., *J. Cell. Mol. Med.* **14**, 2697–701 (2010).

285. S. L. Pereira, A. S. Rodrigues, M. I. Sousa, M. Correia, T. Perestrelo, J. Ramalho-Santos, From gametogenesis and stem cells to cancer: common metabolic themes, *Hum. Reprod. Update* **20**, 924–943 (2014).

RESUME

Après sa découverte en 1948, l'ADN circulant (ADNcir) a été étudié dans divers domaines. Il est devenu un biomarqueur émergent, en particulier en oncologie, un domaine dans lequel plusieurs travaux ont récemment cherché à étudier son intérêt dans le dépistage et la détection précoce du cancer. La première partie de ma thèse a été consacrée à l'étude des caractéristiques quantitatives et structurelles de l'ADNcir, en prenant en compte son origine (ADNcir nucléaire et mitochondrial) et sa structure (fragmentation et profil de taille), pour le dépistage et la détection précoce du cancer. Deux paramètres, le Ref A 67 (concentration totale d'ADNcir nucléaire) et le MNR (Rapport entre la concentration de l'ADNcir mitochondrial et nucléaire), ont été quantifiés par q-PCR dans un modèle murin puis validés dans les milieux des cellules en culture pour évaluer leur potentiel à discriminer un état sain d'un état cancéreux. Ces deux paramètres ont été quantifiés chez l'Homme, en prenant en compte d'autres paramètres quantitatifs et structurels de l'ADNcir, après réajustement en fonction de l'âge, dans le plasma de 289 individus sains, 99 individus à risque de cancer colorectal (CCR) et 983 patients atteints de CCR (n = 791), de cancer du sein (n = 169) et d'autres cancers (hépatocellulaire, pancréatique, ovarien et lymphome) (n = 23). Par une approche d'apprentissage automatique, nous avons combiné ces différents paramètres dans un modèle de prédiction en utilisant des arbres de décision pour la classification des patients sains et cancéreux. Nous avons obtenu des résultats très encourageants, en particulier pour les cancers de stades précoces. Cette méthode semble prometteuse pour une détection précoce et non invasive du cancer. L'ajout d'autres biomarqueurs, comme le profil de taille ou le profil de méthylation de l'ADNcir, pourrait encore en augmenter le potentiel. La deuxième partie de ma thèse a été consacrée à l'étude de la relation entre la quantité d'ADN extracellulaire d'origine nucléaire et mitochondriale dans le milieu de culture d'embryons, et la qualité de ces embryons lors d'une fécondation in vitro (FIV). En effet, il a été montré qu'un embryon libère de l'ADN extracellulaire dans le milieu de culture lors d'une FIV, et que cet ADN pourrait être un biomarqueur prédictif de la qualité de l'embryon et servir comme test génétique préimplantatoire (PGT) non invasif. Nous avons détecté le gène SRY dans le milieu de culture afin de déterminer le sexe de l'embryon, ce qui constitue une information importante dans les cas des maladies génétiques liées au sexe. Nous avons également entrepris de détecter la présence de la mutation Delta F508 du gène CFTR responsable de la mucoviscidose par analyse de l'ADN extracellulaire issu d'embryons à risque, afin d'évaluer son potentiel en tant que PGT non invasif.

ABSTRACT

After its discovery in 1948, circulating DNA (cirDNA) was studied in various fields. It has become an emerging biomarker, particularly in oncology, and several studies have recently sought to investigate its interest in cancer screening and early detection. The first part of my thesis was devoted to the study of the quantitative and structural characteristics of cirDNA, taking into account its origin (nuclear and mitochondrial cirDNA) and its structure (fragmentation and size profile), for the screening and early detection of cancer. Two cirDNA parameters, the Ref A 67 (total nuclear cirDNA concentration) and the MNR (Mitochondrial to Nuclear Ratio), were quantified by q-PCR in a mouse model and further validated in cell culture media to assess their potential to discriminate between a healthy and a cancerous state. These two variables were evaluated by taking into account other quantitative and structural parameters of cirDNA, after age adjustment, in the plasma of 289 healthy individuals, 99 individuals at risk of colorectal cancer (CRC) and 983 patients with CRC (n = 791), breast cancer (n = 169) and other cancers (hepatocellular, pancreatic, ovarian and lymphoma) (n = 23). Through a machine learning approach, we combined these different parameters into a prediction model using decision trees for the classification of healthy and cancer patients. We have obtained very encouraging results, especially for early-stage cancers. This method seems promising for early and non-invasive cancer detection. The addition of other biomarkers such as the size profile of the cirDNA or the detection of methylation markers could further increase its potential. The second part of my thesis was devoted to the study of the relationship between the quantity of extracellular DNA of nuclear and mitochondrial origin in the embryo culture medium, and the quality of these embryos during *in vitro* fertilization (IVF). It has been shown that an embryo releases extracellular DNA into the culture medium during IVF, and that this DNA could be a predictive biomarker of embryo quality and thus be used as a non-invasive preimplantation genetic test (PGT). We detected, as well, the *SRY* gene in the culture medium to determine the sex of the embryo, which is an important information in the case of gender-related genetic disorders. We also tried to detect the presence of the Delta F508 mutation of the *CFTR* gene responsible for cystic fibrosis, by analyzing extracellular DNA from high-risk embryos to assess its potential as a non-invasive PGT.