Detectable clonal mosaicism: underlying mechanisms, clinical implications and genetic counselling

Judith Reina Castillón

TESI DOCTORAL UPF / 2017

THESIS SUPERVISOR

Prof. Luis A. Pérez Jurado

Departament de Ciències Experimentals i de la Salut

Universitat Pompeu Fabra Barcelona

De tots, l'únic cercle que
esdevé traç i
m'endevina
la llum al fons
de l'aparença sostinguda,
és una frase grisa, segurament
plena d'
espais
i
petits exilis.
Poc a poc,
l'abisme
m'obrirà la seva porta
i em deixarà
passar.

ÒSCAR REINA

Als meus pares

i a la Júlia

Agraïments

Sembla mentida que ja hagin passat 5 anys des de l'inici d'aquesta etapa. Ara que s'està acabant, és el moment de pensar en totes les persones que han format part d'ella i realment no sé ni per on començar... (em temo que no podré ser breu). El que sí que puc dir amb total seguretat és que cadascuna d'elles m'ha ensenyat alguna cosa que m'ha ajudat a evolucionar i créixer fins al punt en el que estic ara i només per això ha valgut la pena conèixer-les a totes elles.

En primer lloc, voldria agrair al Luis la seva dedicació, guia i coneixement que he anat rebent al llarg de tots aquests anys. També, l'oportunitat que en el seu moment em va donar de compaginar el doctorat amb el màster d'Assessorament Genètic ja que això m'ha permès fer el que tenia tantes ganes de dur a terme des d'un inici i també rebre una formació en el camp de la genètica tant des del punt de vista de la recerca com des del mèdic i clínic; en definitiva una formació molt més rica i completa.

Jordi, moltes gràcies per la teva ajuda i implicació, per la fluïda i bona relació que hi ha hagut i per les bones paraules que sempre he rebut de tu. Ha estat molt fàcil treballar amb tu i amb els membres del teu equip com la Roser i la Montse.

Moltes gràcies a tota la gent que forma i ha format part del laboratori 422 durant el temps que he estat (espero no deixar-me a ningú): Raquel, Ivon, Clara, Roser, Mariví, Andrés, Marcos, Francesc, Mar, Paula, Marina, Goretti, Tina, Marta, Aïda, Gaby, Armand, Débora, Guillermo, Arturo, Eulàlia, Ricard, Gaudi, Mireia, Èrika, Dora, Ania, Goreti, Montse i Sorina.

Mar, em fa molta gràcia recordar com ens vam conèixer quan eres estudiant de pràctiques i jo, en el primer any de doctorat, intentava ensenyar-te l'apassionant món del *FISH* i les inversions on amb prou feines me'n sortia i que ens va portar a la desesperació, literalment, a les dues més d'un cop. Després, amb un petit *break* entremig, has tornat com a estudiant de doctorat i et vull donar les gràcies per les mil i una coses que hem compartit: les improvisades classes de *zumba* i sessions de perruqueria (de fer trenes del dret i del revés), els ensurts quan vens pel darrere perquè estic amb la música a tot drap, els atacs de riure, les xerrades on ens expliquem la vida mútuament, alguna que

altra tarda de compres que se'ns ha escapat de les mans i les sortides nocturnes. Gràcies també per les persones meravelloses que he conegut a través teu, la Nerea i la Gemma.

Paula, la meva companya dels *ColaCaos* amb *Kellogg's* matiners!! Gràcies per la teva santa paciència, per escoltar-me sempre i calmar-me analitzant cada situació i fent-ne al final un resum objectiu de totes les solucions possibles. Gràcies per les tardes on ens hem quedat les dues soles al laboratori, pels trajectes en metro cap a casa i per totes les converses. També, com no, per les mil nits a *Merlín* (tot i que el ritme ha decaigut una mica en aquest últim any cosa que no pot ser eh...) i per tots els companys de festa increïbles que m'has presentat (Marta, els Paus, Jan, Isa, Molly, Carles...i molts més!).

Marina, gràcies pels consells i ànims en els últims mesos, per calmar els meus atacs d'estrès i pels enèrgics "bon dia!" a les 8:00 del matí.

Raquel, companya de picar *BACs*! Quin horror i quin estrès cada cop que hem hagut de pujar eh? Gràcies per ajudar-me sempre que he tingut algun problema tècnic, i d'altre no tan tècnic... Primer vam començar amb aquella infinitat de *minipreps* de *BACs*, després no em sortien les *maxipreps* fins que no vaig fer-ho amb tu (resulta que s'havia d'agitar en comptes d'invertir, mira per on!) i finalment els blocs d'agarosa i *Pulsed-field* (quin fracàs... sobretot quan la màquina va decidir no funcionar més). Gràcies per tot el que he après de tu i pels moments divertits que hem viscut en el dia a dia amb la resta.

Clara, moltes gràcies pels teus consells i per totes les coses que he après de tu al laboratori, al màster i a SEAGen. Gràcies per ajudar-me en les situacions on no sabia com prioritzar les diferents coses que havia de fer. Tu sempre amb la teva ment tant ordenada i lògica fas que moltes coses semblin fàcils quan els altres les trobem difícils. En definitiva, gràcies per estar sempre disposada a escoltar i a ajudar.

Ivon, vas ser la primera en ensenyar-me com fer *FISH* i a la que he recorregut cada cop que tenia un problema o em sortien resultats inexplicables. Moltes gràcies, no només pel que m'has ensenyat sinó també per ser-hi per qualsevol altre dubte o problema.

Roser, gràcies pels teus consells, per estar sempre disposada a ajudar, sobretot cada cop que he vingut tot estressada amb algun problema,

alguna cosa relacionada amb la tesi o algun email que se'm resistia a l'hora de redactar o respondre. Gràcies també per cobrir-me alguna de les tasques comunes en aquests últims mesos!

Als *qGenomics,* Maria, Jairo, Anna, Manel, Cristina, Sònia, Lluís, Xavi, Mª José i Sílvia, entre d'altres noves incorporacions, que han estat companys de laboratori, d'elefant blanc i de sopars de nadal i de comiat una bona temporada. Sobretot a en Benja i a l'Olaya perquè més d'un cop us he atabalat amb emails demanant-vos dades de la prehistòria i a l'Olaya també perquè, juntament amb la Ivon, em va ensenyar com fer *FISH* al principi i durant un temps va ser víctima de moltes de les meves preguntes i errors amb la tècnica.

El meu dia a dia no hagués estat igual aquest últim any sense dues persones increïbles. Mònica i Carla, gràcies per estar quan us necessito, per saber escoltar, per les abraçades, per les converses maratonianes al mig del passadís o al despatx on només us faig que repetir les mateixes coses (com qui es va donant cops contra la paret) i vosaltres aneu aguantant el tipus, gràcies pels *whatsapp* en els moments indicats i necessaris. Mònica gràcies pels mil consells i pels *Conguitos*! Carla gràcies per embolicar-me en els teus plans i amb la teva gent... l'André, el Simone i molts altres de bioevo... sou fantàstics.

Raquel, sempre batallant amb diferents temes de la FPU! (IRPF, docència, memòries anuals...). Gràcies per ser companya d'aquest viatge i per estar allà quan més ho he necessitat.

Tomàs, gràcies per ser el meu "tutor" al llarg de la carrera, per orientar-me quan tenia dubtes sobre del meu futur i, per interessar-te sobre com em van les coses al llarg d'aquests darrers anys.

Al llarg d'aquesta etapa he tingut la sort de coincidir també amb un grup d'onze persones d'una qualitat humana i professional increïble: Adrià, Albert, Anna, Diana, Èlia, Gisela, Gina, Josep, Luna, Marta i Mònica. A part de ser uns assessors genètics meravellosos, han estat (i seran) uns fantàstics companys de festes nocturnes i diürnes (com les minicolònies a Llançà), barbacoes, caminades per la muntanya, sopars i dinars de nadal, *pongos* i congressos.

Voldria agrair-vos Clara, Mar, Núria, Estela, Anna i Irene tot el que m'esteu ensenyant a les reunions de SEAGen des que formo part de la

junta. Estela, especialment et voldria agrair també la supervisió tan propera que vaig tenir per part teva durant la tesina del màster.

Mar, gràcies per les trobades que anem fent per posar-nos el dia, que malgrat els nostres horaris impossibles sempre busquem una estoneta per xerrar. Gràcies pel teu suport continu.

Núria i Isa, les meves *Dmovers* indignades preferides!! Qualsevol cosa que pugui escriure es queda curta per tot el que heu i esteu fent per mi, en tots els sentits. Gràcies per les tardes de *zumba* i abdominals, pels sopars i les festes, pels cines, per les tardes de compres i de passeig, per les xafarderies... i tot el que vindrà! (que serà molt i molt bo). Moltes gràcies pels *whatsapps* i notes de veu maratonians, per ser-hi sempre i venir quan us he necessitat.

Andrea, dir que he compartit els moments viscuts durant la tesi seria insuficient. Absolutament sempre has estat al meu costat sense dubtar-ho, des que teníem cinc anys, en els bons i en els mals moments en tots els aspectes de la meva vida, personals i professionals. Et saps la meva vida de memòria i jo em sé la teva perquè ens l'hem explicat mútuament i perquè l'hem viscut. Ara comences una nova i emocionant etapa de la teva vida amb una persona increïble, espero i no dubto, continuar al teu costat per viure i gaudir tots els que vinguin.

Per acabar, però evidentment no menys important, mama, gràcies per ser el meu recolzament sempre que ho necessito, per fer-m'ho tot més fàcil quan em veig ofegada i per animar-me a no conformar-me amb el més còmode pel simple fet que ho sigui. Papa, gràcies per motivar-me a continuar endavant, a perseguir el que vull per molt difícil que sigui, i per escoltar-me sempre i intentar entendre tots els problemes que t'explico, que per concrets i específics que siguin tu sempre intentes donar-me solucions. Gràcies també per tota l'ajuda logística en la tesi i pel regal que m'has fet i que demostra el teu talent no suficientment reconegut. Júlia, gràcies per ser-hi sempre i perquè sé que sempre puc recórrer a tu. Avi Jesús i iaia Angelina, sé que va arribar un moment en què us pensàveu que estava estudiant eternament però també sé que us sentíeu orgullosos del que feia. Encara que no hagueu pogut veure com ha acabat aquesta etapa de la meva vida sé que us faria il·lusió saber que: **he après, he crescut i ho tornaria a repetir.**

Moltes gràcies a tots.

Abstract

Somatic genetic mosaicism can be present both in healthy individuals but also in subjects with certain conditions as ageing or cancer. Its detection by SNP array in blood can be used as biomarker of cancer risk improving patients' management and survival in the future. Given the increased cancer risk in Fanconi anemia (FA) patients, we have evaluated the prevalence and evolution of mosaicism in them as well as its relationship with cancer and survival. We also have studied mosaic uniparental disomy (UPD) as a putative protective mechanism in hematologic cancer. Finally, we have explored FA knowledge and follow-up adherence of FA families. Our results suggest blood mosaicism detection as a good less invasive follow-up strategy for cancer prevention in FA and propose UPD as a possible rescue mechanism avoiding hematologic cancer development. Two opposite roles of mosaicism in cancer become evident from our data. FA knowledge and follow-up adherence is satisfactory in FA families thanks to genetic counsellors' actions.

Resum

El mosaicisme genètic somàtic pot està present en individus sans i en altres amb certes condicions com edat avançada o càncer. La seva detecció en sang per *array* d'SNPs pot ser utilitzada com a biomarcador de risc tumoral millorant el tractament i la supervivència dels pacients en un futur. Donat el risc de càncer incrementat en pacients amb anèmia de Fanconi (AF), hem avaluat la prevalença i l'evolució del mosaicisme en ells i la seva relació amb càncer i supervivència. També hem estudiat la disomia uniparental (DUP) en mosaic com a possible mecanisme de protecció en front el càncer hematològic. Finalment, hem explorat el coneixement sobre l'AF i l'adherència al seguiment en famílies AF. Els nostres resultats suggereixen la detecció de mosaicisme en sang com a sistema adequat i menys invasiu de detecció precoç de càncer en AF i proposa la DUP com a possible mecanisme de rescat protector en front el càncer hematològic. Dos rols oposats del mosaicisme envers el càncer esdevenen evidents considerant les nostres dades. El coneixement sobre l'AF i l'adherència al seguiment són satisfactoris en famílies AF gràcies a les accions dels assessors genètics.

Prologue

Genetic mosaicism is a phenomenon whose frequency is usually underestimated because it does not always produce a phenotypic effect. However, it has been associated to numerous health conditions as some Mendelian and mitochondrial disorders, lethal diseases, chromosome abnormalities, miscarriages, cancer or ageing. Thanks to the development of new technologies, as Single Nucleotide Polymorphism (SNP) array among others, the detection of chromosome rearrangements at more and more lower mosaicism level has become a reality leading to the establishment of a more real and accurate mosaicism prevalence and the progress in uncovering the consequences of mosaicism on health.

This thesis expands the knowledge about mosaicism and its implications on cancer by providing new data related to its use as a biomarker of cancer risk in Fanconi anemia (FA) but also its possible role on protecting from cancer development in FA and chronic lymphocytic leukemia (CLL). Moreover, this thesis addresses some issues as follow-up adherence, knowledge and psychological impact of FA in affected families highlighting the necessary function of genetic counsellors in a FA unit.

The introduction gives an overview of the main topics related to the results shown in the thesis. These are a review of mosaicism by detailing the main sources, detection methods and its implications on health; hematologic malignancies, focusing on CLL; a deep description of FA; and finally, the key points to consider in genetic counselling in every of the previous reviewed situations.

The main body of the thesis is divided in three chapters, with chapter 1 divided into two parts, where each one addresses the results related to the corresponding objective.

The discussion is a global interpretation and integration of the results shown in the main body of the thesis and tries to contextualize them to previous published knowledge and to fit them into the field.

Finally, conclusions are a summary of principal findings and remarkable messages derived from the studies performed.

Contents

List of figures

CHAPTER 2

CHAPTER 3

DISCUSSION

List of tables

INTRODUCTION

1. MOSAICISM

Greek and Roman cultures popularized the creation of large mosaic compositions by using small objects with different size, color, and texture. These mosaic creations seem a unified whole where a closer inspection reveals multiple individual small pieces defining the details of the picture. The comparison of human individuals genetic heterogeneity with this artistic expression became evident as early as 1945 (1). Genetic mosaicism is now defined as the coexistence of cells with different genetic composition within an individual caused by postzygotic mutations acquired during the development that are propagated only to a subset of adult cells (2). Mosaicism can be classified as somatic mosaicism, when mutations are only present in somatic cells and cannot be transmitted to the offspring, and germline mosaicism in which mutations affect sexual cells and can be inherited by the next generation (3,4). Mosaicism should be distinguished from two different concepts: *de novo* mutations and chimerism. *De novo* mutations are those genetic alterations detected in the offspring but not in parents and they can occur or not in mosaicism in that individual (5). Otherwise, chimerism is a rare phenomenon consisting of the detection of a mixture of cells with genotypes derived from different germ cells in an individual due to the fusion of two independently conceived zygotes within a single embryo. Dizygotic twin-twin transfusion, iatrogenic microchimerism[1], the persistence of fetal cells in maternal organs as well as the detection of mother's cells in her offspring are examples of chimerism (3) (**Figure 1**).

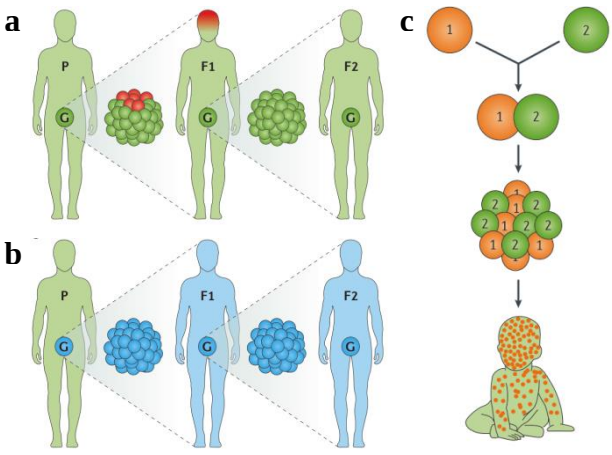


Figure 1.
Scheme of mosaicism and chimerism.
a) Somatic mosaicism
b) Germline mosaicism
c)Chimerism. Pictures adapted from (3).

[1] Iatrogenic microchimerism is the creation of a chimeric line through a medical procedure, such as blood transfusion or organ transplantation (3).

1.1. Sources of genetic mosaicism

Identifying molecular mechanisms responsible of mosaicism is a challenge. However, it is widely accepted that they are closely related to the nature of the mutation that is appearing. Before entering to processes behind genetic mosaicism, a brief revision about genetic rearrangements classification is shown below.

Chromosomal alterations can be classified in numerical and chromosomal structural abnormalities. The first group consists of alterations in the correct number of chromosomes, condition also known as aneuploidy, and includes those cases in which either a chromosome from a pair is missing (monosomy) or more than two chromosomes of a pair are present (trisomy, tetrasomy, etc.). X monosomy (Turner's syndrome), XXY (Klinefelter syndrome), trisomy 21 (Down syndrome), trisomy 13 (Patau syndrome) and trisomy 18 (Edwards syndrome) are the most frequent aneuploidies in humans. Regarding structural alterations group, it includes losses and gains of genetic material (deletions and duplications) usually referred as Copy Number Variants (CNVs); translocations, in which a portion of one chromosome or an entire chromosome is transferred to another one; inversions, where a piece of chromosome changes its orientation after a process of breakage, turn and reattachment; insertions, phenomenon by which a segment of DNA is deleted in its origin place and inserted into another position; rings, when a chromosome breaks and forms a circle leading or not to the loss of genetic material; isochromosomes, situation in which a chromosome losses an arm and the other one is duplicated and attached to the centromere leading to a monosomy of the loss arm and a trisomy of the duplicated one (6). Finally, uniparental disomies (UPDs) are another example of structural chromosome alterations in which two copies of one chromosome or a segment come from the same progenitor. UPDs can be classified in heterodisomies, in which both chromosomes are different inherited from the same parent (from the grandparents) and isodisomies where chromosomes are identical through duplication and lead to a Loss of Heterozygosity (LOH). LOH, due to either monosomy or UPD, can lead to disease secondary to homozygosity for recessive alleles or aberrant patterns of imprinting (7).

There are several mutational mechanisms that have been associated with genetic mosaicism, all of them occurring in somatic cells (2,3,8).

1.1.1. Point mutations

Throughout the human lifespan, intrinsic and extrinsic mutagenic agents in combination with the basal mutation rate (mainly resulting from DNA replication errors and DNA repair defects) can give rise to single-nucleotide variants (SNVs) and small insertions and deletions (indel) in mosaicism (3). Estimation of mutational rate is daring since this differs between tissues and age; however, one of the early estimates was in normal and malignant lymphocyte cells where it was established around 10^{-7}–10^{-8} per base pair per generation (8).

1.1.2. Retrotransposon-induced mutations

Transposable Elements (TEs) are segments of DNA that have the inherent ability to move from one genomic location to another one thanks to an element-encoded protein such as DNA transposase or reverse transcriptase, because of this they are referred as "autonomous TEs". There is a group of non-coding TEs that are considered as "non-autonomous" since they need the autonomous TEs machinery to move across the genome (9). Autonomous TEs are classically divided in two general categories according to their mobilization intermediate. Class I TEs or retrotransposons are TEs that "jump" from one site of the genome to another by "copy-and-paste" mechanism and using an RNA intermediate. Consequently, Class I TEs are usually the major contributors to generate multiple copies from the same sequence in the genome. Otherwise, Class II TEs or DNA transposons are TEs that move across the genome by "cut-and-paste" process leaving a gap in the origin site (**Figure 2**). Because of TEs nature, some diseases can appear as a result of a disruption of a functional gene, incorrect correction of the gap left in the origin site by DNA polymerase and misalignment of chromosomal pairs during cell division because of multiple copies of the same sequence. Some examples are cystic fibrosis, hemophilia, neurofibromatosis (NF1/2) and some cancer disorders as familiar breast and ovarian cancer (BRCA1/2) (a detailed list of retrotransposition events associated with human disease is shown in (9) reference) (9).

Within the Class I, there are two TEs that have been associated with mosaicism: Long Interspersed Nuclear Element 1 (LINE1 also known as L1) and Alu elements. L1 is a retrotransposon without Long Terminal Repeats (non-LTR) and is the only active autonomous TE in humans (9). L1 retrotransposons are active during embryogenesis and can give rise mosaicism (10). Interestingly, published data demonstrates that L1 RNA

transcription can be carried out in male and female mouse and rat germ cells (during spermatogenesis and oogenesis) whereas L1 genome integration could occur during the embryogenesis showing an interesting example of how L1 can create mosaicism during the mammalian development (11). L1 retrotransposition activity has also been detected in both neuronal progenitor cells and adult human brain tissue giving more evidences of L1 and mosaicism relationship (12). Alu elements, in turn, are Short Interspersed Nuclear Elements (SINEs) and are also active during embryonic development. They are non-autonomous retrotransposons since L1-enzyme machinery is needed for Alu movement along the genome, so L1 and Alu mutagenic action as well as their influence on mosaicism could be slightly associated (13).

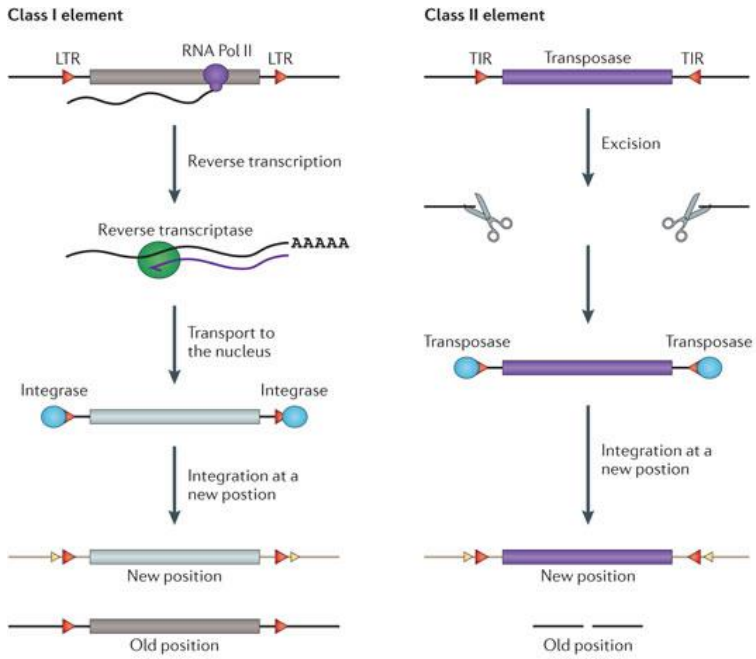


Figure 2. Retrotransposition processes. Class I elements or retrotransposons follow a "copy-and-paste" mechanism by which they are, first, transcribed to mRNA by RNA polymerase II (RNA Pol II); second, converted into cDNA by reverse transcriptase and finally, integrated to the genome by an integrase enzyme. Class II elements or DNA transposons follow a "cut-and-paste" mechanism by which the element is physically excised from the chromosome and reintegrated in a new location of the genome by a transposase enzyme encode by the TE. This process generates a double-strand gap in the old position that needs to be correctly repaired. Picture adapted from (14).

1.1.3. Polymerase slippage and tandem repeat variation

The existence of short homologous DNA sequences within the length of a piece of the genome that will expand one Okazaki fragment (1-2Kb or shorter in human) during DNA replication increases de probability of the slippage of the polymerase during the process leading to the deletion or duplication of the sequence between the homologous regions (15) (**Figure 3**). This phenomenon is particularly common in genome regions with trinucleotide tandem repeats whose expansion is associated with neurological diseases as Huntington's disease, Spinocerebellar ataxia and Fragile X syndrome. Replication machinery slippage along tandem repeats is a source of genetic heterogeneity in terminally differentiated neurons (16). When repeated sequences are not fully identical because of containing mismatches, replication slippage rate is higher when carrying mutations that impair mismatch-repair system function (15). Concordantly, it has been reported that mismatch repair deficiency causes expansion of trinucleotide repeats leading to genetic variability in mouse models of Huntington's disease (17).

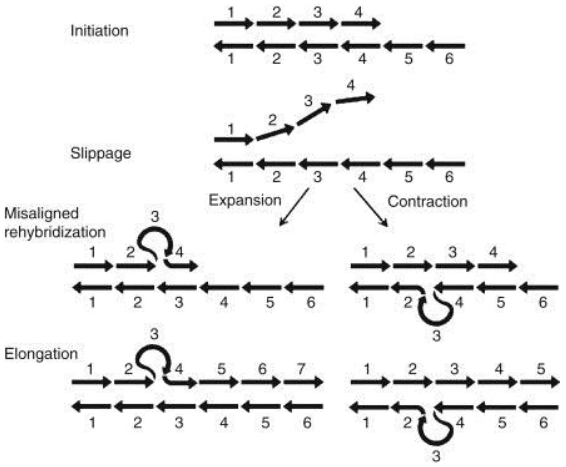


Figure 3. Polymerase slippage mechanism. Tandem repeats can lead to DNA polymerase slippage and, depending on which strand suffers the misalignment hybridization, expansion or contraction of the number of repeats will occur. DNA repeat units are shown by arrows and numbered within each strand. Picture from (18).

1.1.4. Erroneous homologous recombination

Homologous recombination (HR) is the basis of several DNA processes as DNA damage repair, telomere maintenance, chromosome segregation during cell division as well as generation of new allele combinations during meiosis (19). In context of DNA reparation, HR is characterized by using an identical sequence to the altered one to perform the reparation of the damaged region. Concretely, the homologous sequence used as a template to perform the reparation is exactly located at the same position as the damage sequence but in the sister chromatid or homologue chromosome. HR requires around

300pb of homologous sequence in mammalian cells to repair DNA damage successfully. HR repairs Double Strand Breaks (DSBs) mainly through two different models: double-strand break repair (DSBR) pathway (sometimes called the Double Holliday junction model) and the synthesis-dependent strand annealing (SDSA) pathway both of them explained in detail in **Figure 4a** and **b** (15,19). Although HR is one of the majors DNA repairing systems, it also can have deleterious consequences. DSBR pathway can lead to LOH after formation of crossovers between homologous chromosomes and the co-segregation of chromatids carrying the same alleles during mitosis, revealing in this way recessive mutations (19). SDSA pathway, in turn, would not lead to LOH since it avoids crossing-over but in cases where DNA template contains direct repeats, CNVs can arise. DSBR and SDSA pathways repair DNA breaks that consist with two double-stranded ends, however, HR can also be used to repair breaks that consist only in one double-stranded end that generates collapsed or broken replication forks. In this case, the mechanism used is named break-induced replication (BIR) explained in detail in **Figure 4c.** This process normally has no deleterious effects except when the broken end invades a homologue chromosome instead to the sister chromatid leading again to LOH (15).

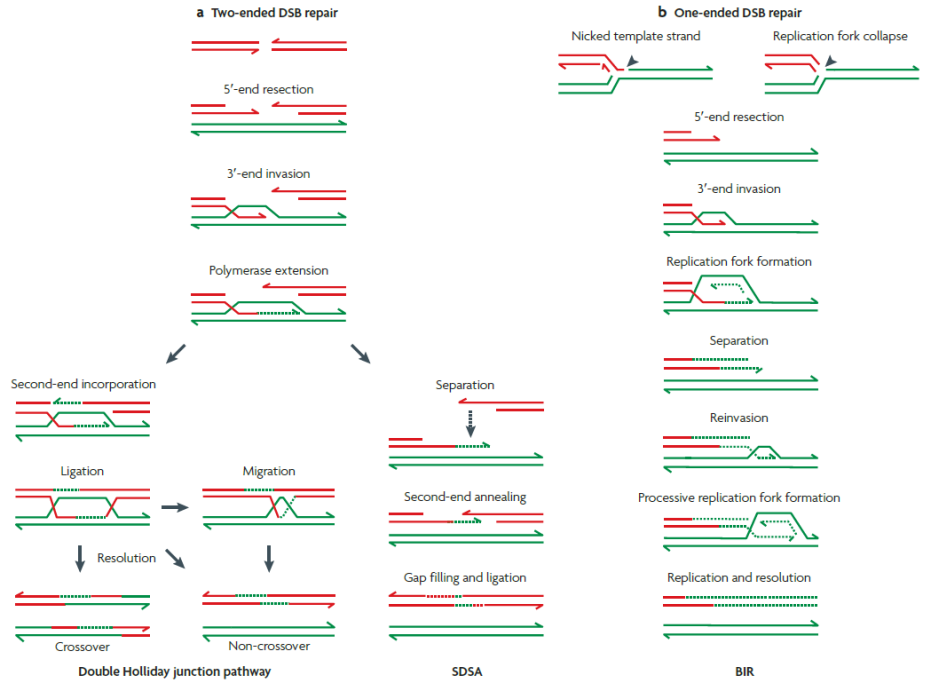


Figure 4. Mechanisms of DSBs reparation through HR. DSBs can be repaired by two different HR-mediated pathways including (**a**) double-strand break repair (DSBR) pathway (or Double Holliday junction model) and (**b**) the

synthesis-dependent strand annealing (SDSA) pathway. In both models, a DSB is resected to provide 3' single-stranded DNA (ssDNA) overhangs and then, these one of the 3' ends invade into homologous sequence forming a D-loop and DNA synthesis starts. At this point, in the DSBR model, the second DBS end is captured to form an intermediate with two Holiday junctions that, after gap-repairing with DNA synthesis and ligation, will be resolved either in a non-crossover or crossover manner depending whether two junctions are resolved in the same or different orientations. An alternative pathway is SDSA model where the invading end is separated from the template together with the new synthesized DNA, then the second end after annealing to the structure by complementarity is extended and ligated generating always a non-crossover repaired product. (**c**) Break-induced replication occurs when a broken (collapsed) replication fork with a nick in a template strand is detected by replicative helicase. BIR is a variation of SDSA pathway but, here, the separated end from the template fails to find a complementary second end to which to anneal and reinvades again into homologous sequence to be further extended in a low processivity replication fork. The process of invasion-extension-separation should be repeated several times until getting a more processive replication fork. Picture adapted from (15).

1.1.5. Erroneous non-allelic homologous recombination

Because of the presence of Low Copy Repeats (LCR) in the genome (or segmental duplication)[2], reparation machinery sometimes can use as a template for reparation a homologous sequence located in a different chromosomal position than the damaged DNA. This is called non-allelic or ectopic homologous recombination (NAHR) and, similarly to HR, can follow both classical HR-mediated DSB repair via a double Holliday junction or BIR (15). Equally to HR, NAHR could lead to structural changes in chromosomes. In the context of NAHR by unequal crossing-over between directly orientated repeats in *trans*, a duplication and its reciprocal deletion affecting the sequence between the repeats appears whereas the crossing-over between inversely orientated repeats in *cis* would lead to an inversion. In case of NAHR by BIR where a broken end invades the homologue chromosome instead of the sister chromatid, a translocation, duplication or deletion can occur (**Figure 5**) (15). Interestingly, NAHR during mitosis in somatic cells has become to be recognized also as a source of LOH in some tumors as NF1 in last years (20).

[2] Segmental duplications or LCR represent 5% of the genome and are defined as blocks of 1-400Kb of DNA that occur in at least two sites along the genome and share >90% of identity (273).

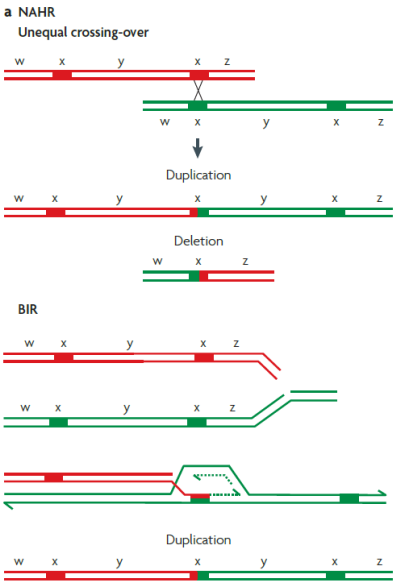


Figure 5. Mechanisms of NAHR reparation. Recombination between homologous sequences located in *trans* generates a duplication and its reciprocal deletion of the sequence between the repeats. NAHR can also occur by BIR when the broken end uses as a template to restart the replication fork an ectopic homologous sequence, then a trans-location, duplication or deletion can appear. Picture adapted from (15).

1.1.6. Erroneous non-homologous end joining

DSB can also be repaired by two mechanisms that do not need homology or need very short microhomologies. These are non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ). NHEJ joins DSB ends directly or, if available, with the guide of microhomologous sequences (15). In the case of error, this mechanism could join two DNA segments from two different regions generating small indels (1-4bp) or other kinds of rearrangement (8). MMEJ uses 5–25bp microhomologous sequences to align broken ends before rejoin them; consequently, MMEJ always results in deletions of the regions flanking the original break. Other chromosome abnormalities such as translocations, inversions and other complex rearrangements have been reported to be associated with MMEJ. There is some controversy of whether MMEJ is a subclass of single-strand annealing (SSA), a pathway that creates deletions after annealing directly repeated sequences over 30bp (15,21) (**Figure 6**).

Two additional processes that usually occur during replication and have the potential to generate complex rearrangements have been reported: Fork stalling and template switching (FoSTeS) and microhomology mediated break induced repair (MMBIR) (8). FoSTeS is a process in which the 3’ end of a DNA strand of a blocked replication fork breaks off the template and aligns with other exposed single-stranded template on another replication fork

with a shared microhomology generating amplicons longer than one fork. MMBIR is very similar to FoSTeS and is based on the BIR mechanism but mediated by microhomology. In both cases, depending on where the annealing reaction takes place versus the collapsed point in the fork some rearrangements can appear: deletions and duplications (when the annealing occurs with the sister chromatid in front of or behind the collapsed position respectively), inversion (if the orientation of the homologous sequences is changed versus the collapsed fork), LOH (when the annealing occurs with the homologous chromosome instead of the sister chromatid) and chromosome translocations (if the microhomology is located in a different chromosome) (15).

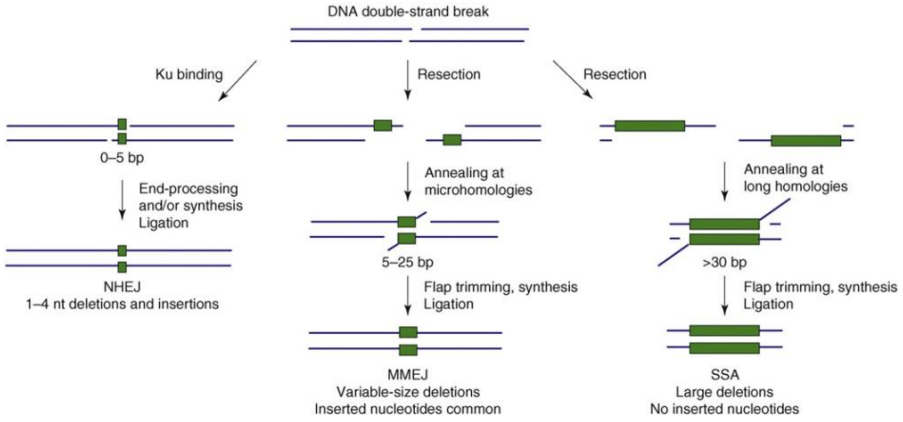


Figure 6. Comparison between NHEJ, MMEJ and SSA pathways in S. cerevisiae. In NHEJ, Ku70-Ku80 heterodimer anneals to DSB to prevent DNA end resection. With the binding of short microhomologies (green boxes), Pol4 can start to fill the gap and ligase IV can end the process with the ligation step. All will result in 1-4 bp deletion and insertion products. Regarding MMEJ and SSA processes, both of them need the resection or unwinding of homologous sequences being shorter in MMEJ (5-25pb) than in SSA (>30pb). A cleavage of 3' flap is mandatory to proceed with the fill and ligation of the gap. As a result of MMEJ and SSA, deleted products will appear whereas insertions will only occur as a result of MMEJ. Picture from (21).

1.1.7. Mitotic mis-segregation

During cell division, mitotic mis-segregation of chromosomes leading to aneuploidy occurs at a frequency of 1:50 to 1:100 per cell division. This kind of errors occurs mainly during the anaphase because of either a non-disjunction of sister chromatids during mitotic anaphase generating one trisomic and one monosomic daughter cell (**Figure 7**) or the failure of one or more chromosomes to be incorporated in a daughter cell at the end of the mitosis leading to monosomic daughter cells which is called anaphase lagging. Additional non-disjunction or

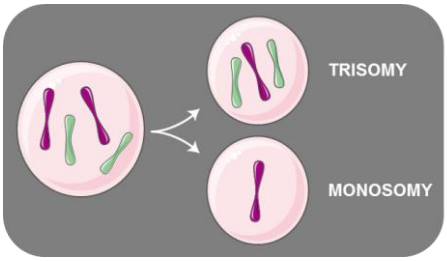


Figure 7. Representation of non-disjunction during mitosis. This generates monosomic and trisomic daughter cells. Isodisomy can appear if a non-disjunction occurs in a trisomic cell. From *Servier Medical Art*.

anaphase lagging events affecting a trisomic cell for a certain chromosome can lead to LOH in the form of UPD for the whole chromosome. The main causes of these two processes are mitotic check-point slippage, abnormal telomere shortening and problems in merotelic spindle attachments (3). The loss or gain of entire chromosomes secondary to their mis-segregation during mitosis is uncommon in normal somatic cells since this involves such dosage effect for many genes that would have an important negative effect for the cell and consequently it would be negatively selected (8). In fact, mosaic aneuploidy is the main cause of miscarriage and entails the major difficult in assisted reproduction. It occurs in more than 10% of human pregnancies and, importantly, this incidence could exceed 50% of pregnancies for women nearing the end of their reproductive lifespan (22). However, some cases of aneuploidies in mosaicism since very early stages of human development have been associated with some diseases; an example of this is Turner's syndrome (23).

1.1.8. Reversion mosaicism

A trisomic or monosomic mosaic cell line can appear because of a postzygotic de novo mis-segregation of chromosomes during mitosis in cells from a normal zygote (above explained) or due to a partial rescue of an aneuploidy zygote due to prezygotic mistakes in parental meiosis I or II. Gametes harboring some aneuploidy after parental meiotic errors can lead to a constitutionally trisomic or monosomic zygote. The loss or duplication of the affected chromosome would lead to the reversion of the aneuploidy and get a normal zygote in mosaicism or not depending on whether the rescue affects all cells or not. UPD can also appear secondary to the duplication of a monosomic chromosome (isodisomy), the loss of a trisomic one (heterodisomy or isodisomy) (3,7) (**Figure 8**) or the fertilization of a nullisomic gamete with a disomic gamete for the same chromosome, process known as gamete complementation. The restoring of the diploid status in the embryonic linage but not in placental cells is relatively common in trisomies rescue giving rise to confined placental mosaicism (CPM) which can lead to placental dysfunction impairing embryo development (3).

The complete or partial reversion of a germline genetic alteration in a subset of somatic cells has been described as a form of mosaicism reversion in some monogenic diseases leading to wild-type (WT) genotype and phenotype restoration (3,7,8), an example is Fanconi anemia which will be further explained in an specific section.

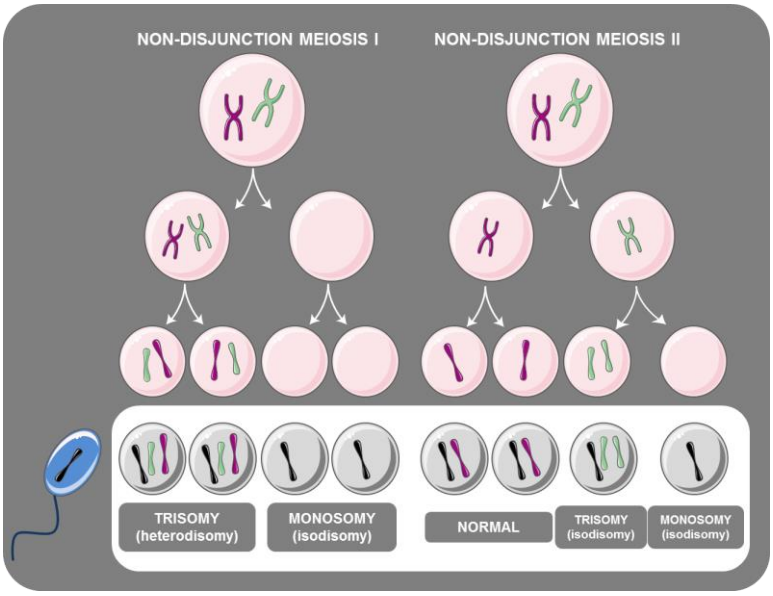


Figure 8. Schematic representation of chromosomes non-disjunction during meiosis I and II. Balanced or unbalanced ova can appear and, after fusing with a haploid spermatozoon, lead to trisomic, monosomic or disomic (normal) zygotes. The possibility of isodisomy or heterodisomy after a rescue of an unbalanced zygote is shown in brackets in each case. The contrary situation in which spermatozoa are unbalanced and ova are haploid is also possible. From *Servier Medical Art.*

1.1.9. Mitochondrial DNA heteroplasmy

Similarly to nuclear DNA, mitochondrial DNA (mtDNA) can accumulate genetic alterations during the lifetime of an individual. Mitochondrial mutations are maternally inherited and usually occur in a subpopulation of mitochondria generating a mosaicism for that mutation at mtDNA level which is known as heteroplasmy. A certain number of affected mitochondria over a threshold is needed to cause phenotype and the final proportion will determine the severity of the disease (24). Since each cell has more than one mitochondria and mtDNA has a higher substitution rate than genomic DNA, a broad spectrum of mtDNA mosaicism associated with age and disease has been described (8,24) (**Figure 9**).

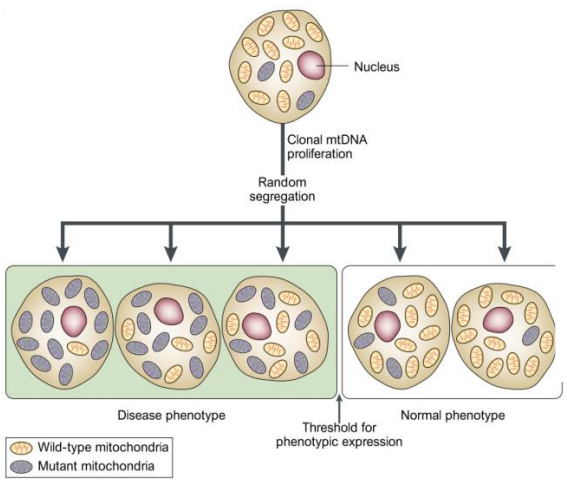


Figure 9. Heteroplasmy phenomenon. Somatic mosaicism for mitochondrial diseases results from the random segregation of mutant and WT mitochondria during mitosis, which can result in daughter cells with different proportions of mitochondrial mutations. Picture from (24).

1.2. Methods for genetic mosaicism detection

1.2.1. General approaches

Genetic mosaicism has been traditionally hard to detect when its percentage is very low and/or the tissue affected is unknown. However, technological advances with higher resolution than traditional techniques let to detect mutations more and more at lower percentage of mosaicism and also in much broad scale (7). There are four main study designs useful for mosaicism and microchimerism detection (3). These are the deep analysis of a single sample in one individual even performing single cell-analysis, the study of different samples (from different tissues or sorted cell populations) in an individual, the evaluation of serially collected samples at different time points during the individual lifetime (25,26) and, finally, the comparative study of monozygotic twins who are derived from a single zygote knowing that whatever genotypic difference detected would be due to post-zygotic variation (25). In the routine medical activity, when a specific genetic alteration in mosaicism is suspected due to a clear phenotype, tissues chosen for the analysis are the once suggested by the syndrome itself if are easy to obtain, that is to say skin in case of patchy pigmentation syndromes as NF1. In cases without phenotypic guidance, mosaicism is studied in more than one easy to obtain sample as blood, buccal, skin and even sperm samples by using sensitive broad scale techniques based on microarray and sequencing technology discussed below (7).

1.2.2. Cytogenetics

Since the early 60's, classic cytogenetics has successfully detected genetic mosaicism both in health and disease. These basically include banded metaphase chromosomes (karyotype) and fluorescence in situ hybridization (FISH) (7). Initially, the karyotype analysis of cells in metaphase enabled determining the total number of chromosomes and identifying the loss or exchange of large DNA portions. With chromosomal banding techniques, distinguishing chromosomes of the same size and even recognizing segmental alterations of a specific chromosome became possible. The efficiency of optic techniques was improved with the introduction of molecular probes labelled with radioisotopes or fluorescence able to hybridize specific genetic regions. FISH uses fluorescently labelled probes to localize and detect the presence or absence of a specific DNA sequence by using fluorescence microscopy. The maximum high-throughput FISH application is the multicolor spectral karyotyping (SKY), typically used to analyze tumor samples, in which some spectrally overlapping probes are mixed to obtain a karyotype with each chromosome “painted” in a different color leading to the detection of structural abnormalities as translocations. Overall, the resolution capacity of these cytogenetic techniques is limited to large events (>5Mb) by the fact that all of them require an optical evaluation through a microscope. This fact together with the low throughput capacity, the unavoidable bias when counting single cells and the induction of aberrations due to in vitro culturing of cells required before the analysis made way to other strategies to detect genetic mosaicism (3,4).

1.2.3. Microarray-based techniques

Since 2005, cytogenetic strategies began to be replaced by microarray-based techniques in the genetic mosaicism field. Microarray-based techniques consist of the use of a set of probes spanning the whole genome and include two different kinds of array: array-based comparative genomic hybridization (aCGH) (also known as molecular karyotype) and single-nucleotide polymorphism (SNP) microarrays (SNP array) (4,7). aCGH was the first technology developed and is based on CGH strategy where differentially labelled genomic DNA from a “test” and a “reference” samples are cohybridized to normal metaphase chromosomes. Ratios of fluorescence intensities obtained at a certain chromosome location on the 'cytogenetic map' are proportional to the copy numbers ratio in that certain genomic region in the test and reference genomes

(27). Because of resolution limitation again attributable to the optic evaluation, CGH evolved to aCGH where metaphases were replaced by array chips containing complementary DNA (cDNA) or oligonucleotides representative of a part or the whole genome (4). Thus, aCGH can analyze genomic CNVs and its resolution is determined by the length of the cloned DNA segments of the array and the distance between them in the genome (27) (**Figure 10**).

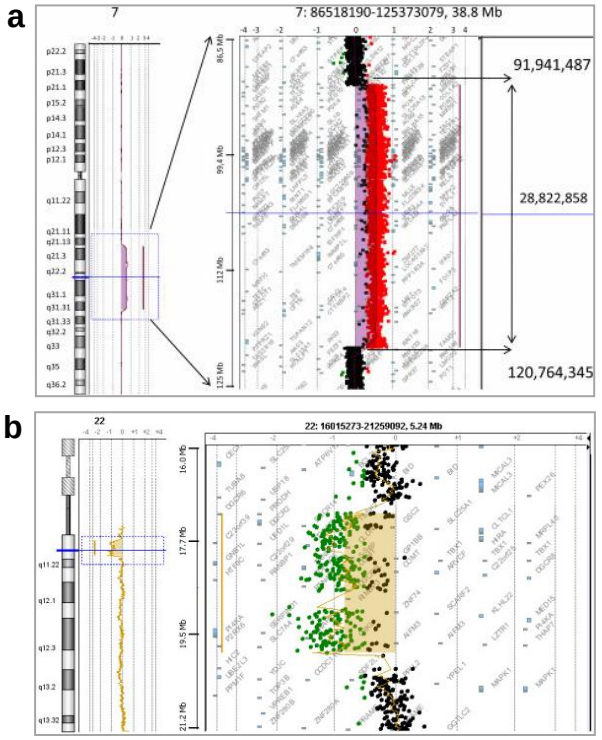


Figure 10. Examples of CNVs detected by aCGH. In both panels, a scheme of the altered chromosome is shown on the left together a zoom in of the rearranged region on the right. Black dots represent similar DNA dosage both in tested and reference DNA whereas red dots (ratios >0) and green dots (ratios <0) represent a gain and a loss of genetic material in the tested DNA versus the reference one respectively. **a)** A 28.8Mb duplication in chromosome 7 (q21.2-31.31) is shown (picture adapted from (28)). **b)** A 2.56 Mb deletion affecting 22q22.12 is shown (picture adapted from (29)).

SNP array are other kind of microarray technology which can detect copy-number-neutral LOH (CNNLOH) besides CNVs by analyzing dosage and allele ratio in a set of SNPs covering the whole genome (7). Several strategies to define mosaic events by using two data tracks from SNP, log R ratio (LRR) and B allele frequency (BAF), have been reported. LRR is a measure of the relative probe signal

intensity and provides data on copy number whereas BAF estimates the frequency of the B allele of a given SNP. In cells without chromosomal events, we expect to detect BAF values of 0 (AA), 0.5 (AB) or 1 (BB) at any locus and LRR near 0. Copy-number and copy-neutral changes will alter BAF, and only copy-number changes will affect LRR. The difference between the observed and expected BAF is denoted as b-deviation (Bdev), so altered regions can also be called by detecting segments with Bdev values different from zero (25,30–37) (**Figure 11**). SNP array resolution depends on the amount of SNPs studied; chromosomal events >2Mb with 7-18% of cellularity can be detected although this threshold can vary depending on the type of rearrangement and the quality of the array (33). Proportion of cells with a mosaic aberration can be calculated with the following reported formulas: % loss = 2Bdev/(0.5+Bdev); % gain = 2Bdev/(0.5-Bdev); % CNNLOH/UPD = 2Bdev (32).

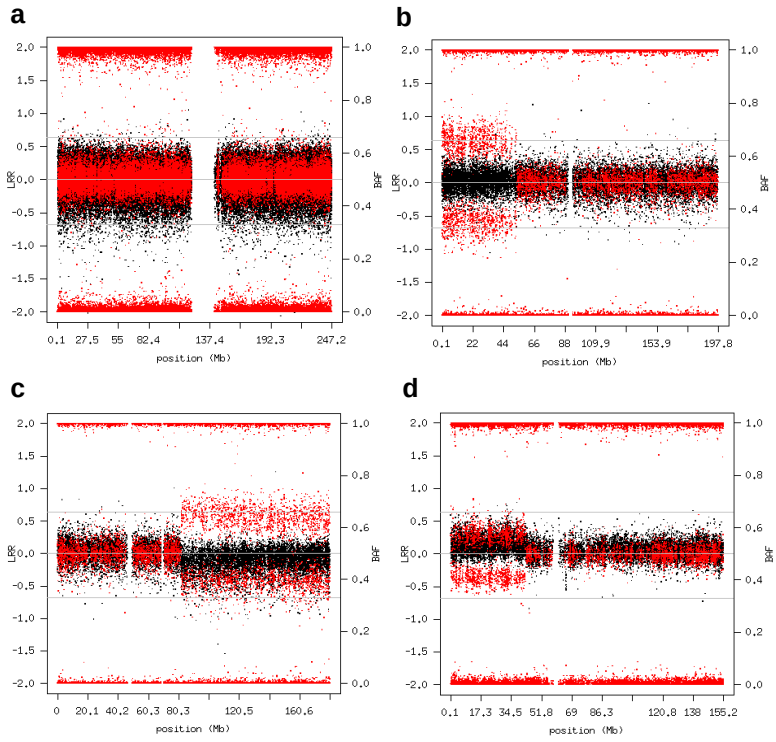


Figure 11. Examples of SNP array plots. a) Normal chromosome; **b)** Terminal mosaic UPD affecting p arm of the chromosome; **c)** Terminal mosaic loss affecting q arm of the chromosome; **d)** Terminal mosaic gain affecting the p arm of the chromosome. In all plots centromere is denoted by a white region without LRR (black dots) and BAF (red dots) signaling. Y axis is double showing a scale for LRR (left) and BAF (right) values. X axis shows chromosome position in Megabases (Mb).

1.2.4. DNA sequencing

Although Sanger-based sequencing can be effective for many applications, it has two main limitations: (I) a limited throughput and (II) the fact that both alleles of an autosomal locus are sequenced at the same time and are shown in the same electropherogram hindering mosaicism detection. Next generation sequencing (NGS) is a high-throughput technique that analyses millions of sequence reads in parallel and leads to analyze the sequence of from the whole genome to target regions; RNA sequencing and genome-wide methylation analysis are other NGS applications (7). Some studies show a complete concordance between Illumina SNP arrays and whole genome NGS at the time of mosaicism detection (38,39). The main advantages of NGS for mosaicism detection are that it does not require genetic mapping data for results analysis and it is an inherently digital assay which reports read counts for each allele as unique counts so, they could be more easily statistically managed to distinguish mosaicism from sequence errors (7). Moreover, NGS has more complete coverage of the genome and is able to detect SNVs, CNVs, CNNLOH and often balanced rearrangements such as inversions or gene fusions. However, there are two main limitations of NGS in mosaicism detection. First is the high cost of this assay compared to SNP array platforms and second, is that bioinformatics pipelines to detect specially CNNLOH in low mosaicism level are not well stablished yet (3).

1.3. Effects of genetic mosaicism

The effect of genetic mosaicism on health is determined by five factors. First, the genomic location of the event can involve housekeeping genes or genes related to cell cycle, having a high probability to cause disease when they do not work properly. Second, the type of the mutation and its negative effect on gene function. Third, the proportion of cells affected by the event, normally it is needed to reach a certain percentage threshold over whom mosaicism has deleterious consequences suggesting that low mosaicism levels can often be tolerated without causing apparent phenotype. Fourth, depending on the tissue/organ affected by the event, mosaicism can have different outcomes. Finally, the time during the development at which the event arises will determine the proportion and distribution of the affected cells so, the phenotype and the severity of the clinical profile (4,5,24).

1.3.1. Clinical manifestations of mosaicism

I. Mosaicism in Mendelian disorders

Postzygotic mosaicism can influence on a phenotype by attenuating or reverting it or unmasking the expression of a mutation that otherwise would be lethal (26). An important fraction of mosaic diseases is that consisting of mosaic forms of same mutations responsible of diseases typically inherited in a Mendelian pattern. In Mendelian disease, mutations are constitutional, so they can be transmitted to the offspring, and its presence in all cells is compatible with life (7). There are continuous new associations between genetic mosaicism and disorders; **table 1** shows some examples of this.

Classification	Disorder
Metabolic disorders	• Tyrosinemia type I • Lesch-Nyhan • Conradi-Hunermann-Happle • Adenosine deaminase deficiency
Immune dysfunction	• Adenosine deaminase deficiency • Wiskott-Aldrich syndrome
Clotting disorders	• Hemophilia A • Hemophilia B
Skeletal disorders	• Marfan Syndrome • Pseudoachondroplasia
Muscle disorders	• Duchenne muscular dystrophy • Congenital myotonic dystrophy
Chromosomal instability	• Bloom syndrome • Fanconi Anemia
Tumors suppressor	• Neurofibromatosis type I • Neurofibromatosis type II • Von Hippel-Lindau disease • Tuberous sclerosis
Skin disorders	• Bullous ichthysiform erythroderma • Incontinentia pigmenti
Endocrine disorders	• Androgen insensitivity
Nervous-system disorders	• Friedreich ataxia

Table 1. Representative list of monogenic disorders in which genetic mosaicism has been reported. This table is an adaptation from (24).

Neurofibromatosis type 1 (NF1) is a classic example of Mendelian disease whose phenotype can be attenuated by somatic mosaicism (26). NF1 is an autosomal dominant inherited tumor predisposition syndrome caused by mutations in *NF1* gene. Although most of patients have germline inherited mutations in *NF1* gene, some individuals have been described to suffer segmental NF1 in which clinical manifestations are limited to certain parts of the body

(including or not gametes). However, variable expressivity oh the phenotype (typical of NF1) should not be confused with mosaicism (7). Fanconi anemia is an example of Mendelian disorder whose phenotype can be restored by mosaicism reversion (3,7,8) (this will further reviewed in an specific section).

II. Lethal disorders that manifest only as mosaicism

There are some monogenic disorders that need to have the mutation causing the disorder in somatic mosaicism to be viable. In these cases, the mutation can be present in germline cells but it will never be transmitted due to its lethality in a constitutional state. McCune-Albright (40) and Proteus (41) syndromes are two well-known examples of lethal disorders that only manifest as somatic mosaicism. McCune-Albright syndrome is characterized by abnormal bone growth, café-au-lait skin spots and endocrine dysfunction and it is caused by gain-of-function mutations in *GNAS1* gene. Proteus syndrome is a disfiguring disease characterized by an irregular overgrowth, risk of tumors and hyperplasia of multiple tissues that is caused by gain-of-function mutations in *AKT1* gene (4).

III. Mosaicism in chromosomal abnormalities

As explained in "Sources of mosaicism" section, chromosome mis-segregation during mitosis can generate somatic mosaic aneuploidies as well as chromosome non-disjunction in meiosis but, here, followed by an aneuploidy rescue in some cells. Trisomies 13, 18 and 21 are commonly detected in somatic mosaicism state in up to 5% of cases of Patau (42), Edwards (43) and Down (44) syndromes respectively. In these cases, a milder phenotype is observed in correlation with trisomic cells fraction detected consisting of, roughly, a lower new-born mortality due to less severe malformations in trisomies 13 and 18 and milder intellectual disability in 21 trisomy. Another relative common form of aneuploidy in somatic mosaicism is Turner's syndrome which consist on the coexistence of 45,X cells with normal karyotype cells that leads to short stature, lymphedema, broad chest, low hairline and infertility with more or less severity depending on the percentage of cells 45,X (45). There is the recent accepted idea that Turner's syndrome is lethal in the constitutional state given the fact that between 1-5% of all pregnancies are 45,X whereas Turner's syndrome occurs in 1 in 1000-2500 live births (3). So, all Turner's syndrome patients would actually be cryptic mosaics where the loss of one X chromosome would have been consequence of mitotic mis-segregation in the embryonic stage. In apparent non-mosaic 45,X individuals, placenta would seem to be the best candidate for the location of the rescue

line (23). Interestingly, an additional cell line with extra X or Y genetic material have been identified in apparently non-mosaic Turner patients by using more sensitivity techniques than classical cytogenetics (46). Giving more complexity to the situation, a 5.7±13.0% increase of diploid 46,XX cells fraction was reported in 70% of tested Turner patients in a 10 years follow-up study suggesting changes of mosaicism level over time (47). Finally, Pallister-Killian syndrome is another example of disease that is only seen in mosaicism. It is caused by an isochromosome 12p which that, by a mechanism still unknown, leads to a wide range of features as craniofacial dimorphism, pigmentary skin anomalies, congenital heart defects and diaphragmatic hernia, hypotonia, intellectual disability and epilepsy (48).

IV. Mosaicism in early embryo and miscarriages

Since 60s, it is well-known that prevalence of chromosome abnormalities is high in spontaneous abortions. Some publications using microarray analysis approaches have revealed that very early embryos (obtained before implantation) harbored a very increased rate of chromosome abnormalities being in mosaicism in 45-70% of studied embryos (49,50). This high rate of chromosomal abnormalities (an important proportion of them in mosaicism) at early development stages provides an explanation for low pregnancy and high spontaneous abortion rates reported in humans (51).

V. Germline mosaicism

There are some autosomal dominant disorders that occasionally show an inheritance pattern that suggest that they primarily manifests as germline mosaic disorders. Osteogenesis imperfecta type II (OI II) is a prototype of this kind of disorders in which it has been reported more than one affected new-born from a healthy couple or with one of the parents with a milder phenotype (52). OI II is caused by mutations in *COL1A1* or *COL1A2* genes and, before detecting germline mosaicism in several parents, it initially was suspected to be autosomal recessive inherited (7). Apparent germline mosaicism has been reported in several diseases in a frequency between 1% (in Apert syndrome) to 11-12% (in Duchenne muscular dystrophy (DMD)) (51,53). In cases were germinal mosaicism is so much prevalent as DMD, it should be considered to give a correct recurrence risk (RR) in a genetic counselling process. A RR due to germline mosaicism has been stablished around 14-20% in DMD non-carriers females when the risk haplotype is transmitted (54).

VI. Mosaicism in mitochondrial diseases

Mitochondrial diseases are typically related to those organs with higher dependence of ATP generated by mitochondria as retina, brain and cardiac and skeletal muscle (24). Although most cases are in heteroplasmy form, a few mitochondrial disorders as mitochondrial dystonia (55) and hypertrophic cardiomyopathy (56) have been reported in an homoplasmy state, it is said, that causative mutation is present in all mitochondria of every cell of one or more tissue usually leading to worse phenotype even higher mortality (24).

VII. Mosaicism in cancer: LOH and Knudson two-hit hypothesis

The most prominent example of somatic mosaicism is cancer and numerous causative mechanisms have been described being LOH the most remarkable one (24). Around 1970, Alfred G. Knudson developed the "two-hit hypothesis" in which he postulated that some tumors as Rb (57) and Tuberous Esclerosis Complex (TSC) (58) need two mutational events to occur. In the inherited forms, a "first hit" would be a dominant inherited mutation through germinal cells that would be in a constitutional state and the "second hit" would be a mutation acquired postzygotically in somatic cells. In nonhereditary forms, both mutations would appear in somatic cells during the development (57) (**Figure 12**).

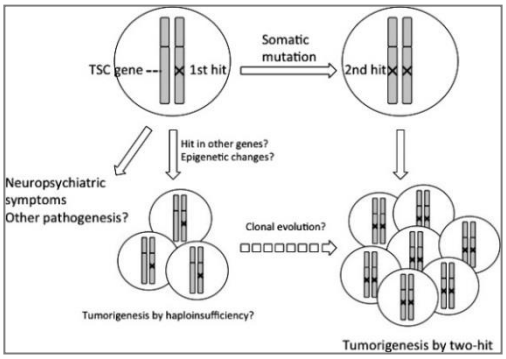


Figure 12. Two-hit hypothesis in TSC. Germline mutations (1st hit) cause neuropsychiatric symptoms during infancy and probably other lesions as cortical tubers without the presence of a 2nd hit. Most of tumorous lesions appear after two hits in TSC genes although tumorigenesis without a 2nd hit has been reported in animal models. Picture adapted from (58).

LOH is the major mechanism associated with somatic tumor suppressor genes inactivation and various mechanisms responsible of this have been described including both DNA copy-number changes and copy-neutral changes. In LOH, the first event is usually a mutation inactivating the first allele whereas de second hit can be acquired by also an inactivating mutation (including promoter

methylation), loss of the allele, gene conversion[3] or UPD of the allele with the 1st hit. LOH by UPD (copy-neutral change) is probably the predominant mechanism by which a second hit can be acquired in cancer and the main mechanisms involved are: (I) segmental UPD through HR and NAHR and (II) whole-chromosome UPD in somatic cells that appears as a result of the rescue of a trisomy or a monosomy secondary to chromosome mitotic or meiotic mis-segregation (these mechanisms are reviewed in section "Sources of mosaicism") (59). LOH through UPD also can lead to oncogene activation by mutation duplication leading to increased gene product and cell proliferation (60).

General increased prevalence of UPD, not only associated to 1st hit homozygosis phenomenon, have been reported in many cancer types both hematologic and solid (**table 2**).

UPD	Disease	Frequency of UPD	Affected Gene	Abnormality
1p	MDS/MPN	8.9%	*MPL*	Mutation
2p	HNPCC	52%	*MSH2*	Mutation/ Deletion
2q	MCL	10-16.7%*	*MAP2*	Deletion
3p	HNPCC	4%	*MLH1*	Mutation
	Colorectal/Esophageal Cancer	1.1%/73.9%	*FHIT*	Deletion
4q	MDS MDS/MPN	3.9-8.7% 8.8%	*TET2*	Mutation
5q	Colorectal cancer	28.6-44.4%*	*APC*	Mutation
6p	Loss of GVLE (leukemia relapse after HSCT)	29.4%	*HLA-A,B,C*	Loss of mismatch
6q	FL; DLBCL; MALT	8%; 3.1%;10.3%	*A20*	Mutation/ Deletion
7q	MDS/MPN	6%	*EZH2*	Mutation
9p	AML; MDS/MPN; MPN PV; ET; PMF	5%; 11%;25-43% 41-80%; 5.9-17% 43.8-67%	*JAK2*	Mutation
	AML; ALL	2.6%; 7.1-29%	*CDKN2A*	Del, Del
	FL; MCL	33%*; 7.1**-60%*		Del, Del
	Esophageal Carcinoma	26.1%**		Deletion
	Ovarian cancer	7.5%**		Deletion
	Glioblastoma	3.3%**		Deletion
	Neuroblastoma	4.3%**		Deletion
	CNS lymphoma	21.1%**		Methylation/ Deletion
	Colorectal cancer	55.6%*		Methylation
9q	BCC	35.7%**	*PTCH*	Mutation
11p	AML; APL	3.2-4.5%**; 6.4%	*WT1*	Mutation

[3] Gene conversion is a unidirectional transfer of genetic material from a "donor" sequence to a highly homologous "acceptor" (274).

	AML	4.7%**	*H19*	Methylation
	Hepatoblastoma	23.5%**	*IGF2, H19*	Methylation
	Rhabdomyosarcoma	33.3%**	*HRAS*	Mutation
	Wilms' tumor	2.5-5.6%**	*CDKN1C, IGF2, H19*	Methylation
	Wilms' tumor	36%**	*WT1*	Mutation
	Beckwith-Wiedemann syndrome	7.2-16.8%**	*CDKN1C, IGF2, H19*	Methylation
11q	MDS/MPN	4.9%**	*CBL*	Mutation
13q	AML	2.3-5.4%**	*FLT3*	Mutation
	CLL	3.6%**	*miR-15a/ 16-1*	Deletion
	MCL	10%*	*RB1*	Deletion
	Breast , Ovarian Cancer	6%**, 23.8%**	*RB1*	Deletion
	Retinoblastoma	59.5%**	*RB1*	Mutation
	Ovarian Cancer	15%**	*BRCA2*	Mutation
17p	MDS, CLL	1.8%, 6.1%	*P53*	Mutation
	FL	19.2%		
	MCL	3.8-10.7%		
	Colorectal cancer	57.1%*		
	Breast cancer	6%**		
	Glioblastoma	3.3%**		
17q	JMML	25-80%**	*NF1*	Mutation
	Ovarian cancer	40%**	*BRCA1*	
19q	AML	0.6-1.6%**	*CEBPA*	Mutation
21q	AML	2.6%	*RUNX1*	Mutation

Table 2. UPD prevalence and affected genes in various cancers. This table, adapted from (60) shows the general increased prevalence of UPD in several malignancies, the gene/s affected in each case and the kind of abnormality responsible of every disease. MDS: Myelodysplastic Syndrome; MPN: Myeloproliferative Neoplasms; HNPCC: hereditary non-polyposis colorectal cancer; MCL: mantel cell lymphoma; GVLE: graft versus leukemia effect; HSCT: hematopoietic stem cell transplantation; FL: follicular lymphoma; DLBCL: diffuse large B-cell lymphoma; MALT: mucosa-associated lymphoid tissue; AML: acute myeloid leukemia; PV: polycythemia vera; ET: essential thrombocythemia; PMF: primary myelofibrosis; ALL: acute lymphoblastic leukemia; CNS: central neural system; BCC: basal cell carcinoma; APL: Acute promyelocytic leukemia; CLL: chronic lymphocytic leukemia; JMML: juvenile myelomonocytic leukemia; Del: deletion; *data from cell line; **data from primary sample.

Due of its remarkable involvement in cancer and its specific importance in this thesis, UPDs' roles in disease are deeply reviewed below. UPDs are not randomly distributed across the genome and there is some kind of relationship between UPDs and gene mutations associated with cancer (61). Accordingly to Knudson's hypothesis, acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) are two examples in which all segmental UPD13q were found to harbor *FLT3* mutations (62), almost all UPD11p had *WT1* mutations (62) and all UPD17q co-

occurred with NF1 mutations (63), in all cases mutations were in homozygosis showing that UPD was leading to 1st hit homozygosis. Large-scale studies have also shown that UPD can cause deletions' homozygosis, for instances, UPD13q can lead to homozygous deletions of miR-15a and miR-16-1 in chronic lymphocytic leukemia (CLL) (64) or homozygosis deletions affecting *Rb* gene in RB disease (65). An interesting case is UPD9p in JAKV617F mutation-positive patients in which JAKV617F homozygosis through UPD will lead to a certain myeloproliferative neoplasm (MPN) (polycythemia vera (PV), primary myelofibrosis (PMF) or essential thrombocytosis (ET)) depending on the JAKV617F mutational burden (60,61,66). UPD do not always appear as a driver mechanism to cancer but also as a consequence of it. Certain cancers display a higher frequency of UPD due to the presence of fragile sites prone to recombination or chromosome instability, examples of this are MUTYH-associated polyposis colon carcinomas (60,67).

Although, UPD is a well-known mechanism that promotes cancer by leading the homozygosis of the mutated allele, it also can promote the homozygosis of the WT allele generating "health clones" in mosaicism and trying to avoid disease development. This "rescue" UPD phenomenon has been described in a severe sporadic skin disease, ichthyosis with confetti (IWC), that can be caused by dominant mutations in *keratin 1* or *10 (KRT1, KRT10)* genes and is characterized by the accumulation of thousands "health" skin spots harboring homozygous WT alleles by UPD17q. The observed high frequency of somatic reversion in IWC patients suggests that revertant clones are under strong positive selection and/or the reversion rate is elevated (68). Similarly, reversion thought mitotic recombination of mutant *TERC* (telomerase RNA component) alleles was detected in six patients from four families affected by dyskeratosis congenita (DC) (69).

Heterodisomy UPD is not expected to cause any abnormality unless genes within the affected region are imprinted. This phenomenon is defined as the monoallelic expression of a gene since one copy is silenced through methylation depending on allele's parental origin. When a UPD occurs in a region with imprinted genes, two active/unmethylated or two inactive/methylated alleles are transmitted to daughter cells leading to loss of imprinting (LOI) or gain of imprinting (GOI) respectively and altering final gene expression (59,61). Certain neurodevelopmental disorders, some of them with increased cancer risk, are associated with imprinting

pattern alterations due to UPDs. Examples of this are Prader-Willi syndrome (PWS) caused by maternal UPD15 (both copies inherited from the mother) in ~25% of cases; Angelman syndrome (AS) were ~2% of cases are due to paternal UPD15; Beckwith-Wiedemann syndrome (BWS) in which paternal UPD11p15.5 is behind ~20% of BWS cases; and finally, Silver Russell syndrome (RSS) in which maternal UPD7 leads to 6-10% of cases (61). In cancer, GOI and LOI phenomenon have been reported, for instances, GOI of *CDKN1C* and *IGF2* genes in hepatocellular carcinoma and LOI of *IGF2* gene in colorectal cancer and Wilms' tumor (59).

1.3.2. Mosaicism in general population

Mosaicism frequency is usually underestimated because it does not always cause a phenotypic effect. SNP array technology applied to studies of GWAS have revealed a more "real" mosaicism frequency (30–32,70). However, resolution of SNP array technique, although being high, is also contributing to the underestimation of mosaicism prevalence. Thus, the term "detectable clonal mosaicism" will be used from now on to refer to those chromosomal mosaic events (CMEs) occurring in a proportion of cells sufficient for their detection.

I. Mosaicism and aging

Several publications have reported an increased prevalence of mosaicism with age in the last years. In 2010, large structural variants (1.5-37Mb) were reported in 1.7% of individuals free of hematologic malignancies (32). Two years later, mosaicism prevalence of large events (Mb) around 3.4% was observed in a healthy cohort between 55-90 years but failing to detect mosaic events in individuals younger than 55 years. Moreover, in this work, the use of a unique cohort of age stratified monozygotic twins studied longitudinally showed that smaller events (of few Kb) also accumulated with age (25). Two parallel studies demonstrated again a frequency of detectable large CMEs (>2Mb) in autosomes increasing with age (<0.5% in individuals <50 years; 2-3% in individuals >75 years) (30,31). These results were in agreement with a recent large analysis of 127000 adults where overall mosaicism (>2Mb) prevalence was detected around 0.73% and an approximate 6-fold increase in CMEs detection was observed in individuals aged 75 years compared to those under 50 years (70). Interestingly, although all these studies reported detectable CMEs falling in the three major categories (losses, gains and CNNLOH or acquired UPD (aUPD)), an unexpected high frequency of CNNLOH/aUPD was reported by three different works: 22% (25), 34% (31), 48% (30); where differences were due to the use of different algorithm in CMEs detection.

II. Mosaicism and loss of sexual chromosomes

Since the 60s, it is known that the somatic loss of Y chromosome (LOY) leading to mosaicism is frequent among aging man (71). The recent analysis SNP array data of blood DNA from a large number of men with ages ≥70 years old revealed mosaic LOY in 20% of them (38,39,72,73) being mosaic LOY the most common aberrant clonal expansion in elderly population. Mosaic LOY in blood cells is more common among current smokers than in non-smokers (72,73) and that effect is both dose dependent (72) and transient, since mosaic LOY disappear from blood after stop smoking (72,73). Interestingly, a germline variant near *TGL1A* gene has been found to increase mosaic LOY suggesting a heritable predisposition to, at least, mosaicism of Y chromosome (73). Mosaic loss of X chromosome (LOX) has also been described in a non-leukemia cohort (around 40000 healthy women or solid tumors patients) with a ~0.25% prevalence, being four times the mean autosomal rate. The frequency of X mosaicism was also shown to increase with age and methylation analyses revealed that the inactive X chromosome is preferentially gained or lost in X mosaic events (74).

1.3.3. Mosaicism as a biomarker

The use of SNP array and NGS has given the opportunity to investigate detectable CMEs prevalence in different tissues as leukocytes, buccal swabs or skin. The detection of an increased prevalence of mosaicism among "healthy" aging population leads to an unavoidable question: can detectable CMEs be used as a biomarker of certain chronic diseases associated with ageing such as cancer, diabetes mellitus and neurodegenerative diseases? (5).

Several works that reported an association between age and mosaicism also detected an increased prevalence of detectable CMEs in cancer patients, being stronger in hematological cancer (OR: 22-30) than in solid tumors (OR: 4), suggesting a potential use of CMEs detection as a biomarker of cancer risk (**Figure 13**) (30,31,34,70). Higher rates of mosaic events have been detected in males (0.98%) compared to females (0.56%) after removing sex-specific cancer and adjusting for ancestry, 5-year age group, and cancer subtype. Elevated male-specific rates of hematologic cancer could contribute to this increased mosaic CMEs prevalence among males.

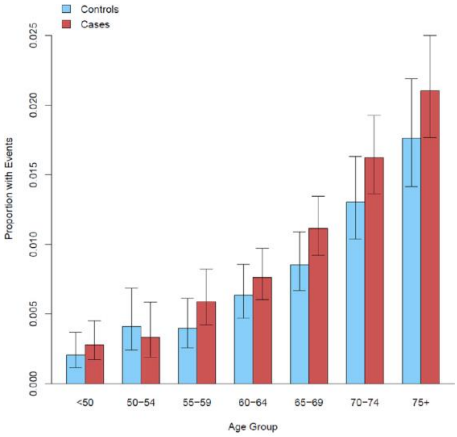


Figure 13. Increasing prevalence of detectable CMEs with age and cancer. By combining Jacobs et al. (30) and Laurie et al. (31) data with new results (70), Machiela et al. show an increasing frequency of mosaicism with age ($p=1.1x10^{-30}$) and a higher prevalence of CMEs among cancer patients (red bars) compared to cancer-free individuals (blue bars). Error bars represent 95% confidence intervals. Picture from (70).

Although the relationship between cancer and somatic clonal mosaicism is proven, we are a bit far from determining specific somatic mosaic events that directly lead to particular diseases. However, mosaic 13q14 and 20q deletions, typically associated with hematologic malignancies, have been frequently found in individuals free of leukemia at time of sample collection. This would suggest that such mosaic deletions in circulating blood or buccal samples could be used as biomarker of the future risk for hematologic malignancies (30–32,70,75,76). In fact, a mosaic 13q14 deletion was even detected in a blood DNA sample collected 14 years before CLL diagnosis (30). Increased frequency of CMEs has also been reported in other conditions associated with aging such as type 2 diabetes specially accompanied with vascular complications (77), coronary heart disease and ischemic stroke (78).

Mosaic LOY has traditionally been considered to have a neutral phenotypic effect being consequence of normal ageing without further ado (71). However, a recent study reported that mosaic blood LOY was associated with a decreased median survival time and with an increased risk of both diagnosis of and mortality secondary to non-hematologic cancer. However, males with and without mosaic LOY showed similar spectrum of cancer (39). Mosaic LOY has also been associated with other solid cancers as prostate (73,79), bladder (73) or colon cancer (79) and linked with hematologic malignancies (80,81), some autoimmune disorders (82) and Alzheimer disease (38). Although mechanisms by which LOY contributes to this wide range of medical conditions are still not completely understood, it seems clear that LOY could be useful as a biomarker for certain age-related diseases and partially explain why men's life expectancy is shorter than women's. No association of LOX with cancer have been reported until now (74).

1.3.4. Mosaicism in human physiology

There are several well-known examples of locus-specific somatic variation in both nuclear and mitochondrial genomes associated with physiological processes. Examples of “physiologic” somatic mosaicism are generation of immune diversity (3), gradual telomere shortening secondary to aging (83), mitochondrial heteroplasmy (24), the presence of fetal cells in mother’s blood for decades (microchimerism) (84), variability among monozygotic twins (24), large-scale structural genetic variants in early stages of embryonic development (50), intra-organ somatic variation (polyploidy in hepatocytes of the adult human liver (85) and postzygotic variation in normal human brain cells (10–12,86)) and inter-organ somatic mosaicism (87).

2. HEMATOLOGIC MALIGNANCIES

Hematologic malignancies constitute a great group of heterogeneous neoplastic conditions that affect cells originated in bone marrow (88). Their incidence, globally, has been reported as 135000 estimated new cases in 2007 whereas the prevalence as >800000 affected persons in 2004 in United States (US) (89). Historically, hematologic malignancies have been divided in three groups according to the organs most involved: leukemia, where cancerous cells (white blood cells, WBC, also known as leucocytes) are present in blood and bone marrow; lymphomas, where malignant cells (lymphocytes) tend to aggregate in lymphatic tissues as lymph nodes; and myelomas, where cancerous cells (mature B lymphocytes also known as plasma cells) accumulate in bone marrow and bones' surface. Broadly, leukemia group is subdivided in "lymphocytic/ lymphoblastic" and "myeloid" depending on whether the affected cells are from lymphoid or myeloid lineage respectively (**figure 14**). Leukemia is also distinguished between acute, when having a rapid progressing with very immature cells, or chronic, when disease progress is slower with cells relatively differentiated (90). Lymphomas are subclassified in Non-Hodgkin's lymphoma (NHL) (90%) and Hodgkin's lymphoma (HL) (10%) where HL is distinguished from NHL by the presence of distinctive cells called Reed-Sternberg cells (giant cells derived from B-cells) (90,91).

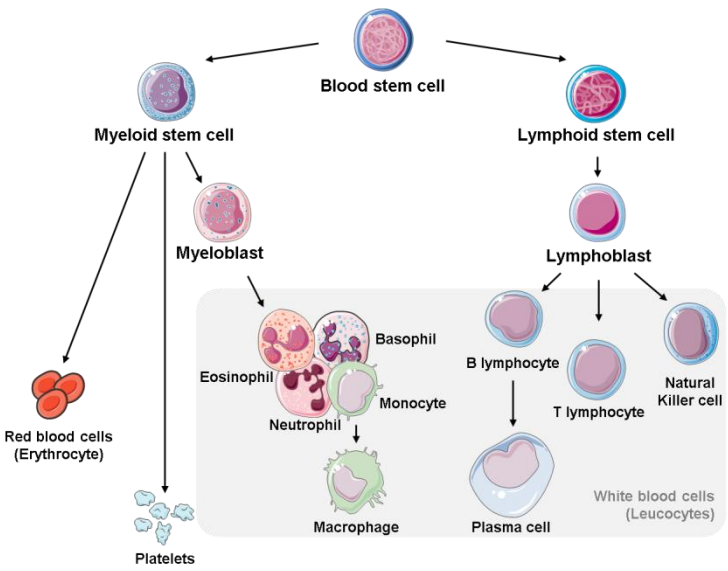


Figure 14. Simplified scheme of blood cells production. Picture performed by using *Servier Medical Art.*

World Health Organization (WHO) is continuously updating hematologic malignancies classification according to their cell lineage, genetic abnormalities and clinical features; a summary from the 2016 update is shown in **table 3** (92,93).

Abbreviated summary of WHO of myeloid neoplasm and acute leukemia classification
Myeloproliferative neoplasms (MPN) Polycythemia vera (PV) Primary myelofibrosis (PMF) Essential thrombocytosis (ET)
Myeloid/lymphoid neoplasms with eosinophilia and rearrangements of *PDFFRA, PDGFRB,* or *FGFR1,* or with *PCM1-JAK2*
Myelodysplastic/myeloproliferative neoplasm (MDS/MPN) Atypical chronic myeloid leukemia (aCML), *BCR-ABL1* Juvenile myelomonocytic leukemia (JMML)
Myelodysplastic syndromes (MDS)
Acute myeloid leukemia (AML) and related neoplasms Acute promyelocytic leukemia (APL)
Blastic plasmacytoid dendritic cell neoplasm
Acute leukemias of ambiguous linage
B-lymphoblastic leukemia/lymphoma (ALL)
T-lymphoblastic leukemia/lymphoma (ALL)
Abbreviated summary of WHO of mature lymphoid, histiocytic and dendritic neoplasms
Mature B-cell neoplasms Chronic Lymphocytic Leukemia/small lymphocytic leukemia (CLL/SLL) Monoclonal B-cell lymphocytosis (MBL) Plasma cell myeloma Mucosa-associated lymphoid tissue (MALT) Follicular lymphoma (FL) Mantel cell lymphoma (MCL) Diffuse large B-cell lymphoma (DLBCL)
Mature T and NK neoplasms
Hodgkin lymphoma
Posttransplant lymphoproliferative disorders (PTLD)
Histiocytic and dendritic cell neoplasms

Table 3. Abbreviated WHO classification of hematologic malignancies. This table is a summary of the main subgroups in which the hematologic malignancies entity is divided by WHO (92,93). Certain examples in each category are shown, selected because of being diseases with a remarkable UPD prevalence (reported in table 2) or due to its importance for this thesis.

2.1. Leukemia

Leukemia is a malignant disease characterized by an uncontrolled proliferation and development of WBC or leucocytes and their precursors in blood and bone marrow (94). Worldwide, leukemia is the 11th most common cancer with around 352000 new cases diagnosed in 2012 representing 2.5% of the total cancer cases. It is more frequent in males than in females, being the 10th (2.7% of all male cancer cases) and the 12th (2.3% of all female cancer cases) most common malignancy respectively (95). Similarly to other cancers, leukemia incidence is strongly related to age, with the highest incidence rates in individuals older than 70 years old (96). However, this is also the most common malignancy among children (~35% of cancers between 0-14 years of age) (88). Leukemia is classically divided into four major types according to the slow or fast disease progress (chronic or acute respectively) and the malignant leucocytes' origin (lymphocytic/lymphoblastic or myeloid): chronic lymphocytic leukemia (CLL), chronic myeloid leukemia (CML), acute lymphocytic leukemia (ALL) and acute myeloid leukemia (AML).

2.1.1. Chronic lymphocytic leukemia (CLL)

CLL is a leukemia subtype characterized by the accumulation of small, mature-appearing neoplastic CD5 positive B-lymphocytes in blood, bone marrow and secondary lymphoid tissues as lymph nodes (97,98). When B-cells with CLL immunophenotype are detected in enlarged lymph nodes but not in blood, the term small lymphocytic lymphoma (SLL) is used, indicating a clinical variant of the same histopathological and molecular entity (99). CLL is the most common type of leukemia in the Western world and is two-fold more prevalent in males than in females. Its incidence has been reported as ~19000 newly diagnosed cases in US in 2016 being equivalent to 1.1% of all new cancer cases (100). CLL frequency increases with age with around 90% of cases occurring from the 5th decade of life (mean age of onset at 70 years) and being inexistent in individuals younger than 30 years (88,101). However, the proportion of young people at very early CLL stages without clinical features is growing up due to more frequent blood testing.

I. Classification

CLL can be divided in two subtypes, with different clinical behavior, depending on whether CLL cells express a mutated or unmutated *immunoglobulin heavy-chain variable region (IGHV)* gene which reflects the stage of the B-cell at which CLL cells appeared. CLL

cells expressing unmutated *IGHV* genes come from undifferentiated B-cells and involve a more-aggressive disease with poorer prognosis than CLL cells expressing mutated *IGHV* which arise from differentiated B-cells (102).

II. Hereditary and environmental factors

Genetic predisposition is the best understood risk factor for CLL since approximately 10% of individuals with CLL report a family history of CLL or a related lymphoproliferative disorder (103). Asian population have significant lower rates of CLL compared to Caucasian-Americans or Europeans and they were maintained low even after migration to USA suggesting that genetic rather environmental factors have a key role in these differences (104). Another factor that would support the genetic influence in CLL is that concordance between twins have been reported to be higher in monozygotic than dizygotic twins (105). An spectacular family known as "Pedigree 14" in 1947 also provided substantive evidence for an inherited predisposition to CLL since three siblings with CLL were initially reported (106) and 10 descendants with CLL were detected in a 57 years longitudinal study (1947-2004) (107). Interestingly, first-degree relatives of CLL patients show 3-8.5 times increased risk of developing CLL (103,108). There is some controversy on whether familiar CLL show 10 years younger age of onset compared to sporadic cases (109) or not (108). Similarly, some studies propose that there is a 15-22 years of anticipation in the age of onset together a more severe phenotype in families with familiar CLL (110) whereas other works disagree (111). No genes when mutated in the germ-line have been reported to unambiguously confer an increased risk of CLL (109). However, genome-wide association and linkage studies have allowed to identify many loci associated to CLL involved in apoptosis and immune pathways (112,113), suggesting that the genetic component of the inherited risk of CLL is probably polygenic mediated by low penetrance alleles; some of them probably may be common (109).

Regarding environmental factors contributing to CLL development, exposure to Agent Orange[4] (114) and to insecticides (115) are considered CLL risk factors. In contrast, little evidence indicates that ionizing radiation (116) and viral infections (98) influence on CLL risk whereas no relationship between blood transfusions (117) and dietary and lifestyle factors (98) with CLL has been reported.

[4] Agent Orange is an herbicide and defoliant widely known for its use by the US military during the Vietnam War from 1961 to 1971 (114).

III. Somatic alterations

Somatic mutations

All cancers arise as a result of the accumulation of somatic genetic changes. However, only few of these changes will be causally implicated in cancer development ("driver mutations") whereas the rest will be consequence of the oncogenic process itself ("passenger mutations") (118). CLL is a heterogeneous disease under genetic point of view but several NGS studies including both whole exome (WES) and genome (WGS) sequencing have identified a wide range of potential CLL drivers: *NOTCH1, ATM, SF3B1, BIRC3*, *CHD2, TP53, MYD88, XPO1, KLKL6*, *POT1, FBXW7, DDX3X, ZMYM3, MAPK1, ZNF292, ARID1A, ZMYM3* and *PTPN11* (119,120). Globally these genes are involved in multiple pathways as B-cell receptor (BCR) signaling, cell cycle regulation, apoptosis, DNA damage response, chromatic remodeling, NF-κB signaling, NOTCH signaling and RNA metabolism (121).

Chromosomal alterations

Around 80% of CLL patients harbor at least one of the four most common chromosomal aberrations in CLL: deletion in 13q14.3 (del13q14.3) (>50%), trisomy 12 (15-20%), del11q (18%, disrupting *ATM* and/or *BIRC3*) and del17p (7%, disrupting *TP53*) (118,122,123). The detection of one of these alterations differentiates patients with "favorable/intermediate prognosis" (those with del13q14.3 or trisomy 12) from those patients with "adverse prognosis" (with del11q or del17p) (118). Del13q14.3 is the most common chromosomal alteration present in >50% of CLL patients. This deletion is mostly monoallelic (76%) but can be biallelic (24%) and occurs both in CLL with somatically mutated and non-mutated IgV genes although it is more prevalent in the former subtype (80% versus 20%) (122–125). Importantly, 13q14.3 region undergoes an epigenetic inactivation of one chromosome randomly, similar to X inactivation, but independent to the parental origin of each chromosome (maternal or paternal chromosomes can equally and randomly be inactivated in each tissue). In this way, the two copies of the critical region replicate asynchronously due to a differential chromatin packaging and genes within the inactivated region show monoallelic expression. Consequently, there is an increased risk of functional nullizygosity if a mutation occurs in the functional allele (126). Deletions in 13q14.3 region can be subclassified depending on whether they include *RB* gene (type II, 20%) or not (type I, 80%), having consequence in patients' prognosis since an accelerated clinical course is usually detected in the former group (127). The minimal deleted region (MDR) in

CLL patients would include the *long non-coding RNA deleted in leukemia (DLEU)-2*, the first exon of the *DLEU1* gene and miRNA-15a and miRNA-16-1 within the intron 4 of *DLEU2* (122) (**Figure 15**).

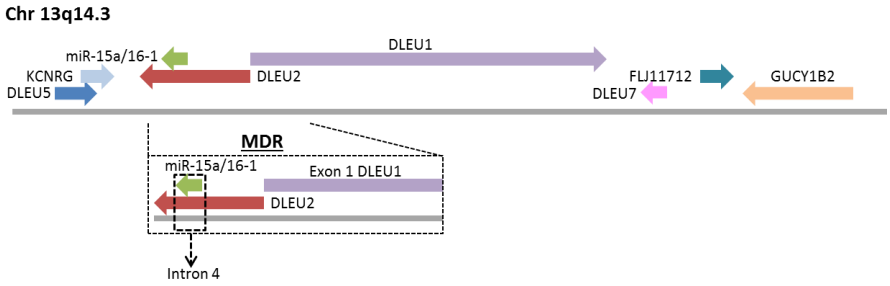


Figure 15. Scheme of 13q14.3 chromosome region and MDR.

The miRNA-15a/16-1 require *DLEU2* promoter for its expression so, any genetic alteration that affects *DLEU2* also affects mRNAs. MDR deletion in mice recapitulates the spectrum of CLL-associated phenotypes observed in humans (122). Moreover, a downregulation of *DLEU2* and miR-15a/16-1 cluster compared to normal B-cells has been proven in CLL patients with del13q14.3 (128).

miRNA alterations

CLL was the first human disease associated with the altered expression of miRNA (98) which are small non-coding RNAs expressed in a tissue specific manner that are able to regulate gene expression by targeting an mRNA or inhibiting its translation. Their deregulation can lead to an alteration of the expression level of many genes and promote the progression of tumors. The miRNA-15a/16-1 function is deleted, altered or downregulated by a deletion or translocation (129) in around 60% of CLL patients (128). The miRNA-15a/16-1 act as a tumor suppressor gene by inhibiting the antiapoptotic genes *BCL2* and *MCL1*, so, reduced expression or loss of these miRNAs enhances the expression of these target oncogenes and promotes cell cycle progression by avoiding apoptosis (130). Deletion of miRNA-15a/16-1 is enough to cause CLL phenotype in mice, however the significantly more aggressive phenotype displayed by the MDR-deleted mice (both in constitutionally or conditionally deleted mice) suggesting that additional genetic elements within the MDR locus contribute to the tumor suppressive function (122). Loss or reduced expression of miRNA as miRNA-29a/b, miRNA-29c, miRNA-34b, miRNA-181b and/or miRNA-3676 have been associated with CLL by increased expression of *TCL1A* gene (98). In contrast, higher expression of miRNA-155 increases

CLL aggressiveness (131). Finally, some works report that tumor-suppressive miRNAs can be secreted into the extracellular media and act in a paracrine way in a cell competitive process showing the importance of microenvironment for tumor formation (132).

Epigenetic changes

The term "Epigenetic" involves any process that alters gene activity without changing the DNA sequence including modifications as methylation, acetylation, phosphorylation, ubiquitination, and sumolyation that can be transmitted to daughter cells (133). Epigenetic pattern most commonly observed in human malignancies is a global hypomethylation together with local promoters hypermethylation (134). General hypomethylation has also been reported in CLL but with poor correlation with gene expression (135). Intra-tumor methylation pattern has also been analyzed and "locally disordered methylation" was detected in CLL cells compared to normal B-cells (136). Aberrant methylation of *BCL2* (137), *NOTCH1* (138) and *ZAP70* (138) as well as of some miRNAs in 13q14.3 region (139) have been reported in CLL.

IV. Diagnosis and prognosis

An important fraction of CLL patients are asymptomatic and they are diagnosed by chance in a routine blood count. However, CLL have a wide range of clinical manifestations as fatigue, involuntary weight loss, excessive night sweets, abdominals fullness with early satiety, increased frequency of infections, anemia (low blood counts of red blood cells (RBCs) or hemoglobin), easy bleedings and lymphadenopathy, hepatomegaly and/or splenomegaly palpable with physical examination which are enlarged lymph nodes, liver and spleen respectively (98). The International Workshop on CLL (iwCLL) 2008 revision of the NCI-96 guidelines established the diagnostic threshold for CLL to a B-cell lymphocyte count of $5.0x10^9$/L rather than absolute lymphocyte count of $5.0x10^9$/L as considered in previous guidelines version, in both cases this elevated count should be maintained at least for 3 months (140,141). Detection of B-cells expressing CD5, CD19 and CD23 with low and lacking levels of CD20 and CD10 respectively is characteristic of CLL cells. By blood smears, CLL cells should be seen as small mature-appearing cells with a big nucleus leading to a narrow border of cytoplasm, condensed chromatin and with indistinguishable nucleoli. These cells can be found admixed with larger and atypical lymphocytes or prolymphocytes which cannot exceed 55% of total blood lymphocytes fraction. Bone marrow biopsy is

often performed showing increased percentage of mature-appearing B-cells and low fraction of myeloid and erythroid cells with normal maturation. Finally, in case of lymph nodes enlargement, a biopsy is mandatory (97,98).

CLL patients are categorized accordingly their clinical features and prognosis by using Rai and Binet staging systems (97). Poor CLL patients' prognosis is determined by male sex, ≥65 years of age, poor performance status due to medical comorbidities, certain CLL cell characteristics as expressing unmutated *IGHV* gene, *ZAP70,* CD49d or CD38, the presence of del11q or del17p usually disrupting *ATM* and *TP53* genes respectively, a karyotype with 3 or more rearrangements, <3.5mg/L of β2-microglobulin in serum, high absolute lymphocyte count >$5.0x10^9$/L and/or late-stage disease at initial presentation (98,102,123).

V. Management

Newly diagnosed patients with asymptomatic early stage CLL should be kept monitored without therapy unless disease progresses. CLL treatment should be initiated when there is evidence of bone marrow failure (pronounced anemia or thrombocytopenia (low blood platelets count) refractory to standard therapies), symptomatic splenomegaly or lymphadenopathy, progressive increased lymphocytosis (high blood lymphocytes count) over 50% for more than 2 months, lymphocyte doubling time[5] of less than 6 months and/or symptoms related to active CLL. Summarizing, the main CLL treatments are chemotherapy (purine analogues and alkylating agents), combination of chemotherapy with immunotherapy (i.e. anti-CD20 monoclonal antibody) which is considered the first-line treatment for patients with good physical conditions since this combination increases progression-free survival compared to chemotherapy alone, and drugs that specifically target those signaling pathways that promote the growth and/or survival of CLL cells (i.e. inhibitors of *BCL-2* and BCR signaling as BTK, PI3K and SYK inhibitors). An allogenic hematopoietic stem cell transplantation (HSCT)[6] should be offered to patients with high risk features such as del17p, with refractory CLL or at the first or second relapse. However, this strategy is considered as one of the last options

[5] Lymphocyte doubling time (LDT) is defined as the period of time needed for lymphocytes to double in number the amount found at diagnosis (275).

[6] Allogenic HSCT is a kind of HSCT in which the donor is genetically similar but not identical to patient (a patient's relative or unrelated individual) (169).

due to problems related to donor availability, advanced patients age, associated toxicities of myelosuppression, graft-versus-host disease[7] and vulnerability to infections. The existence of better tolerated and also efficient BCR signaling and *BCL-2* inhibitors also contributes to make allogenic HSCT less desirable. CLL patients have increased risk of other diseases as infections, autoimmune disorders and secondary cancers not only due to the disease itself but also secondary to the treatments (97,98).

VI. Monoclonal B-cell lymphocytosis

Monoclonal B lymphocytosis (MBL), firstly described in 2002 (142), is an asymptomatic preclinical hematologic condition defined by the presence of less than $5.0x10^9$/L clonal B-cells with a nearly identical immunophenotype as CLL cells in healthy individuals (141,143). MBL is the precursor state for CLL but not all MBLs progress to CLL (144,145). MBL prevalence progressively increases with age being 0.2-0.3% at <40 year, 3.5-6.7% at 40-60 years and 5-9% >60 years. However, when using high sensitivity methods, MBL frequency in individuals older than 60 years is >20% being around 75% in individuals older than 90 years (143). MBLs can be divided in three categories according to their immunophenotype: typical CLL-like MBL (75% of all cases) in which MBL cells have the same immunophenotype as CLL cells; atypical CLL CD5+ in which MBL cells do not fully meet the criteria to be typical CLL cells neither Mantel Cell Lymphoma cells; and finally, CD5(−) non-CLL MBL where cells are characterized by an immunophenotype consistent with marginal zone origin and displays many similarities with marginal zone lymphomas (MZL) (142,146). In turn, CLL-like MBL is subdivided into low-count (LC) and high-count (HC) MBL based on a cut-off value of $0.5x10^9$/L clonal B-cells (147). LC-MBL probably is not a premalignant but an age-related immune senescence condition since typically remains stable and does not progress to CLL whereas HC-MBL is slightly related to CLL at Rai stage 0 (144,147) and has a rate of an annual risk progression to CLL requiring therapy around 1-2% (144). Similarly to CLL, MBL risk factors are family history of CLL, CD38 positivity, certain genetic polymorphisms, unmutated *IGHV* gene, del17p, advanced age, infections and elevated B-cell count (143).

[7] Graft-versus-host disease (GVHD) occurs after allogeneic HSCT and is a reaction of donor immune T-cells against host tissues. About 35%–50% of HSCT recipients will develop acute GVHD (276).

2.1.2. Other leukemias

Although CLL has been explained in detail due to its importance for this thesis, the other kinds of leukemia are shortly reviewed below.

I. Chronic myeloid leukemia (CML)

CML is a leukemia subtype in which there is an abnormal proliferation of myeloid leucocytes in the bone marrow and posterior accumulation in blood (148). It has an incidence of 1-2/100000 adults per year in Europe accounting for 15% of all newly diagnosed leukemias and is considered as an adult leukemia (median age of onset at 60 years) with an increasing frequency with age (incidence of 1/10000 at ≥80 years) (88,148). CML has a biphasic course consisting on the chronic phase, where patients are usually diagnosed, that eventually progress to a terminal and acute leukemia-like phase (also known as blast crisis) that sometimes is preceded by an accelerated phase. CML is characterized by a balanced translocation, t(9;22)(q34;q11.2), known as Philadelphia chromosome in which there is a fusion between the *Abelson* (*ABL1*) gene (9q34) and the *breakpoint cluster region (BCR)* gene (22q11.2) generating the BCR-ABL1 fusion oncogene, a constitutively active tyrosine kinase (TK) that promotes cells' growth and replication (148,149). Thus, the first-line treatment for CML are TK inhibitors whereas allogeneic HSCT is an important option for patients with chronic CML in who at least two TK inhibitors have failed and for all patients in advanced phase disease (148).

II. Acute lymphocytic leukemia (ALL)

ALL is a disseminated malignancy of B- or T-lymphoblast (pre-cursors of B/T-lymphocytes). The reason why ALL is referred as acute lymphocytic leukemia/lymphoma by the WHO classification (**table 3**) is that that ALL have two presentations, the leukemic one that shows a diffuse pattern affecting peripheral blood and bone marrow, and the lymphoma variant which is confined to nodal or extranodal sites without affecting bone marrow. B-ALL is typically presented as pure leukemia whereas T-ALL tend to present lymphomatous masses (150). ALL has an incidence of 2-4/100000 adults per year in Europe (88). However, ALL patients are predominantly children since 60-80% of cases occur at age <20 years with the major peak in the 2-5 years range (88,151). The survival rate of childhood ALL is ~90% thanks to treatment improvements (152). Detecting somatic or germinal genetic rearrangements associated with ALL is a key point to define the best treatment and improve patients' prognosis. Good prognosis

aberrations are del12p or t(12p)/t(12;21)(p13;q22) in B-ALL and t(10;14)(q24;q11) in T-ALL. Intermediate risk would be associated to isolated trisomy 21 or 8, del6q and some translocations as t(1;19)(q23;p13)/*E2A-PBX1.* Patients with t(9;22)(q34;q11)/*BCR-ABL1*, t(4;11)(q21;q23), MLL rearrangements at 11q23, monosomy 7 or hypodiploidy would have the poorest risk with an overall disease-free survival rate of ~25% (151)(150). Treatment of ALL typically lasts 2–2.5 years consisting in 3 phases: remission-induction, intensification (or consolidation), and continuation (or maintenance). Chemotherapy agents' dosage has been optimized on the basis of leukemic cells features, response to therapy and patient pharmacodynamics and pharmacogenomics findings leading to the current high survival rate. CNS-directed therapy consisting on prophylactic cranial irradiation (12–18 Gy) effectively controls disease but its use is limited due to the risk of secondary effects. Allogeneic HSCT is considered for patients at very high risk or persistent disease (153).

III. Acute myeloid leukemia (AML)

AML is an heterogeneous malignant disorder characterized by immature myeloid leucocytes proliferation and bone marrow failure (154). Its incidence is 2-4/100000 adults per year in Europe and is the most common acute leukemia in adults with a median age of onset at 64 years of age. Around 18000 new patients are yearly diagnosed in Europe representing 30% of all leukemia cases in adults and 0.6% of all cancers (88). Based on their cytogenetic profile, AML patients can be stratified into favorable prognosis group (certain translocations, the inversion (inv) inv(16)(p13.1q22) and mutations in *CEBPA* and *NPM1* genes with unmutated *FLT3-ITD*), intermediate-risk (mutations both in *NPM1* and *FLT3-ITD* or unmutated *NPM1* with/without mutated *FLT3-ITD* among others) and adverse-risk group (complex karyotype and certain translocations and inversions). Despite advances in supportive care, AML treatment remains unchanged for the past three decades and it is based on the use of chemotherapy as a first-line approach whereas allogenic HSCT is considered for patients with intermediate or high risk disease. Since elderly difficulties toleration of treatments, age is also considered an adverse prognosis factor itself (154,155).

3. FANCONI ANEMIA

Hematologic malignancies can appear as a unique clinical manifestation or as a part of a syndrome, some examples are Ataxia-telangiectasia (156), Neurofibromatosis (157), Down syndrome (158), Bloom syndrome (159) and Fanconi anemia (FA) deeply reviewed in this section.

The Swiss pediatrician Guido Fanconi was who firstly reported what today we know as Fanconi anemia in three siblings in 1927 (160,161).

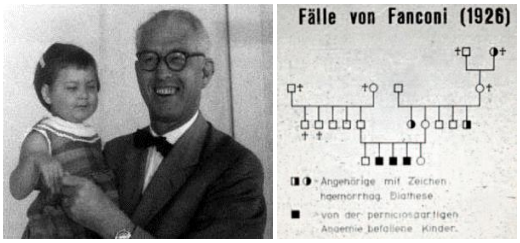


Figure 16. Guido Fanconi and the first FA family described. Half black symbol means "family member with bleeding diathesis" whereas full black symbol represents "children with the condition resembling pernicious anemia". Pictures from (161,162).

3.1. Clinical features and epidemiology

FA is a rare disease with a prevalence of 1-9/1000000 (163) and a ratio of males to females of 1.2:1. This is the most common genetic cause of aplastic anemia[8] and one of the main genetic causes of hematologic malignancy (164). FA clinical manifestations can be broadly defined as a triad of (I) congenital defects and physical malformations, (II) risk of bone marrow failure and (III) increased probability of both hematologic and solid tumors.

3.1.1. Congenital defects and physical malformations

Physical abnormalities are present in around 75% of patients, so since up to 25% of FA individuals do not have any apparent affection, an important fraction of patients remain undiagnosed until hematologic problems appear during the first decade of life or even in the adulthood (165). However, the median age of FA diagnosis is around 6.5 years (166). FA phenotype involves a wide range of systems and organs as nervous, gastrointestinal and reproductive systems, skeleton, skin, organs of sense and some viscera as heart and kidney (**table 4**) (164).

[8] Aplastic anemia or pancytopenia is a deficiency of all kind of blood cells due to the damage of bone marrow and hematopoietic stem cells that reside there (169).

Features of clinical FA phenotype	
Developmental delay and/or intellectual disability/learning problems	10%
Prenatal and/or postnatal short stature; low birth weight	60%
Unilateral/bilateral skeletal malformations of upper limbs	35%
• Thumbs	35%
• Radii	7%
• Hands	5%
• Ulnae	1%
Unilateral/bilateral skeletal malformations of lower limbs	5%
Other skeletal malformations	
• Spine	2%
• Neck	1%
Microcephaly	20%
Facial features: triangular, micrognathia, mid-face hypoplasia	2%
Ophthalmic affection (microphtalmia and visual difficulties)	20%
Hearing loss and abnormal ear shape	10%
Abnormal skin pigmentation (hype/hypopigmentation, café-au-lait spots)	40%
Visceral malformations:	
• Renal	20%
• Male gonads and reduced or absent fertility	25%
• Female gonads and reduced fertility	2%
• Heart	6%
• Gastrointestinal system	5%
• Central nervous system	3%
Endocrine disorders (i.e. reduced fertility or infertility, metabolic disorders)	79%

Table 4. Congenital defects and physical malformations in FA patients. Approximated prevalence of each feature is shown. Data from (164,167).

Although these broad variety of features associated to FA, the most pathognomonic ones would be prenatal and/or postnatal short stature, abnormal skin pigmentation as café-au-lait macules and hypopigmentation, skeletal malformations as hypoplastic thumb and hypoplastic radii, microcephaly and ophthalmic anomalies and genitourinary tract anomalies (**figure 17**) (164).

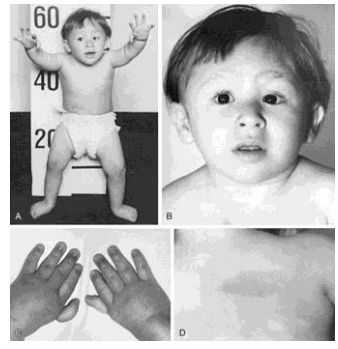


Figure 17. FA physical features. A 3 years old boy with FA exhibits several classic phenotype features including (**a**) short stature and dislocated hips, (**b**) characteristic facial features (broad nasal base, micrognathia, epicanthal folds) and microcephaly, (**c**) thumbs malformation and (**d**) café-au-lait spots with hypopigmented areas. Picture from (168).

3.1.2. Bone marrow failure

Bone marrow failure (BMF) is defined as a decreased production of effective blood cells leading to low blood counts. It can be classified into mild, moderate or severe depending on the degree of cytopenia (deficiency of any blood type) (169). BMF has an average age of onset around 7.6 years, it will affect 75% of FA patients during the first decade of life and its cumulative risk before age of 40 is around 90% (169,170). BMF can remain stable for years but also rapidly evolve to anemia, thrombocytopenia, neutropenia (low blood neutrophil count), aplastic anemia or pancytopenia and ultimately myelodysplastic syndrome (MDS)[9] or acute myeloid leukemia (AML) which are pre-cancerous and cancerous condition respectively of the blood-forming cells in the bone marrow (164,169).

3.1.3. Increased risk of cancer

FA patients have an increased risk of both hematological, secondary to BMF, and solid tumors. Regarding hematological malignancies, FA patients are at high risk of mainly MDS and AML with an cumulative incidence of AML around 15-20% at age 40, and of MDS about 40% by age 50 (171). Overall, FA individuals have a 500-fold increased relative risk of AML (171–173). ALL and lymphomas are generally rare in FA patients and would be restricted to a certain subgroup (FANCD1/BRCA2) (174,175). Regarding to solid tumors, these are the first FA manifestation in those patients without neither birth defect nor BMF. The cumulative incidence of developing any nonhematologic neoplasm is 28% by 40 years of age (170). Most common solid tumors are Head and Neck Squamous Cell Carcinoma (HNSCC) and gynecological cancers (mainly vulvar and cervical), whereas the less frequent ones but with higher incidence than in general population are gastrointestinal cancers (in esophagus and in liver), breast, kidney and brain cancers (169). HNSCC encompasses a wide range of tumors that usually start in the squamous cells that line the moist, mucosal surfaces of the oral/nasal cavity, pharynx and larynx. FA patients have a 500-700 fold increased risk of HNSCC compared to general population (164,176) and around 14% of FA individuals who survive 40 years of age will be diagnosed of HNSCC (176). FA patients tend to be diagnosed of HNSCC earlier (20-40 years) and in advanced stages compared to general population (50-60 years). FA is also characterized by higher proportion of tumors in the oral

[9] MDS is a group of bone marrow disorders characterized by abnormal production of immature and defective blood cells. It is considered a "pre-cancerous" condition (169).

cavity and also by less efficiency of HNSCC treatment. FA women have also a several thousand-fold higher risk for vulvar cancer and a 100-fold increased risk of cervical cancer compared with general population, with an age of onset at 25 and 27 years respectively (173). FA patients, mainly subgroup with *BRCA* mutations, would also have increased risk of breast cancer although few cases have been reported (177,178). Second primary cancers'[10] incidence is higher in FA (60%) compared to general population (30%) and they are typically located in the skin and genitourinary tract and lung and esophagus respectively (169).

3.2. Genetics of FA

FA is characterized by being genetically heterogenic since a total amount of 18 genes have been directly associated with FA during the last 25 years. However, following the criteria of at least two patients with the typical FA phenotype (including the triad malformations, BMF and cancer risk), a total of 15 genes are actually classified as bona fide FA genes whereas the 3 remaining genes are considered as FA-like genes since they cause a FA-like syndrome but without BMF (164,169,179,180). *FANCM* was initially considered as a FA gene but it has been excluded from the list in last publications since only one patient has been reported with biallelic mutations in *FANCM* gene until now (181). Moreover, this individual was shown to also carry biallelic mutations in *FANCA* gene which were also present in a sibling and expression of *FANCM* failed to complement the genetic defect of this patient (182). Finally, the detection of loss of function *FANCM* variants in two healthy Finish individuals without hematologic alterations would go against to classify *FANCM* gene as a bona fide FA gene (183). Thus, *FANCM* gene has been included in the category of FA-associated genes (consisting of 13 genes) together with several Fanconi anemia associated proteins (FAAP) and other FA-core complex interacting proteins (deeply described below) which when mutated would mimic FA phenotype but no biallelic mutations in these genes have still been reported in FA patients (179,180). Similarly, biallelic mutations in *FANCU/XRCC2* (184) and in *FANCV/MAD2L2* (185) genes have been described only in one individual without no more evidences of its causative role in FA disease leading to their classification as FA-associated genes (**table 5**).

[10] Second primary cancer refers to the presence of an additional, unrelated cancer in someone who was previously diagnosed with another type of cancer (169).

Bona fide FA genes	
Gene	**Frequency**
FANCA	60-70%
FANCB/FAAP95	2%
FANCC	14%
FANCD1/BRCA2	3%
FANCD2	3%
FANCE	3%
FANCF	2%
FANCG/XRCC9	10%
FANCI/KIAA1794	1%
FANCJ/BRIP1	2%
FANCL/PHF9	<1%
FANCN/PALB2	<1%
FANCP/SLX4	<1%
FANCQ/XRF/ERCC4	<1%
FANCT/UBE2T	<1%

FA-like genes	
Gene	**Frequency**
FANCO/RAD51C	<1%
FANCR/RAD51	<1%
FANCS/BRCA1	<1%

FA-associated genes	
Gene	**Frequency**
FANCM/FAAP250	<1%
FANCU/XRCC2	<1%
FANCV/MAD2L2	<1%
FAAP10/MHF2	<1%
FAAP16/MHF1	<1%
FAAP20/C1ORF86	<1%
FAAP24/C19ORF40	<1%
FAAP100/C17ORF70	<1%
BOD1L	<1%
UHRF1/NP95/ ICBP90	<1%
USP1	<1%
UAF1/P80/KIAA1449 /WDR48	<1%
FAN1/ KIAA1018	<1%

Table 5. Bona fide FA, FA-like and FA-associated genes. Frequencies of FA patients with mutations in each gene are shown. Table adapted from (164,169,179,180).

FA patients are classified in different complementation groups depending on which gene is truncated, even without knowing the specific mutation/s (169). FA mutations are inherited in an autosomic recessive manner except of those in *FANCB* gene which are X-linked (186) and those in *FANCR/RAD51* gene which are autosomic dominant (187). Thus, for recessive mutations, the frequency of carriers is 1/300 (169) being lower in North America (1/181) and Israel (1/93) (188) and in some populations with founder mutations as Ashkenazi Jewish, northern Europeans, Afrikaners, black South African and Spanish Gypsy (<1/100) (164,169).

Very few strict genotype-phenotype correlations have been established in FA. However, generally talking, null variants, which result in the complete loss of gene's function, are considered to lead more severe phenotype than hypomorphic variants, which partially reduce gene's function (164); but some exceptions to this have been reported in *FANCA* mutations (189). **Table 6** shows a summary of the main genotype-phenotype correlations including genes with reasonably well-documented associations.

Gene	Congenital Anomalies	BMF	AML	Brain/Wilms' tumors
FANCA null	+	+	+	
FANCA hypomorphic	-	Later	Later	
FANCB/FAAP95	+	+		
FANCC (IVS5+4A>T or IVS4)	+	+	+	
FANCC (c67delG)	-	Later	Later	
FANCE	+			
FANCF	+	+		
FANCG /XRCC9		+	+	
FANCL /PHF9	+	+		
FANCD2	+	+		
FANCD1/BRCA2	++		++	++
FANCN /PALB2	++		++	+

Table 6. Genotype-phenotype correlations in FA. The prevalence of each feature in each group is expressed as (-) for those manifestations at lower rate compared to what is generally observed in FA patients, (+) for those with a relatively increased frequency compared to other genotypes and (++) for those with a very high frequency compared to other groups. Empty fields express that prevalence of this feature does not differ from the observed in other groups. Later onset of certain diseases is indicated in some groups as "later". BMF: bone marrow failure; AML: acute myeloid leukemia. Table adapted from (169).

Generally, there is no increased risk of disease among FA carriers but there are some remarkable exceptions. Carriers of monoallelic mutations in *FANCD1* (190), *J* (191,192), *M* (193,194), *N* (195), *O* (196) and *S* (197) genes have an increased risk of breast and ovarian cancer. Heterozygous mutations in *FANCO* are also associated with HNSCC (198,199) and in *FANCN,* with pancreatic cancer (200). Finally, biallelic pathogenic mutations in *FANCQ* gene are associated with Xeroderma Pigmentosum (201), Cockayne syndrome (201) and XFE progeroid syndrome (202).

3.3. FA-BRCA pathway

FA genes codify for proteins that are involved in the FA/BRCA pathway **(Figure 18)** which have the main function of repairing interstrand DNA crosslinkings (ICLs). ICLs are DNA lesions that inhibit essential cellular processes such as DNA replication and transcription, so they must be repaired or bypassed for cell survival (180). Otherwise, unsolved ICLs can lead to tumors extremely sensitive to ICL-generating chemotherapies. ICLs can be caused by either exogenous factors as chemotherapy (203) or endogenous factors resulting from metabolism of alcohol, cigarette smoke and dietary fat

as acetaldehyde and malondialdehyde (204). In FA patients, ICLs cannot be repaired properly due to the defective function of one of the FA proteins generating a generalized chromosomal instability and the accumulation of DNA breakings; in fact, this is a diagnostic hallmark of FA (explained in detail in *Diagnosis* section). Moreover, since chemotherapy and radiotherapy are ICLs-inducing agents, FA tumors are treated with surgery (205–207).

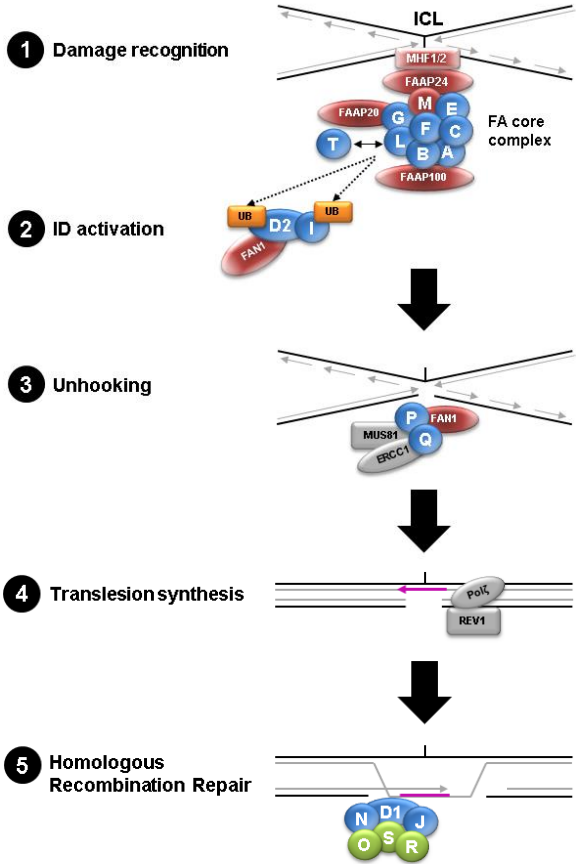


Figure 18. FA-BRCA pathway. (**1**) FANCM-FAAP24-MHF1/2 complex recognizes an ICL in the replication fork and recruits the FA core complex. (**2**) FANCL and FANCT activate the ID (I-D2) complex through monoubiquitylation. (**3**) This will lead to ICL cleavage and unhooking by FANCP, FANCQ, FAN1 and other proteins. (**4**) Translesion polymerases such as REV1 and Polζ will extend the leading DNA strand above and past the unhooked ICL to produce a substrate processed by a (**5**) successive Homologous Recombination upon the action of downstream FA proteins. Proteins in blue are coded by bona fide FA genes while products of FA-like genes are shown in green and those from FA-associated genes in red. Proteins not included in the "FA genes" classification (**table 5**) are in grey (179,180).

3.4. Diagnosis

3.4.1. Chromosome fragility test

FA disease can be suspected by detecting its characteristic physical congenital abnormalities, at least in ~75% of patients. Despite of this, FA diagnosis requires several testing procedures explained below. The first-line diagnosis approach for FA is the chromosome fragility test which consists on cytogenetic detection of lymphocytes hypersensitivity to DNA ICLs agents as diepoxybutane (DEB) or mitomycin C (MMC) (208). Normal cells can repair most of ICLs whereas FA patients' cells are unable to repair them and an accumulation of chromosomal breaks and complex rearrangements such as radial figures are accumulated (169) (**figure 19**).

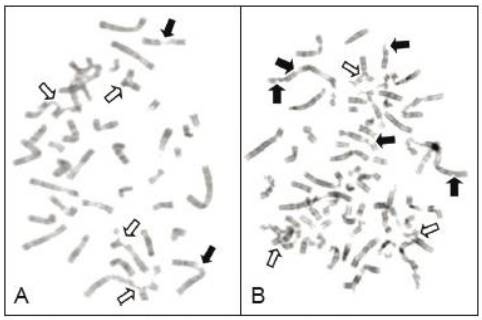


Figure 19. Chromosome fragility. This picture shows two lymphocytes' karyotype after DEB treatment from two different FA patients (A and B). Black arrows indicate chromosome breakages and white arrows, chromosomal rearrangements. Picture from (209).

The types and rates of breakages and rearrangements should be quantified together with the distribution of breakages among cells or the average number of rearrangements per cell with and without radial figures. The test is considered positive when >90% of lymphocytes in the culture show an increased breakage rate. If the test result is considered negative but the FA clinical evidence is strong, skin fibroblast testing is recommended since 20-25% of FA patients undergo somatic mosaicism of FA mutations. Some patients harbor a cell population with biallelic FA mutations (sensitive to ICL-inducing agents) together with a healthy cell population (less sensitive to ICL-inducing agents) leading to an overall percentage of normal blood cells ranging from 50 to 100% (see *Mosaic reversion* section). Chromosome fragility test can be considered as equivocal or inconclusive when percentage of cells with breakages is highly altered but not enough to be considered positive or the breakage pattern is different form the typically seen in FA. Here, blood mosaicism or other chromosome instability syndrome such as Bloom syndrome (159), ataxia-telangiectasia (156), Nijmegen breakage syndrome (210) and Seckel syndrome (211) and should be considered (169).

3.4.2. Cell cycle analysis

Lymphocytes, after ICL agents' treatment, stop at G2 cell cycle phase to repair DNA damage before entering to mitosis. FA lymphocytes show ≥40% lymphocytes retained in G2 compared to non-FA cells since they have more damage to repair. Therefore, some laboratories measure cell cycle kinetics instead of or in conjunction with chromosome fragility test in peripheral blood lymphocytes after DEB or MMC treatment to diagnose FA (212).

3.4.3. Complementation viral analysis and FANCD2 Western blot

Once chromosomal fragility and/or cell-cycle arrest have been proved, determining which FA gene is truncated can be addressed by infecting the unknown FA cells (blood, skin or bone marrow cells) with retroviral vectors expressing WT FA proteins. The FA protein able to rescue the chromosome instability will be indirectly determining which FA gene is truncated in that patient (213). Since not all FA proteins can be tested by this approach, Western blot to detect unmodified FANCD2 or monoubiquitinated FANCD2 protein can be performed to determine whether the truncated FA gene is up- or downstream FANCD2 modification respectively (214).

3.4.4. Mutational analysis

After a positive chromosome fragility tests together or not with additional test above explained, identifying the specific genetic mutation/s behind FA is important for: (I) establishing an appropriate medical management and follow-up according to genotype-phenotype correlations; (II) ensuring that any potential bone marrow donor does not have the same FA mutations; (III) determining patients' prospects to participate in future pharmacologic or gene therapies; (IV) identifying undiagnosed affected siblings and carriers at risk of having affected offspring or even cancer; and (V) performing genetic counselling leading to family planning by prenatal diagnosis (PD) and preimplantacional genetic diagnosis (PGD) to ensure unaffected descendants (169). FA genetic and allelic heterogeneity makes the mutational analysis extremely difficult and time consuming by traditional techniques as Sanger sequencing gene by gene (179). NGS strategies including WES or FA genes targeted sequencing to detect sequence changes (215,216) in combination with array technology (aCGH and aSNP) (217) and targeted Multiplex Ligation-dependent Probe Amplification (MLPA) (218) to detect deletions are the currently strategy to find both known and new pathogenic mutations in FA patients.

3.5. Medical follow-up

The increased knowledge about FA molecular mechanisms together with its ever increasing early detection and therapies improvements have established the survival beyond 18 years of age around 80% of FA patients (168) and an overall median survival greater than 30 years (171). So, there is an every-increasing FA adult population that needs a well established adult medical care system (169). Since FA is a highly clinical heterogenic disease, FA patients need a multi-disciplinary care and follow-up across several medical and surgical specialties. The Spanish guideline *Guía Básica para el Diagnóstico y Seguimiento de Pacientes con Anemia de Fanconi (2012)* from the *Red Nacional para la Anemia de Fanconi* establishes how FA adult patients' follow-up should be orchestrated (205). Depending on clinical manifestations, FA patients will require to be controlled by more or less specialties as traumatology for skeleton malformations, dermatology to control skin macules, endocrinology to monitor growth and endocrine dysfunctions, etc. However, specialists that should be attended by all FA patients are explained in detail blow.

3.5.1. Hematologic follow-up

The risk of any hematologic problem before the age of 25 is around 90% and for MDS/AML is around 20% (205). The average age of onset of BMF is 7-8 years whereas 14 years for AML (169,170,205). Considering this, Spanish guidelines recommends a Complete Blood Count (CBC) in peripheral blood every 3-4 months and a bone marrow aspirate (BMA) every 1-1.5 years to early detect any suspicion of MDS or AML from the time of FA diagnosis (205). BMA have to be studied cytogenetically by G-banding[11] analysis and FISH assay to detect and characterize chromosome rearrangement in bone marrow cells. The detection of clonal abnormalities herald the emergence of cancer or precancerous condition since both MDS and AML are associated to them and their clonal expansion (169). Gains in 1q and 3q and the whole loss of chromosome 7 represent 75% of the clonal abnormalities found in FA patients (219). Chromosome 3q and 7 rearrangements are associated with higher progression to MDS/AML and poorer prognosis whereas 1q gains alone would not seem to predict MDS/AML development (220–223). The proliferation and differentiation pattern of hematopoietic progenitors should be assessed

[11] G or Giemsa-banding is a cytogenetics technique which marks the chromosomes with colored bands and is used to elicit the unique and characteristic staining pattern of each chromosome (169).

in bone marrow by studying their ability to form colonies in a semisolid medium (Colony Forming Cell (CFC)) (224). Morphology of bone marrow cells may suggest early transformation of MDS into AML but FA bone marrow exhibits basal signs of dysplasia that should be differentiated from true forms of MDS/AML (169). CD34+ analysis should also be carried out since its expression is associated with poorer prognosis MDS/AML patients (225) Finally, bone marrow analysis should include a myelogram which is an x-ray examination of the spinal canal after administration of a contrast dye to detect abnormalities affecting the spinal cord such as tumors (226). In cases of both CBC persistently altered without apparent cause and BMA altered, blood counts and BMA are recommended to be performed every 1-2 months and every 2-3 months respectively (169).

3.5.2. Head and neck follow-up

FA patients have a 500-700 fold increased risk of HNSCC than general population (164,176). The accumulated risk at the age of 40 is around 25% and the average age of onset is 31 years without any reported case before age of 10. Importantly, HNSCC risk increases 4.4-fold after the fifth year from HSCT probably due to the GvHD (227,228). Periodic head, neck and buccal vigilance should be started at the age of 10 in non-transplanted patients and after the HSCT independently of patient's age. HNSCC prevention consists of an accurate examination of the buccal cavity, nasopharynx, oropharynx, hypopharynx and larynx every 6 months. In case of malignant or suspicious findings taking a biopsy and doing follow-up every 2-3 months is recommended. After successfully treated carcinoma, an annual radiography should be done additionally every year. Dental revisions twice per year since 1-1.5 years of age are also recommended to do. A direct contact between dentist and oral specialists is required to coordinate patient head-neck-oral follow-up (169,205).

3.5.3. Gynecologic follow-up

FA girls have a 200-400 fold increased risk of cervical, vulvar and vaginal cancer and the age of onset is at 25 years on average. A potential increased risk of breast cancer is also suspected. Spanish guidelines recommend beginning gynecological follow-up at ages of 16 or with the first menstruation and a breast follow-up from age of 20-25 or the first menstruation. Gynecological follow-up should include

Papanicolaou (Pap) test[12], an accurate genital physical examination, the Human Papiloma Virus (HPV)[13] test and an endocrine follow-up yearly (205). Lesions detected should be biopsied and, in case of dysplasia (precancerous condition), medical controls including periodic biopsies should be done every 4-6 months (169). Clinical breast examinations and mammography alternating with magnetic resonance imaging (MRI) should be performed twice a year. However, there should be a balanced risk-benefit on doing mammography screening considering FA radiation sensitivity (169,229).

3.6. Management and prevention

Similarly to medical follow-up, FA management can require a wide range of medical actions including surgery for physical and visceral malformations, cancer treatment and actions to address sensorial dysfunctions, growth and height retardation and endocrine disorders.

3.6.1. Hematologic management

HSCT is currently the only curative treatment for FA although it does not cure the non-hematologic problems. Ideally, it should be from a healthy relative with identical HLA to increase probability of success and reduce morbi-mortality. In most cases the donor is a healthy relative but without identical HLA or an unrelated person reducing survival probability, in part, because of the increased risk of GvHD (169), which also will contribute to a 4.4-fold increased risk of HNSCC (228). However, outcome of transplanting patients with HLA-compatible unrelated donors is acquiring survival rates more and more comparable to transplantations with HLA-matched siblings thanks to fludarabine administration and protocols improvement (169). PGD to ensure an unaffected descendant with HLA compatibility to the affected sibling would also be a good strategy to increase HSCT success (230). Independently of the donor, HSCT is a complex process due to the pre-transplant conditioning and the difficulty of reverting leukemia once is developed (162,206). Androgens, cytokines, blood product transfusion and investigational trials as gene therapy (231) are alternative to HSCT.

[12] Pap test or cervical cytology detects precancerous and cancerous lesions by studying cells scraped from the opening of the cervix under a microscope (169).

[13] HPV is a sexual transmitted infection that can cause genital warts as well as cervical cancer and other types of malignancies (further explained in *Gynecological cancer management and prevention* section) (169).

3.6.2. HNSCC management and prevention

In general population, HNSCC are treated with surgery or radiation when they are detected at early stages whereas with surgery followed by radiation with or without chemotherapy or concomitant treatment with chemoradiation in advanced-stage tumors (169). In FA patients HNSCC treatment is hampered by several factors as (I) FA HNSCC tend to be more aggressive and in advanced stages than general population; (II) healthy FA cells are highly sensitive to DNA ICLs-inducing agents as radiotherapy and chemotherapy; and (III) HNSCC FA cells are not sensitive as non-cancerous cells to DNA-crosslinking agents and do not respond to sub-therapeutic doses of radiation. Thus, surgery is the unique option for HNSCC in FA patients. Since more than 90% of HNSCC cases in general population are related to tobacco (232), alcohol (233) and HPV sexually transmitted (234,235), FA patients are strongly encouraged to avoid all these risk factors together with maintaining proper oral hygiene and sexual conduct.

3.6.3. Gynecological cancer management and prevention

Similarly to HNSCC, optimal genital warts and dysplasia treatment is surgical excision or ablation whereas vulvar lesions may be treated with immune modulating drugs (169). All FA female patients are encouraged to undergo vaccination against HPV since its role in genital warts and cervical cancer is proven although its involvement in genital tract cancer is more controversial. Gardasil® and Cervarix® are two HPV vaccines approved for being used between the ages of 9-26. Gardasil® is effective against HPV 6 and 11, virus subtypes responsible of 90% of cases of genital warts, and HPV 16 and 18 which are implicated in about 70% cervical cancers. Cervarix® is also effective against HPV 16 and 18 but does not protect against genital warts (236,237). Vaccination should be done after age 9 (169). Although it remains unclear the long-term effectiveness, there are some signs that Gardasil® and Cervarix® remain effective for at least 5 and 6.4 years respectively (236). HPV contribution to HNSCC in male and female FA patients remains controversial since some studies support the causative relationship (238,239) whereas others disagree (169,240,241).

3.7. Mosaic reversion

Around 20-25% of FA patients undergo somatic mosaicism of FA mutations (169). This phenomenon was firstly described after detecting two coexisting populations of lymphocytes sensitive and insensitive to ICL-inducing agents respectively in the same individual (242). Structure and function of a FA defective gene can be restored by several processes that generate a subpopulation of cells heterozygous for FA mutation. Genic conversion or intragenic recombination of the mutated alleles would be two examples which involve homologous recombination and should be considered as a potential reversion mechanism in compound heterozygous patients. While gene conversion is a unidirectional and non-reciprocal transfer of genetic material from a “donor” sequence to a highly homologous “acceptor”, intragenic crossover involves the reciprocal transfer of genetic material between a donor and acceptor sequences (24). Back or reverse mutation occurring, by chance, during DNA replication would lead to the change of a pathogenic mutation into a WT sequence generating cell clones with a restored translation of the WT protein (243). Finally, second-site compensatory deletions or insertions could be another possibility for FA mutation reversions. Here, a spontaneous compensating mutation either up- or downstream of a pathogenic frameshift mutation would result in the restore of the reading frame (244). Individuals with mosaic reversion have been reported to be more hematological stable during, at least, an observation period of 3-6 years. This would suggest a proliferative advantage of the reverted cell lineages indicating that mosaic reversion in FA seems to be a “natural therapy” for hematologic function; however, it would not guarantee a general better clinical course (245).

4. GENETIC COUNSELLING

Genetic counseling is the process of helping people to understand and adapt to the medical, psychological and familial implications of a certain genetic condition. This process is performed by a genetic counsellor and includes (I) interpretation of family and medical histories to assess the chance of disease occurrence or recurrence both in patients and relatives; (II) education about disease inheritance, molecular testing, management, prevention, reproductive options, social and medical resources and actual state of related research; and finally, (III) counselling to promote informed choices and adaptation to the new situation (246).

4.1. Mosaicism

In genetic counselling sessions of apparently *de novo* diseases, recurrence risk associated to germ-line mosaicism is traditionally mentioned as a "remaining" risk of <1% (247). The underestimation or even the no consideration of germ-line mosaicism possibility can induce to significant errors and should be clearly considered when phenotypically healthy parents have a second affected child with a certain disease and when facing disorders with known and reported germ-line mosaicism incidences as OI II (16%) (248), DMD (11-12%) (53) or tuberous sclerosis (6%) (249). The real recurrence risk will be determined by the proportion of germ cells carrying the mutation, thus, studying germ-line mosaicism percentage is highly recommended to give a more accurate genetic counselling for each case. However, when the mother is the possible carrier, using empiric recurrence risks instead of measuring her mosaicism percentage is usually recommended since, in a risk-benefit balance, the accuracy gained in the genetic counselling will not be justified by the important disadvantages associated to the oocytes collection.

4.2. Hematologic malignancies

Hematologic malignancies, and particularly the leukemia subgroup, are a great and heterogenic group of diseases that include both sporadic and hereditary cases. Depending on the disorder, the proportion of hereditary cases can vary. For example, in CLL,

around 10% of patients present positive familiar history for the same or related diseases indicating a possible hereditary genetic basis (103). Thus, in a genetic counselling process of a hematologic malignancy as CLL, the genetic counsellor should (I) explain the frequency of hereditary versus sporadic cases, (II) try to determine whether that patient is on the first or second group based on family and personal histories and possible genetic studies performed, and finally, (III) base the genetic counselling on that. Contrary to most of patients believe, facing a probable familiar case does not mean that a unique mutation is responsible of the disease in affected individuals. CLL, as many other hematologic malignancies, are polygenic and associated to alleles with different penetrance; the genetic counsellor will have to deal with all these issues.

4.3. Fanconi anemia

When a hematologic malignancy appears as a part of a syndrome, the genetic counselling should be based on the genetic basis of that syndrome. An example of this would be leukemia as a clinical manifestation of Fanconi anemia.

4.3.1. Family and personal history

When a genetic counsellor is in front of both a possible or diagnosed FA case, a complete patient's family history has to be collected paying particular attention to any FA-related clinical features as well as cancer cases (especially leukemia, HNSCC and cancers of cervix, vulva, anus, breast, ovaries and prostate) deaths, miscarriages and infertility (169). Relatives' medical information provided by the patient or companions during the session must be supported by medical reports. Determining family's ethnicity is also a key point to stablish the most appropriated genetic testing strategy since some FA mutation are "founder" mutations at higher frequency in certain ethnicities. A complete medical history from the patient must also be collected.

4.3.2. Inheritance of FA

FA has predominantly an autosomal recessive inheritance so, parents of an affected child are obligate FA carriers and, in each pregnancy, will have 25% of risk of having another affected child whereas a 50% of risk of having a healthy but FA carrier descendant

and 25% of chance of a healthy non-carrier child. On this direction, healthy proband's sibs will have 66% of risk of being FA carriers whereas proband's offspring will be obligate FA carriers (164). It is important to say that FA men are usually azoospermic whereas FA women have a reduced fertility and reproductive lifespan. These facts together with the shorten life expectancy until the last years have actually led to very few pregnancies from FA patients (169). Finally, each sib of the proband's parents would be at 50% of risk of being a FA carrier. In case of being carriers, they would have 50% of having FA carrier children in case of a non-carrier partner (164).

Around 2% of FA cases are caused by mutations in *FANCB* gene and are inherited in an X-linked recessive manner. Here, affected patients will be males and their healthy fathers will not have the disorder nor will be hemizygous for *FANCB* mutation. Thus, FA males' mothers will be obligate carriers and will have a 50% of chance to transmit the pathogenic variant in each pregnancy. All female descendants will be healthy with 50% of risk of being FA carriers whereas male offspring will have 50% of chance of being affected or unaffected. Theoretically, male children from a FA male would be healthy non-carriers whereas all girls would be obligate FA carriers; however, as above-mentioned, FA males are infertile. Finally, proband's maternal aunts may be at risk of being *FANCB* carriers and aunts' offspring, depending on their gender, may be at risk of being carriers or affected (164).

Less than 1% of FA cases are caused by mutations in *FANCR/RAD51* gene which would follow an autosomic dominant inheritance. To date, reported probands with *RAD51*-related FA harbored *de novo* heterozygous mutations suggesting low recurrence risk for other family members. However, germ-line mosaicism in parents should be considered. Descendants of these patients would have 50% of being affected but only one *RAD51*-related FA have reached the adulthood but without reported offspring (164).

4.3.3. Genetic testing

As mentioned in a previous section, FA molecular diagnosis should end with the identification of mutation/s responsible of the disease and it can be addressed in different ways depending on the context. In case of knowing the mutation/s responsible of FA because of a previous relative affected by the disease, targeted mutation analysis can be performed in a suspect new patient or

carrier or in the context of a PD or PGD. In case of unknowing the mutation/s behind FA, single gene sequencing have been the strategy used traditionally (179). However, genetic and allele FA heterogeneity have caused the increasing trend of using panel testing or WES by NGS technology in combination with array technology (aCGH and aSNP) and MLPA to detect CNVs (215–218). However, single gene sequencing can be useful to test a partner of a fertile FA patient or FA carrier interested on PD or PGD.

A genetic testing result is considered positive when the pathogenic variant/s are found, negative when the known pathogenic/s variants in that family are not found in the individual analyzed or in-conclusive when one or both mutations have not been identified with the technique used. Additionally, with NGS and array technologies, variants of unknown significance (VOUS) can be detected and incidental findings regarding other pathologies as well. Determining FA mutations have some benefits (mentioned in a previous section) but also some risks and limitations that deserve to be discussed with patients and their relatives in a pre-test genetic counselling session (**table 7**) together with all possible genetic testing outcomes (169).

Benefits
• Establishing an appropriate medical management and follow-up according to genotype-phenotype correlations.
• Ensuring that any potential bone marrow donor does not have the same FA mutations.
• Determining patient's prospects to participate in future pharmacologic or gene therapies.
• Identifying undiagnosed affected siblings and carriers at risk of having affected offspring or breast and ovarian cancer (D1, J, M, N, O or S), HNSCC (FANCO) or pancreatic cancer (FANCN).
• Family planning by PD and PGD to ensure unaffected descendants.
• Relieving anxiety.
Risks
• Positive genetic testing results could harm patients at work and/or social level or in terms of health and life insurances.
• Revealing previously unknown family relationships (i.e. non-paternity).
• Altering family dynamics.
• Creating anxiety, distress and feelings of guilty due to the test results.
• In case of NGS and arrays, detecting VOUS and/or incident findings related to other diseases that would create ethical dilemmas.
Limitations
• Some mutations do not give information about management guidance.
• Genetic testing can be inconclusive by finding only one or no mutations.
• Finding mutations does not enable to predict future medical complications.

Table 7. Benefits, risks and limitations of FA genetic testing. Adapted from (169).

4.3.4. Psychosocial issues and follow-up coordination

A part form medical issues, FA can involve problems related to psychosocial aspects. Children with FA are in a potential risk of parental over-protectiveness and no fully participating in childhood activities at school, sport and leisure which can isolate them and cause a delay in their personal and social development (169). In a 25 year follow-up, more adverse general and mental health, functional impairment and activity limitations as well as lower rates of marriage, college graduation, employment and health insurance were reported among ALL survivors compared to healthy siblings (250) showing a clear need of psychosocial support and guidance in patients with chronic diseases since childhood.

In a 13 years study, mood and adjustment disorders, deliriums, aggressiveness and anxious/depressed and withdrawn symptoms were diagnosed in FA patients associated to FA-related aspects as chronic and genetic childhood condition, physical and/or neuro-developmental disability, cancer predisposition syndrome and exposure to therapeutic androgens and pediatric HSCT (251). Additionally, apparently healthy siblings of FA patients have been reported to feel containment, invisibility, worry and despair showing that unrecognized psychosocial issues also exist for them (252). FA parents also have shown to experience stress, uncertainty and active surveillance throughout the course of the illness (253). Thus, psychosocial care should be addressed not only for patients but also for their relatives. Solving disease-related doubts, offering psychological support, informing about socioeconomics issues and helping to establish contact with familiar support groups would be the main pillars of a complete psychosocial care of a FA family.

Although no specific transition programs from childhood medical care to adult follow-up are established in FA, there is the clear evidence that an anticipated and coordinated transition process benefites patients and their families thanks to the experience with well stablished transition programs as in cystic fibrosis (254–256). Transition of health care is particularly important to avoid overaccumulation of FA patients in pediatric services, lead FA patients be treated by adequate adult specialists and contribute to patients' adult independence in terms of assuming responsability of their healthcare (169). However, the moment of this transition should not be abrupt and should be personalized although it tends to be done in the late teenage years (254). Genetic counsellor would be a suitable figure to guide this medical care and follow-up transition as well as to orchestrate all coming multidisciplinary care.

OBJECTIVES

OBJECTIVES

The main goal of this thesis is to contribute to the knowledge of genetic clonal mosaicism implications on cancer as well as to deepen in mechanisms underlying this biological phenomenon. The generosity of around 200 families affected of Fanconi anemia in giving samples and access to clinical information together with the collaboration with other groups have been essential to conduct most of the aims of this work.

In order to address this goal, we stablished the following specific objectives:

1. To study the prevalence, evolution and late or early embryonic origin of detectable clonal mosaicism in blood and/or saliva samples of individuals affected by a chromosome instability syndrome such as Fanconi anemia as well as to evaluate clonal mosaicism detection as a possible biomarker of cancer risk in this population.

2. To deepen in uniparental disomy as a source of mosaicism and its putative dual role as a contributing or protecting factor in cancer development, focusing in three different hematologic malignancies or disorders as Fanconi anemia, chronic lymphocytic leukemia and polycythemia vera.

3. To study the knowledge, perceptions and medical follow-up adherence and, definitely, the role of the genetic counsellor in Fanconi anemia families as well as to show the necessity of less invasive follow-up strategies for cancer prevention.

CHAPTER 1

Detectable clonal mosaicism in Fanconi anemia, analysis of events' embryonic origin and their evolution along time

Judith Reina-Castillón, Roser Pujol, Marcos López-Sánchez, Benjamín Rodríguez-Santiago, Miriam Aza-Carmona, Juan Ramón González, José A. Casado, Juan A. Bueren, Julián Sevilla, Isabel Badel, Albert Català, Cristina Beléndez, María Ángeles Dasí, Cristina Díaz de Heredia, Jean Soulier, Detlev Schindler, Luis Alberto Pérez-Jurado and Jordi Surrallés.

In agreement with previous publications, results presented in this work support the idea of using detectable clonal mosaicism as an early marker of tumor development, particularly in a chromosome instability syndrome characterized by generally increased cancer risk since early childhood such as Fanconi anemia. The study of mosaicism evolution in serial samples of a reduced group of patients annexed in this chapter is the first stone of an indispensable study to finally prove the utility of mosaicism as a reliable tool for early cancer detection in the routine medical follow-up of patients affected by not only Fanconi anemia but also by other chromosome instability syndromes.

Part I

Reina-Castillón J, Pujol R, López-Sánchez M, Rodríguez-Santiago B, Aza-Carmona M, González JR, et al. Detectable clonal mosaicism in blood as a biomarker of cancer risk in Fanconi anemia. Blood Adv. 2017 Jan 24;1(5):319–29. DOI: 10.1182/bloodadvances.2016000943

Part II

Extension of the study of detectable clonal mosaicism in Fanconi anemia, analysis of events' embryonic origin and their evolution along time

Judith Reina-Castillón, Roser Pujol, Marcos López-Sánchez, Benjamín Rodríguez-Santiago, Miriam Aza-Carmona, Juan Ramón González, José A. Casado, Juan A. Bueren, Julián Sevilla, Isabel Badel, Albert Català, Cristina Beléndez, María Ángeles Dasí, Cristina Díaz de Heredia, Jean Soulier, Detlev Schindler, Luis Alberto Pérez-Jurado and Jordi Surrallés.

In preparation

ABSTRACT

Our previous data demonstrated an increased prevalence of detectable chromosomal mosaic events (CMEs) by Single-Nucleotide Polymorphism (SNP) array in blood of Fanconi anemia (FA) patients related to a higher cancer risk and poorer prognosis. We present a similar study but with a final sample size of 167 individuals. A total of 160 blood and 7 saliva FA samples were analyzed by SNP array at the end of both studies. We detected 70 CMEs (4.4-159Mb in size) in 26/167 patients (15.6%), 13 of them with multiple CMEs. Most frequent events were gains at 1q (n=8) and 3q (n=8), events at 6p (n=6) and UPD at 3p (n=5). Compared to 15743 age-matched controls, FA patients had 170-180 fold increased risk of harboring detectable CMEs in blood or saliva ($p<4.2x10^{-13}$). Prevalent and incident cancer were 5.6 times more common in CME carriers ($p=1.3x10^{-4}$) leading to a 3 fold increased exitus rate (p=0.014). The age-adjusted hazard risk (HR) of having cancer was almost 4 times higher in CMEs carriers than in those without CMEs. The longitudinal study along 5y follow-up period on average of 28 FA patients did not reveal any new mosaic case but showed some fluctuations of mosaicism percentage in CMEs previously detected. Saliva analysis of 6 FA patients with CMEs in blood revealed an early embryonic origin of these events. Finally, a possible new mosaic individual was detected after observing a mosaic 20q deletion in saliva obtained 6y after a blood sample free of CMEs. Therefore, our data goes in favor of the potential use of detectable CMEs as a biomarker of cancer risk in FA.

INTRODUCTION

Detectable chromosomal mosaic events (CMEs) have been strongly associated with aging and increased risk of cancer in general population, with special evidence in case of hematologic malignancies. These studies also have estimated that the risk of hematological cancer is 10 fold higher for mosaic than for non-mosaic individuals (1–3). The publication shown in the first part of this chapter (8) demonstrates that the cancer predisposition disorder Fanconi anemia (FA) (4,5) manifests a high rate of CMEs detectable by SNP array in blood, even at early age, and that CMEs could be indicative of a higher risk of cancer and shortened survival in the coming years.

We found that two thirds of FA patients with both CMEs in blood and hematologic complications harbored mosaic 1q gains, 3q gains and/or chromosome 7 monosomy; rearrangements typically assocated with hematologic malignancies (6). This data together with the fact that two mosaic rearrangements were found both in blood and solid tumor sample in a FA patient would be in agreement with the idea that CMEs can lead tumor development or arise as consequence of tumor clonal expansion (8).
In order to early detect hematologic problems, FA clinical guidelines recommend a strict follow-up of these patients by annual bone marrow sampling and its cytogenetic analysis to detect oligoclonal chromosomal rearrangements associated to preneoplastic or early neoplastic processes (7). However, these kinds of controls are really difficult to perform due to their invasiveness, especially when patients' hematology is stable, and the development of less invasive but equally efficient screening strategies would be very useful for both physicians and affected families.

Given the high prevalence of CMEs found in blood samples of FA patients and the documented association of detectable CMEs with hematologic and also, but slighter, with solid cancer, it is logical to propose that early detection of CMEs could contribute to clinical decision-making. Thus, periodic SNP array analysis of DNA from easy to obtain samples, as blood or saliva, could be incorporated in the regular FA follow-up to identify those patients at higher risk of cancer who should not avoid bone marrow aspirates and would need closer vigilance.

Since prospective data on sequential sampling of a large cohort would be required to obtain firmer conclusions and better

define the sensitivity and specificity of the method in the context of a disease follow-up as well as the most appropriate sampling intervals, here we provide an extension of the study previously published by us (8) performed in a larger FA sample set together with the evaluation of embryonic origin and evolution along time of some CMEs.

MATERIALS AND METHODS

***Subjects*.** We studied mosaicism prevalence in a new FA sample set (FA cohort II) consisting of a total of 94 samples (68 peripheral blood (PB) and 26 saliva samples) from 37 new FA patients and 21 FA patients belonging to the sample set already studied and published by Reina-Castillón et al. (8). Note that this 94 sample set included serial samples (from 28 individuals) and additional saliva samples (from 18 subjects). Genomic DNA was extracted from PB and saliva by using a standard phenol-chloroform extraction method or the Puregene DNA Isolation Kit (Qiagen Inc, US).

***Genotype data and mosaicism detection in FA cohort II*.** DNA from a total of 94 samples was analyzed by SNP array using Illumina Infinium HumanCore BeadChip (250K) platform. Genetic mosaic rearrangements were inferred from SNP array data by using Mosaic Alteration Detection (MAD) algorithm implemented in R Genomic Alteration Detection Analysis (R-GADA) software (9) which detects mosaic alterations larger than 2-0.5 Mb in length by using both B-allele frequency (BAF) and log relative probe intensity ratio (LRR). Visual inspection of SNP array plots let us to detect additional mosaic rearrangements undetected by MAD algorithm due to too low or high mosaicism level. The proportion of cells harboring each event was calculated as specified by Reina-Castillón et al. (8).

Statistical Analysis. To estimate CMEs frequency, we used data from the same age-matched controls datasets used in the first phase of the study (8). Statistical tests for 2x2 tables comparing CME prevalence for FA individuals and controls were based on Fisher exact test due to the small numbers. P-values (p), Odds Ratios (ORs) and Confidence Intervals at 95% were computed using Fisher test R function (10). Kaplan-Meier (KM) estimator (11) was used to study survival function from lifetime data as previously

described (8). A total amount of 161 individuals were included in the estimation of survival function when considering prevalent cancer cases (6 subjects were excluded due to incomplete medical record information) whereas 140 FA patients were included in the analysis when considering only incident cancer cases after sample collection (20 cases excluded due to having cancer at time of sampling and 1 case due to dying also at sampling). Age-adjusted cancer risk was computed by using Cox proportional hazard model.

RESULTS

High CMEs prevalence in FA related to increased cancer risk and poorer survival.

DNA from peripheral blood (n=30) or saliva (n=7) was analyzed from a group of 37 new FA patients by SNP array technique. The mean age of subjects was 15.08 years of age (y) (range: 0-48y) and, as expected, they were mainly classified in the FA-A complementation group (FA-A: 70.3%, FA-D2: 8.1%, FA-E: 2.7%, FA-G: 10.8% and unassigned group: 8.1%). CMEs were detected in 7 out of 37 (18.9%) individuals by using MAD software (9). We observed a total amount of 15 events in 7 individuals with 3 cases of 3-5 events per individual. Only 2 out of 15 events (13.3%) were interstitial and we detected 4 uniparental disomies (UPD), 5 copy-losses and 6 copy-gains; including a complete chromosome 7 monosomy, a 3q tetrasomy and two unbalanced translocations (**Figure S1, Table S1**).

Considering this new data and our previous study (n=130) (8), we obtained a final FA sample set of 167 FA patients. The mean age was 14.6y (range: 0-50y) and most individuals belonged to FA-A subgroup (66.5%) (FA-C: 0.6%, FA-D1: 3%, FA-D2: 10.2%, FA-E: 2.4%, FA-G: 5.4%; FA-J: 2.4%, FA-Q: 1.2%, unassigned group: 8.3%). With the analysis of peripheral blood DNA in 160 patients and saliva DNA in 7 participants, we detected CMEs in 26 out of 167 individuals (15.6%). When separating subjects in age ranges, we observed a mosaicism prevalence of 7% among those patients below 18y and of 34.6% among those between 19-50y. By considering mosaicism prevalence reported in age matched controls (<18y: 0.04%, 19-50y: 0.31%) (8), we detected that FA children and adolescents (<18y) have 181.9 times (OR=181.9, CI95%= 47.6-695.2, $p=4.2x10^{-13}$) higher

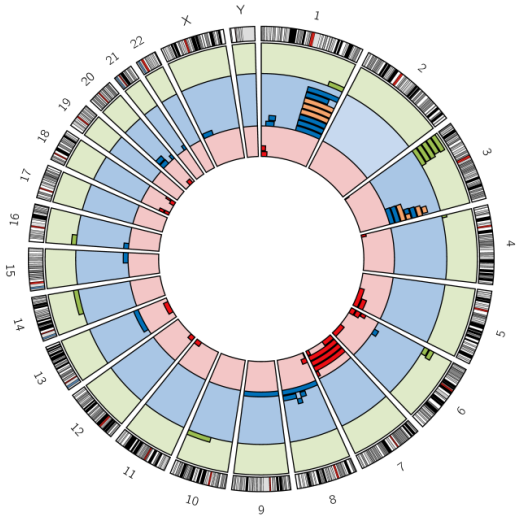


Figure 1. Genomic distribution of 70 CMEs detected in 26/167 FA patients in blood or saliva. The circular plot shows the chromosomal location of each of the 70 mosaic events detected (red bars: losses; blue bars: gains; orange bars: multi-copy gains; green bars: copy-neutral events or UPDs). Most frequent CMEs were at 1q (8 gains and 1 UPD), 3p (5 UPDs), 3q (8 gains), 6p (2 UPDs, 1 gain and 3 losses) and chromosome 7 (3 monosomies, one 7p loss and two 7q losses).

risk of harboring CMEs than age matched controls and 171.3 fold increased risk (OR=171.3 CI95%=86.0-341.2, $p=4.9x10^{-30}$) in case of adult patients (19-50y). Importantly, there were 3 individuals from the first study (FA110, FA664, and FA681) that, after revising MAD data, were reclassified as subjects "with CMEs". We detected a mosaic 1q (gain)-18p (loss) unbalanced translocation, an UPD3p and another UPD3p respectively (**Figure S1, Table S1**). Overall, we detected 70 CMEs in 26 FA patients (**Figure 1**) where 13 of them presented more than one CME (16 events in one individual was the maximum detected). We detected different kinds of rearrangements including segmental copy-neutral losses of heterozygosity due to UPDs (n=12), segmental losses (n= 23), segmental single (n=23) or multi-copy gains (n=7), entire chromosome monosomies (n=3) and trisomies (n=2). Most CMEs included the telomeres (53/70, 75.7%) whereas 7.1% (5/70) spanned centromeres and only 17.2% (12/70) were interstitial.

Contemplating both prevalent and incident cancer cases, clinical data from 167 FA patients **(table 1/S3 from (8); table 1/S2)** revealed a global 5.6 fold increased risk of cancer (OR=5.6, CI95%=2.3-13.6, $p= 1.3x10^{-4}$) among FA patients with CMEs (15/26, 57.7%) compared to those without CMEs (27/139, 19.4%, 2 no informative cases). As previously observed (8), we corroborated a 40 fold augmented risk of prevalent and incident cancer with the increasing amount of CMEs (cancer in subjects with multiple CMEs: 12/13; cancer in subjects with one CME: 3/13) (OR=40, CI95%=3.5-447.05, $p= 9.8x10^{-4}$). By using the Kaplan-Meier algorithm (11), we observed that FA patients with

Sample	Genetic group	CMEs (size in Mb)	Start (Mb)	End (Mb)	Cell fraction (%)	Age at DNA sampling	Age at BMF	Cancer diagnosis (age)	Years from DNA sampling to	Evolution (age)
FA255	D2	UPD 3p (44-49)[†]	pter	44-49	5.1	20	19	No	-	Alive (21)
FA29	Unknown	UPD 1q (47.0)	202.1	qter	19.8	20	No	No	-	Alive (28)
FA383	D2	UPD 3p (46.4)	pter	46.4	8.8	23	14	No	-	Alive (24)
FA573	G	Loss 13q (62.1)	32.0	94.1	29.9	24	23	No	-	Alive (26)
FA825 (BS)	Unknown	Gain 1q (102.5)	146.6	qter	70.2	27	No	Esophageal tumor (27)	0	Alive (28)
		Gain 3q (25.4)*	172.3	197.8	56.0					
		Loss 6p (15.1)	pter	15.1	42.0					
		UPD 14q (72.7)[†]	35.0	qter	76,8					
		Gain 15q (41.2)	61.1	qter	36.9					
FA842	A	Gain 1q (96.6)**	152.5	qter	70.6	18	18	MDS (18)	0	Alive (19)
		Loss 8p (24.2)**	pter	24.2	83.1					
		Gain 3q (61.1)	136.6	qter	72.4					
FA860	A	Loss 7 (159.0)[•]	pter	qter	28.6	10	5	MDS (10)	0	Alive (10)
		Loss 18p (13.4)**	pter	13.4	29.0					
		Gain 20p (29.5)**	pter	29.5	27.2					
FA110	A	Gain 1q (104.7)[†,**]	cen	qter	14,8	18	4	No	-	Alive (22)
		Loss 18q (15.1)[†,**]	pter	cen	13,6					
FA664	D2	UPD 3p (49.9)[†]	pter	49.9	9,6	15	No	No	-	Alive (20)
FA681	D2	UPD 3p (NA-49.8)[†]	pter	NA (49.8)	6,7	19	19	No	-	Alive (23)

Table 1. Features of the 7 new and 3 reclassified FA individuals with CMEs. This table is an extension of table 1 of the main publication of this chapter (8). FA110, FA664 and FA681 were individuals belonging to the first study's phase (8) and have been reclassified as individuals with CMEs after a reanalysis of MAD output.

pter: p arm terminal; qter: q arm terminal; cen: centromere; [†]: rearrangements not detected by MAD but by visual inspection; [•]: Monosomy; *: Gains of more than one extra copy (tetrasomy); **: rearrangements probably involved in an unbalanced translocation; "NA" (not available): refers to those coordinates that are imprecise due to low mosaicism percentage; Mb: megabases; BMF: bone marrow failure; MDS: myelodysplastic syndrome: -: unfilled fields due to the absence of a cancer diagnosis.

CMEs in blood or saliva had a global shortened cancer-free time after sampling compared to those patients without CMEs (HR=4.6, CI95%=2.4-8.8, p= $2.2x10^{-6}$). Given that the mean age was different between both groups of FA patients (CMEs: 21.8y; No CMEs: 13.3y), we performed an age-adjustment and we obtained that the age-adjusted hazard risk (HR) of having cancer was 3.7 times higher in individuals with CMEs than in individuals without CMEs (HR=3.7, CI95%=1.9-7.0, p= $9.0x10^{-5}$) (**figure 2A**). We observed the same results when repeating the analysis to consider only incident cancer cases by excluding FA patients with cancer (n=20) or exitus (n=1) at the time of sample collection (HR=4.7, CI95%=1.9-11.6, p= $8x10^{-4}$). In this case, the age-adjusted HR of having cancer was almost 4 times higher in CMEs individuals than subjects without CMEs (age-adjusted HR=3.9, CI95%=1.5-9.9, p= $4.5x10^{-3}$) (**figure 2B**). Focusing only in cancer appearing after sample collection (incident cases), we detected 18 individuals without cancer at the time of sample collection among 26 FA patients with CMEs. Seven of these 18 subjects developed cancer (4 hematologic tumors and 3 Squamous Cell Carcinoma (SCC)) during an average follow-up of 4.61y (0.3-10y) after sample collection **(table 1 from (8); table 1)**. Among the 141 FA patients without CMEs, we detected 122

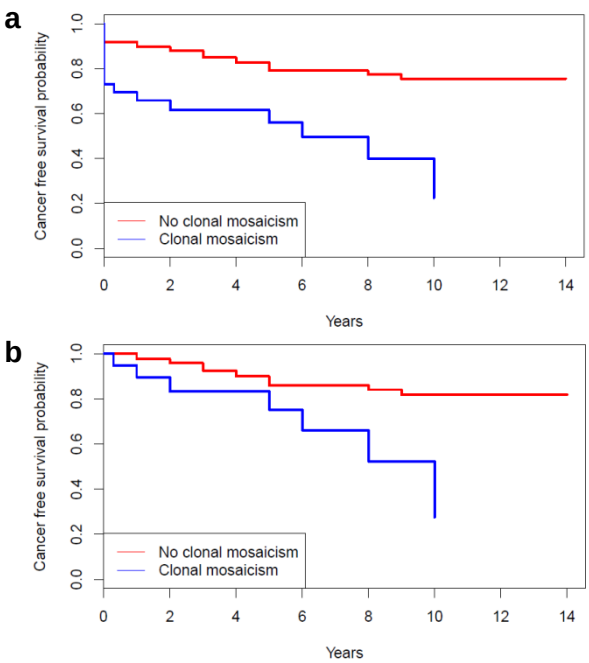


Figure 2. Kaplan-Meier (KM) plots show shortened cancer free time in CMEs carriers compared to non-carriers. a) KM performed considering all cancer cases (prevalent and incident cancers after sampling) and doing an age-adjustment (n= 161, 6 no informative cases, 42 events of cancer. HR=3.7, CI95%=1.9-7.0, p= $9.0x10^{-5}$) **b)** KM performed considering only incident cancer cases after sampling and doing an age-adjustment (n=140, 6 no informative cases and 21 exclusions, 22 events of cancer. HR=3.9, CI95%=1.5-9.9, p=$4.5x10^{-3}$).

(7 excluded cases for incomplete information) who did not have cancer at the time of sample collection and 15 of them developed a malignancy during an average follow-up of 3.73y (1-9y) (8 individuals with hematologic malignancies, 3 cases of SCC, 1 patient with both myelodysplastic syndrome (MDS) and SCC and 3 patients with other tumors) (**table S3 from (8); table S2**). Consequently, there was a statistically significant 4.5 fold increased risk of developing cancer after mosaicism detection (OR=4.5, CI95%=1.5-13.5, $p=9.3x10^{-3}$). Curiously, when considering hematologic and solid incident cancer cases separately, we observed a little bit higher risk of solid incident cancer after mosaicism detection (OR=5.9, CI95%=1.2-28.9, p=0.045) than hematologic malignancy (OR= 3.6, CI95%=0.9-13.2, p=0.065) but with poor statistical significance due to low sample size (CMEs carriers (n=18): 3 solid and 4 hematologic incident cancer cases; no-CMEs carriers (n=122): 4 solid and 9 hematologic incident cancer cases). Regarding survival, we detected a global increased exitus rate among CMEs carriers (12/26, 46.2%) compared to those without mosaicism (30/136, 22.1%, 5 no informative cases) showing and increased risk of death in the former group (OR=3.02; CI95%=1.26-7.23; p= 0.014). Since there was not dead subjects in FA cohort II, the proportion of cancer-related deaths was maintained higher in FA patients with CMEs group compared to no-CMEs group (OR=15.5, CI95%=1.7-137.7, $p= 4.7x10^{-3}$).

Reverse mosaicism protects from hematologic malignancy.
Reverse mosaicism consist on the reversion of a mutated allele into a wild-type form and is described in up to a 20-25% of FA patients (24). This would act as a "natural therapy" by leading to a reduction of chromosome fragility and hematologic problems (37,38). Reverse mosaicism can be identified when less than 50% of cultured cells have chromosomal aberrations after diepoxybutane treatment. With this, we detected that 35/167 (20.9%) FA patients had reverse mosaicism. Three of them carried one CME in blood (and saliva in one case) and one of these three individuals was additionally diagnosed of cancer (vulvar SCC). Importantly, only one of these 35 patients with reverse mosaicism had an hematologic malignancy (MDS).

Longitudinal study of 28 FA patients reveals no new mosaic cases but confirms CMEs previously detected.
Two to four peripheral blood samples serially obtained along a vigilance period of almost 5 years on average (0.42-15y) were studied by SNP array for

mosaicism detection in 26 out of 167 FA subjects. Additionally, two serial saliva samples from 2 out of 167 FA patients were both obtained during a 2 years follow-up period and were also studied to evaluate mosaicism evolution. A total of 24 out of 28 individuals serially studied were no-CMEs carriers and maintained this condition after around 5 years of follow-up. CMEs detected in 4 out of 28 FA patients evolutionary studied were maintained along the time with some slight variations regarding mosaicism level. Importantly, no cancer diagnosis before neither after DNA collection was reported among 28 individuals (**Figure 3, Table 2, Figure S2**).

CMEs detected in 6 cases had an early embryonic origin.
Saliva DNA samples were studied by SNP array for CMEs detection in 18 out of 167 FA patients. A total of 6 out of 18 FA individuals studied at both blood at saliva level were identified as CMEs carriers in blood in the previous transversal study. Interestingly, we detected that all CMEs present in these 6 subjects had an early embryonic origin since they were detected in both blood

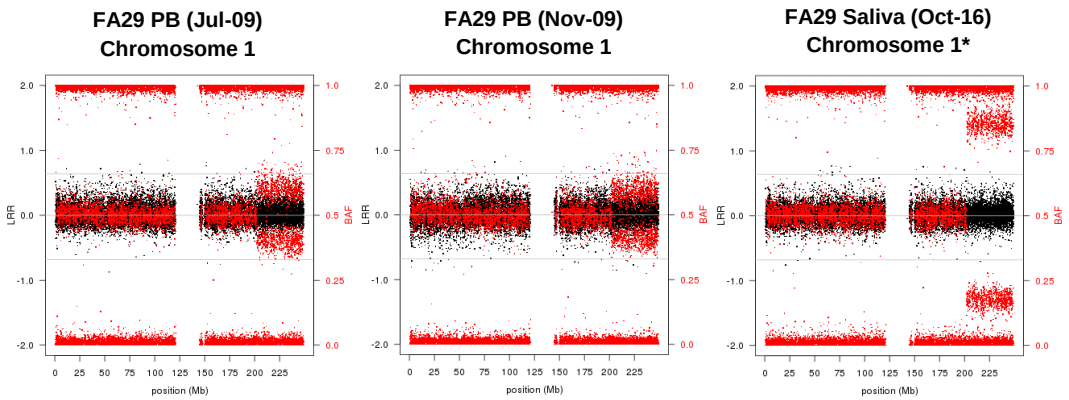


Figure 3. SNP array plots of mosaic UPD 1q detected in blood of one FA patient serially studied and also in saliva. SNP array plot images show the LogR ratio (LRR) indicated with black dots (scale on the left side) and B-allele frequency (BAF) indicated with red dots (scale on the right side). LRR is a measure of relative probe intensity ratio and provides copy number data whereas BAF is an estimation of the frequency of the B allele of a given SNP in the cell population. LRR and BAF in combination were used to distinguish between normal cells (BAF at 0 (AA), 0.5 (AB) or 1 (BB) and LRR at 0) from copy-number changes (LRR and BAF are altered) and copy-neutral changes as UPDs (LRR at 0 and BAF is altered). The tissue analyzed (PB: peripheral blood, Saliva) and the date of sample collection are specified in each individual. Rearrangements not detected by MAD algorithm but detected after visual inspection are indicated with "*". Jul: July; Nov: November; Oct: October.

FA patient	Tissue	Sample 1		Sample 2		Sample 3		Sample 4		Time of follow-up
		Date sample collection	Mosaicism (%)	Date sample collection	Mosaicism (%)	Date sample collection	Mosaicism (%)	Date sample collection	Mosaicism (%)	
FA103	PB	2002	No	2004	No	-	-	-	-	2
FA110	PB	2013	Gain 1q (15%)†** Loss 18p (14%)†**	2014	Gain 1q (18%)** Loss 18p (17%)**	2016	Gain 1q (12%)†** Loss 18p (12%)†**	-	-	3
FA125	PB	2002	No	2015	No	2016	No	-	-	14
FA126	PB	2002	No	2014	No	2016	No	-	-	14
FA164	PB	2003	No	2008	No	-	-	-	-	5
FA281	PB	2006	No	2007	No	-	-	-	-	1
FA311	PB	2005	No	2014	No	-	-	-	-	9
FA328	PB	2006	No	2014	No	2015	No	-	-	9
FA438	PB	2009	No	2012	No	2014	No	2016	No	7
FA537	PB	2010	No	2016	No	-	-	-	-	6
FA55	PB	2001	No	2016	No	-	-	-	-	15
FA558	PB	2012	No	2014	No	2016	No	-	-	4
FA574	PB	2012	No	2015	No	-	-	-	-	3
FA630	PB	2013	No	2013	No	2015	No	-	-	2
FA652	PB	2014	No	2015	No	-	-	-	-	1
FA82	PB	2001	No	2014	No	2016	No	-	-	15
FA655	PB	2013	No	2014	No	2016	No	-	-	3
FA712	PB	2014	No	2016	No	-	-	-	-	2
FA739	PB	2015	No	2015	No	2016	No	-	-	1
FA573	PB	2015	Loss 13q (30%)	2016	Loss 13q (33%)	-	-	-	-	1
FA29	PB	2009	UPD 1q (20%)	2009	UPD 1q (18%)	-	-	-	-	0,42
FA379	PB	2014	No	2016	No	-	-	-	-	2
FA771	PB	2015	No	2016	No	-	-	-	-	1
FA772	PB	2015	No	2016	No	-	-	-	-	1
FA807	PB	2016	No	2016	No	-	-	-	-	0,5
FA848	PB	2016	No	2017	No	-	-	-	-	1
FA664	Saliva	2014*	UPD 3p (14%)	2016	UPD 3p (19%)	-	-	-	-	2
FA536	Saliva	2014	No	2016	No	-	-	-	-	2

Table 2. A total of 28 out of 167 FA individuals were serially studied along 5 years of follow-up on average. After transversally analyzing 167 FA patients for CMEs detection (sample 1), 28 of them were studied evolutionarily through the analysis of 1-3 more samples collected serially along an average 5y follow-up period (0.42-15y). With this longitudinal analysis we did not detect any individual acquiring CMEs with the pass of the time but evolution of CMEs previously detected could be observed. PB: peripheral blood; †: rearrangements not detected by MAD but by visual inspection; *: is the unique sample that does not correspond to the once used for the transversal analysis; **: rearrangements probably involved in an unbalanced translocation; -: unfilled fields due to the absence of sample collection.

and saliva samples (**figure 3, figure S3**). Surprisingly, we detected one individual (FA535) that did not carry any CME in blood but, after 6 years follow-up, he presented an interstitial deletion affecting the long arm of chromosome 20 in saliva almost parallel with a MDS diagnosis (**figure S3, table S3 from (8)**).

Altogether, our results confirm an increased CMEs prevalence among FA patients associated with higher cancer risk and mortality. Serial samples analysis does not reveal new mosaicism cases but shows CMEs' evolution across the time. Finally, the study of saliva DNA confirms the early embryonic origin of the tested CMEs and uncovers a possible new CME carrier.

DISCUSSION

CMEs have been widely associated with aging and cancer risk, especially with hematologic malignancies, becoming a potential biomarker to identify those individuals at higher risk of cancer (1–3,12–14). This would be particularly useful in risky populations with a baseline higher risk of neoplasia like aging people or patients with a chromosome instability syndrome such as Fanconi anemia (4,5). The development of effective, economic and less invasive screening approaches to early detect cancer, as CMEs detection by SNP array in blood or saliva, would have a positive impact on health care. It would let to improve cancer patients' prognosis (due to an earlier diagnosis) as well as their adherence to medical follow-up (with less invasive screening tests). Fanconi anemia would be an example of this since it requires periodic bone marrow aspirates to attempt an early leukemia diagnosis as much as possible (7). Considering this, we wanted to determine whether CMEs detection in an easy to obtain sample as blood or saliva would actually be a feasible screening approach to identify those FA patients that would be at a higher risk of cancer, a part from their intrinsic increased risk, and should follow a closer medical vigilance without avoiding bone marrow aspirates.

The work here presented is an extension of the study published by Reina-Castillón et al. (8) shown in the first part of the chapter. We increased the total FA samples size from 130 to 167 FA patients. As expected, we corroborated that FA patients have and increased prevalence of CMEs in blood in agreement with previous publiccations that associate mosaicism with hematologic malignan-

cies (1–3,12–14). Concordantly with previous knowledge that associate mosaicism with age (1,2,12–17), mosaicism prevalence was higher among FA patients between 19-50 years of age (34%) compared to those below 18y (7%). Since this difference due to the age was also seen in age-matched control groups (<18y: 0.04%, 19-50y: 0.31% (8)), the overall increased risk of mosaicism in both groups was very similar (~170-180 fold-increase).

Regarding the kind of rearrangements most frequently detected, we observed that an important proportion of them were gains in 1q, gains in 3q or monosomies in chromosome 7 which are typically associated with hematologic malignancies and poorer prognosis (6). In this context, we observed that 90% (10/11) of CME carriers with MDS or acute myeloid leukemia (AML) harbored, at least, one of these rearrangements in blood and/or saliva. Regarding other genome regions, we detected another individual presenting a 6p terminal deletion with the interstitial breakpoint located at Major Histocompatibility Complex (MHC) loci. With this, we reinforce the idea exposed in in the first part of the chapter in which we proposed that MHC region is a hotspot for CMEs in FA patients. We overall detected 6 CMEs (2 UPDs, 3 losses and 1 gain) with a breakpoint within this region in a total of 5 patients where 4 of them presented a cancer diagnosis (2 solid tumors and 2 AML) in parallel or after mosaicism detection. Finally, we observed 5 individuals from *FANCD2* complementation group carrying a terminal UPD 3p with very similar, if not the same, interstitial breakpoint. These results will be extensively discussed in **chapter 2** since they are the starting point of one of the main messages of the thesis.

Concordantly with our previous results (8), we corroborated an augmented (5.6 fold increase) global risk of malignancy and a shortened survival secondary to cancer after DNA sample collection in CMEs carriers when compared with FA patients without CMEs. Prevalent and incident cancer rate also depended on the number of CMEs detected, being higher (40 fold increase) with the augment of the total amount of events observed. Additionally, considering only incident cancer cases after sampling, we observed a reduced cancer free time in FA patients with CMEs after sample collection showing a tendency of more cancer diagnosis after mosaicism detection (4.5 fold increase).

Focusing on three new cases detected with both cancer and clonal mosaicism, we identified five CMEs in saliva from one patient (FA825) who was diag-

nosed of esophageal cancer in parallel with sample collection. Given the proximity between buccal cavity (from where saliva was taken) and esophagus under a physical and embryologic origin point of view, we propose that CMEs detected in saliva could be related somehow with the diagnosed esophageal neoplasia. However, analyzing tumor DNA for detecting the same CMEs observed in saliva would be crucial to stablish whatever potential mechanistic relationship between CMEs and tumor development. Moreover, it would be interesting to detect saliva CMEs also in blood DNA (obtained at the same time of the tested saliva) but at similar or lower mosaicism percentage than CMEs in saliva. This would demonstrate, first, an early embryonic origin of mosaic events and, second, that findings in saliva sample were not secondary to its contamination with blood CMEs carrying cells. Similarly, the detection of higher mosaic percentages of CMEs in tumor compared to blood would also go against a possible tumor contamination with blood, cell free DNA or lymphocytes carrying CMEs. On this direction, in the first phase of our study (8), we detected two CMEs in both blood and tumor (anal SCC) samples of one FA patient (FA13) and at different mosaicism percentage, suggesting a possible mechanistic link between clonal mosaicism and tumor development. Our observations, together with previous published data showing similar proportion of cells carrying CMEs both in the blood and bladder mucosa of 4 bladder cancer patients (12), would be in agreement with the idea of that CMEs detected in easy to obtain samples could be also used as a biomarker for solid cancer risk.

Concordantly with estimations of mosaicism reversion prevalence in FA around 20-25% (24), a total of 20.9% FA patients presented a mosaic reversion of one of their constitutional *FANC* mutations. The proportion of CMEs carriers among them was low (3/35) and only one of them was diagnosed of cancer (vulvar SCC). In agreement with publications that suggest that reverse mosaicism would act as a “natural therapy” reducing chromosome fragility and hematologic problems risk (37,38), diagnosis of hematologic malignancy was reported only in one out of these 35 patients with reverse mosaicism. Moreover, we propose that FA mutation reversion could become visible following the same clonal process as any CME: after clonal expansion of a preexisting reverted wild-type clone due to a selective advantage.

One of the weak points of our previous work was the lack of the study of serial samples during a follow-up period to monitor

mosaicism evolution and detect new CMEs and new cancer diagnosis related to them. The longitudinal analysis of 28 individuals did not reveal any new CME but showed some fluctuations of mosaicism percentage of CMEs previously detected along the time. Thus, we observed no new cancer cases secondary to CMEs acquired in an average 5y follow-up.

Finally, we had available saliva DNA from 18 FA individuals (6 of them identified as CMEs carriers). The detection of the same CMEs both in blood and saliva in all 6 mosaic carriers revealed an early embryonic origin of these events. Importantly, mosaicism percentage of events in saliva were similar or higher than the observed in blood in all 6 cases, demonstrating that finding those CMEs also in saliva was not secondary to its contamination with blood cells carrying CMEs but by the presence of those rearrangements in cells from buccal cavity. In 2/6 cases (FA29 and FA681), saliva and blood samples were obtained with more than 1 year of difference. Saliva was taken 7 and 3 years after blood sample collection respectively and, in both individuals, it showed an increase of clonal cell fraction harboring the mosaic event. With this information, we could not determine whether CME mosaicism increased only in saliva or in both tissues along the time. Similarly, we detected one FA patients that in 2010 did not carry any CME in blood but, after 6 year follow-up, he presented an interstitial mosaic deletion 20q in saliva coinciding with the diagnosis of a MDS. This data could suggest an individual with acquired clonal mosaicism along the time with possible consequences on cancer development. However, because of the gap between blood and saliva collection, we could not obtain a firm conclusion. In case of finding the 20q mosaic deletion in a blood sample from the same time that saliva was collected (2016), we would have identified a new CME carrier after 6y of follow-up and we also would corroborate the early embryonic origin of the event. Contrary, in case of not finding the 20q mosaic deletion in a blood sample from 2016 and corroborating the presence of the rearrangement in saliva from 2010, we only could affirm a late embryonic origin of the mosaic event specifically in saliva.

To conclude, our results from the first (8) and second study support the idea of that detection of CMEs in blood or saliva by SNP array could be used as an early biomarker of cancer, at least in FA patients subgroup. This screening strategy could be incorporated in the cronic FA patients' follow-up to highlight those cases that would require additional controls or, at least, not avoiding the recommended ones.

ACKNOWLEDGMENTS

We are grateful to the FA families, the Asociación Española de Anemia de Fanconi and Ralf Dietrich for their collaboration and continuous support, as well as the contribution of the Spanish FA Clinical Network taking care of these patients (**table S4 from (8)**). Grant Support: J.R. is supported by a fellowship from the Spanish Ministry of Education (FPU13/ 00782), and M.A-C. by a Juan de la Cierva postdoctoral fellowship (JCI-2011-10660). J.S. team is funded by the Generalitat de Catalunya (SGR0489-2009;SGR317-2014), the ICREA-Academia program, the Marató de TV3 (project 464/ C/2012), the Spanish Ministry of Health (projects FANCOSTEM and FANCOLEN), the Spanish Ministry of Science and Innovation (projects CB06/07/0023, SAF 2012-31881 and SAF2015-64152-R), the European Commission (EUROFANCOLEN project HEALTH-F5-2012-305421, and FANCODRUG project H2020-703521), the Fanconi Anemia Research Fund Inc, and the European Regional Development (FEDER). L.A.P-J. team is funded by the Spanish Ministry of Health (FIS-PI1302481 cofunded by FEDER), the Generalitat de Catalunya (SRG1468-2014), the ICREA-Academia program, and also acknowledges support from the Spanish Ministry of Economy and Competiveness “Programa de Excelencia Maria de Maeztu” (MDM-2014-0370). CIBERER is an initiative of the Instituto de Salud Carlos III, Spain. D.S. is grateful to the Schroeder-Kurth Fund for financial support.

REFERENCES

1. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet*. 2012;44(6):651–8.

2. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet*. 2012 Jun;44(6):642–50.

3. Schick UM, McDavid A, Crane PK, Weston N, Ehrlich K, Newton KM, et al. Confirmation of the reported association of clonal chromosomal mosaicism with an increased risk of incident hematologic cancer. *PLoS One*. 2013;8(3):e59823.

4. Bogliolo M, Surralles J. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. *Mol gen*. 2015 Aug 6;33:32–40.

5. Wang AT, Smogorzewska A. SnapShot: Fanconi anemia and associated proteins. *Cell*. 2015 Jan 15;160(1-2):354–354.

6. Mehta PA, Harris RE, Davies SM, Kim M-O, Mueller R, Lampkin B, et al. Numerical chromosomal changes and risk of development of myelodysplastic syndrome--acute myeloid leukemia in patients with Fanconi anemia. *Cancer Genet Cytogenet*. 2010 Dec;203(2):180–6.

7. Fanconi Anemia Research Fund. Inc. Fanconi Anemia: Guidelines for Diagnosis and Management. Eiler ME, Frohnmayer D, Frohnmayer L, Larsen K, Owen J, editors. 4th ed. 2014;421.

8. Reina-Castillón J, Pujol R, López-Sánchez M, Rodríguez-Santiago B, Aza-Carmona M, González JR, et al. Detectable clonal mosaicism in blood as a biomarker of cancer risk in Fanconi anemia. *Blood Adv*. 2017;1(5):319–29.

9. González JR, Rodríguez-Santiago B, Cáceres A, Pique-Regi R, Rothman N, Chanock SJ, et al. A fast and accurate method to detect allelic genomic imbalances underlying mosaic rearrangements using SNP array data. *BMC Bioinformatics*. 2011 Jan;12(1):166.

10. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

11. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc*. 1958;53(282):457–81.

12. Rodríguez-Santiago B, Malats N, Rothman N, Armengol L, Garcia-Closas M, Kogevinas M, et al. Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *Am J Hum Genet*. 2010 Jul 9;87(1):129–38.

13. Forsberg LA, Rasi C, Razzaghian HR, Pakalapati G, Waite L, Thilbeault KS, et al. Age-related somatic structural changes in the nuclear genome of human blood cells. *Am J Hum Genet*. 2012 Feb 10;90(2):217–28.

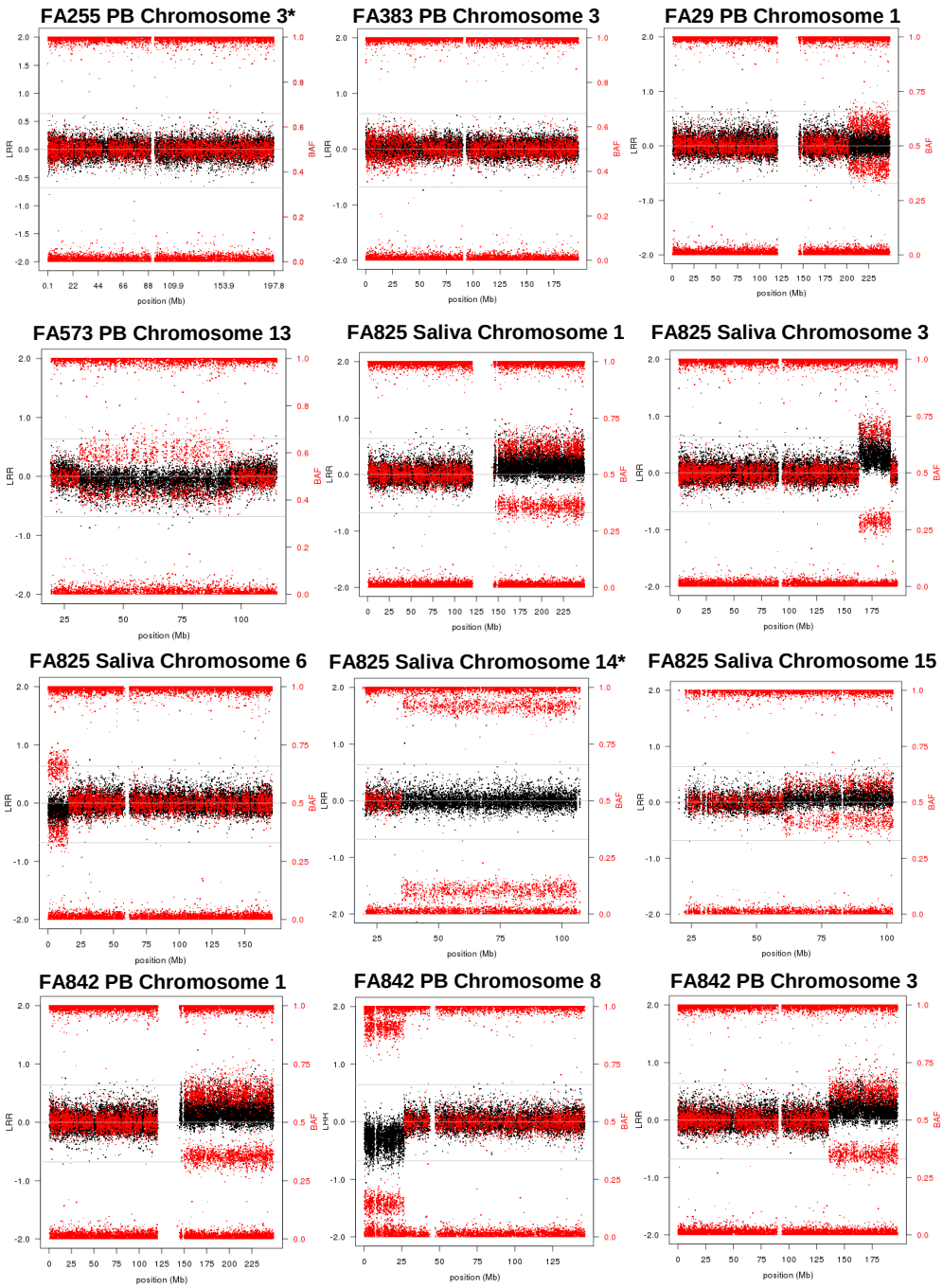
14. Machiela MJ, Zhou W, Sampson JN, Dean MC, Jacobs KB, Black A, et al. Characterization of Large Structural Genetic Mosaicism in Human Autosomes. *Am J Hum Genet*. 2015 Mar 5;96(3):487–97.

15. Machiela MJ, Zhou W, Caporaso N, Dean M, Gapstur SM, Goldin L, et al. Mosaic chromosome 20q deletions are more frequent in the aging population. *Blood Adv*. 2017;1(6):380-385.

16. Machiela MJ, Chanock SJ. The ageing genome, clonal mosaicism and chronic disease. *Curr Opin Genet Dev*. 2017 Feb 6;42:8–13.

17. Score J, Chase A, Forsberg LA, Feng L, Waghorn K, Jones A V, et al. Detection of leukemia-associated mutations in peripheral blood DNA of hematologically normal elderly individuals. *Leukemia*. 2015 Jul 28;29(7):1600–2.

SUPPLEMENTARY INFORMATION



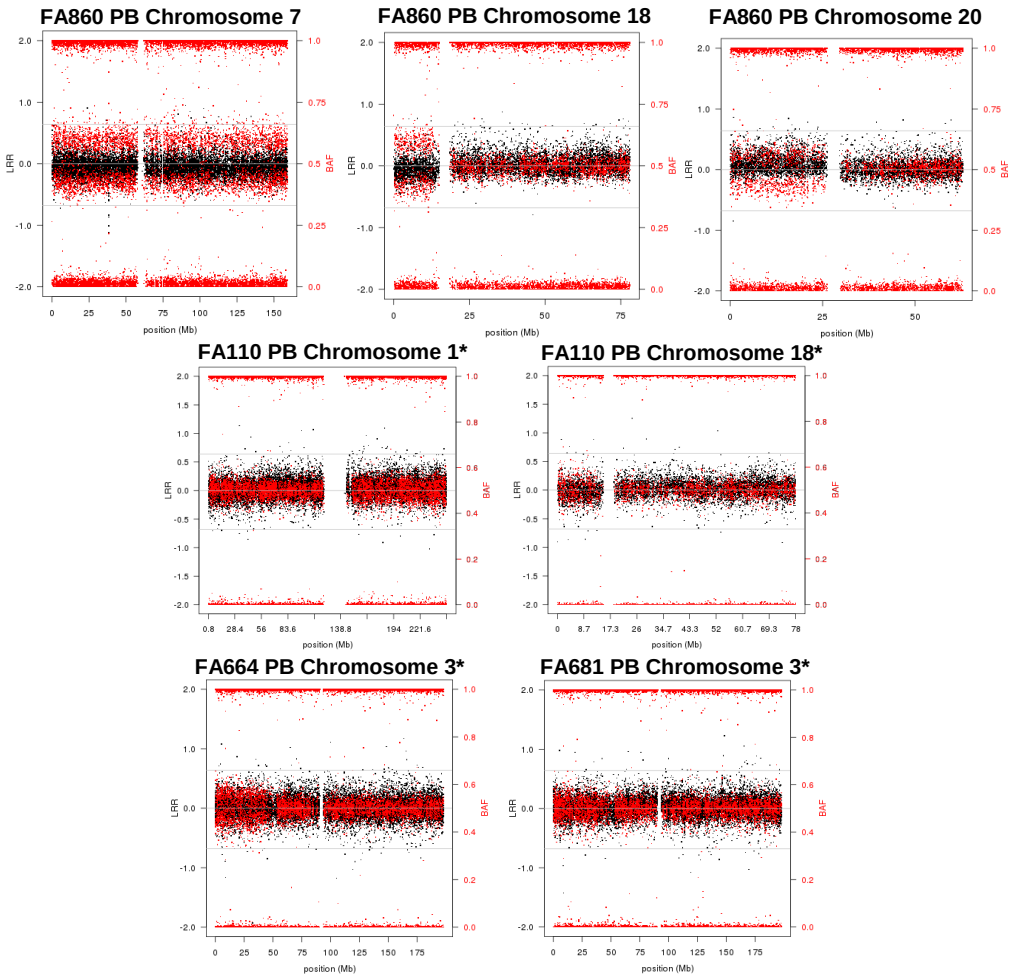


Figure S1. SNP array plots for CMEs detected in 7 new and 3 reanalyzed FA patients. SNP array plot images show the LogR ratio (LRR) indicated with black dots (scale on the left side) and B-allele frequency (BAF) indicated with red dots (scale on the right side). LRR is a measure of relative probe intensity ratio and provides copy number data whereas BAF is an estimation of the frequency of the B allele of a given SNP in the cell population used for validations. LRR and BAF in combination were used to distinguish between normal cells (BAF at any locus is either 0 (AA), 0.5 (AB) or 1 (BB) and LRR is 0) from copy-number changes (LRR and BAF are altered) and copy-neutral changes as UPDs (LRR is 0 and BAF is altered). The tissue analyzed in each individual is specified (PB: peripheral blood, Saliva). Rearrangements not detected by MAD algorithm but detected after visual inspection are indicated with "*". FA110, FA664 and FA681 were individuals belonging to the first study's phase (8) and have been reclassified as individuals with CMEs after a reanalysis of MAD output.

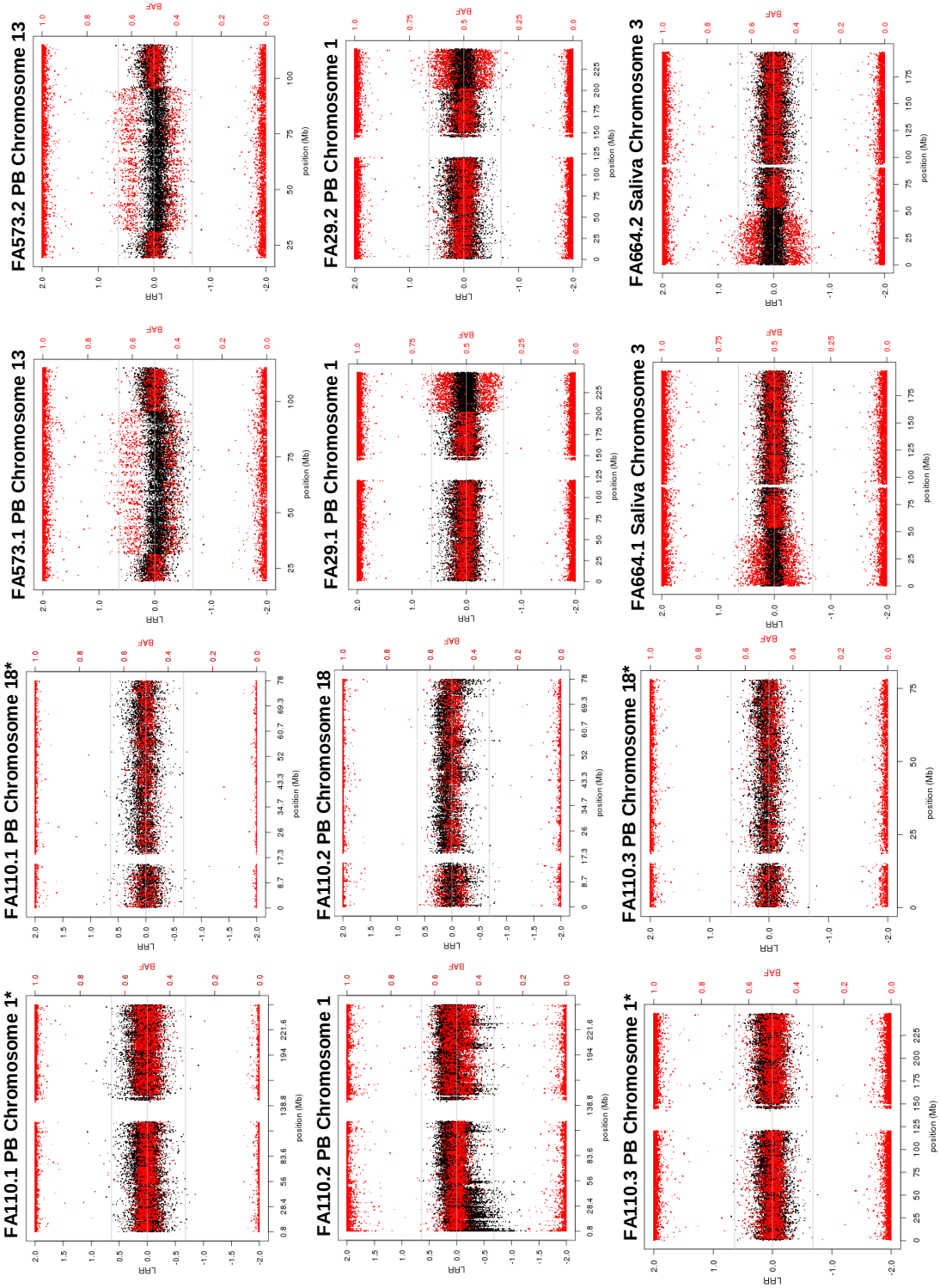
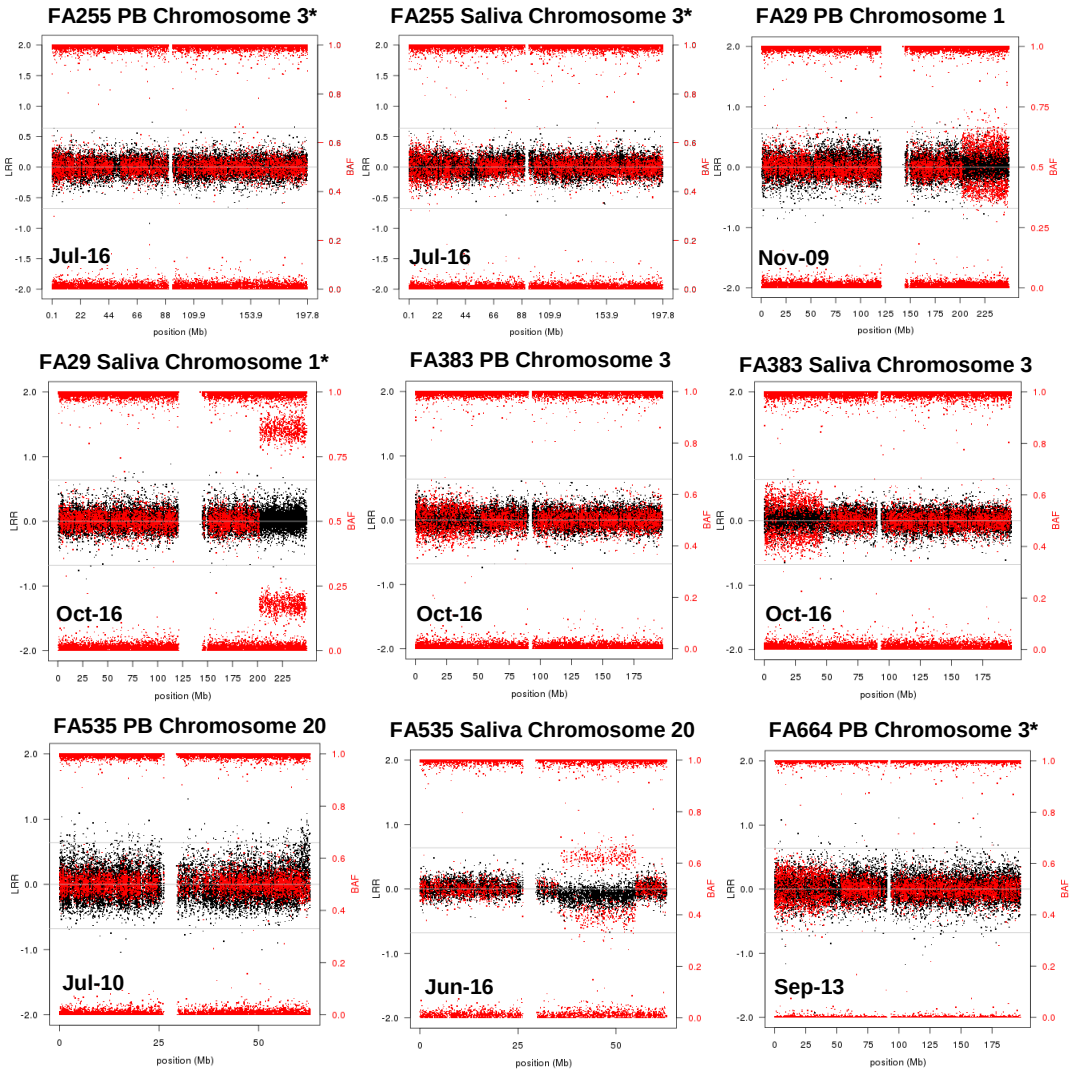

FA573.2 PB Chromosome 13
FA29.2 PB Chromosome 1
FA664.2 Saliva Chromosome 3
FA573.1 PB Chromosome 13
FA29.1 PB Chromosome 1
FA664.1 Saliva Chromosome 3
FA110.1 PB Chromosome 18*
FA110.2 PB Chromosome 18
FA110.3 PB Chromosome 18*
FA110.1 PB Chromosome 1*
FA110.2 PB Chromosome 1
FA110.3 PB Chromosome 1*
LRR
BAF
position (Mb)

Figure S2. SNP array plots for CMEs detected in 4 FA individuals serially evaluated along a 5y follow-up period on average. SNP array plot images show the LogR ratio (LRR) indicated with black dots (scale on the left side) and B-allele frequency (BAF) indicated with red dots (scale on the right side). Each sample is specified with a number corresponding to the nomenclature in table 2 (i.e. Sample 1 from FA110 in table 1 is FA110.1 in figure S2). The specific tissue analyzed is also indicated. PB: peripheral blood; *: rearrangements not detected by MAD but by visual inspection-



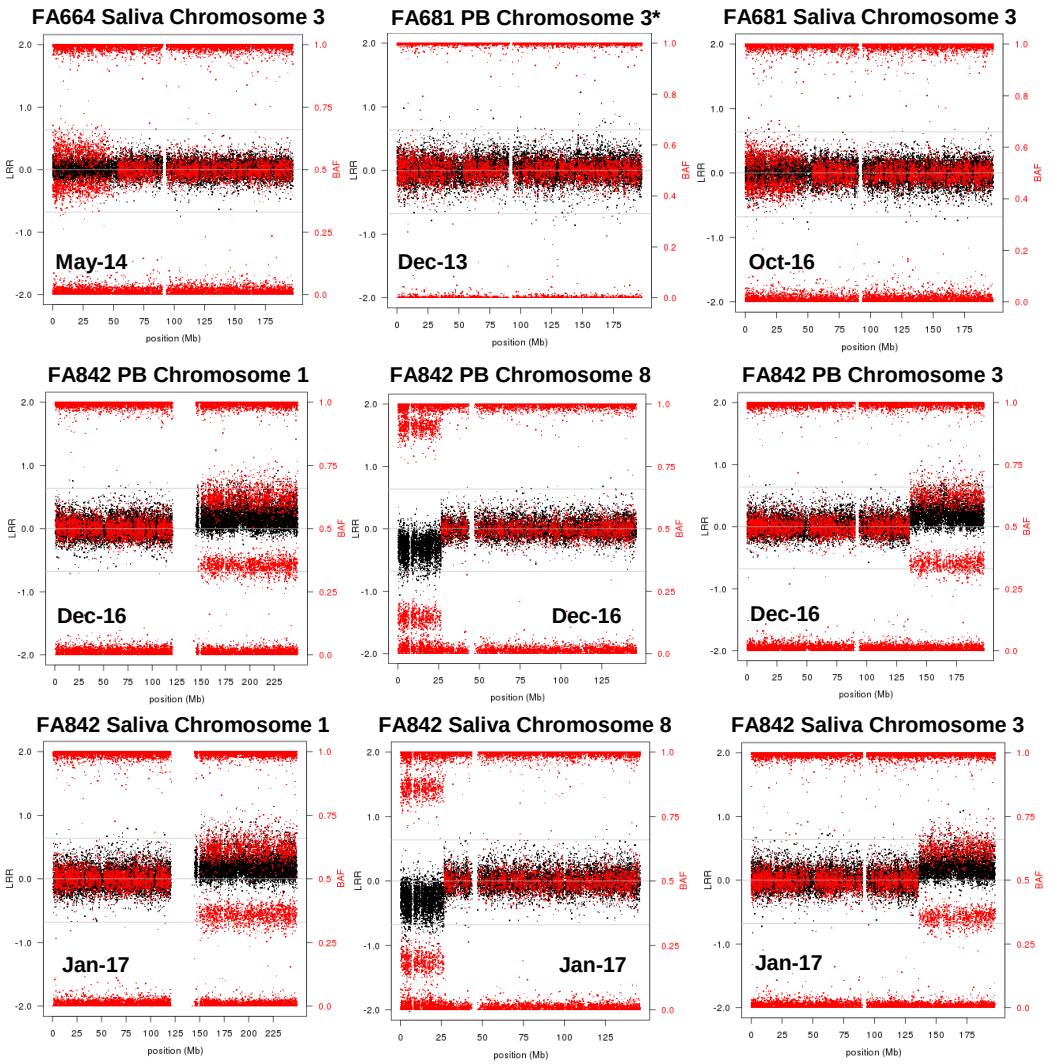


Figure S3. Blood and saliva SNP array plots for CMEs detected in 7 FA individuals. SNP array plot images show the LogR ratio (LRR) indicated with black dots (scale on the left side) and B-allele frequency (BAF) indicated with red dots (scale on the right side). The specific tissue analyzed (PB: peripheral blood, Saliva) together with date of sample collection are indicated. Blood sample was the once included in the previous transversal analysis. In cases of having more than one blood sample analyzed for CMEs detection, here we show the results from the blood with date collection more similar to saliva sample date collection. Similarly, in case of more than one saliva analyzed, we show the one with date collection more similar to blood's date collection. *: rearrangements not detected by MAD but by visual inspection. Jul: July; Oct: October; Jun: June; Sep: September; Dec: December; Jan: January (months are followed by years).

FA patient	Tissue	Date sample collection	Number of events	Type of event	Chr	Start SNP probe (bp)	End SNP probe (bp)	Size (bp)	LRR	Bdev	Cellullarity (%)
FA255	PB	Jul-16	1	UPD[ǂ]	3p	pter	NA (44000000-49000000)	NA (44000000-49000000)	0,01	0,03	5,10
FA29	PB	Jul-09	1	UPD	1q	202130754	qter	47081461	0,02	0,10	19,80
FA383	PB	Oct-06	1	UPD	3p	66894	46605447	46538553	0,03	0,04	8,80
FA573	PB	Dec-15	1	Loss	13q	32008008	94121685	62113677	-0,09	0,09	29,93
FA825	Saliva	Jul-16	5	Gain	1q	cen	qter	103226830	0,18	0,13	70,27
				Gain*	3q	171386231	191833747	20447516	0,31	0,18	56,01
				Loss	6p	pter	16133986	16133985	-0,13	0,13	42,02
				UPD[ǂ]	14q	34994335	qter	72293328	0,001	0,38	76,88
				Gain	15q	60241509	qter	42157122	0,09	0,08	36,97
FA842	PB	Dec-16	3	Gain**	1q	cen	qter	103226830	0,18	0,13	70,64
				Loss**	8p	pter	25421567	25421566	-0,30	0,36	83,12
				Gain	3q	135933264	qter	61900494	0,20	0,13	72,48
FA860	PB	Feb-17	3	Loss[•]	7	pter	qter	159119220	-0,06	0,08	28,67
				Loss**	18p	pter	cen	14308620	-0,06	0,09	29,06
				Gain**	20p	pter	cen	29558463	0,09	0,06	27,27
FA110	PB	Oct-13	2	Gain[ǂ,**]	1q	cen	qter	104737887	0,03	0,03	14,82
				Loss[ǂ,**]	18p	pter	cen	15102421	-0,04	0,04	13,57
FA664	PB	Sep-13	1	UPD[ǂ]	3p	pter	49956628	49956627	0,01	0,05	9,62
FA681	PB	Dec-13	1	UPD[ǂ]	3p	pter	NA (49836707)	NA (49836707)	0,01	0,04	6,70

Table S1. SNP array data for CMEs detected in 7 new and 3 reanalyzed FA patients. This table shows SNP array data for each rearrangement detected. Mosaicism level was calculated for each event by using the following formulas: %UPD = 2Bdev*100; %Loss = ((2Bdev)/(0.5+Bdev))*100; %Gain = ((2Bdev)/(0.5-Bdev))*100; %MultiGain = (Bdev)/(0.5-Bdev))*100. B-deviation (Bdev) value indicates the difference between the observed and expected BAF so, altered regions can be also called by detecting segments with Bdev values different from zero. FA110, FA664 and FA681 were individuals belonging to the first study's phase (8) and have been reclassified as individuals with CMEs after a reanalysis of MAD output. Chr: chromosome; bp: base pair; LRR: logR Ratio; PB: peripheral blood; Jul: July; Oct: October; Dec: December; Feb: February; Sep: September; UPD: Uniparental Disomy; pter: p arm terminal; qter: q arm terminal; cen: centromere; [ǂ]: rearrangements not detected by MAD but by visual inspection; [•]: monosomy; *: gains of more than one extra copy (tetrasomy); **: rearrangements probably involved in an unbalanced translocation; "NA" (not available): refers to those coordinates that are imprecise due to low mosaicism percentage. Coordinates are in hg19 assembly.

Sample ID	Tissue analyzed	Genetic Group	Age at DNA sampling	Years from DNA sampling to cancer diagnosis	Cancer diagnosis (age)	Evolution (age)	Inclusion in Kaplan-Meier analysis
FA310	Saliva	A	37	0	SCC tongue (37)	Alive (38)	Yes
FA850	PB	A	36	0	Solid tumor (NI) (36)	Alive (36)	Yes
FA705	PB	A	48	0	Colon cancer (48)	Alive (50)	Yes
FA707	PB	A	16	2	MDS (18)	Alive (19)	Yes
FA235	Saliva	A	26	-	No	Alive (27)	Yes
FA379	PB	A	17	-	No	Alive (20)	Yes
FA395	Saliva	A	14	-	No	Alive (15)	Yes
FA536	Saliva	D2	12	-	No	Alive (15)	Yes
FA628	Saliva	A	12	-	No	Alive (13)	Yes
FA654	PB	A	21	-	No	Alive (22)	Yes
FA655	PB	A	3	-	No	Alive (7)	Yes
FA708	PB	A	3	-	No	Alive (6)	Yes
FA712	PB	A	5	-	No	Alive (8)	Yes
FA739	PB	A	4	-	No	Alive (6)	Yes
FA746	PB	A	13	-	No	Alive (14)	Yes
FA766	PB	A	7	-	No	Alive (9)	Yes
FA771	PB	G	5	-	No	Alive (7)	Yes
FA772	PB	G	1	-	No	Alive (3)	Yes
FA778	PB	A	6	-	No	Alive (8)	Yes
FA779	PB	A	0,9	-	No	Alive (2)	Yes
FA796	PB	E	8	-	No	Alive (9)	Yes
FA800	PB	A	41	-	No	Alive (42)	Yes
FA806	PB	A	3	-	No	Alive (4)	Yes
FA807	PB	A	5	-	No	Alive (6)	Yes
FA818	Saliva	Unassigned	9	-	No	Alive (10)	Yes
FA829	PB	A	36	-	No	Alive (37)	Yes
FA830	PB	A	6	-	No	Alive (7)	Yes
FA846	PB	G	6	-	No	Alive (7)	Yes
FA847	PB	A	10	-	No	Alive (11)	Yes
FA848	PB	A	5	-	No	Alive (6)	Yes

Table S2. Features of FA patients with no detectable CME at the time of sample collection from the FA cohort II (30/37). This table is an extension of table S3 of the main publication of this chapter (8). ID: identification; PB: peripheral blood; MDS: myelodysplastic syndrome; SCC: squamous cell carcinoma; -: unfilled fields due to the absence of a cancer diagnosis; NI: no information about the kind of tumor.

CHAPTER 2

Somatic uniparental disomy in the rescue of hematologic disorders

Judith Reina-Castillón, Marcos López-Sánchez, Roser Pujol, Montserrat Peiró, Olaya Villa, Benjamín Rodríguez-Santiago, Tonu Esko, Francisco X. Real, Stephen J. Chanock, Jordi Surrallés, Luis Alberto Pérez-Jurado

In preparation

Uniparental disomies (UPD), generated by meiotic and/or mitotic rearrangements with ulterior clonal selection, have classically been associated with disease by unmasking pathogenic recessive mutations or epigenetic marks leading to the complete loss function of some gene/s through homozygosis. This detrimental output fits with the well-known "two-hit hypothesis" developed by Alfred G. Knudson in 1971 for cancer development. However, UPDs can also have an opposite role by selecting the more functional allele in a given mutated locus, then acting as a protective mechanism in disease development. This work provides evidence that UPD's rescue role is common for some hematologic disorders and malignancies. The identification of UPDs should follow the characterization of its either protective or detrimental role in order to better define the risk and prognosis of disease and the molecular mechanisms involved.

Somatic uniparental disomy in the rescue of hematologic disorders

Judith Reina-Castillón, Marcos López-Sánchez, Roser Pujol, Montserrat Peiró, Olaya Villa, Benjamín Rodríguez-Santiago, Tonu Esko, Francisco X. Real, Stephen J. Chanock, Jordi Surrallés, Luis Alberto Pérez-Jurado

ABSTRACT

Uniparental disomy (UPD) leading to copy-neutral Loss of Heterozygosity (cnLOH) is considered the underlying mechanism of several disorders and many cancers. It has also been reported as a rescue mechanism for some lethal dominant mutations by selecting the clone with the wild-type (WT) allele. Here, we show that clonal somatic UPD can be selected to rescue hematologic disorders or malignancies as Chronic Lymphocytic Leukemia (CLL) associated with 13q14.3 deletions and Fanconi Anemia (FA) type D2, but also can lead to hematologic diseases as CLL and Polycythemia Vera (PV) associated with *JAK2* V617F mutation. A meta-analysis of 70144 non-leukemia individuals revealed a mosaic UPD13q-del13q14.3 co-occurrence in blood with a prevalence of 20.51% (8/39). By using mathematic approaches and simulations, we detected 6/8 individuals (7/9 samples) harboring an UPD13q acting as a second hit mechanism leading to del13q14.3 homozygosis. Two of these six individuals (three samples) had an additional UPD13q acting as a rescue UPD, surely, in one of them. In the two remaining individuals, the unique UPD13q detected probably acted as a rescue mechanism in one case and no conclusive results were obtained for the other one. Among 10 FA patients with compound heterozygous *FANCD2* mutations, we detected 5 with mosaic UPD3p including the *FANCD2* gene in blood and saliva; the allele selected allele in the UPD carried the more benign mutation. Finally, we observed that an UPD9p that encompassed the *JAK2* gene carrying V617F mutation was probably selecting the truncated allele being the cause of the diagnosed PV. Thus, this work gives an evidence that UPD can act as a rescue mechanism against cancer development besides the classical second-hit one that originate an important proportion of cancers.

INTRODUCTION

Loss of Heterozygosity (LOH) is the major mechanism associated with cancer disease due to somatic tumor suppressor genes inactivation. This is known as the "two-hit hypothesis" developed by Alfred G. Knudson in which he postulated that some tumors as Retinoblastoma (Rb) (1) and Tuberous Sclerosis Complex (TSC) (2) need two mutational events to occur. The first event is usually a mutation inactivating the first allele that can be inherited or not whereas de second hit appears always postzygotically and can be acquired by an inactivating mutation, loss of the allele, gene conversion or uniparental disomy (UPD) of the allele with the 1st hit. Thus, UPD is a source of mosaicism and the moment at which both hits arise during individual's development will determine the final percentage of cells affected by both hits. UPD is probably the most prevalent mechanism by which a second hit can appear and, besides by inactivating a tumor suppressor gene, it can also lead cancer through oncogene activation (3). Segmental UPD13q, UPD11p, UPD17p leading to homozygous mutations in *FLT3*, *WT1* and *NF1* genes respectively (4,5) have been associated with development of acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS). UPD17p leading to homozygous mutations in *TP53* gene have been frequently seen in chronic lymphocytic leukemia (CLL) and some lymphomas (6). An specific interesting case is UPD9p in JAK2V617F (c.1849G>T) mutation-positive patients in which somatic mutation homozygosis will lead to a certain myeloproliferative neoplasm (MPN) or other (polycythemia vera (PV), primary myelofibrosis (PMF) or essential thrombocytosis (ET)) depending on the JAK2V617F mutational burden (3,7,8). In summary, UPD9p is found in 48-80% PV patients, ~50% of cases of PMF and 6-18% of patients with ET. It also has been associated with ~5% AML cases that occur de novo but up to ~57% of AML derived from chronic ET, PV, or PMF (6). UPD affecting chromosome 17 or 13q have been also described in high prevalence in serous ovarian cancer and the former have been associated with shorter overall survival highlighting the need of further functional studies on candidate genes in UPD regions that could have some prognostic relevance (9). UPD also can lead to cancer by generating deletion homozygosis; CLL and Rb are two examples of this (10,11).

Besides acting as a second hit mechanism promoting the homozygosis of a genetic mistake, UPD can also promote homozygosis of the wild-type (WT) allele in a given heterozygous clone for a genetic alteration generating "health clones" in mosaicism and trying to avoid disease development. This "rescue" UPDs phenomenon has been described in ichthyosis with confetti (IWC) disease which is a severe sporadic skin disorder caused by dominant mutations in *keratin 1* or *10 (KRT1, KRT10)* genes and characterized by the accumulation of thousands of "health" skin spots harboring homozygous WT alleles thanks to UPD17q. Clones homozygous for the mutant allele are not usually detected suggesting that they are negatively selected or unviable. The high frequency of somatic reversion in IWC patients suggests that reverted clones are under strong positive selection due to a selective advantage and/or the reversion rate is elevated (12,13).

As above mentioned, CLL has been reported to appear sometimes as consequence of deletion homozygosis secondary to a UPD that acted as a second hit mechanism. CLL is the most common leukemia of adults in Western world and is defined as a malignancy of mature B-lymphocytes in which there is an accumulation of CD5-positive monoclonal B-cells in peripheral blood (more than 5×10^9/L) and/or in primary and/or secondary lymphoid tissues (14). Some works suggest that 5% of elderly individuals have a detectable ($<5\times10^9$/L) clonal B-cell population that resembles a CLL tumor cell, which is named monoclonal B-cell lymphocytosis (MBL), where only approximately 1% of cases will progress to CLL whereas the rest will remain asymptomatically (15,16). There are some genomic aberrations associated with CLL as deletions in 13q14.3 (55%) and in 11q (18%), trisomy of 12q (16%), deletions in 17p (7%) and in 6q (7%) (17). The most commonly deleted genomic region in CLL, 13q14.3, is known to undergo epigenetic silencing in one chromosome randomly during normal development in a similar way to X inactivation mechanism. However, genes of this region have been found to be silenced either on the maternal or paternal chromosome copy so imprinting has been excluded. Concordantly, many genes of 13q14.3 region have been shown to be monoallelically expressed in B and T cells, so there is an increased risk of functional nullizygosity if a deletion or mutation occurs in the functional allele (18). In fact, deletions of 13q14.3 region associated with CLL can be

monoallelic (76%) or biallelic (24%) depending on whether the first hit abolishes the functional allele or the other (19). The minimal deleted region (MDR) in 13q14.3 deletion (del13q14.3) has been identified as the region containing the *deleted in leukemia* 2 (*DLEU2*) gene and the first exon of *DLEU1* gene. Importantly, the miRNA (miR)-15a/16-1 cluster, which is thought to be a negative regulator of proliferation, is located within the intron 4 of *DLEU2* gene (19,20). Machiela et al. recently reported a frequency of 0.073% of mosaic del13q14.3 (including the MDR) among non-leukemia individuals (healthy individuals and patients with solid cancer) without significant differences between both subgroups (21). As previously reported (22,23), they found an association between age and mosaicism since they detected an accumulation of mosaic del13q14.3 with age. The well-known relationship between del13q14.3 and CLL together with the increasing frequency of both mosaic 13q14.3 loss and MBL/CLL with age would be in agreement with two different scenarios: (a) 13q14.3 mosaic deletions could be early MBL/CLL biomarkers or (b) del13q14.3 could appear because of the impairment in genomic maintenance capacities associated to aging leading or not to MBL/CLL development (21). The first option proposed would be consistent with our previous work in which we demonstrated that early detection of chromosomal mosaic events in blood seemed to be early marker for cancer in Fanconi anemia patients (24).

Fanconi anemia (FA) is a rare chromosomal instability syndrome characterized by the presence of congenital defects and physical malformations in around 75% of patients, risk of bone marrow failure and increased probability of both hematologic and solid tumors as Head and Neck Squamous Cell Carcinoma (HNSCC) and some gynecological tumors among others (25). FA is both clinically and genetically heterogeneous since upon 15 FA, 3 FA-like and 13 FA-associated genes have been described until now; where *FANCA* gene (FA gene) is mutated in 60-70% of cases (25–28). Mutations in these genes are inherited in an autosomic recessive manner except of those in *FANCB* gene that are X-linked (29) and those located in *FANCR/RAD51* gene which are autosomic dominant (30). All these genes cooperate in the FA/BRCA pathway responsible of repairing Interstrand DNA Cross-linkings (ICLs) and when its function is impaired, chromosome fragility and DNA breakages accumulation appear, being the diagnostic hallmark of these patients.

Although UPD is a thoroughly and well-known mechanism underlying several kinds of cancer by acting as a second hit mechanism, very few is known about its role in preventing cancer as a "rescue" system by selecting the wild-type allele. Thus, the present study is trying to elucidate the functional role of UPD in the context of two diseases occurring with leukemia as CLL and FA in order to contribute to the knowledge of UPD "rescue's role".

MATERIALS AND METHODS

***Subjects*.** We studied a total amount of 70144 non-CLL individuals (healthy controls and patients with solid tumors) whose blood or saliva DNA was analyzed by SNP array technology for clonal mosaicism detection. This cohort was obtained by merging data from Jacobs et al. (57699 subjects) (23) with new unpublished data from a dataset of 12455 subjects. These individuals belonged to the Estonian Gene Expression Cohort (EGCUT, www.biobank.ee) which comprises a large cohort of 53000 samples of the Estonian Genome Center Biobank, University of Tartu (31). Moreover, we studied a CLL cohort of 722 patients belonging to two different datasets previously published by Gunnarsson et al. (369 subjects) (32) and Edelmann et al. (353 subjects) (33) where, again, blood samples of these individuals were analyzed by SNP array. In this work, we also reanalyzed 130 FA patients reported in a previous publication from Reina-Castillón et al. whose blood DNA was studied by SNP array (FA cohort I) (24) together with a new FA cohort (FA cohort II) consisting of a total subset of 94 samples (68 peripheral blood (PB) and 26 saliva samples) from 37 new FA patients and from 21 FA patients already studied in the FA cohort I (serial samples). Finally, blood DNA from a PV patient belonging to the EGCUT project obtained at the time of the diagnosis (V52014) and also 3 years before (V38104) was also studied.

***Genotype data and mosaicism detection in EGCUT and FA cohort II*.** DNA was obtained from 12445 EGCUT subjects who were genotyped using three different arrays: 7166 samples with the OmniX and 5278 samples with lower resolution arrays (2641 with Human370CNV & 2637 with Metabochip). In case of FA cohort II, a total amount of 68

peripheral blood and 26 saliva samples were analyzed by SNP array using Illumina Infinium HumanCore BeadChip (250K) platform. Genetic mosaicisms were detected from SNP array data by using MAD algorithm (34). The method is able to detect mosaic alterations of ≥2-0.5 Mb in length using both B-allele frequency (BAF) and log relative probe intensity ratio (LRR). The fixed B-deviation (Bdev) value (deviation from the expected BAF value of 0.5 for heterozygous SNPs) is used to call an allelic imbalance by using a segmentation procedure that detects those regions with a Bdev value different from 0. Some false positive alterations were detected, mainly due to bad quality arrays. Therefore, a manual curation via visual inspection and by analyzing variability of LRR and BAF mean values in the segment was performed by two independent investigators. Visual inspection was key to detect 13q14.3 deletions (<2Mb) in regions affected by long 13q UPDs in CLL and non-CLL cohort as well as to detect UPD3p previously not detected by MAD algorithm in some FA samples due to the low mosaicism level.

Statistical Analysis. Based on Fisher exact test, due to the small numbers, and doing 2x2 tables, we determined the statistical significance for the following comparisons: (I) del-13q14.3-UPD13q co-occurrence significance versus del-13q14.3 prevalence in non-leukemia group (II) del13q14.3-UPD13q co-occurrence in non-leukemia versus leukemia individuals, (III) deletion-UPD co-occurrence in chromosome 13q versus other genomic regions typically altered in leukemia (2p, 3p, 5q, 9q, 11q and 20) (35) in non-CLL individuals, (IV) deletion-UPD co-occurrence in chromosome 13 in non-leukemia subjects versus patients with Rb and (V) statistical significance of finding UPD3p only among *FANCD2* FA patients. P-values (p), Odds Ratios (ORs) and Confidence Intervals at 95% were computed using Fisher test R function (36).

Definition of number of UPDs and deletions in chromosome 13. For each non-leukemia sample with mosaic del13q14.3 and UPD13q (Case 234, Case 962, V10627/V52467 (two samples from the same individual but obtained with 3 years of difference), V10178, Sample 138, Sample 191, Sample 416 and Sample 5175), we studied SNP array plots in detail to stablish the number of rearrangements affecting chromosome 13 (1 UPD and 1 deletion, 2 UPDs and 1 deletion, 1 UPD and 2 deletions or 2 UPDs and 2 deletions) since we realized that there were some samples that had not only a simple UPD

13q plus a del13q14.3. Visual analysis of SNP array plot was enough to determine the presence of one or two UPDs whereas statistical comparisons between LRR of deleted regions and LRR of similar segments in undeleted regions of chromosome 13 were required to stablish the presence of one or two deletions (data not shown).

Estimation of cell fraction subtypes in non-CLL samples. SNP array data (Bdev, LRR and the presence/absence of BAF split) for each event and sample was used to calculate the number of chromosomes affected by each rearrangement and, with this, we determined percentages of every cell fraction in two scenarios: (I) UPD and deletion coexisting in same cells and (II) UPD and deletion being in different clones. Some specific formulas were used in the intermediate steps of these estimations. These were (I) a previously reported formulae $\%\mathrm{UPD} = (2 * \mathrm{Bdev}) * 100$ (37) that allowed us to obtain the percentage of cells harboring a UPD and (II) the equation $\mathrm{Chromosomes\ with\ deletion}=(2 - 2 * \mathrm{e}^{1.5*\mathrm{LRR}}) * 100$ that converted the LRR of a deletion into the number of chromosomes affected by that rearrangement. This numerical expression defines an exponential relationship between LRR and ploidy and was developed by our group after using unpublished data from our group and the app notes from Illumina (Technote cytoanalysis https://www.illumina.com/Documents/products/technotes/technote_cytoanalysis.pdf; Appnote CGH, https://www.illumina.com/Documents/products/appnotes/appnote_cgh.pdf; Technote CNV Algorithms, https://www.illumina.com/documents/products/technotes/technote_cnv_algorithms.pdf). We considered as possible those scenarios in which we could obtain a total sum of percentages of all cell fractions equal to 100% and all SNP array parameters were according with our prediction.

Simulator of SNP array plots. In order to give strength to our predictions in every non-CLL case, we used a simulator developed by our group that reproduces the expected SNP array plot for a given combination of cell fraction percentages. So, those models where cell fraction percentages were able to reproduce the original SNP array plot were considered as more probable for the corresponding sample. The simulator requires shiny package of R-Studio (38) and it is available in: https://github.com/dwalinybalin/chr13sim

Fluorescence-activated cell sorting (FACS) and microsatellite analysis. B-cells (CD5+, CD19+) of Case 234, V10627,

V52467 and V10178 were separated from the rest of hematologic cells by flow cytometry. DNA from both fractions was extracted with Puregene DNA Isolation Kit (Qiagen Inc, US) in every sample. Microsatellite analysis (D13S272, D13S263) was performed in all fractions and in PB. PCRs were done by standard protocols and relative quantifications of alleles were obtained with Genescan ABI3100 genotyper (Applied Biosystems, US). PB and bladder cancer DNA of Case 962 was also analyzed by microsatellite analysis (D13S272, D13S263) as above mentioned. DNA from 3 blood control samples, FA664 (PB and saliva), FA681 (PB, saliva and Fibroblast (FB)), FA820 (PB, FA664/681's mother), FA821 (PB, FA664/681's father), FA383 (PB, saliva and FB), FA826 (PB, FA383's mother), FA255 (PB and saliva), FA866 (PB, FA255's mother) and FA867 (PB, FA255's father) were analyzed to study D13S1263 microsatellite in order to determine the relative amount of paternal versus maternal allele at 3p region affecTed by UPD.

cDNA synthesis and qPCR. RNA was isolated from B and hematologic non-B cells of Case 234 using Trizol according to manufacturer's instructions (Invitrogen, US). cDNA was obtained with random hexamers and Superscript II reverse transcriptase according to manufacturer's instructions (Invitrogen, US). *DLEU2* expression was quantified by qPCR (**table S1**) and Power SYBR® Green Master Mix (Applied Biosystems, US). Results were analyzed by using a method described by Pfaffl et al. (39).

FISH. BAC probes targeting 13q14.3 region and hybridizing in 13q (as a control) were used to quantify the proportion of cells homozygous or heterozygous for the short deletion in V52467. BACs were isolated with standard alkaline lysis (Qiagen GmbH, Germany), were labeled with SpectrumRed /Green-dUTP by nick translation (Abbott Molecular, US) and hybridized to lymphocytes' fixed nuclei from V52467 as previously reported (40). A total amount of 100 interphase nuclei were analyzed with Olimpus-BX51 fluorescence microscope and images were captured with the Cytovision software (Applied Imaging Systems, UK).

Sanger sequencing. DNA from V38104 (PV patient), FA664 (PB and saliva), FA681 (PB, saliva and FB), FA821 (PB), FA383 (PB, saliva and FB), FA826 (PB), 110243 (PB), FA255 (PB and saliva), FA866 (PB), FA867 (PB) were analyzed by Sanger sequencing using capillary electro-

phoresis (ABI 3730xl DNA Analyzer systems (96 capillary instrument), GATC Biotech AG (Headquarter), European Genome and Diagnostics Centre, Germany) to detect *FANCD2* mutations. Primers were designed (PRIMER 3 application) (http://bioinfo.ut.ee/primer3-0.4.0/) and used in standard PCR conditions (**table S1**).

The ratio between WT and mutated alleles' relative peak height was used to determine the selected allele by the UPD. In the FA family carrying c.226_228delA mutation, a total of 11 heterozygous points were tested and the overall amount of points with a selected WT or mutated allele was counted.

RESULTS

Previous work from our group showed that mosaic UPD including almost the entire 13q arm (UPD13q) was found in blood DNA from two unrelated bladder cancer patients who also had a deletion in 13q14.3 region (del13q14.3) partially affecting the MDR (37) (**figure 1**). The co-occurrence of two chromosomal events in two unrelated individuals without diagnosed hematologic problems would suggest a mechanistic or functional link between both events that could ultimately have some effect on MBL/CLL development since del13q14.3 is a well-known event associated with this disease. In order to validate this co-occurrence, we performed a meta-analysis in which we evaluated the prevalence of mosaic UPD13q and del13q14.3 detected by SNP array in PB or saliva DNA in a sample set of 70144 subjects from general population but without a clinical CLL diagnosis. We detected that 39 out of 70144 (0.055%) individuals had mosaic del13q14.3 and 8 out of these 39 (20.51%) subjects with the deletion also had an UPD13q showing a co-occurrence between UPD13q and del13q14.3 higher that what would be expected by chance (Fisher exact test, $p=1.3x10^{-18}$). This co-occurrence was also detected in a sample set of 722 CLL patients since around 58% (419/722) of them had del13q14.3 and 21 out of 419 (5.01%) had a UPD13q besides the deletion (Fisher exact test, $p=1.53x10^{-82}$). Interestingly, the co-occurrence was found to be 5 times higher in the non-CLL group than in the CLL one (OR:4,89; CI95%: 2,0-11,94; p=0.0015). The co-occurrence of UPD13q and del13q14.3 in non-leukemia cohort was even

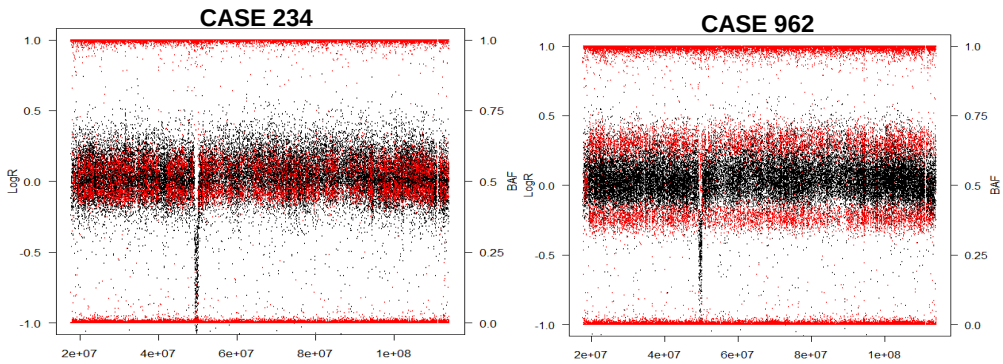


Figure 1. Two examples of non-leukemia individuals with mosaic del13q14.3 and UPD13q detected in blood by SNP array. Here we present cases 234 and 962, initially reported by Rodriguez-Santiago et al., in which the UPD13q encompassing the 13q14.3 deleted region is denoted. Each picture show two parameters: the LogR ratio (LRR) indicated with black dots (scale on the left side) and B-allele frequency (BAF) indicated with red dots (scale on the right side). LRR is a measure of the relative probe intensity ratio and gives copy number information whereas BAF is an estimation of the frequency of the B allele of a given SNP in a cell population. Both values, LRR and BAF, in combination let us to distinguish between normal cells (LRR is 0 and BAF is either 0 (AA), 0.5 (AB) or 1 (BB)) from copy-number changes (LRR and BAF are altered) and copy-neutral changes as UPDs (LRR is 0 and BAF is altered). The difference between the observed and expected BAF is denoted as b-deviation (Bdev), thus, altered regions can be also called by detecting segments with b-deviation values different from zero. SNP array plots coordinates are in hg18 assembly.

strongly supported by the fact that we did not find any UPD13q no affecting 13q14.3 region nor a single case of UPD overlapping with a deletion in other chromosome region typically deleted in leukemia patients (2p, 3p, 5q, 9q, 11q and 20) (35) (Fisher exact test, $p=2x10^{-6}$).

After validating UPD13q-del-13q14.3 coexistence, we focused on elucidating whether the relationship between both events was mechanistic or functional. To address this point, we analyzed in detail both rearrangements detected by SNP array in the 8 non-leukemia subjects (9 samples since 2 of them belonged to the same individual) (**figure S1** and **table S2**). We observed several issues that gave us some information important to consider for our purpose. Firstly, in all cases, the UPD13q was longer than del13q14.3 (but always including it) suggesting that the relationship between both rearrangements is not mecha-

nistic but functional since in case of being mechanistic, we would expect similar breakpoints in both events. Secondly, 7 out of 8 samples had heterozygous signaling in the deleted region with BAF values around 0.5 suggesting an equilibrium between A and B alleles in that region (we did not consider Case 234 since a ~1.5Mb homozygosity no informative region including 13q14.3 was detected). When a germline heterozygous deletion occurs, we expect extreme BAF values (1 or 0) for the SNP within the deleted region since no heterozygosis exist there. Otherwise, mosaic heterozygous deletions draw intermediate BAF values (between 1 and 0.5 and between 0.5 and 0) leading to a "BAF split". So, the detection of BAF values around 0.5 in 13q14.3 deleted region in 7/8 samples indicates that there are equivalent amounts of both maternal and paternal allele at that genomic region and two possible explanations could be given to this: (a) the proportion of cells heterozygous for the deletion are too low to affect the overall A/B ratio and generate a BAF split or (b) if heterozygous deleted cells are not negligible, cells with UPD13q are providing the "double dosage" of the same allele lost in cells with heterozygous del13q14.3 leading to an overall equal levels of the paternal and maternal alleles at the deleted region. Finally, we detected that Samples 138, 191, 416 and 5175 had one UPD13q and one del13q14.3 whereas Cases 234 and 962 harbored one UPD and two different deletions in size, V10178 had two different UPDs and one deletion and V10627/V52467 showed two different UPDs and two deletions. Overall, we actually detected eleven del13q14.3 and ten UPD13q in eight individuals.

To go further in the functional UPD13q-del13q14.3 relationship, we wanted to determine whether both rearrangements coexisted in the same cell fraction or were located in different clones in these 8 non-leukemia individuals (9 samples). The first scenario would suggest that the UPD would be acting as a second hit mechanism leading to deletion homozygosis whereas in the second situation the UPD13q of the functional allele of 13q14.3 region would be trying to compensate the general 13q14.3 deficit generated by deleted cells. In order to determine the most feasible situation for each sample, we considered the number of deletions and UPDs detected together with the SNP array parameters (Bdev, LRR and presence/absence of BAF split) for each event to estimate the percentage of all cell fractions in each scenario. Therefore, in cases in which the total sum of cell fractions was equal to 100% and all SNP array

parameters were in agreement, that scenario was accepted as the most probable. All calculations were done by assuming the simplest situation with the minimal number of cell fractions and mutational events (**table 1 and table S3**).

With these assumptions, we observed that all samples with one UPD13q and one del13q14.3 (samples 138, 191, 416, 5175) had both rearrangements together in the same clones suggesting UPD13q as a second hit mechanism leading to del13q14.3 homozygosis. We affirmed this based on not only for obtaining a total cell fraction equal to 100% under this supposal but also due to two facts. Firstly, we were able to discard, in fact in all 9 samples, the possibility of having only heterozygous deleted cells (and not homozygous) coexisting with other clones harboring the UPD alone because the total LRR drop observed in the deleted region could only be explained by the coexistence of both homozygous and heterozygous deleted cells. Moreover, supposing this situation led to an unreal total sum of cell fractions over 100% (data not shown). Secondly, focusing on these four samples, the existence of homozygous deleted cells and the fact that UPD is the most probable mechanism to generate them would be in favor with the idea of having both rearrangements in the same cell clones. Interestingly, V10178 sample would be in the same situation but having an extra fraction of cells harboring another UPD13q alone (longer than the other UPD) but affecting the same allele deleted in other fractions. Moreover, mosaicism level of this longer UPD13q (18%) was very similar to two extra UPDs detected in chromosomes 3 (17%) and 17 (22%) showing a similar origin in time (**table 1**, **table S3-panel A**).

Regarding samples V10627 and V52467, they belonged to the same individual but the latter was obtained three years after the former. Both samples showed two different deletions 13q14.3 and UPDs 13q (four events in total) and, after considering their SNP array data, we could determine that at least one of these UPDs was acting as a second hit mechanism. The reason why we did not obtain conclusive results for the remaining UPD was that we found two different explanations that perfectly matched with SNP array data of both samples (finding the same explanation matching with both samples was mandatory since they came from the same subject). The first compatible option (**table 1**, **table S3-panel B option A**) consisted of having some cells heterozygous for the short deletion and others for the

	A%	B%	C%	D%	E%	F%	G%	H%	I%	J%	TOTAL%	UPD
S.138	50,3	17,7	-	-	-	32,0	-	-	-	-	100	Del hom
S.191	48,5	9,5	-	-	-	42,0	-	-	-	-	100	Del hom
S.416	53,8	2,2	-	-	-	44,0	-	-	-	-	100	Del hom
S.5175	28,5	1,5	-	-	-	70,0	-	-	-	-	100	Del hom
V10178	13,5	31,5	-	-	-	37,0	-	-	-	18,0	100	Del hom + Rescue
V10627	47,6	8,9	13,2	-	-	-	17,8	12,5	-	-	100	Del hom + Del hom
	31,5	-	29,3	-	8,9	-	17,8	-	-	12,5	100	Del hom + Rescue
	20,8	-	40,0	-	8,9	-	-	12,5	17,8	-	100	Del hom + Rescue
V52467	4,3	40,5	37,6	-	-	-	10,5	7,1	-	-	100	Del hom + Del hom
	30,7	-	11,2	-	40,5	-	10,5	-	-	7,1	100	Del hom + Rescue
	23,8	-	18,1	-	40,5	-	-	7,1	10,5	-	100	Del hom + Rescue
Case 962	3,4	39,6	-	33,0	-	-	-	-	24,0	-	100	Rescue*
Case 234	6,5	40,3	-	43,3	-	-	-	-	9,9	-	100	Rescue
	26,3	20,4	-	43,3	-	9,9	-	-	-	-	100	Del hom

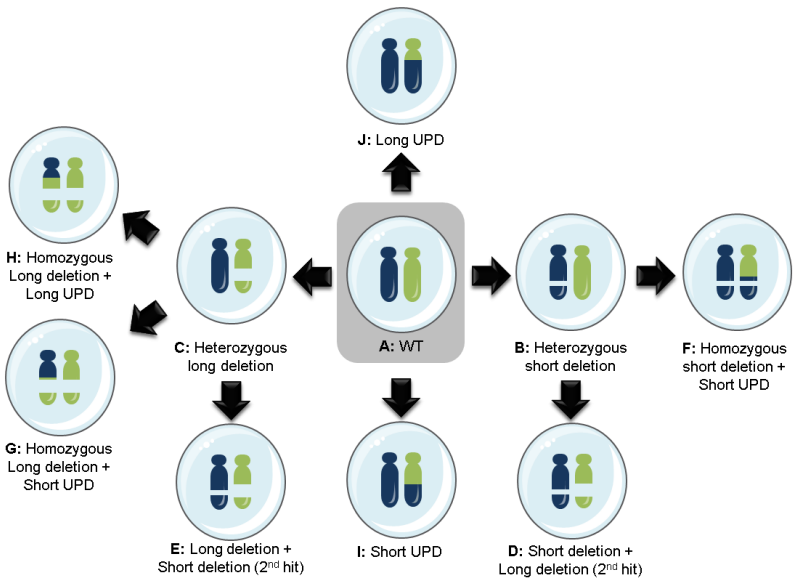


Table 1. UPD13q leading del13q14.3 homozygosis and rescue UPD13q were found, at least, in 7/9 and 1/9 samples respectively. In the upper part, percentages of different cell fractions observed in every sample are presented together with the final conclusion derived from the data (last column). In table S3, details of all calculations are specified. At the bottom, a detailed description of all cell fractions found in the whole sample set is shown as well as their corresponding letter in the table. Overall, we detected, at least, one UPD13q

acting as a second hit mechanism leading to del13q14.3 homozygosis in 7 out of 9 samples. Additionally, among these 7 samples, there was an additional UPD13q alone, and possibly acting as a rescue UPD, surely in 1 sample (V10178) and possibly in 2 samples (V10627, V52467) belonging to the same individual. The 2 remaining cases deserve to be considered apart because, in Case 234, either having the UPD13q alone or in the same clone with del13q14.3 was compatible with SNP array data; and, in Case 962, the only possible option was that UPD13q was acting as a rescue mechanism but we considered this data as possible and not conclusive since this involved a marked BAF split in the segment deleted only in the long deletion that was not actually clearly detected in the original SNP array plot (marked with *). Del hom: homozygous deletion; WT: wild-type; -: empty field since that sample does not harbor that cell fraction.

longest one, so, both deletions located in different cell clones. These cells would coexist with clones homozygous for the longest deletion, some cells, thought the long UPD and, others, through the short one. In this situation both UPDs would be acting as a second hit mechanism to lead long deletion homozygosis. The other possibility was those in which some heterozygous cells for the longest deletion (first hit) could acquire a second hit deletion (the short one) generating partial homozygous deleted cells. Regarding UPDs, one of them would be leading to long deletion homozygosis in another subset of cells whereas the other UPD would be alone selecting the same allele affected by the long deletion in other cells to avoid BAF split. Both short and long UPD could be alone or acting as a second hit mechanism, so both combinations were possible. Moreover, we knew that the longest deletion was the first hit and the one that got the homozygosis since in the oldest sample (V10627) there were almost no cells with the short deletion. Additionally, both V10627 and V52467 presented a deletion in the short arm of chromosome 2 whose cellularity was between 40-50% (**table 1, table S3-panel B options B**).

Finally, there were 2 out of 9 samples (Case 962 and Case 234) that presented two different deletions in 13q14.3 and one UPD13q. In Case 962, the only feasible option would be considering the coexistence of clones with UPD13q alone, clones heterozygous for the short deletion and clones harboring both short and long deletion together. However, with this, we obtained a notable BAF split for the region only deleted in the long deletion in our prediction compared to what was slightly insinuated in the SNP array plot. Interestingly, this individual presented an UPD17p with a very similar cellularity (26%) compared to

UPD13q (24%) suggesting a common origin in time. Regarding Case 234, we did not have information about the presence or absence of BAF split in the common region of both deletions due to an overlapping 1.5 Mb homozygous region. Thus, having the UPD both alone or causing a deletion homozygosis would be feasible by considering SNP array information (**table 1**, **table S3-panel C**).

Overall, we detected, at least, one UPD13q acting as a second hit mechanism leading to del13q14.3 homozygosis in 7 out of 9 samples. Additionally, among these 7 samples, there was a sample (V10178) and possibly two more (V10627, V52467) with an additional UPD13q possibly acting as a rescue mechanism. We also detected 1 out of 9 samples with a unique UPD13q possibly acting as a rescue UPD and one with ambiguous results since either having the UPD13q alone or in the same clone with del13q14.3 were compatible (**table 1**). Moreover, we were able to reproduce the original SNP array plot with the inferred cell fractions percentages for all samples by using a simulator which gave more strength to our findings and conclusions (**figure S2**).

By microsatellite assay, we observed that Case 234, as expected, was homozygous in the deleted region and UPD13q was present in total blood fraction. Interestingly, B-cells also presented the UPD13q and no D13S272 signaling in agreement with almost null levels of *DLEU2* expression detected in this cell fraction by quantitative PCR (qPCR). Regarding hematologic non B-cells fraction, it was almost non-existent since it only represented 7% of the total sample obtained from the cell sorting. Microsatellite assay did not work in hematologic non B-cells fraction of Case 234 but normal *DLEU2* expression levels were detected by qPCR (**figure S3a,b**). Altogether, these data would suggest that deletion would be present only in B-cells but would not clarify which scenario of the ones compatible with SNP array data would be the most feasible one: (a) coexistence of clones with UPD13q alone, clones heterozygous for the short deletion and clones harboring both short and long deletion or (b) coexistence of clones heterozygous for the short deletion, clones harboring both short and long deletion and clones homozygous for short deletion through the UPD. In Case 962, microsatellite assay showed an UPD13q and a rescue of the Loss of Heterozygosity (LOH) secondary to del13q14.3 in PB whereas bladder cells only harbored the del13q14.3 indicating an early embryonic origin of this event (**figure S3a**). Microsatellite ana-

lysis of DNA from samples V10627 and V52467 (belonging to the same subject) revealed again an UPD13q and a rescue of the LOH resulting from del13q14.3 both in total blood and in hematologic non B-cells fraction. Accordingly with SNP array plots, a decreasing UPD13q mosaicism level with time was detected with microsatellite analysis. Unfortunately, this assay did not work in B-cells fraction from these two samples because of technical problems (**figure S3a**). Additionally, a FISH assay revealed that lymphocyte fraction of peripheral blood in V52467 sample was composed by 88% of cells homozygous for 13q14.3 deletion, 6% of heterozygous deleted cells and 6% of non-deleted cells (WT and UPD cells) demonstrating the presence of both heterozygous and homozygous deleted cells in the cellular fraction of blood in this patient (B and non B cells). Since SNP array analyzed the whole blood whereas FISH assay studied only blood cellular fraction, percentages related to cell fractions' cellularity derived from each method were not comparable between them (**figure S3c**). Finally, D13S272 and D13S263 microsatellites were no informative for sample V10178 (data not shown).

With the reanalysis of 130 FA patients whose blood DNA was studied by SNP array and published by us (24) together with the study by SNP array of a new sample set of 68 blood and 26 saliva samples from 37 new FA patients and 21 FA patients already studied in the previous publication, we detected a total amount of 5 FA patients (two of them siblings) harboring a mosaic UPD3p both in blood and saliva except in case of one individual where only peripheral blood was studied. Thus, UPD3p was detected in a total of 10 samples (5 peripheral blood and 5 saliva samples) (**figure S4, table S4**). All UPD3p were terminal encompassing *FANCD2* gene and had a very similar right breakpoint within or near a homozygous region located at 50 Mb position of the short arm of chromosome 3 suggesting a hotspot for postzygotic mitotic rearrangements (**figure 2a**). Importantly, germinal compound heterozygous mutations in *FANCD2* gene were responsible of FA in all these five cases with mosaic UPD3p. Considering that UPD3p was actually observed in 5/10 compound heterozygous D2 patients and any FA patient from other complementation group harbored a UPD3p ($p= 1.1x10^{-5}$), some functional relationship between UPD3p and *FANCD2* gene mutations was suggested.

To go deeper in this aspect, we performed Sanger sequencing and microsatellite analysis of FA samples with mosaic UPD3p in order to determine the allele

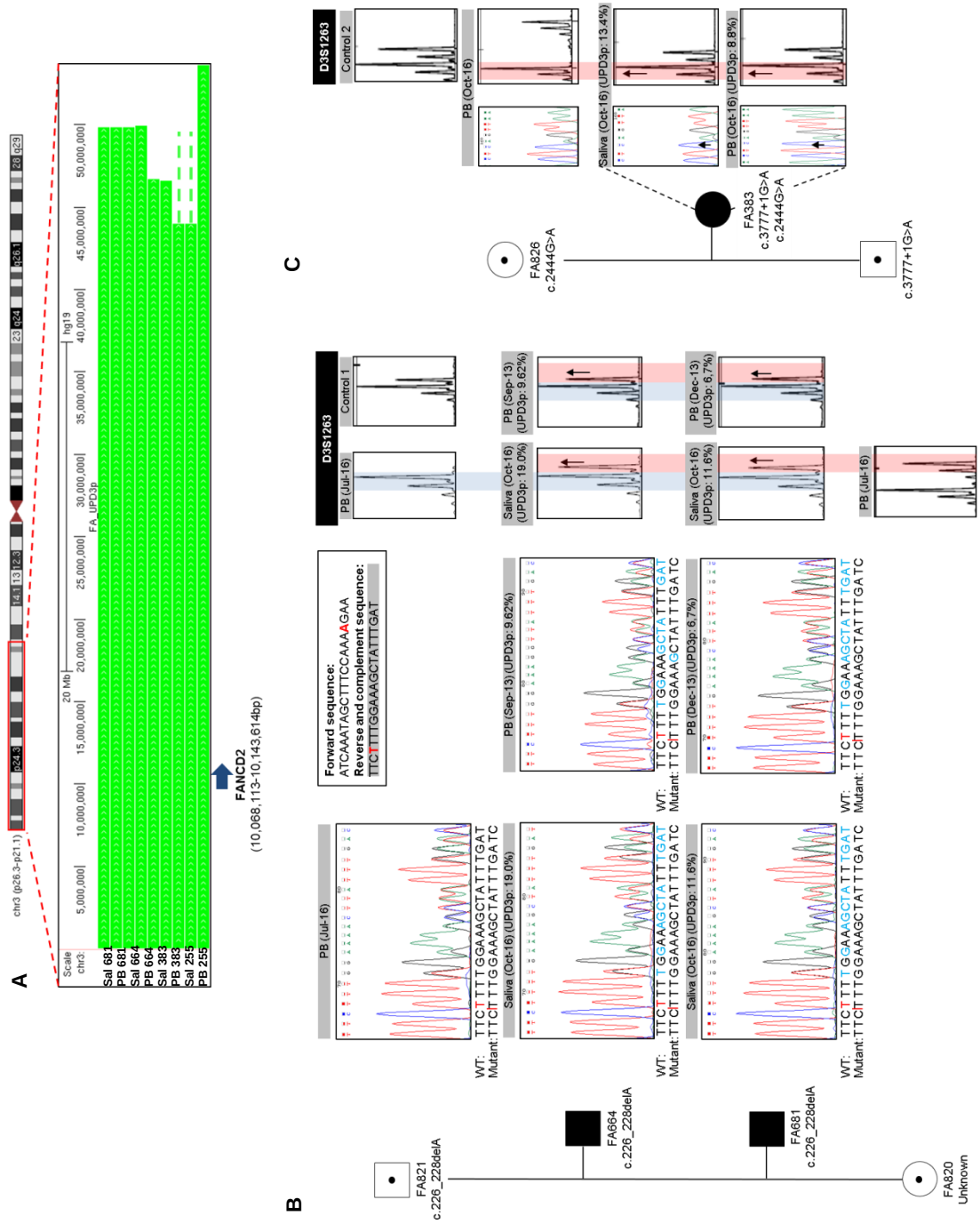

A
chr3 (p26.3-p21.1)
FA_UPD3p
Scale
chr3:
hg19
Sal 681
PB 681
Sal 664
PB 664
Sal 383
PB 383
Sal 255
PB 255
FANCD2
(10,068,113-10,143,614bp)
B
C
D3S1263
Control 1
Control 2
PB (Jul-16)
Saliva (Oct-16) (UPD3p: 19.0%)
PB (Sep-13) (UPD3p: 9.62%)
Saliva (Oct-16) (UPD3p: 11.8%)
PB (Dec-13) (UPD3p: 6.7%)
PB (Oct-16)
Saliva (Oct-16) (UPD3p: 13.4%)
PB (Oct-16) (UPD3p: 8.8%)
Forward sequence:
ATCAAATAGCTTTCCAAAAGAA
Reverse and complement sequence:
TTCTTTTGGAAAGCTATTTGAT
WT:
Mutant:
FA821
c.226_228delA
FA664
c.226_228delA
FA681
c.226_228delA
FA820
Unknown
FA826
c.2444G>A
FA383
c.3777+1G>A
c.2444G>A
c.3777+1G>A

Figure 2. Two examples in which UPD3p is selecting the less truncating FANCD2 mutated allele. A) UPDs 3p found in both blood and saliva tissues in 4 patients and only in blood of 1 patient are shown as green bars (UPD3p found in a second saliva sample from FA664 obtained in May 2014 is not shown). Dashed lines in FA255 samples indicate unprecise coordinates from SNP array data due to the low UPD mosaicism level. *FANCD2* gene and its coordinates are also indicated showing that UPD3p is encompassing this gene in all samples. UPD3p coordinates were obtained by using SNP array data but the homozygous region at 50Mb found in all samples could be masking the real right breakpoint of all these rearrangements. **B)** Here we present a family with two FA children where only the paternal mutation was known (c.226_228delA, marked in red). Sanger sequencing and microsatellite analysis revealed a selection of the WT allele for c.226_228delA mutation (maternal allele) in all samples with mosaic UPD3p from both siblings and with more clarity in those samples with higher mosaicism level of UPD. In Sanger sequencing panels, the relative height of WT and mutant peaks was measured for 11 different points. By comparing the WT/mutant height ratio in each specific position in the tested sample versus the control one (FA821 PB Jul-16), we determined the allele selected by the UPD in all 11 positions (marked in blue). All tested samples, which harbored a mosaic UPD3p, presented a global selection of the maternal WT allele for c.226_228delA mutation. In microsatellite assay, D3S1263 (chr3:11,417,247-11,617,535bp, hg19) analysis revealed that mosaic UPD3p caused an allelic imbalance in all samples and favored the selection of the maternal allele (pink shadowed), that is to say the WT allele for c.226_228delA mutation (highlighted in pink), in agreement with Sanger sequencing results. Again, this selection was more evident in those samples with higher percentages of UPD3p. These results would suggest that the unknown mutation is probably less truncating than the frameshift mutation already known (c.226_228delA) and selected by UPD3p. Note that for all samples, sample collection date and percentage of UPD3p are indicated next to the sample type. Results regarding a second saliva sample from FA664 obtained in May 2014 are shown in **Figure S5. C)** Another family is presented in this picture where a FA child harbored a missense (c.2444G>A, maternal) and an exon skipping (c.3777+1G>A, paternal) mutations. Since no father sample was available, Sanger sequencing and microsatellite analysis were only performed on mother's and patient's samples. Concordantly with previous family, the UPD3p was selecting the a priori less truncating mutated allele (c.2444G>A) that corresponded to the maternal (pink shadowed) one both in blood and saliva samples. Quantification of Sanger sequencing and microsatellite peaks' height is shown in **table S5**. WT: wild-type; PB: peripheral blood; Sal: Saliva; Oct-16: October 2016; Sep-13: September 2013; Dec-13: December 2013; Jul-16: July 2016. Coordinates are in hg19 assembly.

selected by the UPD3p. Considering that missense mutations are usually more benign than exon skipping, frameshift and nonsense mutations, we detected that the UPD3p was selecting the allele harboring the less *FANCD2* truncating mutation in 10 out of 10 tested samples suggesting a possible rescue function of this UPD3p in FA patients (**figure 2b and c, figure S5, table S5**).

Finally, a healthy 65 yo individual with a mosaic UPD9p detected in blood from 2009 (21.2%) showed a higher proportion of cells with this rearrangement (60%) plus several additional chromosomal aberrations 3 years after the first analysis suggesting a malignant proliferation. After this, the individual was actually diagnosed of PV. By Sanger sequencing, we observed that, in 2009 blood DNA, this patient had the typically somatic PV-associated mutation JAK2V617P. Considering that this mutation is post-zygotically acquired and that heterozygous germinal mutations show a variant allele fraction around 50%, we would expect a variant allele fraction lower than 50% for JAK2V617P mutation. The fact of finding almost 1 to 1 proportion of WT versus mutated allele by Sanger sequencing suggested a selection of the mutated allele by the UPD leading to PV (**figure 3**).

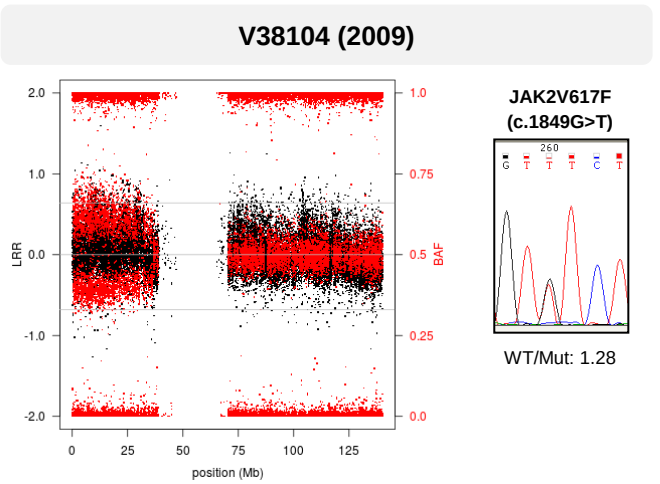


Figure 3. An individual with JAK2V617P mutation harbored a mosaic UPD9p three years before PV diagnosis acting as a second hit mechanism. SNP array analysis detected a mosaic UPD9p in a healthy individual who, 3 years later, accumulated extra chromosomal events (data not shown) together with an increased cellular fraction of cells with UPD9p. After this, this individual was diagnosed of PV. Sanger sequencing and relative peaks' ratio in the initial blood DNA sample (2009) revealed almost 1 to 1 proportion of WT versus mutated allele suggesting a UPD9p selecting the mutated allele and leading to its homozygosis. WT: wild-type; Mut: mutant.

DISCUSSION

Postzygotically acquired UPD has classically been considered as the underlying mechanism of the origin of some cancers by leading to tumor suppressor genes inactivation or oncogene activation through LOH that reveal the presence of recessive point mutations (1–5). Deletion homozygosis through UPD have also been described to cause certain cancer types as CLL and Rb (10,11). However, less is known about the possible rescue function of UPD in which the WT allele would be selected instead of the mutated one although some examples of this have been reported as some severe sporadic skin disorders (12,13). Here we show that UPD can act as a rescue mechanism besides the traditional "second-hit" one, by leading the selection of the most benign allele in a given heterozygous mutated/ deleted locus in the context of two different diseases (CLL and FA) with a shared clinical feature: the presence of hematologic alterations/malignancy.

The random detection of two unrelated non-leukemia individuals with a 13q14.3 deletion, typically associated with CLL, together with a UPD13q in 2010 (37) suggested a possible mechanistic or functional relationship between both events and, ultimately, with some effect on CLL development. The meta-analysis of 70144 individuals without any diagnosis of leukemia and studied by SNP array for mosaicism detection in blood revealed a prevalence of UPD13q-del13q14.3 around 20.51% which would confirm the initially suspected selection of this co-occurrence in the general population. This co-occurrence was also detected with a prevalence of 5.01% in a group of 722 CLL patients. Finding a higher co-occurrence of both events in the non-leukemia group compared to the leukemia one could be explained somehow by the fact that 13q14.3 deletions found in non-CLL group were smaller than in CLL group (data not shown) having less probability to cause cancer and consequently having more chance to be detected together with UPD13q during more the time. Considering that UPD13q and del13q14.3 found in our sample sets were very different in size in all cases (data not shown) together with the fact that the co-occurrence rate detected in non-leukemia group (20.41%) was not significantly higher than the published prevalence of UPD-deletion co-occurrence in other genome region and in the context of other cancer disease as Neurofibromatosis 2 (NF2) (46.6%) (41), we would conclude that the UPD13q-del13q14.3 co-

occurrence would not be explained by a mechanistic but by a functional relationship among both rearrangements.

To further characterize this possible functional relationship, we used mathematic approaches together with a simulator developed by our group and we detected that 7/9 tested non-leukemia individuals' blood samples harbored an UPD13q acting as a second hit mechanism leading to del13q14.3 homozygosis in agreement with the classical view of UPD as a source of rearrangements' homozygosis associated sometimes to cancer development. However, despite the well-known causal relationship between del13q14.3 and CLL, detection of del13q14.3 in homozygosis in individuals undiagnosed of leukemia does not automatically mean that the disease is present in them because monoclonal B-cell expansion sometimes harboring del13q14.3 in peripheral blood of asymptomatic persons has long been known. This condition, known as MBL, is specially associated with aging and does not necessary involve progression to cancer stages (CLL) (15,16). Additionally, 3/7 individuals presented another extra UPD13q different in size to the one acting as a second hit mechanism. We found, surely in 1 sample and possibly in 2 samples, that this UPD was selecting the WT allele corresponding to the same allele deleted in the cell fraction with del13q14.3 and acting as a rescue mechanism of the global 13q14.3 deficit. To really compensate 13q14.3 deficit, we propose that the WT allele selected by the UPD, and consequently the allele deleted in the other cell fraction, should correspond to the unmethylated and functional 13q14.3 allele. Finally, one individual out of the eight analyzed would harbor either a rescue or a second hit acting UPD whereas another individual probably would harbor a unique UPD acting as a rescue mechanism but results were not conclusive since our simulator was not able to reproduce with total fidelity the original SNP array plot. Unfortunately, any experimental assay that we performed was able to clarify those cases with two possible scenarios (V10627, V52467 and Case 234). Curiously, we detected two samples (V10178 and Case 962) with a mosaic UPD in chromosome 17p and, in V10178, a mosaic UPD also in chromosome in 3p. Interestingly, cellularity of these rearrangements was very similar to UPD13q mosaicism level observed in each corresponding sample suggesting a common origin in time.

In the context of FA, we identified 5 patients harboring a

mosaic UPD3p encompassing the *FANCD2* mutated gene. Considering that these 5 FA individuals were compound heterozygous for *FANCD2* mutations, the less truncating mutation was the one selected by the UPD3p. Interestingly, UPD3p found in all 5 FA patients had a right breakpoint coinciding with an homozygous region known to be typically homozygous in Africa, Middle East and Europe populations (42) suggesting a new mechanism of UPD where the two identical DNA strands due to the homozygosity could favor homologous recombination and UPD emergence.

Finally, we detected a PV patient that probably developed the disease due to JAK2V617F mutation selection through a somatically acquired UPD9p (encompassing *JAK2* gene), being another example of the well-known UPD role on causing cancer through LOH and mutation homozygosis.

Our findings suggest an important evidence of the rescue function that UPDs can play. Moreover we would distinguish between two quite different rescue mechanisms. In CLL, we have described a rescue UPD mechanism where cellular microenvironment would play a key role. Considering that secretory tumor-suppressive miRNAs, as miRNA-15a/16-1, can act as a death signal in a cell competitive process/paracrine way suggesting that cellular microenvironment plays a critical role in cancer development (43), we propose that del13q14.3 affecting the functional allele would lead to a microenvironment defective in miR-15a/16-1/*DLEU1*/*DLEU2* function and, hence, a proliferative and anti-apoptotic media suitable for tumor formation. Considering that UPD is a rather common event at embryonic stages but it does not use to show up due to the low proportion of cells with the rearrangement, we hypothesize that the existence of a microenvironment deficient in 13q14.3 function would promote a positive selection of a putative clone with UPD13q of the functional non-deleted allele in order to rescue the global MDR deficit. Similarly, in FA, a clone harboring a UPD selecting the most benign/less truncating allele would be positively selected, in this case, to improve global gene function. Thus, besides the classical well-known involvement of UPD in tumorigenesis by LOH secondary to second hit mechanism, UPD can also play a remarkable role on going against tumor formation by being involved in a rescue mechanism that would have as an ultimate goal of compensating an specific genetic deficit associated to disease.

ACKNOWLEDGMENTS

L.A.P-J. team is funded by the Spanish Ministry of Health (FIS-PI1302481 co-funded by FEDER), the Generalitat de Catalunya (SRG1468-2014), the ICREA-Academia program, and also acknowledges support from the Spanish Ministry of Economy and Competiveness "Programa de Excelencia Maria de Maeztu" (MDM-2014-0370). CIBERER is an initiative of the Instituto de Salud Carlos III, Spain. D.S. is grateful to the Schroeder-Kurth Fund for financial support. J.R. is supported by a fellowship from the Spanish Ministry of Education (FPU13/00782). B.R-S and L.A.P-J are employee and scientific advisor, respecttively, of qGenomics S.L.

REFERENCES

1. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971 Apr;68(4):820–3.
2. Hino O, Kobayashi T. Mourning Dr. Alfred G. Knudson: the two-hit hypothesis, tumor suppressor genes, and the tuberous sclerosis complex. *Cancer Sci*. 2017 Jan;108(1):5–11.
3. Makishima H, Maciejewski JP. Pathogenesis and consequences of uniparental disomy in cancer. *Clin Cancer Res*. 2011 Jun 15;17(12):3913–23.
4. Bullinger L, Krönke J, Schön C, Radtke I, Urlbauer K, Botzenhardt U, et al. Identification of acquired copy number alterations and uniparental disomies in cytogenetically normal acute myeloid leukemia using high-resolution single-nucleotide polymorphism analysis. *Leukemia*. 2010 Feb 17;24(2):438–49.
5. Flotho C, Steinemann D, Mullighan CG, Neale G, Mayer K, Kratz CP, et al. Genome-wide single-nucleotide polymorphism analysis in juvenile myelomonocytic leukemia identifies uniparental disomy surrounding the NF1 locus in cases associated with neurofibromatosis but not in cases with mutant RAS or PTPN11. *Oncogene*. 2007 Aug 23;26(39):5816–21.
6. Wang L, Wheeler DA, Prchal JT. Acquired uniparental disomy of chromosome 9p in hematologic malignancies. *Exp Hematol*. 2016 Aug;44(8):644–52.
7. Lapunzina P, Monk D. The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer. *Biol Cell*. 2011 Jul;103(7):303–17.

8. Kralovics R, Passamonti F, Buser AS, Teo S-S, Tiedt R, Passweg JR, et al. A Gain-of-Function Mutation of *JAK2* in Myeloproliferative Disorders. *N Engl J Med*. 2005 Apr 28;352(17):1779–90.

9. Tuna M, Ju Z, Smid M, Amos CI, Mills GB. Prognostic relevance of acquired uniparental disomy in serous ovarian cancer. *Mol Cancer*. 2015 Feb 3;14(1):29.

10. Lehmann S, Ogawa S, Raynaud SD, Sanada M, Nannya Y, Ticchioni M, et al. Molecular allelokaryotyping of early-stage, untreated chronic lymphocytic leukemia. *Cancer*. 2008 Mar 15;112(6):1296–305.

11. Mol BM, Massink MPG, van der Hout AH, Dommering CJ, Zaman JMA, Bosscha MI, et al. High resolution SNP array profiling identifies variability in retinoblastoma genome stability. *Genes, Chromosom Cancer*. 2014 Jan;53(1):1–14.

12. Choate KA, Lu Y, Zhou J, Elias PM, Zaidi S, Paller AS, et al. Frequent somatic reversion of KRT1 mutations in ichthyosis with confetti. *J Clin Invest*. 2015 Apr 1;125(4):1703–7.

13. Choate KA, Lu Y, Zhou J, Choi M, Elias PM, Farhi A, et al. Mitotic Recombination in Patients with Ichthyosis Causes Reversion of Dominant Mutations in KRT10. *Science*. 2010 Oct 1;330(6000):94–7.

14. Sagatys EM, Zhang L. Clinical and laboratory prognostic indicators in chronic lymphocytic leukemia. *Cancer Control*. 2012 Jan;19(1):18–25.

15. Landgren O, Albitar M, Ma W, Abbasi F, Hayes RB, Ghia P, et al. B-cell clones as early markers for chronic lymphocytic leukemia. *N Engl J Med*. 2009 Feb 12;360(7):659–67.

16. Rawstron AC, Bennett FL, O'Connor SJM, Kwok M, Fenton JAL, Plummer M, et al. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *N Engl J Med*. 2008 Aug 7;359(6):575–83.

17. Döhner H, Stilgenbauer S, Benner A, Leupolt E, Kröber A, Bullinger L, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*. 2000 Dec 28;343(26):1910–6.

18. Mertens D, Wolf S, Tschuch C, Mund C, Kienle D, Ohl S, et al. Allelic silencing at the tumor-suppressor locus 13q14.3 suggests an epigenetic tumor-suppressor mechanism. *Proc Natl Acad Sci U S A*. 2006 May 16;103(20):7741–6.

19. Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T, et al. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell*. 2010 Jan 19;17(1):28–40.

20. Klein U, Dalla-Favera R. New insights into the pathogenesis of chronic lymphocytic leukemia. *Semin Cancer Biol*. 2010 Dec;20(6):377–83.

21. Machiela MJ, Zhou W, Caporaso N, Dean M, Gapstur SM, Goldin L, et al. Mosaic 13q14 deletions in peripheral leukocytes of non-hematologic cancer cases and healthy controls. *J Hum Genet*. 2016 May;61(5):411–8.

22. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet*. 2012 Jun;44(6):642–50.

23. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet*. 2012;44(6):651–8.

24. Reina-Castillón J, Pujol R, López-Sánchez M, Rodríguez-Santiago B, Aza-Carmona M, González JR, et al. Detectable clonal mosaicism in blood as a biomarker of cancer risk in Fanconi anemia. *Blood Adv*. 2017;1(5):319–29.

25. Alter BP KG. Gene Reviews: Fanconi Anemia. University of Washington, Seattle; 1993-2017. Last reviewed: 23 February 2017.

26. Bogliolo M, Surralles J. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. *Mol gen*. 2015 Aug 6;33:32–40.

27. Ceccaldi R, Sarangi P, D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol*. 2016 May 5;17(6):337–49.

28. Fanconi Anemia Research Fund. Inc. Fanconi Anemia: Guidelines for Diagnosis and Management. Eiler ME, Frohnmayer D, Frohnmayer L, Larsen K, Owen J, editors. 4th ed. 2014;421.

29. Meetei AR, Levitus M, Xue Y, Medhurst AL, Zwaan M, Ling C, et al. X-linked inheritance of Fanconi anemia complementation group B. *Nat Genet*. 2004 Nov 24;36(11):1219–24.

30. Ameziane N, May P, Haitjema A, van de Vrugt HJ, van Rossum-Fikkert SE, Ristic D, et al. A novel Fanconi anaemia subtype associated with a dominant-negative mutation in RAD51. *Nat Commun*. 2015 Dec 18;6:8829.

31. Metspalu A. The Estonian Genome Project. *Drug Dev Res*. 2004;62:97–101.

32. Gunnarsson R, Mansouri L, Isaksson A, Goransson H, Cahill N, Jansson M, et al. Array-based genomic screening at diagnosis and during follow-up in chronic lymphocytic leukemia. *Haematologica*. 2011 Aug 1;96(8):1161–9.

33. Edelmann J, Holzmann K, Miller F, Winkler D, Buhler A, Zenz T, et

al. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood*. 2012 Dec 6;120(24):4783–94.

34. González JR, Rodríguez-Santiago B, Cáceres A, Pique-Regi R, Rothman N, Chanock SJ, et al. A fast and accurate method to detect allelic genomic imbalances underlying mosaic rearrangements using SNP array data. *BMC Bioinformatics*. 2011 Jan;12(1):166.

35. Machiela MJ, Zhou W, Sampson JN, Dean MC, Jacobs KB, Black A, et al. Characterization of Large Structural Genetic Mosaicism in Human Autosomes. *Am J Hum Genet*. 2015 Mar 5;96(3):487–97.

36. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

37. Rodríguez-Santiago B, Malats N, Rothman N, Armengol L, Garcia-Closas M, Kogevinas M, et al. Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *Am J Hum Genet*. 2010 Jul 9;87(1):129–38.

38. RStudio Team (2015). RStudio: Integrated Development for R.. RStudio, Inc., Boston, MA.

39. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001 May 1;29(9):e45.

40. Puigdecanet E, Espinet B, Villa O, Florensa L, Besses C, Serrano S, et al. Detection of abnormalities of PRV-1, TPO, and c-MPL genes detected by fluorescence in situ hybridization in essential thrombocythemia. *Cancer Genet Cytogenet*. 2006 May;167(1):39–42.

41. Hadfield KD, Smith MJ, Urquhart JE, Wallace AJ, Bowers NL, King AT, et al. Rates of loss of heterozygosity and mitotic recombination in NF2 schwannomas, sporadic vestibular schwannomas and schwannomatosis schwannomas. *Oncogene*. 2010 Nov 25;29(47):6216–21.

42. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet*. 2012 Aug 10;91(2):275–92.

43. Kosaka N, Iguchi H, Yoshioka Y, Hagiwara K, Takeshita F, Ochiya T. Competitive interactions of cancer cells and normal cells via secretory microRNAs. *J Biol Chem*. 2012 Jan 6;287(2):1397–405.

SUPPLEMENTARY INFORMATION

Figure S1

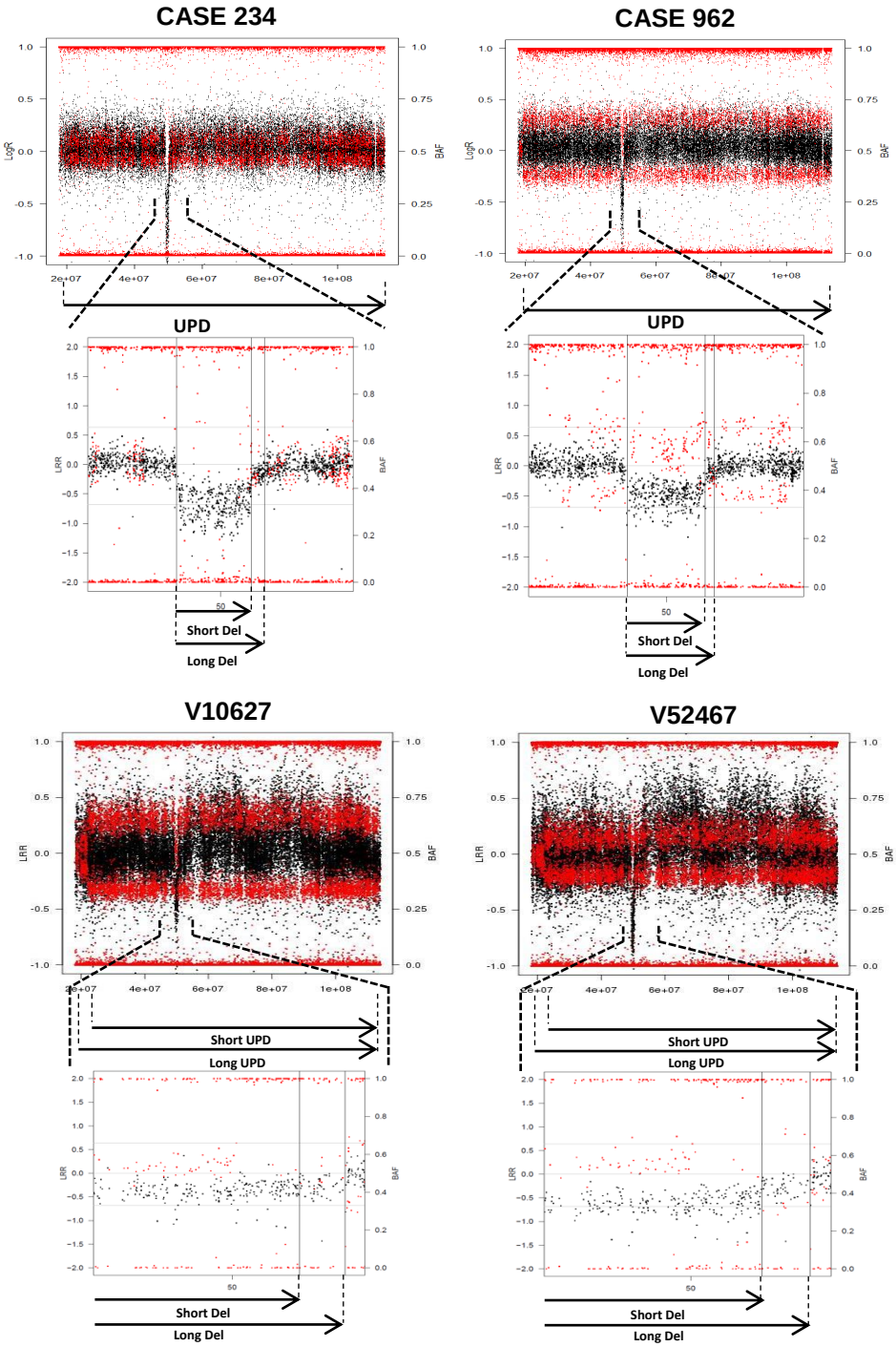


Figure S1 *(cont.)*

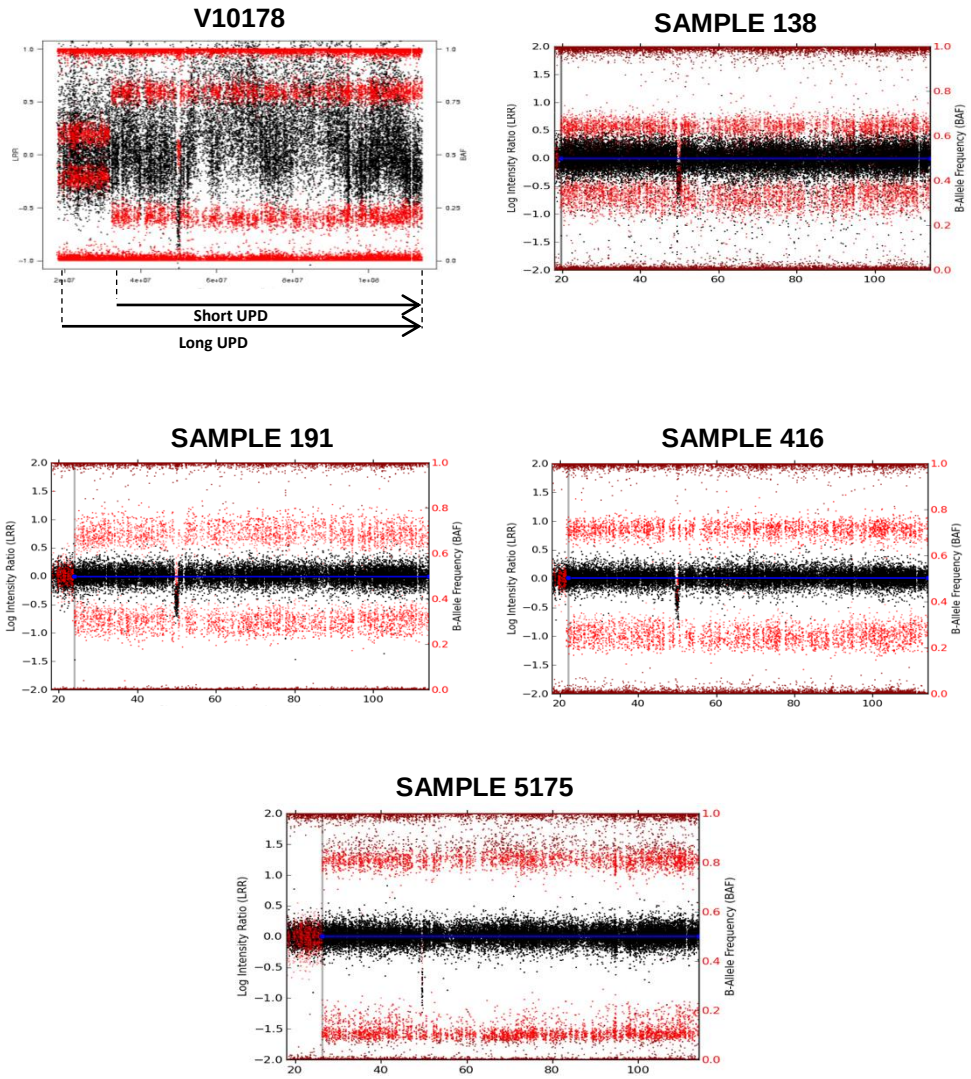
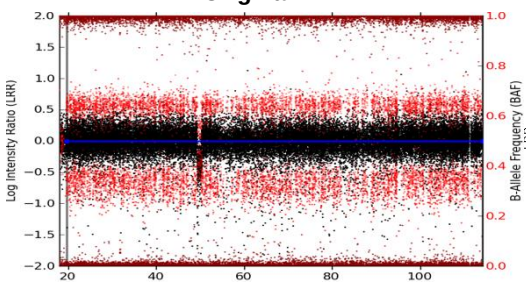
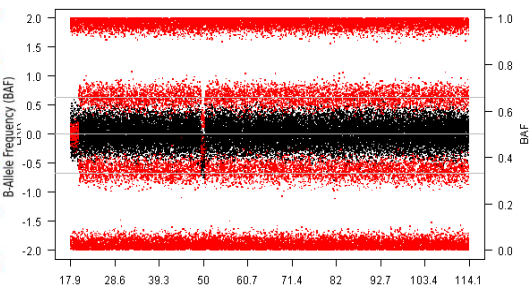


Figure S1. SNP array plots of 8 individuals with mosaic UPD13q and del13q14.3. A total number of 8 out of 39 (20.51%) individuals with a solid tumor or cancer free and harboring a deletion (del) in 13q14.3 had also a uniparental disomy (UPD) 13q. SNP array plots of long arm of chromosome 13 are shown for each individual with a zoom of the deleted region when required. V10627 and V52467 are samples belonging to the same individual, however, V52467 is a sample taken three years after the first collection (V10627). Overall, we detected that Samples 138, 191, 416 and 5175 had one UPD13q and one del13q14.3 whereas Cases 234 and 962 harbored one UPD and two different deletions, V10178 had two different UPDs and one deletion and V10627/V52467 showed two different UPDs and two deletions. SNP array plots coordinates are in hg18 assembly.

Figure S2

A) SAMPLE 138

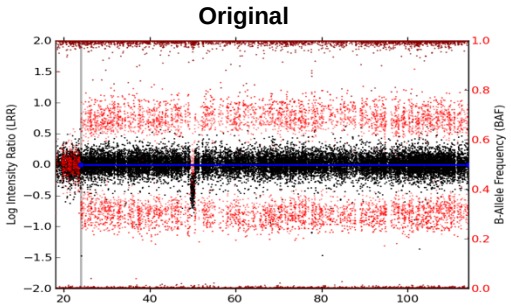


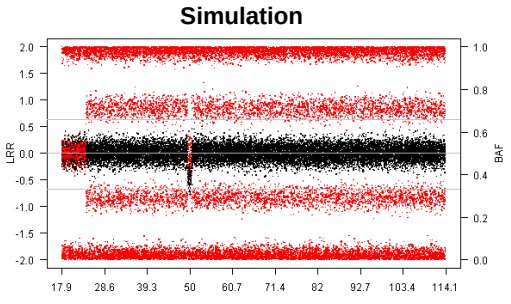


Bdev UPD	LRR Del	Bdev Del
0.16	-0.35	0.07

%UPD-Del Hom	%Del Het	%WT
32.00	17.69	50.31

B) SAMPLE 191

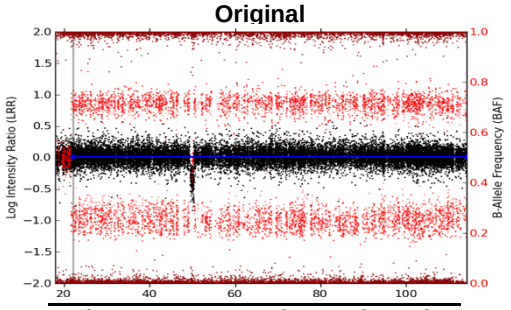


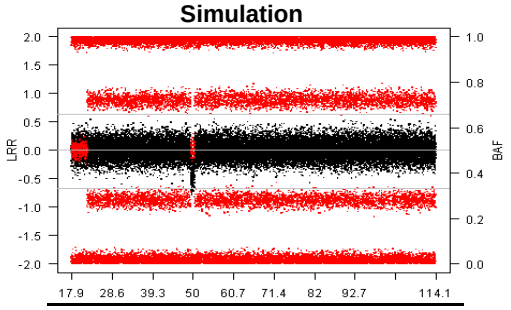


Bdev UPD	LRR Del	Bdev Del
0.21	-0.42	No

%UPD-Del Hom	%Del Het	%WT
42.00	9.48	48.52

C) SAMPLE 416





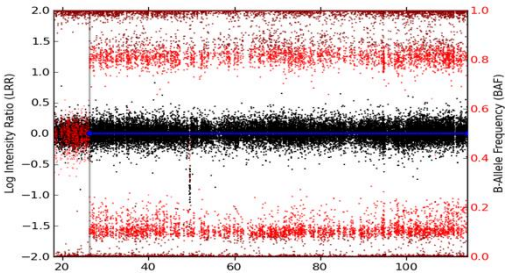
Bdev UPD	LRR Del	Bdev Del
0.22	-0.40	No

%UPD-Del Hom	%Del Het	%WT
44.00	2.24	53.76

Figure S2 *(cont.)*

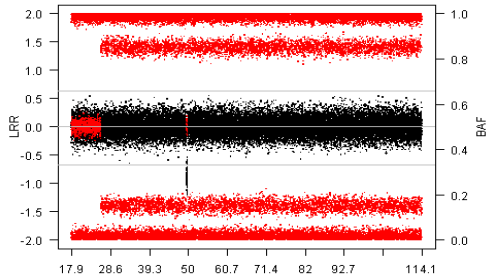
D) **SAMPLE 5175**

Original



Bdev UPD	LRR Del	Bdev Del
0.35	-0.82	No

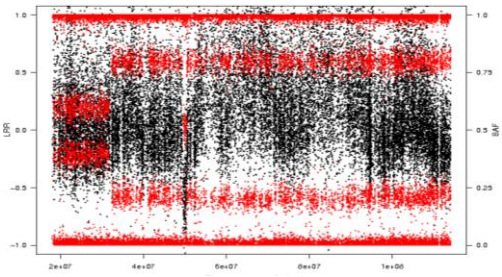
Simulation



%UPD-Del Hom	%Del Het	%WT
70.00	1.54	28.46

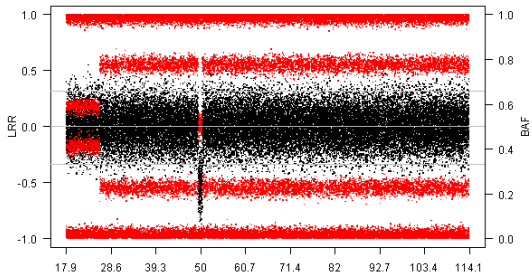
E) **V10178**

Original



Bdev UPD1	Bdev UPD2	LRR Del	Bdev Del
0.28	0.09	-0.50	No

Simulation

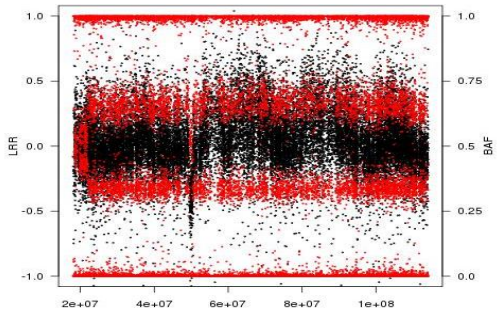


%UPD Short-Del Hom	%UPD Long	%Del Het	%WT
37.00	18.00	31.53	13.47

Figure S2 *(cont.)*

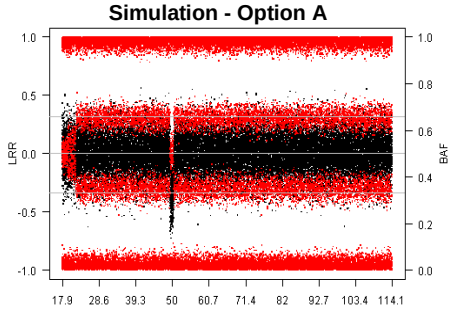
F) **V10627**

Original

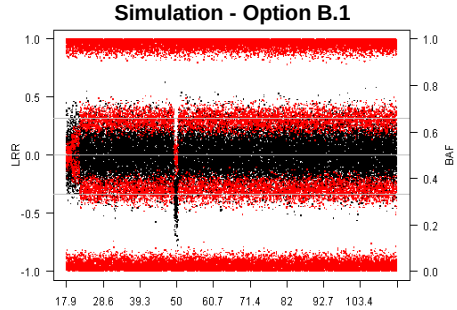


SNP ARRAY DATA					
Bdev UPD1	**Bdev UPD2**	**LRR Del1**	**Bdev Del1**	**LRR Del2**	**Bdev Del2**
0.15	0.06	-0.36	No	-0.31	No
OPTION A					
%UPD Short-Del Long Hom	**%UPD Long-Del Long Hom**	**%Del Short Het**	**%Del Long Het**	**%WT**	
17.84	12.50	8.88	13.20	47.57	
OPTION B.1					
%UPD Short-Del Long Hom	**%UPD Long**	**%Del Short-Long (2nd H)**	**%Del Long Het**	**%WT**	
17.84	12.50	8.88	29.32	31.46	
OPTION B.2					
%UPD Long-Del Long Hom	**%UPD Short**	**%Del Short-Long (2nd H)**	**%Del Long Het**	**%WT**	
12.50	17.84	8.88	40.0	20.78	

Simulation - Option A



Simulation - Option B.1



Simulation - Option B.2

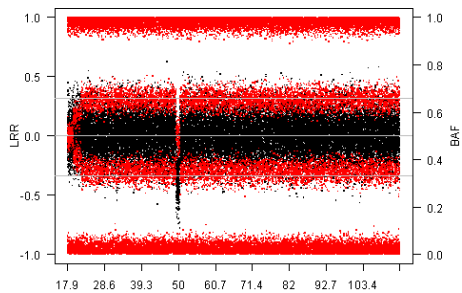
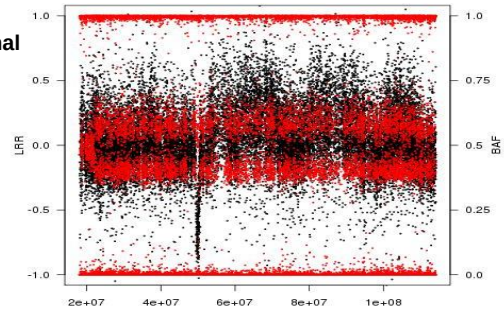


Figure S2 *(cont.)*

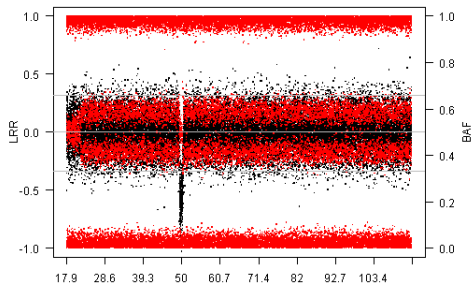
G) **V52467**

Original

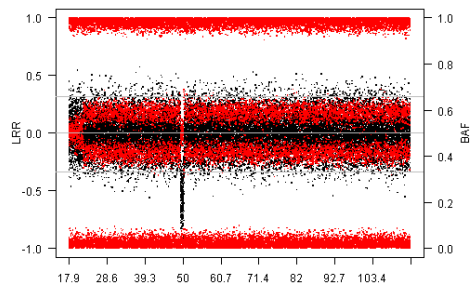


SNP ARRAY DATA					
Bdev UPD1	**Bdev UPD2**	**LRR Del1**	**Bdev Del1**	**LRR Del2**	**Bdev Del2**
0.09	0.04	-0.56	No	-0.30	0.19
OPTION A					
%UPD Short-Del Long Hom	**%UPD Long-Del Long Hom**	**%Del Short Het**	**%Del Long Het**	**%WT**	
10.50	7.06	40.52	37.60	4.32	
OPTION B.1					
%UPD Short-Del Long Hom	**%UPD Long**	**%Del Short-Long (2nd H)**	**%Del Long Het**	**%WT**	
10.50	7.06	40.52	11.20	30.72	
OPTION B.2					
%UPD Long-Del Long Hom	**%UPD Short**	**%Del Short-Long (2nd H)**	**%Del Long Het**	**%WT**	
7.06	10.50	40.52	18.08	23.84	

Simulation - Option A



Simulation - Option B.1



Simulation - Option B.2

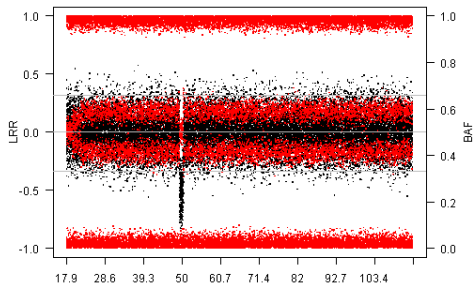
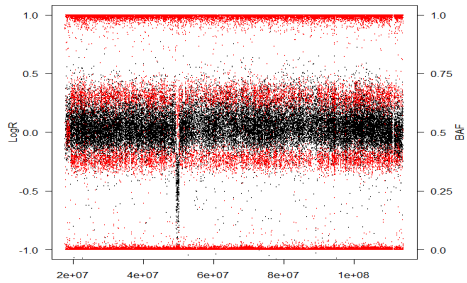


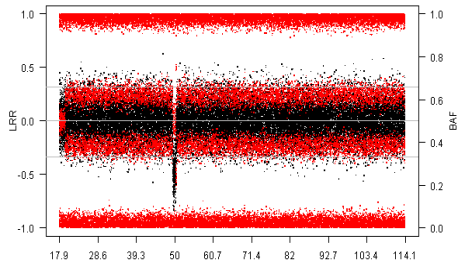
Figure S2 *(cont.)*

H) CASE 962

Original



Simulation

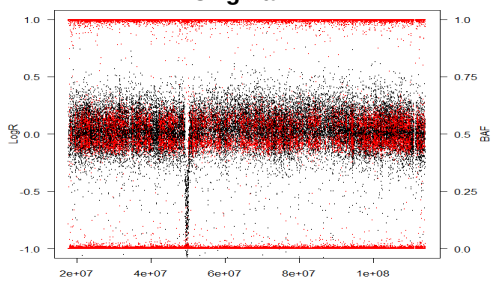


Bdev UPD	LRR Del1	Bdev Del1	LRR Del2	Bdev Del2
0.12	-0.5	No	-0.12	0.10

%UPD alone	%Del Short Het	%Del Hom (2nd hit)	%WT
24.00	39.63	32.95	3.42

I) CASE 234

Original

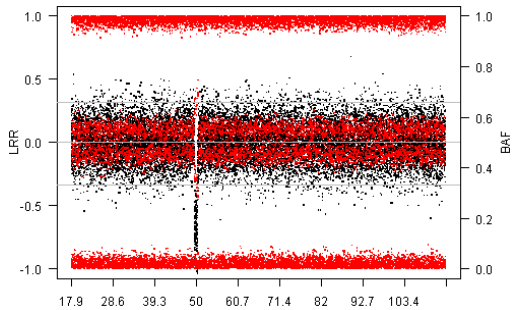


SNP ARRAY DATA				
Bdev UPD1	LRR Del1	Bdev Del1	LRR Del2	Bdev Del1
0.05	-0.67	NA (LOH)	-0.16	0.07

OPTION A			
%UPD alone	%Del Short Het	%Del Hom (2nd H)	%WT
9.92	40.27	43.33	6.48

OPTION B			
%UPD-Del Short Hom	%Del Short Het	%Del Hom (2nd H)	%WT
9.92	20.43	43.33	26.32

Simulation - Option A



Simulation - Option B

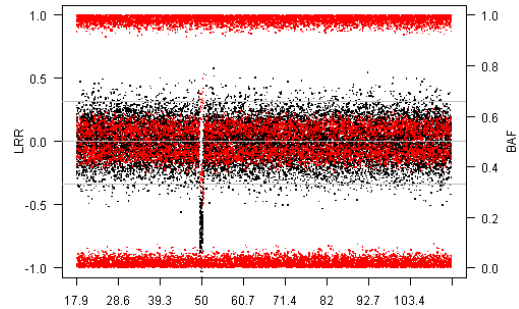


Figure S2. Simulated SNP array plots very similar to original plots were obtained in all tested samples. In each panel (one panel per sample), we included a picture corresponding to the original plot, a table with SNP array parameters (LRR, Bdev) used in the analysis, a table with estimated cell fractions percentages and the plot obtained with the simulation. **A to E and H)** For every scenario tested in all samples, we were able to reproduce the original SNP array plot indicating that the way in which we link all rearrangements and the cellularities estimated for each cell fraction were feasible in these 6 samples. **F,G,I)** Note that we show more than one simulated plot since, as explained in the main text, we had more than one feasible possibility to test. In case of F and G panels, simulations did not give more strength to any of the tested options versus the other one. Regarding **I** panel, this also shows almost identical simulated plots for two different supposals.

An important clarification regarding table nomenclature: (I) "del1" is referring to deleted segment in the plot with highest level of mosaicism (higher LRR value) whereas "del2" is referring to the deleted segment in the plot with lower level of mosaicism (lower LRR value). The same reasoning is valid for "UPD1" vs "UPD2". (II) "del short" refers to a deletion corresponding to the segment del1 in the plot whereas "del long" corresponds to a deletion formed by del1 segment plus del2 segment. The same reasoning is valid for "UPD short" vs "UPD long".

Plot coordinates are in the hg18 assembly. Bdev: B deviation; UPD: Uniparental Disomy; LRR: Log R Ratio; Del: Deletion; Hom: Homozygous; Het: Heterozygous; WT: wild-type; 2 nd H: second hit; %: percentage.

Figure S3

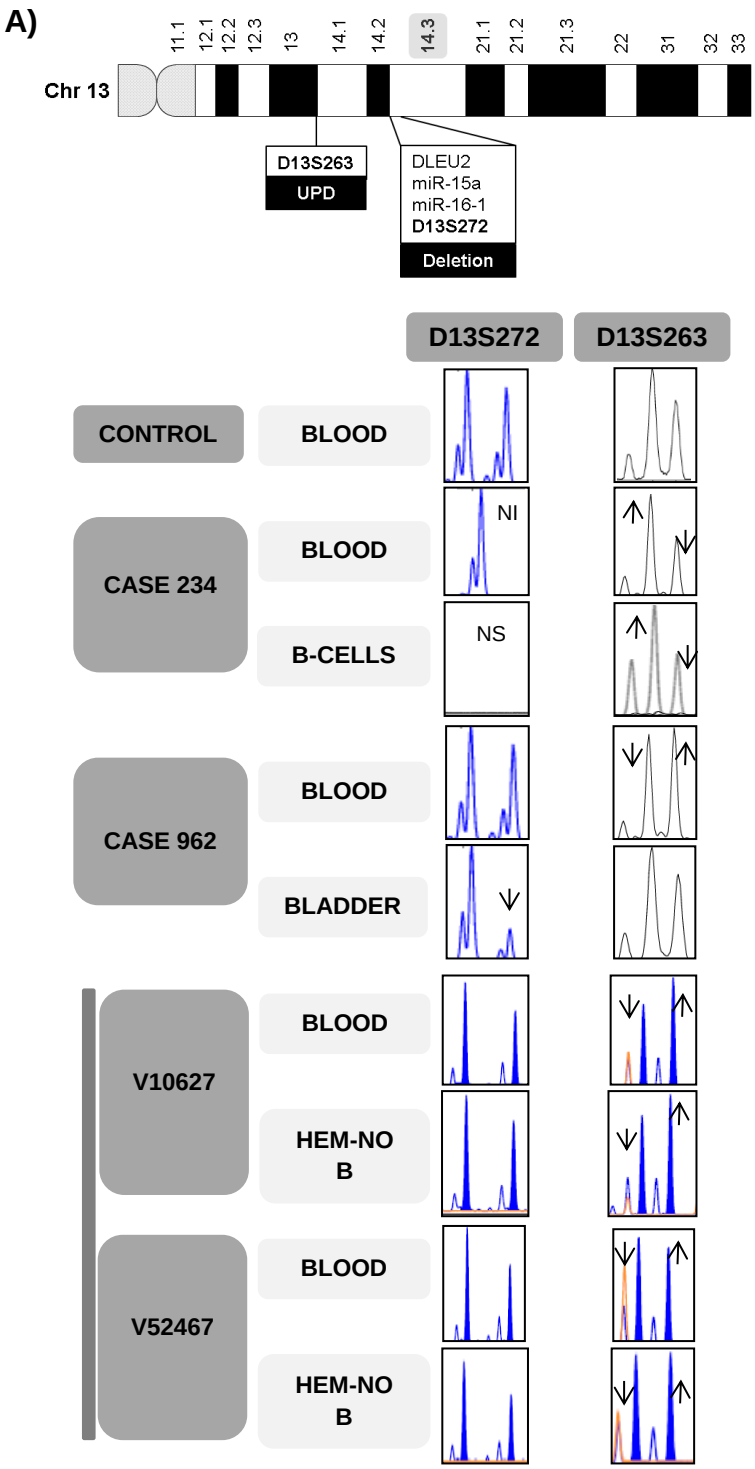

A)
Chr 13
11.1
12.1
12.2
12.3
13
14.1
14.2
14.3
21.1
21.2
21.3
22
31
32
33
D13S263
UPD
DLEU2
miR-15a
miR-16-1
D13S272
Deletion
D13S272
D13S263
CONTROL
BLOOD
CASE 234
BLOOD
NI
B-CELLS
NS
CASE 962
BLOOD
BLADDER
V10627
BLOOD
HEM-NO B
V52467
BLOOD
HEM-NO B

Figure S3 *(cont.)*

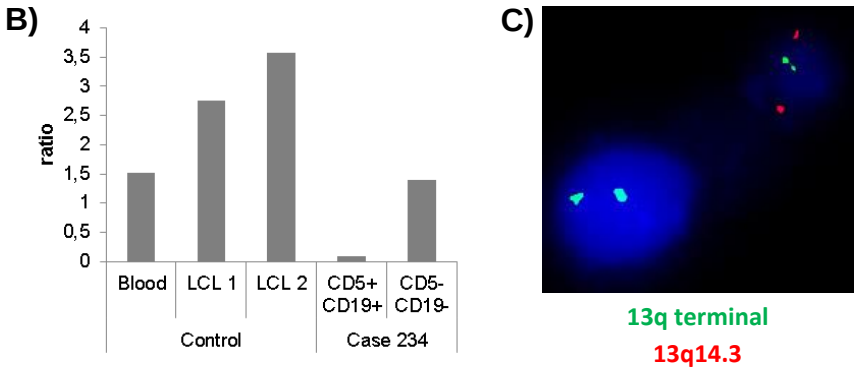


Figure S3. Results of experimental assays. In 4 out of 9 samples (Case 234, V10627, V52467 and V10178), B-cells (CD5+, CD19+) were separated from the rest of hematologic cells and DNA was extracted from each cell fraction to perform microsatellite analysis together with total blood DNA. In Case 962 (a bladder cancer patient), DNA coming from both peripheral blood and bladder was also available to be analyzed by microsatellite approach. **A)** Microsatellites D13S263 (within UPD region) and D13S272 (within the deleted region) were selected for the study. The expected pattern for both microsatellites is shown in a control blood sample. Arrows present along the picture indicate those cases with an allelic imbalance. Regarding Case 234 we detected, as expected, the UPD13q in total blood and an homozygous region overlapping with the deleted 13q14.3 region. In B-cells, we also observed UPD13q and no D13S272 signalling (no associated to technique failure). Hematologic non B-cells fraction was almost non-existent and not enough to obtain results from microsatellite assay. In Case 962, we observed an UPD13q and a rescue of the LOH secondary to del13q14.3 in peripheral blood whereas cells from bladder only harboured del13q14.3. In samples V10627 and V52467 (belonging to the same subject), we detected again an UPD13q and a rescue of the LOH resulting from del13q14.3 both in total blood and in hematologic non B-cell fraction. Here, microsatellite assay in B-cells fraction did not work in these two samples because of technical problems. D13S272 and D13S263 microsatellites were no informative for sample V10178 (data not shown). **B)** Quantitative PCR was performed on RNA isolated from both B and non-B cells of Case 234 to measure *DLEU2* expression. We detected almost null levels of *DLEU2* expression in B-cells fraction, in agreement with microsatellite results, and normal expression in hematologic non B-cells compared to a control samples. The Y axis of the graph shows a ratio which indicates the level of *DLEU2* expression relativized to a control gene (*AGPAT1*) and also to a control sample (the lymphoblastoid cell line "LCL1" for B-cells fraction and the "blood" for hematologic non B-cells fraction). **C)** FISH assay revealed that lymphocyte fraction of peripheral blood in V52467 sample was composed by 88% of cells homozygous for 13q14.3 deletion, 6% of heterozygous deleted cells and 6% of non-deleted cells (WT and UPD cells).

Chr: chromosome; UPD: uniparental disomy; HEM-NO B: hematologic non B-cells; NI: No informative; NS: no signalling; LCL: Lymphoblastoid cell line.

Figure S4

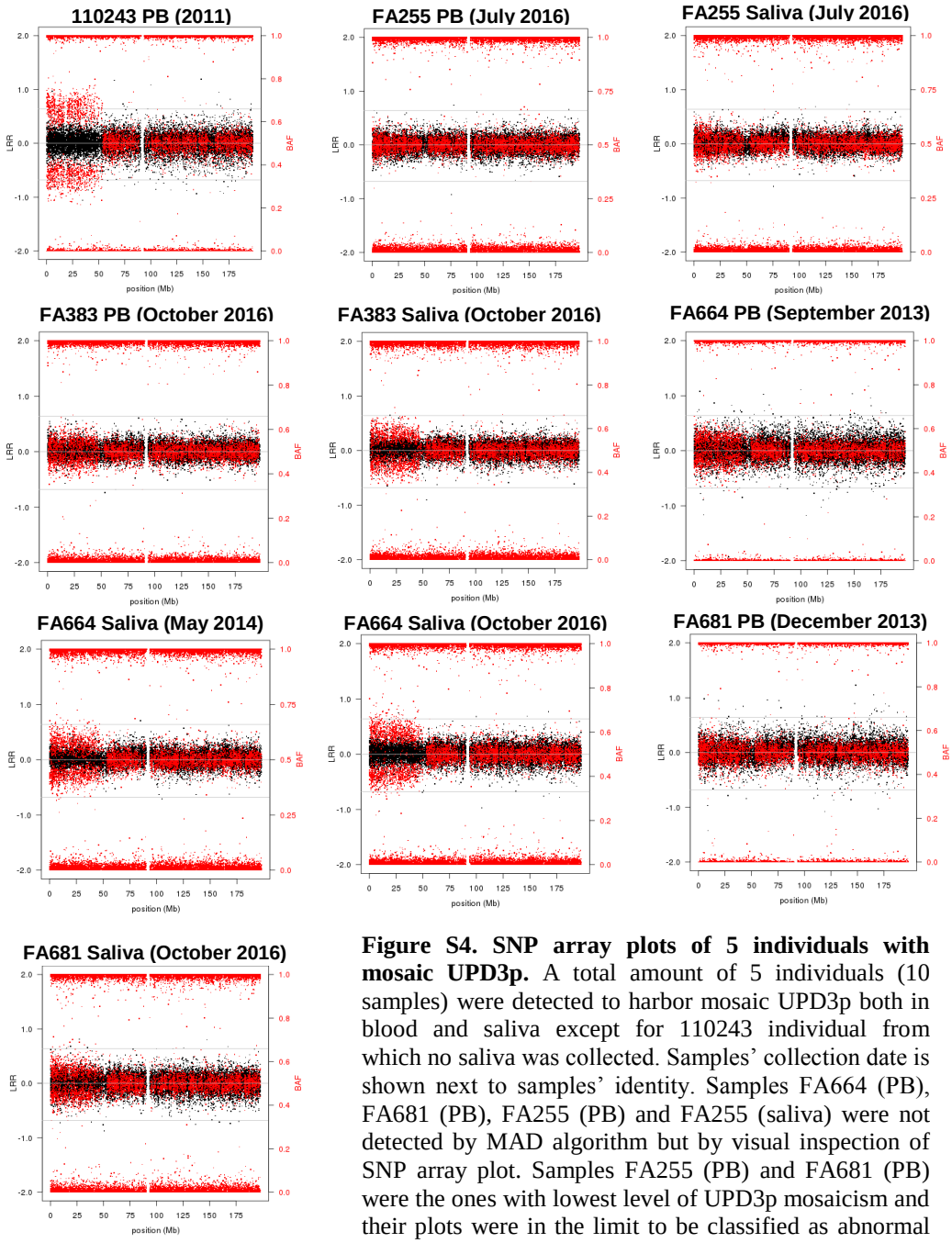


Figure S4. SNP array plots of 5 individuals with mosaic UPD3p. A total amount of 5 individuals (10 samples) were detected to harbor mosaic UPD3p both in blood and saliva except for 110243 individual from which no saliva was collected. Samples' collection date is shown next to samples' identity. Samples FA664 (PB), FA681 (PB), FA255 (PB) and FA255 (saliva) were not detected by MAD algorithm but by visual inspection of SNP array plot. Samples FA255 (PB) and FA681 (PB) were the ones with lowest level of UPD3p mosaicism and their plots were in the limit to be classified as abnormal plots. PB: peripheral blood. Coordinates are in hg19 assembly.

Figure S5

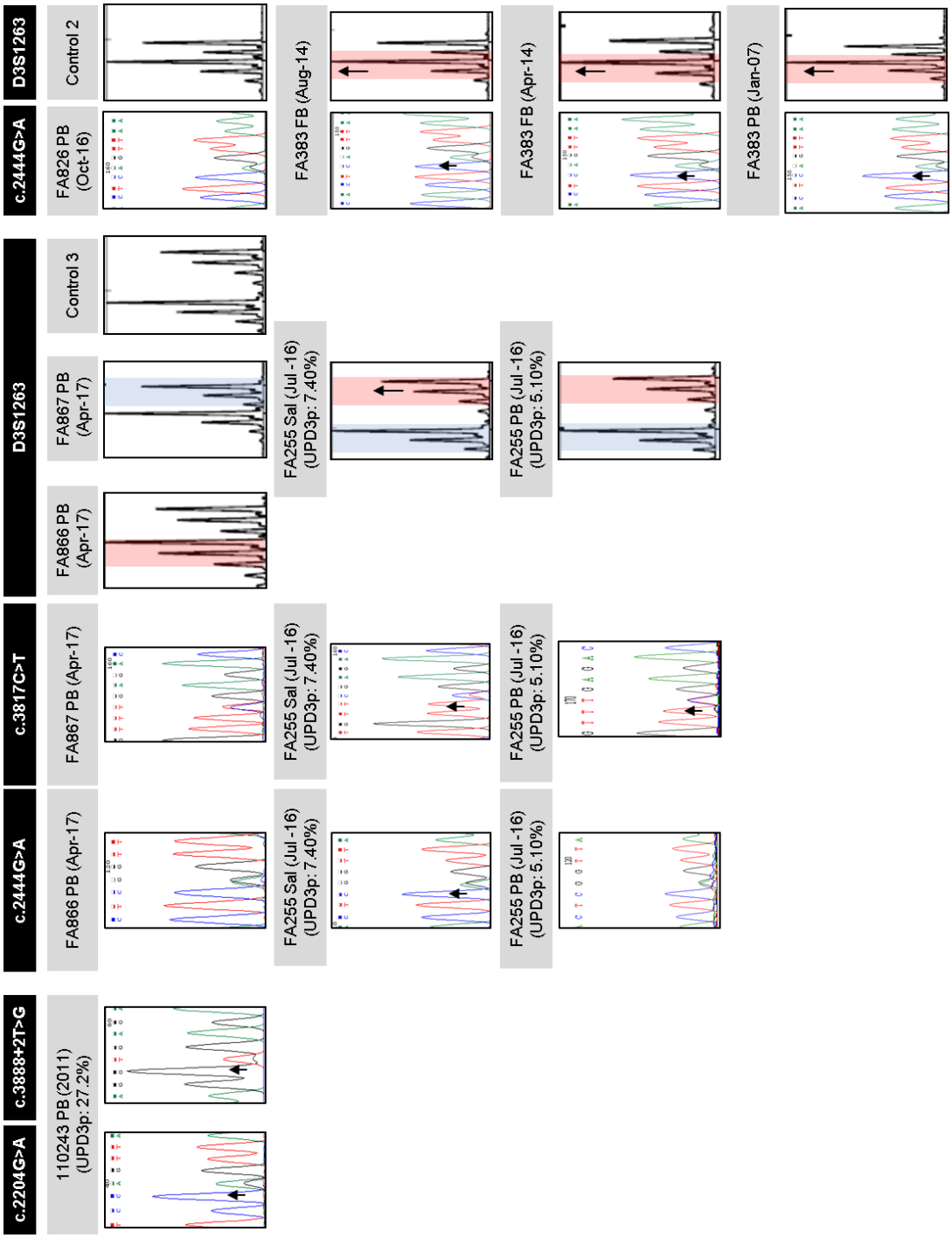

c.2204G>A
c.3888+2T>G
c.2444G>A
c.3817C>T
c.2444G>A
D3S1263
D3S1263
c.2444G>A
D3S1263
110243 PB (2011) (UPD3p: 27.2%)
FA866 PB (Apr-17)
FA867 PB (Apr-17)
FA866 PB (Apr-17)
FA867 PB (Apr-17)
Control 3
FA826 PB (Oct-16)
Control 2
FA255 Sal (Jul-16) (UPD3p: 7.40%)
FA255 Sal (Jul-16) (UPD3p: 7.40%)
FA255 Sal (Jul-16) (UPD3p: 7.40%)
FA383 FB (Aug-14)
FA255 PB (Jul-16) (UPD3p: 5.10%)
FA255 PB (Jul-16) (UPD3p: 5.10%)
FA255 PB (Jul-16) (UPD3p: 5.10%)
FA383 FB (Apr-14)
FA383 PB (Jan-07)

Figure S5 *(cont.)*

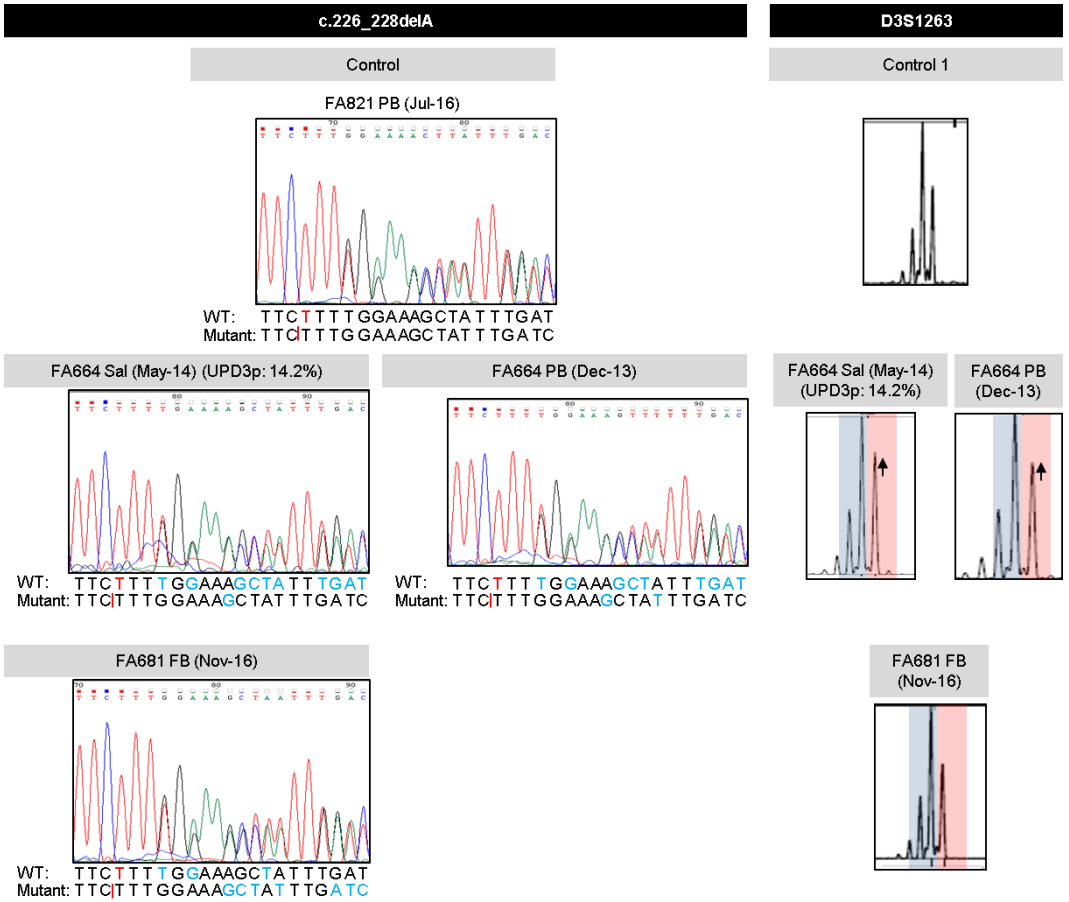


Figure S5. Sanger sequencing and microsatellite analysis of peripheral blood, saliva and fibroblast samples from 5 FA individuals with mosaic UPD3p. Sanger sequencing and microsatellite analysis were performed to study the relative amount of WT versus mutant allele in one or both *FANCD2* mutations detected in all 5 FA individuals with mosaic UPD3p observed by SNP array. Here we present those results not shown in **figure 2** of the main text. Regarding 110243 individual, we had no available parents' samples and we only could study *FANCD2* mutations in peripheral blood DNA of the patient obtained in 2011. Despite of not having parents' results as controls, we could observe by Sanger sequencing that the apparently most benign mutation (c.2204G>A, missense) was being selected by the UPD3p against the most theoretically truncating one (c.3888+2T>G, exon skipping). In case of FA255 individual, Sanger sequencing and microsatellite analysis again revealed that the maternal (pink shadowed) allele carrying the *a priori* less truncating mutation (c.2204G>A, missense) together with the WT allele for the supposing most invalidating mutation (c.3817C>T, nonsense) was selected by the mosaic UPD3p to the detriment of

the paternal (blue shadowed) allele carrying the most invalidating mutation (c.3817C>T, nonsense) and the WT allele for the most benign one (c.2204G>A, missense). These results were clear in saliva sample where the percentage of cells with UPD was a little bit higher whereas in blood sample the evidence of c.2204G>A mutation selection was slight due to UPD's low mosaicism level. In FA383 patient, apart from analyzing peripheral blood and saliva samples harboring mosaic UPD3p from 2016 (shown in **figure 2**), we analyzed other samples previously obtained (two fibroblast (FB) samples from 2014 and one peripheral blood (PB) sample from 2007). Here we observed the same tendency of selection of the maternal (pink shadowed) allele carrying the most benign mutation (c.2444G>A) in FB and PB samples suggesting that UPD3p selecting the *a priori* less truncating mutation was already present at that time and in tissues with different embryonic origin. Finally, in FA664/681 family, we also detected the UPD3p selecting the maternal (pink shadowed) allele WT for c.226_228delA mutation in saliva and peripheral blood of FA664 obtained on May of 2014 and December of 2013 respectively in agreement with the observed in saliva and blood taken on October 2016 and September 2013 respectively (**figure 2**). The analysis of FA681 fibroblasts (obtained in 2016) showed an absence of UPD3p in that tissue. Numerical quantification of Sanger sequencing and microsatellite peaks' height is shown in **table S5.** Estimated mosaicism levels of UPD3p detected by SNP array are indicated in brackets in the corresponding samples. WT: wild-type; PB: peripheral blood; FB: fibroblast sample; Sal: saliva; Jul-16: July 2016; Oct-16: October 2016; Aug-14: August 2014; Apr-14; April 2014; Jan-07: January 2007; May-14: May 2014; Dec-13: December 2013; Nov-16: November 2016.

Table S1

Location	Sequence 5'-3'	Mutation	Sample
DLEU2 Exon 2	Fw: TGTAGCAGAGAACCAATTCTGG Rv: GTAATTATCCACCTTGGTAAAAG	-	Case 234
JAK2 Exon 12	Fw: GGGTTTCCTCAGAACGTTGA Rv: TCATTGCTTTCCTTTTTCACAA	c.1849G>T (Missense)	V38104 (PB)
FANCD2 Exon 4	Fw: ACTTGGGTTTTTAGAGAAGGAAAAC Rv: TCCAAATTCTAGAGTTATGCTGGC	c226_228delA (frameshift)	FA664 (PB, Sal) FA681(PB, Sal) FA821 (PB)
FANCD2 Exon 24	Fw: CAGATTGGTGTCTCCGCTGT Rv: ATTCCTCCCCACATACACCA	c.2204G>A (missense)	110243 (PB)
FANCD2 Exon 26	Fw: AGACATCTCTCAGCTCTGGATAA Rv: GCTTTATTGCCTCACTACTGGC	c.2444G>A (missense)	FA383 (PB, Sal) FA826 (PB) FA255 (PB, Sal) FA866 (PB)
FANCD2 Exon 38	Fw: TGGATGCACTGGTTGCTACA Rv: AGGATGGGTAGGGGGACAAA	c.3817C>T (nonsense)	FA255 (PB, Sal) FA867 (PB)
FANCD2 Intron 39-40	Fw: AAGAGGCTGGAGTGCTCAAA Rv: CATCCATTGCCTTCCCTAAA	c.3888+2T>G (exon 40 skipping)	110243 (PB)

Table S1. Primers used for qPCR and Sanger sequencing analysis. Gene and exon/intron location of *primers* as well as their specific sequence is indicated in the first and second column respectively. Mutations' notation (column 3) and samples analyzed in each case (column 4) are also shown. PB: peripheral blood; Sal: Saliva.

Table S2

		SNP ARRAY DATA						
		UPD13q						
Sample_ID	Study	UPD1_Start SNP probe	UPD1_End SNP probe	UPD1_LRR	UPD1_Bdev	UPD2_Start SNP probe	UPD2_End SNP probe	UPD2_LRR
Case 234	Rodríguez-Santiago, B et al.	19058717	115109878	0	0,05	-	-	-
Case 962	Rodríguez-Santiago, B et al.	20656439	115109878	0	0,12	-	-	-
V10627	EGCUT	23037704	115109878	0	0,15	20655065	23037704	0
V52467	EGCUT	23037704	115109878	0	0,09	20655065	23037704	0
V10178 (+)	EGCUT	27102000*	115103529	0	0,28	19058717	27102000*	0
Sample 191 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	25084316	115103150	0	0,21	-	-	-
Sample 138 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	20700859	115105020	0	0,16	-	-	-
Sample 416 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	23228060	115105020	0	0,22	-	-	-
Sample 5175 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	27102000*	115109878	0	0,35	-	-	-

(continued)

		SNP ARRAY DATA							
		UPD13q	DELETION 13q14.3						
Sample_ID	Study	UPD2_Bdev	Del1_Start SNP probe	Del1_End SNP probe	Del1_LRR	AB_signal	BAF_split	Del1_Bdev	Del2_Start SNP probe
Case 234	Rodríguez-Santiago, B et al.	-	50530603	51478716	-0,67	No (LOH)	-	-	51478716
Case 962	Rodríguez-Santiago, B et al.	-	50641176	51542209	-0,50	Yes	No	-	51542209
V10627	EGCUT	0,06	50497033	51383086	-0,36	Yes	No	-	51383086
V52467	EGCUT	0,04	50497033	51383086	-0,56	Yes	No	-	51383086
V10178 (+)	EGCUT	0,09	50601999*	51551999*	-0,50	Yes	No	-	-
Sample 191 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	-	50601999*	51551999*	-0,42	Yes	No	-	-
Sample 138 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	-	50601999*	51551999*	-0,35	Yes	Yes	0,07	-
Sample 416 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	-	50601999*	51551999*	-0,40	Yes	No	-	-
Sample 5175 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	-	50601999*	50901999*	-0,82	Yes	No	-	-

(continued)

		SNP ARRAY DATA										
		DELETION 13q14.3					OTHER EVENTS					
Sample_ID	Study	Del2_End SNP probe	Del2_LRR	AB_signal	BAF_split	Del2_Bdev	Event	Chr	Start SNP probe	End SNP probe	LRR	Bdev
Case 234	Rodríguez-Santiago, B et al.	51850888**	-0,16	Yes	Yes	0,07	-	-	-	-	-	-
Case 962	Rodríguez-Santiago, B et al.	51679044**	-0,12	Yes	Yes	0,10	UPD	17	3587	19657531	0	0,13
V10627	EGCUT	51563406	-0,31	Yes	No	-	Del	2	36801999	38521908	-0,15	0,15
V52467	EGCUT	51563406	-0,30	Yes	Yes	0,19	Del	2	36801999	38521908	-0,21	0,20
V10178 (+)	EGCUT	-	-	-	-	-	UPD	3	61495	8572373	0	0,09
							UPD	17	8547	11531649	0	0,11
Sample 191 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	-	-	-	-	-	-	-	-	-	-	-
Sample 138 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	-	-	-	-	-	-	-	-	-	-	-
Sample 416 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	-	-	-	-	-	-	-	-	-	-	-
Sample 5175 (+)	Jacobs, K.B. et al.& Laurie, C.C et al.	-	-	-	-	-	-	-	-	-	-	-

Table S2. SNP array data for mosaic UPD13q and del13q14.3 in eight non-leukemia individuals. In this table, we show data related to SNP array analysis of 8 non-leukemia individuals (samples V10627 and V52467 belong to the same subject but they were obtained with 3 years of difference). Columns A and B indicate samples' identity and the study to which they belong to. From C to J columns, SNP array data related to UPD13q is shown. Concretely, we describe the start and end coordinates (the first and last SNP probe included in the event respectively), the LogR ratio (LRR) and B-deviation (Bdev). Note that, for samples with two different mosaic UPD levels (V10627, V52467, V10178), all these information is shown for each segment. Regarding 13q14.3 deletion (columns K to V), we indicate the starting and ending coordinates (the first and last SNP probe included in the event respectively) as well as the LRR. We also describe the detection or not of BAF signaling in that region as well as the presence or not of BAF split, indicating Bdev value when required. Note that there were 4 samples (Case 234, Case 962, V10627and V52467) in which we detected two different deletions at 13q14.3 with different breakpoints, so, we have indicated the parameters above mentioned for both deletions. Finally, we also show extra mosaic rearrangements affecting chromosomes different to chromosome 13 that were detected in some cases (Case 962, V10627, V52467 and V10178). We indicate the kind of event, the chromosome affected, starting and ending coordinates, LRR and Bdev. All coordinates are expressed in hg19 assembly.
Note that UPD1 is referring to the DNA segment affected by the UPD with highest level of mosaicism whereas UPD2 is the one with lowest percentage in the SNP array plot. The same criterion has been used to define Del1 and Del2.

(+) indicates those samples from whom we did not have SNP array raw data and most parameters were directly inferred from the plot. *indicates those coordinates that are approximated derived from the plot. **indicates those coordinates that cannot be precisely defined because of low percentage of mosaicism in that altered region. - indicates those unfilled fields because it made no sense due to information mentioned in other columns. ID: identity; EGCUT: Estonian Gene Expression Cohort; UPD: Uniparental Disomy; Del: deletion; LRR: Log R Ratio; Bdev: B-deviation; BAF: B Allele Frequency; Chr: chromosome; LOH: Lost of Heterozygosity

Table S3

PANEL A									
	SNP array data								
Sample_ID	Chr	UPD1 Bdev	UPD2 Bdev	Del LRR	Del Bdev	UPD Short Chr	UPD Long Chr	Del Chr	Total No deleted Chr
Sample 138 (+)	13	0,16	-	-0,35	0,07	64,00	-	81,69	118,31
Sample 191 (+)	13	0,21	-	-0,42	No	84,00	-	93,48	106,52
Sample 416 (+)	13	0,22	-	-0,40	No	88,00	-	90,24	109,76
Sample 5175 (+)	13	0,35	-	-0,82	No	140,00	-	141,54	58,46
V10178 (+)	13	0,28	0,09	-0,50	No	74,00	36,00	105,53	94,47
	3	0,09	-	-	-	-	-	-	-
	17	0,11	-	-	-	-	-	-	-

(continued Panel A)

PANEL A							
	Cellularity						
Sample_ID	% Extra event	%UPD Short-Del Hom	%UPD Long	%Del Het	%WT	Total%	Predicted Bdev in Del	UPD (Conclusion)
Sample 138 (+)	-	32,00	-	17,69	50,31	100,00	0,07	Del hom
Sample 191 (+)	-	42,00	-	9,48	48,52	100,00	No	Del hom
Sample 416 (+)	-	44,00	-	2,24	53,76	100,00	No	Del hom
Sample 5175 (+)	-	70,00	-	1,54	28,46	100,00	No	Del hom
V10178 (+)	-	37,00	18,00	31,53	13,47	100,00	No	Del hom + Rescue
	17,00	-		-	-	-	-	-
	22,00	-		-	-	-	-	-

PANEL B												
	SNP array data											
Sample_ID	Chr	UPD1 Bdev	UPD2 Bdev	Del1 LRR	Del1 Bdev	Del2 LRR	Del2 Bdev	UPD Short Chr	UPD Long Chr	Del Short Chr	Del Long Chr	Total No deleted Chr
V10627	13	0,15	0,06	-0,36	No	-0,31	No	35,68	25,00	8,88	73,88	117,23
	2	-	-	-0,15	0,15	-	-	-	-	41,23	-	-
V52467	13	0,09	0,04	-0,56	No	-0,30	0,19	21,00	14,12	40,52	72,72	86,76
	2	-	-	-0,21	0,20			-	-	54,22		-

OPTION A (2 deletions in different clones; 2 UPDs leading to long deletion homozygosis). Cellularity.											
Sample_ID	Chr	% Extra event	%UPD Short-Del Long Hom	%UPD Long-Del Long Hom	%Del Short Het	%Del Long Het	%WT	Total %	Predicted Bdev in Del1	Predicted Bdev in Del2	UPD (Conclusion)
V10627	13	-	17,84	12,50	8,88	13,20	47,57	100,00	No	No	Del hom+ Del hom
	2	41,23	-	-	-	-	-	-	-	-	-
V52467	13	-	10,50	7,06	40,52	37,60	4,32	100,00	No	0,15	Del hom+ Del hom
	2	54,22	-	-	-	-	-	-	-	-	-
OPTION B.1 (2 deletions in the same clone; UPD long alone; UPD short leading to long deletion homozygosis).Cellularity.											
Sample_ID	Chr	% Extra event	%UPD Short-Del LongHom	%UPD Long	%Del Long + Short (2nd hit)	%Del Long Het	%WT	Total %	Predicted Bdev in Del1	Predicted Bdev in Del2	UPD (Conclusion)
V10627	13	-	17,84	12,50	8,88	29,32	31,46	100,00	No	No	Del hom+ Rescue
	2	41,23	-	-	-	-	-	-	-	-	-
V52467	13	-	10,50	7,06	40,52	11,20	30,72	100,00	No	0,14	Del hom+ Rescue
	2	54,22	-	-	-	-	-	-	-	-	-
OPTION B.2 (2 deletions in the same clone; UPD short alone; UPD long leading to long deletion homozygosis).Cellularity.											
Sample_ID	Chr	% Extra event	%UPD Long-Del LongHom	%UPD Short	%Del Long + Short (2nd hit)	%Del Long Het	%WT	Total %	Predicted Bdev in Del1	Predicted Bdev in Del2	UPD (Conclusion)
V10627	13	-	12,50	17,84	8,88	40,00	20,78	100,00	No	No	Del hom+ Rescue
	2	41,23	-	-	-	-	-	-	-	-	-
V52467	13	-	7,06	10,50	40,52	18,08	23,84	100,00	No	0,15	Del hom+ Rescue
	2	54,22	-	-	-	-	-	-	-	-	-

Table S3 *(cont.)*

PANELC										
	SNP array data									
Sample_ID	Chr	UPD Bdev	Del1 LRR	Del1 Bdev	Del2 LRR	Del2 Bdev	UPD Chr	Del Short Chr	Del Long Chr	Total No deleted Chr
Case 962	13	0,12	-0,50	No	-0,12	0,10	48,00	72,58	32,95	94,47
	17	0,13	-	-	-	-	-	-	-	-
Case 234	13	0,05	-0,67	NA (LOH)	-0,16	0,07	19,84	83,60	43,33	73,07

(continued Panel C)

OPTION A (UPD alone). Cellularity.							
%UPD alone	%Del Short Het	%Del short+ long (2nd hit)	%WT	Total %	Predicted Bdev in Del1	Predicted Bdev in Del2	UPD (Conclusion)
24,00	39,63	32,95	3,42	100,00	0,04	0,24	Rescue
26,00	-	-	-	-	-	-	-
9,92	40,27	43,33	6,48	100,00	0,14	0,10	Rescue
OPTION B (UPD leading to short deletion homozygosis).Cellularity.							
%UPD -Del Short Hom	%Del Short Het	%Del short+ long (2nd hit)	%WT	Total %	Predicted Bdev in Del1	Predicted Bdev in Del2	UPD (Conclusion)
9,92	20,43	43,33	26,32	100,00	0,13	0,10	Del Hom

Table S3. **Estimation of cell fractions' cellularity under the possible supposals for each sample**. **Panel A)** This panel is referred to those 5 samples where we found a unique and conclusive scenario. The first part of the panel, SNP array data, includes SNP array parameters (LRR, Bdev) for each rearrangement and the number of chromosomes affected by each event for all samples. To calculate the number of chromosomes affected by UPD, we calculated UPD cellularity by using the formula *%UPD = 2*Bdev*100* (see methods) and multiplying per two the result since each cell have two chromosomes affected by the UPD. By using the equation $Chromosomes\ with\ deletion = (2 - 2 * e^{1.5*LRR}) * 100$ (see methods), we could calculate the number of deleted chromosomes and, considering this and a total amount of 200 chromosomes, we could determine the number of non-deleted chromosomes. Regarding the second part of the panel, we show estimated percentages related to cellularity of each cell fraction under the situation that we were proving. Cellularity of UPDs (alone or coexisting with homozygous deletion) were calculated by using the formula above mentioned or, what is the same, by dividing per two the number of chromosomes affected by the UPD. Percentages of cells heterozygous for the deletion were calculated by considering that each heterozygous cell have 1 deleted chromosome and the following formula: *%delhet= Chromosomes deleted in heterozygosis = total deleted chromosomes - chromosomes deleted in homozygosis by UPD*. The proportion of WT cells was inferred with the equation *%WTcells = (Non-deleted chromosomes - chromosomes non-deleted in heterozygous cells)/2*. In case of V10178, cellularity of two extra events (UPDs) was calculated also using the above mentioned formula. Finally, we calculated the proportion of non-deleted B alleles versus the total non-deleted chromosomes to infer de Bdev expected in the deleted region. As indicated in the table, we obtained null Bdevs (below 0.05) in all cases without BAF split in the original plot and the same Bdev as the original one in the unique case with BAF split.

Considering this, together with the fact that the total sum of cell fractions was 100% in all samples, we conclude that having the UPD13q acting as a second hit mechanism leading to deletion homozygosis was feasible in all these 5 samples whereas an extra rescue UPD13q was probable in one sample. **Panel B)** This panel contains exactly the same information as the previous one but regarding to two samples belonging to the same individual in which we observed several possible scenarios. As explained in the main text, SNP array data of two these samples were compatible with (option A) having cells heterozygous for each deletion (separately) coexisting with two different cell fractions undergoing one UPD each one acting as a second hit mechanism for the longest deletion, or (option B) having cells heterozygous for the longest deletion (which is the first hit) together with homozygous cells harboring both long and short deletion (2nd hit), cells homozygous for the long deletion through one UPD and cells with the other UPD alone; being possible two different combinations: (B.1) short UPD as a second hit mechanism and long UPD alone, (B.2) one the other way around. The number of chromosomes affected by each rearrangement as well as percentages for each cell fraction was calculated similarly to the panel A. For both samples, option A and B (B.1 and B.2) were possible since total sum of percentages was 100% and predicted Bdev for all deleted regions was the same as expected. **Panel C)** This panel shows the same information as the previous ones but regarding Cases 234 and 962 for which we did not obtain conclusive results. In Case 234, since no information regarding BAF split in the common region between short and long deletion was available due to a LOH, we could consider as possible both Option A (UPD alone) and B (UPD leading short deletion homozygosis). In both options, the total cellularity was 100% and the predicted Bdev for the deleted segment only in the long deletion was similar to the predicted one. In Case 962, the only feasible option was A although we obtained too high Bdev value for region exclusively deleted in the long deletion compared to what was suggested by SNP array data.

An important clarification regarding table nomenclature: (I) “del1” is referring to deleted segment in the plot with highest level of mosaicism (higher LRR value) whereas “del2” is referring to the deleted segment in the plot with lower level of mosaicism (lower LRR value). The same reasoning is valid for “UPD1” vs “UPD2”. (II) Consequently, “del short” refers to a deletion corresponding to the segment del1 in the plot whereas “del long” corresponds to a deletion formed by del1 segment plus del2 segment. The same reasoning is valid for “UPD short” vs “UPD long”.

(+) indicates those samples from whom we did not have SNP array raw data and most parameters were directly inferred from the SNP array plot. ID: identity; Chr: Chromosome; Bdev: B deviation; UPD: Uniparental Disomy; LRR: Log R Ratio; Del: Deletion; Hom: Homozygous; Het: Heterozygous; WT: wild-type; 2nd hit: second hit; %: percentage; NA: not available; LOH: Lost of Heterozygosity; -: empty cells (no calculation was done); del hom: homozygous deletion; green cells: these results are in agreement with the final conclusion; red cells: these results are in conflict with the final conclusion.

Table S4

		SNP ARRAY DATA						
Sample_ID	**Study**	**Karyotype event**	**Start SNP probe (bp)**	**End SNP probe (bp)**	**Size (bp)**	**LRR**	**Bdev**	**% mosai-cism**
110243	FA cohort I (Reina-Castillón, J *et al*.)	UPD3p	pter	53613751	53613750	0,03	0,14	27,2
FA255 (PB)*	FA cohort II	UPD3p	pter	NA (44000000-49000000)	NA (44000000-49000000)	0,01	0,03	5,1
FA255 (Sal)*	FA cohort II	UPD3p	pter	NA (44000000-49000000)	NA (44000000-49000000)	0,02	0,04	7,4
FA383 (PB)	FA cohort II	UPD3p	pter	46605447	46605446	0,03	0,04	8,8
FA383 (Sal)	FA cohort II	UPD3p	pter	46711135	46711134	0,02	0,07	13,4
FA664 (PB)*	FA cohort I (Reina-Castillón, J *et al.*)	UPD3p	pter	49956628	49956627	0,01	0,05	9,6
FA664 (Sal)	FA cohort II	UPD3p	pter	49836707	49836706	0,02	0,10	19,0
FA664 (Sal2)	FA cohort II	UPD3p	pter	49836707	49836706	0,02	0,07	14,2
FA681 (PB)*	FA cohort I (Reina-Castillón, J *et al*.)	UPD3p	pter	NA (49836707)	NA (49836707)	0,01	0,04	6,7
FA681 (Sal)	FA cohort II	UPD3p	pter	49836707	49836706	0,02	0,06	11,6

Table S4. SNP array data for mosaic UPD3p detected in both blood and saliva of 4 FA patients and in blood of 1 additional subject. SNP array data regarding UPD3p detected in each FA sample is shown in this table, including starting and ending coordinates, event's size, LogR ratio (LRR), B-deviation (Bdev) and percentage of mosaicism calculated by using the formula *%UPD = 2*Bdev*100*. Samples marked with a "*" are those in which we detected the UPD3p by visual inspection of SNP array plots and not by MAD algorithm; thus, because of the low mosaicism level, UPD coordinates are not accurate. FA664 and FA681 are siblings. "Sal2" refers to Saliva sample from FA664 obtained in May 2014 whereas "Sal" sample was saliva taken in October 2016. PB: peripheral blood; Sal: saliva; bp: base pairs; NA: not available; pter: p terminal. Coordinates are in hg19 assembly.

Table S5

		D3S1263 A1/A2	WT/Mut 1	WT/Mut 2	WT/Mut 3	WT/Mut 4	WT/Mut 5	WT/Mut 6	WT/Mut 7	WT/Mut 8	WT/Mut 9	WT/Mut 10	WT/Mut 11	WT	Mut	=
	Control 1 D3S1263	1,65	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Control 2 D3S1263	1,30	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Control 3 D3S1263	1,52	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	FA820 PB (Jul-16)	1,35	-	-	-	-	-	-	-	-	-	-	-	-	-	-
c226_228delA	FA821 PB (Jul-16)	NI	0,84	0,58	1,14	1,09	0,80	1,00	1,00	1,00	1,20	1,25	0,75	-	-	-
c226_228delA	FA664 Saliva (Oct-16)*	1,24	1,21	0,89	1,50	1,43	1,11	1,86	1,41	1,55	1,70	1,88	1,10	11	0	0
c226_228delA	FA664 Saliva (May-14)*	1,28	1,11	1,00	1,00	1,57	1,00	1,57	1,19	1,27	1,67	1,71	0,91	10	1	0
c226_228delA	FA664 PB (Dec-13)	1,41	1,21	0,82	1,00	1,19	0,82	1,41	0,89	1,30	1,40	1,35	0,91	9	2	0
c226_228delA	FA664 PB (Sep-13)*	1,39	0,89	0,94	0,80	1,54	1,00	1,36	1,06	1,00	1,58	1,47	0,85	9	1	1
c226_228delA	FA681 Saliva (Oct-16)*	1,38	1,00	0,70	1,33	1,31	1,00	1,38	1,25	1,27	1,40	1,56	0,95	11	0	0
c226_228delA	FA681 FB (Nov-16)	1,61	0,88	0,64	0,84	0,80	0,77	1,10	0,82	1,00	1,13	1,10	0,71	3	7	1
c226_228delA	FA681 PB (Dec-13)*	1,43	0,88	0,67	1,18	1,18	0,91	1,20	1,10	1,11	1,29	1,30	0,80	11	0	0
c.2444G>A	FA826 PB (Oct-16)	1,81	0,63	-	-	-	-	-	-	-	-	-	-	-	-	-
c.2444G>A	FA383 Saliva (Oct-16)*	1,91	0,22	-	-	-	-	-	-	-	-	-	-	-	-	-
c.2444G>A	FA383 PB (Oct-16)*	1,72	0,33	-	-	-	-	-	-	-	-	-	-	-	-	-
c.2444G>A	FA383 FB (Aug-14)	1,42	0,50	-	-	-	-	-	-	-	-	-	-	-	-	-
c.2444G>A	FA383 FB (Apr-14)	1,46	0,43	-	-	-	-	-	-	-	-	-	-	-	-	-
c.2444G>A	FA383 PB (Jan-07)	1,57	0,38	-	-	-	-	-	-	-	-	-	-	-	-	-
c.2444G>A	FA866 PB (Apr-07)	1,46	1,06	-	-	-	-	-	-	-	-	-	-	-	-	-
c.2444G>A	FA255 Saliva (Jul-16)*	1,45	0,78	-	-	-	-	-	-	-	-	-	-	-	-	-
c.2444G>A	FA255 PB (Jul-16)*	1,50	1,08	-	-	-	-	-	-	-	-	-	-	-	-	-
c.3817C>T	FA867 PB (Apr-07)	1,33	0,78	-	-	-	-	-	-	-	-	-	-	-	-	-
c.3817C>T	FA255 Saliva (Jul-16)*	1,45	2,00	-	-	-	-	-	-	-	-	-	-	-	-	-
c.3817C>T	FA255 PB (Jul-16)*	1,50	0,94	-	-	-	-	-	-	-	-	-	-	-	-	-
c.2204G>A	110243 (2011)	-	0,4	-	-	-	-	-	-	-	-	-	-	-	-	-
c.3888+2T>G	110243 (2011)	-	3,8	-	-	-	-	-	-	-	-	-	-	-	-	-

Table S5. Quantification of Sanger sequencing and microsatellite data. This table shows relative peaks' height ratio for both D3S1263 microsatellite and Sanger sequencing results for all tested samples shown in **figures 2** and **S5**. In microsatellite analysis, A1 refers to that allele with lower size and A2 corresponds to the allele with major size. In Sanger sequencing peaks ratio quantification of family FA664/681, red columns show the total count of points out of the 11 tested in which there was a selection of the WT or mutated allele or no selection of none of them. "*" samples are those in which UPD3p was detected by SNP array. PB: peripheral blood; NI: no informative; Jul: July; Oct: October; Dec: December; Sep: September; Nov: November; Aug: August; Apr: April.

CHAPTER 3

Fanconi anemia: evaluation of the medical follow-up and the role of genetic counsellor

J.Reina-Castillón*, E.Carrasco*, N.Gadea, I.Esteban, S.Torres, D.Valcárcel, C.Centeno, C.Bescós, J.Lorente, R.Pujol, J.Surrallés, C.Díaz, J.Balmaña

*Both authors contributed equally to this manuscript

In preparation

Certain characteristics inherent to Fanconi anemia disease as its genetic basis, heritability, chronicity, incurability and the associated cancer and death risks deserve special consideration from health professionals both at the time of disease diagnosis and during its chronic follow-up in order to ensure as much as possible the acceptance of the illness by patients and their relatives and minimize the associated psychological burden. Moreover, facilitating a well-orchestrated medical follow-up will help to the general wellness of these families. This work shows, for the first time in Fanconi anemia field, that the periodic contact of Fanconi anemia families with a genetic counsellor favors the adherence to adult medical follow-up, keeping family and patients well-informed about the disease and the covering, upon a certain extent, of psychological necessities.

Fanconi anemia: evaluation of the medical follow-up and the role of genetic counsellor

J.Reina-Castillón*, E.Carrasco*, N.Gadea, I.Esteban, S.Torres, D.Valcárcel, C.Centeno, C.Bescós, J.Lorente, R.Pujol, J.Surrallés, C.Díaz, J.Balmaña

*Both authors contributed equally to this manuscript

ABSTRACT

Fanconi anemia (FA) is a rare disease both clinically and genetically heterogeneous characterized by congenital defects and increased risk of bone marrow failure (BMF) and hematologic and solid tumors that requires multidisciplinary medical care. We have evaluated (I) the quality of FA adult follow-up in Hospital Vall d'Hebron (HVH) and (II) the role of a genetic counsellor in a FA referral unit. We also have improved a clinical data registry used regularly in FA adult follow-up (III). We took advantage of 21 FA patients' medical record to study FA adult follow-up (aim I). We evaluated 50 individuals (patients and parents) through a questionnaire to assess FA knowledge, psychological impact and reasons for adhesion or not to medical care in families regularly attended by genetic counsellors (or similar specialists) (aim II). Finally, we performed an update of a clinical data registry model (aim III). Regarding adult FA medical controls, an underfollowing in Hematology and overfollowing in Gynecology were detected together with low frequency of dentist follow-up, Bone Marrow Aspirates (BMA), Human Papilloma Virus (HPV) test and gynecological hormonal follow-up. Specialists' coordination in the same center, receiving visits reminders and scheduling controls in the same day increase adherence to follow-up. Information and psychological support from genetic counsellors contribute to satisfactory FA knowledge both in patients and parents although some concepts need to be reinforced. Mothers are more worried about FA manifestations and they badly deal with conclusive molecular diagnosis. Finally, we present an actualized clinical data register to be used in FA regular follow-up. Coordination of specialists and visits favors global accomplishment of Spanish guidelines recommendations for FA diagnosis and management although some variations were detected in Hematology and Gynecology controls' frequency and some tests. Genetic counsellors contribute to a satisfactory FA knowledge by keeping families informed and providing psychological support to them.

INTRODUCTION

Fanconi anemia (FA) is a rare disease with a prevalence of 1-9/1000000 characterized by a triad of clinical manifestations: (I) congenital defects and physical malformations, (II) risk of bone marrow failure (BMF) and (III) increased probability of both hematologic and solid tumors (1). Physical abnormalities affect to ~75% of FA patients and include a wide range of congenital defects and physical abnormalities as pre and/or postnatal developmental delay, skeletal malformations, microcephaly, some typical facial features as triangular face, microphtalmia and visual difficulties, ears abnormal shape and possible hearing loss, skin hyper or hypopigmentation, visceral malformations, reduced or absent fertility and learning problems even intellectual disability (1–4). Regarding to BMF, it appears around 7 years old (yo) and affects 98% of patients at the age of 40. Bone marrow dysfunction can cause a variety of health problems including hematologic pre-cancerous and cancerous condition as myelodysplastic syndrome (MDS) and acute myelogenous leukemia (AML) respectively among others (1,5). Finally, FA patients have an increased risk of hematological, as above-mentioned, and solid tumors where the most common ones are Head and Neck Squamous Cell Carcinoma (HNSCC) and gynecological cancers (mainly vulvar and cervical), whereas the less frequent ones are gastrointestinal cancers (in esophagus and in liver), breast, kidney and brain cancers (3–5).

FA is both clinically and genetically heterogeneous since 18 causative genes (or complementation groups) and up to 13 FA-associated genes have been described until now. FA causative genes are subdivided in 15 bona fide *FANC* genes (*A* (60-70%), *B/FAAP95, C, D1/BRCA2, D2, E, F, G/XRCC9, I/KIAA1794, J/BRIP1, L/PHF9, N/PALB2, P/SLX4, Q/ERCC4 and T/UBE2T*) and 3 FA-like genes (*O/RAD51C*, *R/RAD51* and *S/BRCA1*) (1,5–7). FA genes codify for proteins that cooperate in the FA/BRCA pathway responsible of correcting interstrand DNA cross-linkings (ICLs). When this system does not work, increased number of DNA breakings in hematopoietic cells is detected after their exposure to ICL-inducing agents as diepoxybutane (DEB) or mitomycin C. In fact, this is known as chromosome fragility tests and is the first molecular diagnostic test usually performed in patients suspicious of having FA (3,4,10,11). Mutations in FA genes are inherited in an autosomic recessive manner

except those located in *FANCB* which are X-linked (8) and those in *FANCR/RAD51* gene which are probably autosomic dominant since, up to now, only two cases with heterozygous *de novo* pathogenic variants have been reported (9). Considering the autosome recessive inheritance in most of genes, the frequency of carriers in general population is 1/300 and <1/100 in some populations with founder effect (1,5). There is no evidence of increased risk of disease among FA carriers with the exception of carriers of mutations in *FANCD1*, *J*, *M*, *N*, *O* and *S* genes who have an increased risk of breast and ovarian cancer; in *FANCO* gene, who are more prone to develop HNSCC and in *FANCN,* with more prevalence of pancreatic cancer (2–5).

Thanks to a variety of factors, including increased recognition of FA diversity, better scientific understanding of FA underlying mechanisms, improvement of supportive care options and earlier detection of the disease, approximately 80% of patients with FA survive beyond 18 years of age. In fact, the median survival of FA patients is now greater than 30 years. Thus, FA is no longer an exclusively pediatric disease; there is an every-increasing FA adult population that needs an adult medical care system well stablished (5). Going on this direction, FA is a complex disease with high clinic heterogeneity that needs well-orchestrated multi-disciplinary care and follow-up across several medical and surgical specialties. There are guidelines that try to stablish how has to be FA adult patients' follow-up in terms of specialists and periodicities. According to the Spanish guideline *Guía Básica para el Diagnóstico y Seguimiento de Pacientes con Anemia de Fanconi (2012)* from the *Red Nacional para la Anemia de Fanconi,* for the prevention of hematologic cancers, it must be considered that the risk of developing any hematologic problem before 25 yo is around 90% and for MDS or myeloid leukemia is around 20%. Moreover, while the average age of BMF is around 7 yo, MDS and myeloid leukemia appear at the age of 14 on average. Thus, in general it is recommended to perform Complete Blood Count (CBC) in peripheral blood every 3-4 months and a myelogram, cytogenetics, CD34+ analysis and Colony Forming Cell (CFC) analysis in bone marrow aspirate (BMA) every 1-1.5 years since FA diagnosis (3). In cases of CBC and BMA altered, blood counts and BMA are recommended every 1-2 and 2-3 months respectively (5). For the prevention of HNSCC it is important to note that the accumulated risk at the age of

40 is around 25%, the average age of onset is at 31 yo and the fact that FA patients have 500 to 700 fold higher risk of HNSCC than the general population. Moreover, there are no reported cases before the age of 10 and the risk of developing HNSCC increases after the fifth year from Hematopoietic Stem Cell Transplantation (HSCT). So, the routine follow-up starts at the age of 10 in non-transplanted patients and after the HSCT independently of patient's age. This medical control consists on an accurate examination of the buccal cavity, nasopharynx, oropharynx, hypopharynx and larynx every 6 months, but in case of malignant or suspicious findings as lichen planus, leukoplakia and/or erythroplakia taking a biopsy and doing follow-up every 2-3 months is recommended. In case of carcinoma detection, an annual radiography should also be done yearly (3,5). Moreover, for HNSCC prevention is important to ensure good oral hygiene as well as nor tobacco neither alcohol exposure. In order to do a correct dental care, 2 dental revisions every year since 1-1.5 years of life are highly recommended for an early detection of buccal lesions, suspicious ulcerations or leukoplakia (3). For gynecological cancer prevention, it must be considered that the estimated risk of cervical, vulvar and vaginal cancer is 200 to 400 fold higher in FA patients than general population and the age of onset is at 25 years of age on average. So, an annual gynecological follow-up should begin at the age of 16 or with the first menstruation including Papanicolaou test (cytology), an accurate genital physical examination, the Human Papilloma Virus (HPV) test and an endocrine follow-up (3). Breast follow-up consisting on mammography alternated with magnetic resonance imaging (MRI) every six months is also recommended starting at 20-25 yo or with the first menstruation. All female patients with FA should undergo vaccination against HPV after the age of 9 since this virus can cause genital warts and cervical cancer (5). Finally, guidelines do not stablish the frequency of visits for other specialist, but regular visits in traumatology, endocrinology, dermatology and neurology should be considered according to patients' needs.

FA patients' follow-up is especially complex because it requires multiple visits in different specialists, so it should be performed preferably in hospitals with specialists from all areas. Nowadays, there is not published data related to patients' and parents' adherence to FA medical follow-up neither reasons why they do a complete or incomplete care. Thus, the first hypothesis of this work was that doing medical controls in a

hospital with a multidisciplinary team will favor a correct follow-up but other factors as psychological aspects can be contraproductive. Our second hypothesis was that the wide range of clinical manifestations, the complex medical care required, the genetic origin and heritability of FA could have a psychological impact in FA families, but it could be mitigated by giving clear information to them. Ideally, there should be a professional in all FA units, for example a genetic counsellor, responsible for giving information related to the disease and supervising FA patients' medical care. The last hypothesis of this project was that a good registry of FA follow-up and findings would lead to a better disease control as well as generate knowledge about risks of developing second malignancies and establish protocols for early detection and medical manage of them.

Altogether, the aims of this work were to (I) determine whether FA patients controlled in Hospital Vall d'Hebron (HVH) perform a medical follow-up in agreement with Spanish guidelines' recommendations (II) assess the need of the genetic counsellor figure in a FA referral unit by analyzing FA knowledge, psychological impact of certain aspects and reasons for adherence or not to follow-up in families currently attended by these professionals or others with same functions; and finally (III) design a clinical data registry to be used periodically by genetic counsellors who attend FA patients.

MATERIALS AND METHODS

This project, approved by the Ethics Committee for Clinical Research of HVH, was performed based on a cohort of 28 FA patients consisting of (I) 22 adult FA patients (>16 yo), individuals registered in the adult follow-up system (Unitat d'Alt Risc i Prevenció de Càncer (UARPC), HVH) and (II) 6 pediatric FA patients, individuals registered in the pediatric follow-up system (Unitat d'Oncologia i Hematologia Pediàtriques, HVH). The UARPC is coordinated by an oncologist expert in genetics and a genetic counsellor and UOHP by pediatricians with a wide knowledge in hematology and genetics.

Study of adult FA medical follow-up in HVH. We analyzed different parameters to determine the proportion of adult patients with a good medical FA control in terms of specialists attended, frequency of visits and

clinical tests performed in each specialty. We used 21 adult patients' medical record information (1 individual was discarded since he died before starting adult follow-up) to identify differences between guidelines' recommendations and the registered "real" medical follow-up. Statistical analysis (t-student test) was performed with SPSS software and significant differences were quantified by considering 95% confidence intervals ($p<0.05$ was considered significant).

Assessing the need of the genetic counsellor figure in a FA referral unit. Through a questionnaire **(Supplementary 1)**, we determined the knowledge and the psychological impact of some concepts related to FA both in patients (>16 yo) and parents currently visited by genetic counsellors from UARPC or professionals with similar functions (pediatric hematologists perform genetic counsellor' roles in pediatric services). We also evaluated the opinion of these families towards FA follow-up and the reason/s why they do or not a complete medical care. Patients' and parents' participation was voluntary and confidential through an informed consent **(Supplementary 2).** We used two questionnaires specifically designed for patients and parents respectively, however, both surveys addressed the same questions but where adapted to each familiar role. Questionnaires consisted of three parts: (I) demographic data collection; (II) questions to evaluate FA information reception and familiarization with the term "genetic counselling"; and finally (III) 11 questions to assess FA knowledge, psychological impact and the opinion versus medical follow-up. Survey's questions had closed answers (yes/no or multiple choice), were intensity scales from 1 to 5 and, in one case, the question was based on Six-item State-Trait Anxiety Inventory (STAI6). Survey answers were evaluated by using SPSS statistic software (chi square and Fisher exact tests) to obtain valuable conclusions. Questionnaires were done face to face as the preferable system but some surveys were done by telephone and in one case by mail. Significant differences were quantified by considering 95% confidence intervals ($p<0.05$ was considered significant).

Designing a register of clinical data to use periodically for evaluation of medical follow-up and manage of FA patients. We started from an initial version of the clinical registry document developed by the UARPC service and, by using recent publications, we did an updated version.

RESULTS

Characteristics of FA cohort

Our cohort consisted of 28 FA patients (23.7 (3-36) yo on average, 57% male and 43% female) where the death rate was 11% (3/28) at the moment of the study. We detected 11% consanguineous parents from gypsy ethnicity **(Supplementary 3)**. On average, disease diagnosis was performed at the age of 8.4 and, as expected, 17/28 (60.7%) of patients were diagnosed before the age of 8. Fragility test was reported in 24/28 (85.7%) patients whereas no information was available in four pediatric subjects. Positive fragility was detected in 16/24 (66.7%) individuals and intermediate fragility, indicating somatic mosaicism, was observed in 7/24 (29.2%) cases (no information was available in 1 case). Complementation group study was done in 20/28 (71.4%) of patients reveling, as expected, that most patients (13/20, 65%) had mutations in *FANCA* gene whereas in 2/20 (10%) of cases mutations were detected in *FANCD1*, 2/20 (10%) in *FANCD2* and in 3/20 (15%) of subjects no mutations were already found. FA patients are highly recommended to avoid cancer agents as tobacco, alcohol, sun exposure without protection and unsafe sexual conducts. Around 40% (11/27) of FA patients older than 14 yo were registered as habitual smokers (the excluded patient died at 3 yo and we had no information of 4 pediatric patients). A total amount of 10/11 habitual smokers were adult patients regularly visited by UARPC in order to coordinate their medical controls. Importantly, half of them stopped smoking habit after some visits in this unit showing UARPC important role in terms of health education. A total of 19/28 (68%) patients developed malignant lesion/s along their life and 5/19 (26.3%) twice. The average age of onset of the first malignancy was around 13.9 yo and in 16/19 (84%) of cases it was a hematologic problem. We talk about “hematologic problems” instead of “hematologic cancer” because patients generally do not arrive in a malignance stage since they receive HSCT before. So, patients that had received HSCT or were in a situation that could require it were included in the “hematologic problems” category. As expected, the average age of onset of a second malignancy was later (27.6 yo) and HNSCC was the most frequent type (3/5, 60%). Finally, 13/28 (46%) FA patients received HSCT along their life at 9.7 yo on average. Contrary as expected, we did not detect increased prevalence of solid tumors among transplanted individual (data not shown).

Evaluation of FA follow-up

To determine whether adult FA patients' follow-up was in agreement with guidelines' recommendations in HVH, we considered only FA patients of our cohort with reported adult medical care (21/28, 75%) (average age: 26.2 yo). Firstly, we studied the regular attendance to each specialty by calculating the number of visits that each patient should had done with each specialists. For this, we considered their actual age and at FA diagnosis, whether they received HSCT or not (since it determines the beginning of head and neck controls) and guidelines' recommendations in terms of starting age and frequency of controls for each specialty. This expected number of visits was compared with the number of registered visits to obtain a value representative of the real following done. Those individuals that performed ≥25% of the expected visits were considered to have had medical follow-up in that specialty. Overall, we detected 3/21 (14.2%) individuals without Otorhinolaryngology controls and 7/21 (33.3%) without Maxillofacial medicine care. Importantly, we actually detected 2 cases that never had had head and neck regular vigilance since no Otorhinolaryngology either Maxillofacial controls were reported. In case of dentist, we could not obtain any conclusion since we had no information about attendance in most of cases (14/21, 66.6%). However, among those with reported visits in dentist, none of them went regularly. Finally, all patients went/had gone periodically to Hematologist, Gynecologist and UARPC (in this service, we detected 1 individual who died before starting UARPC visits). To evaluate the loss of follow-up, we calculated the number of patients expected to have medical controls during last two years in each specialty by excluding patients with never follow-up, death patients and cases in which we did not have actualized information because of external follow-up. Then we compared this estimation of "expected patients" with the number of patients who actually performed visits in last two years in every area. With this, we detected only 1 individual (one of those without head and neck regular vigilance) that also lost the rest of medical care (Hematology, Gynecology and UARPC). Additionally, one individual lost Maxillofacial controls but had correct Otorhinolaryngology medical care and other individual lost UARPC controls due to economic problems for travelling but maintained the rest of the controls in her home city. Finally, dentist was excluded from this analysis since no enough data was available **(Figure 1a)**. Regarding to the

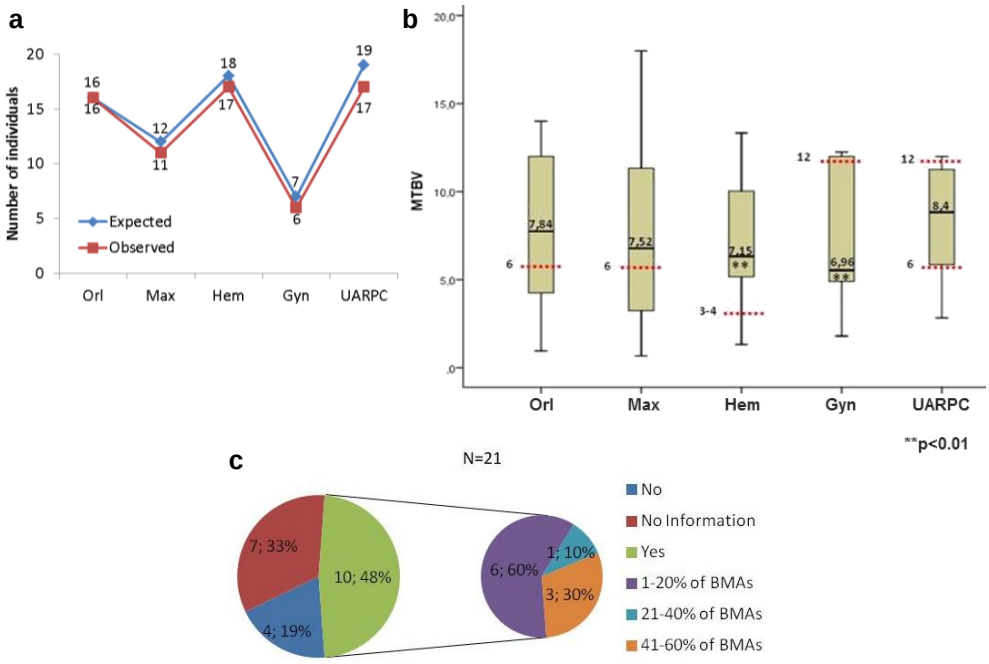


Figure 1. Evaluation of FA adult follow-up in HVH. a) The number of patients (among those with reported follow-up) who performed visits during last two years for each specialty is shown. "Expected" data (blue line) refers to the number of expected patients having follow-up in last two years by excluding patients with never follow-up in each specialty, death patients and cases in which we did not have actualized information because of external follow-up. "Observed" data (red line) shows individuals with registered follow-up in last two years. We overall detected loss of total medical care in 1 individual (one of those who never had had head and neck regular vigilance). Additionally, 1 individual lost Maxillofacial (Max) controls but had correct Otorhinolaryngology (Orl) medical care and 1 subject lost UARPC controls due to economic problems for travelling but maintained basic controls in her home city. Dentist was again excluded from this analysis since no enough data was available. **b)** Mean time between visits (months) (MTBV) are here reported. Otorhinolaryngology (Orl) visits and Maxillofacial medicine (Max) ones did not differ statistically significant from the recommended frequency (6 months). However, an underfollowing of Hematology (Hem) controls (p=0.001) overfollowing of Gynecology (Gyn) visits (p=0.002) were detected. UARPC is an special case because visits are scheduled every 6-12 months depending on patients' characteristics, for this, the mean time between visits obtained was in in this range. **c)** Only 10/21 (48%) of FA patients had, at least, 1 reported BMA. A total of 6/10 (60%) of patients had performed only between 1-20% of the expected BMAs (calculated considering patients' actual age, their age at FA diagnosis and guidelines' recommendation) whereas the remaining 4 cases performed 20-60% of the expected BMAs.

frequency of visits, we calculated the average months between visits for each specialty and then we compared them to the recommended frequencies. Time between visits in Otorhinolaryngology and Maxillofacial medicine did not differ statistically significantly from recommendations (one visit every 6 months in both cases). On the contrary, Hematology and Gynecology visits were done every 7.15 and 6.96 months on average respectively and these values differed statistically (p=0.001 and p=0.002 respectively) from the recommended periodicities (every 3-4 and 12 months in each specialty); suggesting an underfollowing in Hematology but an overfollowing in Gynecology. Finally, UARPC is a special case since visits are scheduled every 6 or 12 months on average (even more frequently at the beginning of FA diagnosis) depending on patients' severity, family needs, distance between patients' home city and HVH and availability to travel among others. Consequently, mean time between visits obtained was in between 6-12 months **(Figure 1b)**. Finally, we wanted to study whether clinical tests of some specialties were done at the recommended frequency. We first focused on BMAs in Hematology that should be done every 1-1.5y since FA diagnosis. We detected that only 10/21 (48%) of FA patients had reported, at least, 1 BMA. Among them, BMAs detected malignant findings in in 4/10 (40%) of cases. We also calculated the number of BMA that each patient should had done considering their actual and at FA diagnosis and guidelines' recommendations of 1 BMA every 1-1.5y since diagnosis moment. We detected that 6/10 (60%) had performed only between 1-20% of the expected BMAs **(Figure 1c)** showing a clear underfollowing through at BMA level. Since patients' attendance to scheduled BMAs was 100% in 8/10 FA patients, 50% in one individual and 25% in another case, the low follow-up through BMAs could be attributable to an underscheduling of this test by doctors. Secondly, regarding gynecological controls, genital examination, cytology (Papanicolau test), HPV test and hormonal following are highly recommended to be performed yearly since the age of 16 or the menarche. Here, we detected that gynecological exploration was regularly performed in all patients whereas one 22 yo girl (1/10, 10%) was uncontrolled by cytology justified for sexual inactivity and in 2 cases (20%) no information was available due to external follow-up. Regarding to the HPV test, we confirmed it in 6/10 (60%) of cases whereas we did not have information in 4/10 (40%) cases

whose gynecological controls were external to the HVH. Finally, we found only 1/10 (10%) reported case of endocrine follow-up at gynecological level. Here, we considered that patients with gynecological following in HVH but without reported endocrine controls did not actually have a hormonal follow-up (5/10, 50%), whereas in cases of external gynecological controls we put it on doubt (4/10, 40%). Altogether, these observations would suggest an underfollowing at HPV test and hormonal level. We also noticed that 8/10 (80%) of FA female patients had received HPV vaccine whereas 1/10 (10%) had not and in 1 case we had no available information because of undergoing external controls. Finally, 16/21 (76.2%) of patients performed medical follow-up with other specialists. Dermatology (9/16, 56.25%), traumatology (8/16, 50%) and endocrinology (8/16, 50%) were the most frequent ones.

Assessing the need of a genetic counsellor in FA.

By using a questionnaire (**Supplementary 1**)**,** we evaluated the knowledge and the psychological impact regarding certain concepts related to FA in adult patients (>16 yo) and parents regularly attended by genetic counsellors (in UARPC) or other health professionals with similar roles (in pediatric unit) to probe that receiving regular support from these professionals lets to improve disease knowledge and follow-up adherence as well as mitigate psychological impact of the disease. After identifying the real available FA cohort by excluding deceased patients and their parents, deceased parents and counting only once those parents with more than one affected child; we observed a total participation of 18/25 (72%) of patients (P), 13/21 (61.9%) of fathers (F) and 19/23 (82.6%) of mothers (M) with an average age of 24.0, 54.8 and 48.5 yo respectively. Firstly, we collected some demographic data (**Supplementary 4**) and we asked about whether individuals had received FA information. We detected that patients (12/18, 66.6%) considered little less to have received FA information than their parents (F:10/13, 76.9%; M:6/19, 84.2%) but the differences were not statistically significant. Moreover, among those who had received information, around 90-100% of them claimed having understood all concepts. Among different health professional proposed, both patients and parents chose hematologist as the main first source of FA information (P:33.3%, F:46.67%, M:50%). However, other sources were indicated as oncologist (12.5-27.7%), general pediatrician (8.3-13.3%), genetic counsellor

(5.5-13.3%) and others as FA association, internet or scientists. We also asked for the familiarization with the term "genetic counselling" and we detected that mothers better recognized this term (13/19, 68.4%) than patients (10/18, 55.5%) and fathers (6/13, 46.1%) but no statistical significance was obtained. Among those that knew the concept, we wonder whether they had ever been in a genetic counselling session or not. Here, we obtained similar percentages of attendance in all groups (P: 70%; F:66.6%, M:46.1%). A question asking for FA age diagnosis was performed and we found that around ~30% of individuals of all groups were not able to get the right age of diagnosis (success rate P: 66.6%, F:71.4%, M:70%).

Regarding FA's frequency, we observed that most of participants correctly knew that FA is a rare disease (P:14/18 (77.7%), F:13/13 (100%) and M:6/19 (84.2%)). We also were interested on evaluating patients' and parents' knowledge about clinical manifestations of FA. We gave them a list of different features related or not (and mixed) to the disease and we request them to indicate all the correct options. We observed that all groups had a success rate around 55-60% (P:56.3%, F:58.9%, M:60%) which was calculated by counting the number of correct answers and penalizing extra incorrect answers given. We also noticed that 5/7 correct options (BMF, leukemia, HNSCC, gynecological cancer and motor /skeletal problems) were easily identified by all groups whereas 2/7 (hearing and vision difficulties) were cited less. There was an unexpected statistically significant high rate of heart attack indication among fathers compared to patients (p= 0.012). We also detected that most individuals who did not mark gynecological cancer were male patients (5/8) or were fathers (6/6, p=0.02) or mothers (6/7, p=0.019) of male patients **(Figure 2a)**. Here after, we wanted to analyze the level of anxiety and worrying that FA manifestations previously mentioned caused to participants. We observed that patients (7/13, 53.8%) were globally less worried about FA manifestations than their parents (F:9/13, 69.2%; M:17/19, 89.4%), especially compared to their mothers (p=0.0018) **(Figure 2b)**. HNSCC was the most worrying FA clinical manifestation in three groups (P:83.3%, F:100%, M:93.3%). Related to these, we asked them for the intensity of worrying and anxiety and we detected a tendency of more acuteness in mothers' feelings than in patients' or fathers' since mothers' group indicated the maximum level of intensity of worrying/anxiety (5 in a 1-5 scale) more frequently (7/17), than the other groups (F:2/9,

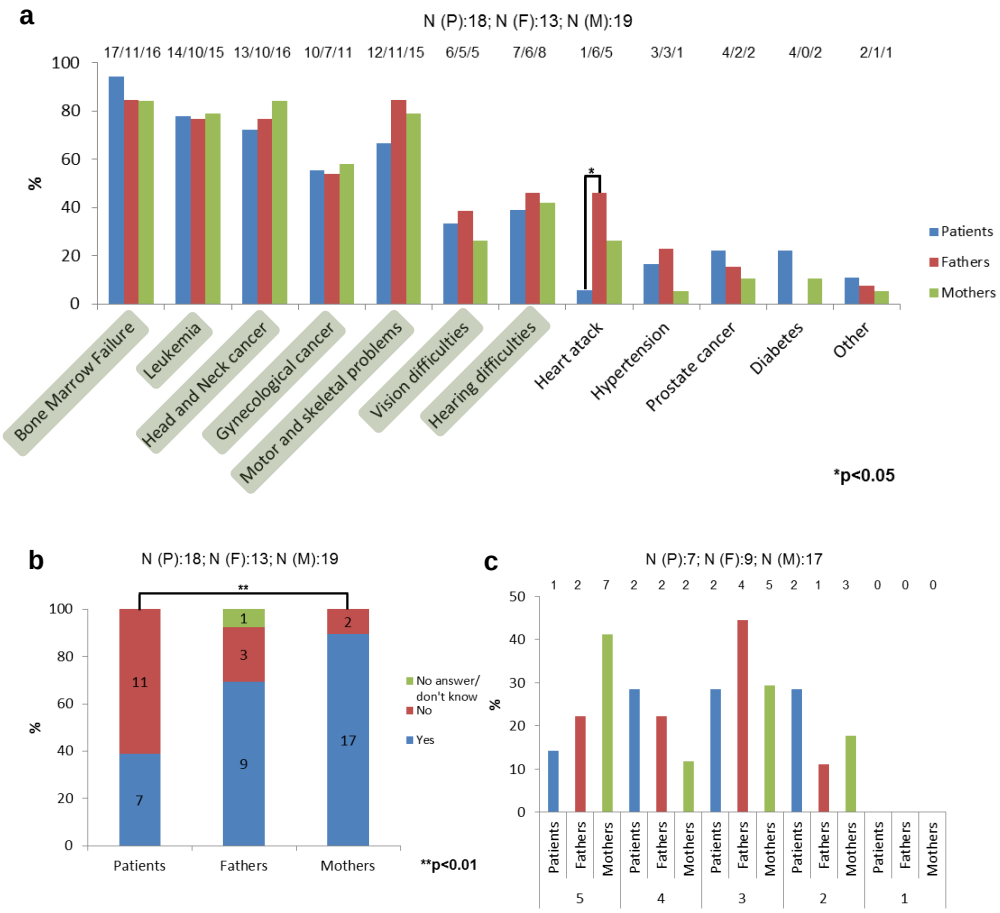


Figure 2. Evaluation of knowledge and psychological impact of FA manifestations. a) A mixture of clinical manifestations associated (highlighted in grey) or not with FA disease were listed to participants and we detected an overall success rate around 55-60% which was calculated by counting the number of correct answers and penalizing the extra given incorrect answers. Five of the seven correct options were easily identified by all groups whereas two of them (hearing and vision difficulties) were cited with more difficulties. *p= 0.012. **b)** FA associated manifestations caused less worrying and anxiety in patients than in their parents, especially compared with mothers' group (**p=0.0018). **c)** In a 1-5 scale where 1 meant "I don't feel any anxiety nor worrying" and 5 meant "I feel so anxiety and worrying that my daily life is disrupted", a tendency of more acuteness in mothers' feelings than in patients' or fathers' was detected.

P:1/7) **(Figure 2c)**. A question for evaluating the knowledge of FA genetic origin was also formulated. Around 80-100% of individuals (P:83.3%, F:91.3%, M:100%) correctly indicated that FA has a genetic origin. In this context, we asked for mutational studies results in every participant in order to know how far this information was well understood and remembered by both patients and parents. Interestingly, we found that around 30-50% of participants did not know the right molecular status reported in their medical record (success rates: P:70.5%, F:53.8%, M: 64.2%) and surprisingly, patients indicated their molecular results correctly more often than their parents. Additionally only 2/6 patients, 1/3 fathers and 2/5 mothers who affirmed that the mutation was identified in their case and it was actually found, could say correctly the name of the FA gene altered. We also asked for the feelings towards knowing or not the genetic cause of FA through a STAI6-based question. To evaluate this, we separated the answers into two groups depending on whether they belonged to individuals who affirmed knowing or not their FA mutation and we calculated a score for positive feelings and another one for negative feelings towards their situation. We detected that patients (p=0.03) and fathers (p=0.02) perceived positively the fact of knowing FA genetic cause whereas mothers more negatively. Among those who affirmed not knowing the specific FA genetic cause in their family, parents perceived it negatively whereas patients more positively **(Figure 3)**. Another item that we wanted to evaluate was the individuals' knowledge regarding recurrence risk (RR) of FA. Here we detected an important level of confusion since, considering that all participants had FA mutations known to undergo autosomic recessive inheritance or have no mutations identified, we obtained a high proportion of individuals who considered that

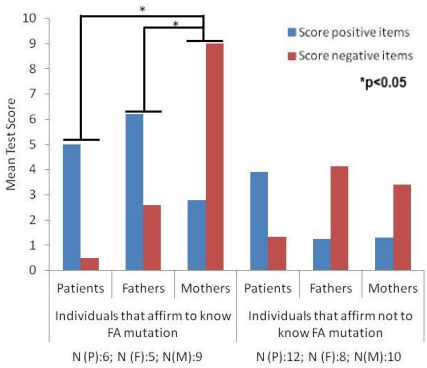


Figure 3. Feelings towards mutational analysis results. Score for positive and negative feelings were separately obtained for individuals who knew FA genetic cause and those who did not. We detected that patients (p=0.03) and fathers (p=0.02) perceived positively knowing FA genetic cause whereas mothers negatively. Parents tolerated worst not knowing FA mutation compared to their children.

FA's RR is 50% instead of 25%. Concretely 9/18 (50%) of patients, 4/13 (30.7%) of fathers and 8/19 (42.1%) of mothers answered that RR of FA was 50% whereas 4/18 (22.2%) of patients, 7/13 (53.8%) of fathers and 9/19 (47.3%) of mothers indicated the correct answer (25% of RR) **(Figure 4)**.

Finally, we performed some questions to evaluate patients' and parents' opinion about FA follow-up as well as their reasons for adherence or not to it. We detected that 18/18 (100%) patients affirmed having FA follow-up at the moment of responding the survey whereas 13/14 (92.8%) of fathers and 20/20 (100%) mothers claimed that their child was doing follow-up at that moment (note that the total number of the responding fathers and mothers was 14 and 20 instead of the real sample size 13 and 19 respectively since here we counted twice those parents of two FA siblings that had to give answers for each children). We compared individuals' answers with registered medical controls and we obtained 94% of matching in patients' answers, 100% in fathers' and 95% in mothers'. Here, we detected that the individual mentioned in the first aim of this work that never had Head and Neck follow-up and lost the rest of controls in the last two years and her mother answered this question inconsistently with the medical record. Among those that affirmed that they or their child had a FA follow-up, we detected a high satisfaction towards control's frequency (P: 4.5, F:4.3; M:4.6 on average in a 1-5 scaled question). Then, we asked whether they considered to attend all recommended specialists and we obtained that 2/18 (11.1%) of patients and 2/20 (10%) of mothers communicated an incomplete follow-up whereas 13/13 (100%) of fathers affirmed a complete follow-up in their child. Here, we wanted to determine reasons for adherence among those that affirmed a

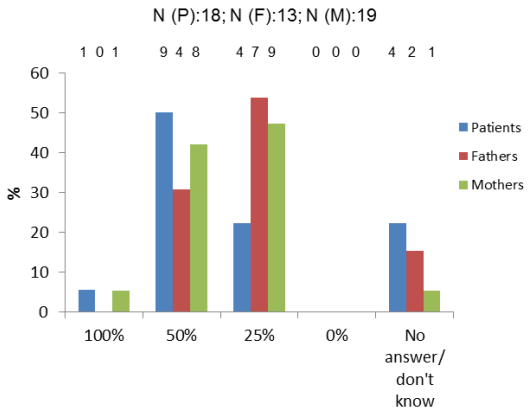


Figure 4. Confusion in FA recurrence risk. Important level of confusion was detected in FA recurrence risk question since a great proportion of responders from three groups indicated 50% as FA RR (P: 9/18, 50%; F: 4/13, 30.7%; M: 8/19, 42.1%).

complete follow-up and the reasons for no adherence among those with uncomplete or null following. As observed in **table 1**, the most voted reason for adhesion was "having doctors in the same center" for patients (12/16, 75%), "receiving reminders from the hospital" for fathers (12/13, 92%) and "coordinating visits in the same day" for mothers (14/18, 77%). We only had 2 patients with incomplete or null follow-up and we could not obtain any conclusion whereas in case of parents we detected that the most voted reason for no adherence was "stress and anxiety due to so many visits" indicated in 3/3 (100%) of parents.

Reasons for adherence to FA follow-up			
	Patients (N:16)	**Fathers (N:13)**	**Mothers (N:18)**
Doctors in the same center	**12 (75%)**	9 (69.2%)	12 (66.6%)
Visits on the same day	11 (68.7%)	7 (53.8%)	**14 (77.7%)**
Reminders from the hospital	10 (62.5%)	**12 (92.3%)**	12 (66.6%)
Short distance between home and hospital	2 (12.5%)	4 (30.7%)	7 (38.8%)
No specific reasons	1 (6.2%)	0 (0%)	0 (0%)
Prevention of health problems concerns about health	4 (25%)	5 (38.4%)	7 (38.8%)
Positive perception of hospital/doctors	1 (6.2%)	0 (0%)	2 (11.1%)
Parental pressure	1 (6.2%)	0 (0%)	0 (0%)
Social aids	0 (0%)	2 (15.3%)	0 (0%)
Time to travel	0 (0%)	1 (7.6%)	0 (0%)
Money to travel	0 (0%)	1 (7.6%)	0 (0%)
FA association	0 (0%)	1 (7.6%)	0 (0%)

Reasons for no adherence or completely loss of FA follow-up		
	Patients (N:2)	**Mothers (N:2) and Fathers (N:1)**
Long distance between home and hospital	1 (50%)	1 (33.3%)
Poor work flexibility to attend medical appointments	1 (50%)	1 (33.3%)
Missing visits due to the lack of reminders from hospital	0 (0%)	0 (0%)
Changes in visits' date by the hospital without informing the patient	1 (50%)	1 (33.3%)
Stress and anxiety due to so many visits	0 (0%)	**3 (100%)**
*Respect of my son/daughter's decision of not attending so many controls	0 (0%)	2 (66.6%)
No specific reasons	0 (0%)	0 (0%)
Negative perception of hospital/doctors	1 (50%)	2 (66.6%)
Perception of having good health without needing so many controls	0 (0%)	1 (33.3%)
Visits on different days	0 (0%)	2 (66.6%)

Table1. Reasons for adherence and no adherence to FA follow-up. Several options were given to the surveyed individuals and they could select all options that fit with their opinion. Extra answers no considered in the survey and proposed by participants are indicated in italics. The number of individuals (patients, fathers and mothers) who indicated each option is presented in every cell together with its corresponding percentage. The most voted options are highlighted in green in each subgroup.

***Design a clinical data registry to be used periodically by genetic counsellors who attend FA patients*.**

One of the main hypothesis of this work was that genetic counsellors could have a key role on improving the coordination of FA medical follow-up and manage and filling patients' and parents' needs at psychological and knowledge level. In this context, we created a tool that could help this professional to play these roles. This tool consisted on a clinical data register initially developed by UARPC that we updated and completed (**Supplementary 5**). In rough outlines, this document consists of 4 parts. The first one collects general information related to consanguinity, anthropometric measurements and patient exposure to cancer risk factors as tobacco, alcohol and unsafe sexual relationships. Secondly, there is a list of clinical manifestations in which each feature found in the patient should be marked. Here, physical malformations and congenital defects, BMF and hematopoietic complications and solid tumors are included. Thirdly, a register of performed molecular tests is also present, including fragility test and complementation study among others. Finally, there is a part designed to collect all information regarding medical controls (visits' frequency and findings in each specialty) and, recently added, a register for HPV vaccination and BMAs (frequency and findings).

With these new tool, a better manage and follow-up of FA patient should be possible as well as the generation of knowledge about risks of developing second malignancies.

DISCUSSION

FA is a minority disease that usually has an early onset and a wide range of clinical manifestations including increased risk of hematologic and solid tumors. During the infancy, the main risk of these patients is the development of hematologic problems whereas no head and neck tumors have been reported before the age of 10 (with the exception of children who receive HSCT in whom the risk of malignancy in buccal cavity and adjoining regions is increased). Gynecological lesions are also restricted to FA girls that reach the adulthood (3,5). Accordingly, we detected the same tendencies in our cohort since hematologic complications were the main problems detected during the early infancy (84%) whereas HNSCC (60%) in the early adulthood.

Patients younger than 16 yo are globally controlled by pediatric hematologists (who also perform oral cavity basic controls) and could be occasionally attended by other specialists depending on their needs. However, around 16 yo the transition to adult system controls should be done in order to initiate regular visits to adult services of Otorhinolaryngology, Maxillofacial medicine, Hematology and Gynecology (Dentist's controls are supposed to be initiated at the 1-1.5 yo). The transition from pediatric to adult follow-up system is an important issue to be carefully addressed in chronic and multiorganic illnesses. Although, there are no specific transition programs for young adults with FA, there is a clear evidence that an anticipated and coordinated transition process benefits patients and their families (12–14). Specific transition programs have been developed for patients with cystic fibrosis (CF), diabetes, juvenile idiopathic arthritis and sickle cell anemia. Transition of health care is particularly important for (I) avoiding the over accumulation of FA patients in pediatric services, (II) ensuring that FA patients are controlled by adult specialists that are not available in pediatric units and (III) helping young adults to develop independence and assume a personal responsibility for their healthcare (5). The moment at which adult transition is done is another key issue to be considered. Some publications show that the most successful transitions are those initiated during the late teenage years and those in which families receive education about future issues (12). However, the timing of care transition should be individualized according to each situation and independent on age. Some publications have identified several barriers

to the adult care transition also applicable to FA context: (I) lack of an organized completed history of the chronic illness which could be solved by using an structured medical registry tool as the once developed by us (**supplementary 5)**, (II) lack of continuing healthcare insurance coverage in the young adult, (III) physician, patients and parents reticence (IV) differences in pediatric versus adult approaches to disease chronicity and (VI) concerns about the knowledge base, experience and quality of care that will be offered by adult medical specialists in childhood -onset diseases (12,14,15). Thus, we found that 4/6 pediatric patients of our cohort should have already done the transition to adult care considering only their age at the time of this study, but, as above mentioned, this is not the only aspect to consider.

Because of the increasing life expectancy of FA patients, the development of both follow-up transition protocol and an appropriate adult care system is becoming the top priority. For this, the first aim of this work was to determine whether adult FA patients controlled in HVH underwent a follow-up according to the Spanish guidelines' recommendations (3) which, in fact, were stablished according to the American guidelines from *Fanconi Anemia Research Foundation, Inc* (**supplementary 6,** (5)). Firstly, we globally detected that the dentist specialty was the most disregarded one since very poor information about its controls was available. Secondly, we observed that most of FA patients went/had gone periodically to all specialties whereas only two individuals had never been controlled by otorhinolaryngology either maxillofacial specialist. Importantly we detected only one case of loss of follow-up in all specialties which was not explained by the exitus of the patient nor care controls in other center **(Figure 1a)**. Regarding the mean time between visits, we detected statistical significate underfollowing in Hematology and overfollowing in Gynecology but a proper frequency in the rest of specialties. However, looking to the mean time between visits of all specialties, we postulate that HVH has the tendency to schedule all visits coordinated every ~7 months probably to reduce patients' hospital stays and improve their quality life. UARPC specialty should be considered a part from the rest since periodicity of visits varies based on patients' needs between 6-12 months **(Figure 1b)**. Moreover, UARPC follow-up is especially important in those families with mutations in FA genes associated to an increased cancer risk in carriers. So, an accurate genetic coun-

selling and an extension of genetic studies to other risky relatives should be offered to promote cancer prevention strategies. Regarding screening tests, we detected very poor hematologic follow-up through BMAs **(Figure 1c)** probably because of an underscheduling of this test by doctors since we detected a good patients' attendance rate to BMAs. Considering the invasiveness of this test and, consequently, patients' reticence to do it, doctors probably try to reduce BMAs to the bare minimum without harming patients' health. Gynecological controls were quite complete except for HPV testing and hormonal controls which were the most disregarded testing strategies. We also found one FA patient without cytology controls justified by being sexual inactive at the age of 22. Since Spanish guidelines recommend to do annual cytology from 16 years of age or the first period and American guidelines specify that this screening test should be done annually but since the age of 18 in sexually inactive patients (**Supplementary 6,** (5)), this patient (22 yo) should start cytology tests in brief in order to properly prevent gynecological complications. As part as the gynecological care, patients should receive HPV vaccination series beginning at the age of 9 to prevent HPV infection and potentially mitigate HPV-associated cancers although it remains unclear its long-term effect (5). The role of HPV in gynecological malignancies is clear but HPV specific contributions to HNSCC in both male and female FA patients remain controversial since some studies suggest that HPV may be a major contributor to HNSCC development in FA patients (16,17), whereas other studies dispute these results (18,19). Consequently, there are two opposite thought: (I) all patients with FA should consider HPV vaccination to prevent both gynecological tumors and HNSCC; (II) only female FA patients are recommended to receive HPV vaccination to prevent gynecological cancer since relationship cause-effect between HPV and HNSCC is still not proved enough. In our cohort, we detected that 80% of female FA patients had received the vaccine whereas none of male FA patients were vaccinated (data not shown). Importantly, after this work, male vaccination was started in UARPC in agreement with publications that support HPV contribution to HNSCC in both male and female FA patients (16,17).

Adherence to chronic and multidisciplinary follow-up is a serious concern which should be given a remarkable attention by health professionals, especially during adolescence and young adulthood where behaviors that increase the risk of cancer are

more prevalent and they could tend to be more irresponsible in terms of health care (5). FA patients are highly recommended to avoid cancer agents as tobacco, alcohol, sun exposure without protection and unsafe sexual conducts (2,10,11). Thus, we detected a high rate of habitual smokers (39.3%) in our cohort but 50% of those habitually visited by UARPC stopped smoking in last two years, showing the important role of this unit in terms of health education.

Apart from health education, disease knowledge and emotional care are two other important aspects for medical care adhesion. Regarding to the first one, it has been demonstrated that patients' and parents' disease knowledge on chronic onset childhood illnesses is important to ensure adherence to treatments and follow-up. Some examples of this have been published: better treatment adhesion in CF patients, better disease control in asthma patients, better maintenance of remission in Crohn's disease following nutritional education, and positive correlation between children's understanding of their illness and its treatment in a pediatric oncology unit. Focusing on CF example, it has been proven that misconceptions, gaps and errors in CF knowledge could result in a non-adherence to treatment impacting on the progression and outcome of the disease (20). In this context, we studied the knowledge in FA families usually attended by genetic counsellors in UARPC (HVH) or other health professionals (in pediatric unit) with similar functions about basic FA concepts. Through a survey, we detected a good perception regarding receiving general information about FA (mainly in parents subgroup) and the main first source of these information cited by all groups was the hematologist which it was the expected result considering that these patients are usually diagnosed by this specialist at early ages. Importantly, although almost all respondents who received FA information affirmed that they had understood it, we globally detected poor knowledge (<50% of individuals got the right answer) in recurrence risk of FA and intermediate knowledge (50-66.6% of individuals got the right answer) in FA manifestations and familiar molecular status concepts. In contrast, we detected good knowledge (>66.6% of individuals got the right answer or the same information reported in medical record) in concepts as FA frequency, age at FA clinical diagnosis and FA genetic origin (**Supplementary 7**). In general, no knowledge differences were detected between patients and parents except for FA recurrence risk question where

parents succeeded more than their children. We neither found sex differences in knowledge questions in parents. Overall, we detected a quiet good FA knowledge in both patients and parents (high knowledge in 3/6 items) showing that their regular contact with genetic counsellors or other health professional with similar role had positive effects, but some effort should be done to solve the identified errors, gaps and misconceptions in FA manifestations, FA recurrence risk as well as the real molecular analyses results in each family. Confusion detected in these aspects could somehow been attributed to the fact that some participants were not actually regularly seen by a genetic counsellor proper but by a pediatrician playing similar functions leading maybe to differences in the way of transmitting the information. Quiet lower intelligence quotient sometimes found in some FA patients could have also contributed to this confusion in some FA concepts.

Another remarkable aspect related to medical care adhesion is disease psychological impact. We detected that mothers were more worried and with more intensity about FA manifestation than fathers and patients. But for all of them, HNSCC was the most worrying FA clinical manifestation (**Figure 2b and c**). We also observed a general lower worrying towards FA manifestations in patients compared to their parents, this could be explained by the fact that they usually do not feel that they are “sick” and they are not really conscious about their real risks. In this sense, genetic counsellor should work on this to ensure patients’ healthy lifestyle and follow-up adhesion. Mothers also were worst capable to live knowing FA mutations compared with fathers and patients whereas all parents tolerated worse than their children the fact of not knowing FA mutation. Here, we detected that mothers had worrying in both situations (knowing or not knowing the genetic cause) whereas fathers would prefer knowing genetic cause and patients are comfortable in both situations **(Figure 3)**. Moreover we detected that, despite of communicating molecular analyses results to families, they forget it overtime. Considering that most of parents that did not know FA mutation showed high worrying and anxiety, we should ensure that all families with identified mutations know this information to avoid at maximum this anxiety linked to the lack of information.

Finally, we detected a general positive opinion towards FA adult medical follow-up frequency. The main reason for adherence was having doctors in the same center (for patients),

receiving reminders from the hospital (for fathers) and concentrating visits in the same day (for mothers) whereas the most indicated reason for no adherence according to parents whose children did incomplete controls was the high anxiety and stress associated to frequent controls (100%) (**table 1**).

Although we obtained remarkable results from this study and initial objectives could be accomplished, some limitations should be mentioned. Regarding the evaluation of FA adult follow-up in HVH, the main problem was the incomplete medical records (especially in those from patients with external or pediatric controls) and this fact could diminish the strength of our results. Regarding to the assessment of the need of a genetic counsellor in a FA unit, the main difficulty was that our initial available cohort was small because of the low frequency of the disease and the fact that we were restricted to HVH patients. Consequently, we had to reach high rates of participation in order to perform the project. Our final sample size was not enough to do certain statistical analysis as determining main reasons for no adhesion to FA follow-up among patients without follow-up or with incomplete medical care since we only had two individuals in this group. Finally, we had another important limitation which was the fact that we could not perform all questionnaires face to face and most of them were done by phone because of participants' availability. We did 77.7% (patients), 53.8% (fathers) and 63.1% (mothers) of surveys by phone. This difference in the method could introduce bias in final results, however, we read-aloud all questions to participants always in the same way in order to add homogeneity to the process and obtain valuable and comparable results. Altogether our results suggested, firstly, that HVH globally performs FA follow-up according to Spanish guidelines' recommendations but adapting the visits' frequency of Hematology and Gynecology to frequency of visits in Head and Neck specialists probably to reduce hospital stays. However, HVH should put some effort to improve the following through BMAs, HPV test and gynecological hormonal follow-up. Secondly, our results would demonstrate the importance of having a genetic counsellors in FA units since we detected a global satisfactory level of FA knowledge, decreasing rates of habitual smokers and high adherence to follow-up among families attended regularly by these professionals (or similar). However, some FA concepts and emotional needs should be paid special attention by these professionals since we detected quite uncovering.

ACKNOWLEDGMENTS

This work has been developed in UARPC of Hospital Vall d'Hebron under the direction of Dra. Judith Balmaña and the genetic counsellor Estela Carrasco who have been supervising step by step the development of the project. The rest of members of the unit, genetic counsellors Irene Esteban and Neus Gadea, have also given their support.

REFERENCES

1. Alter BP KG. Gene Reviews: Fanconi Anemia. University of Washington, Seattle; 1993-2017. Last reviewed: 23 February 2017.
2. Kee Y, D'Andrea AD. Molecular pathogenesis and clinical management of Fanconi anemia. *J Clin Invest*. 2012 Nov 1;122(11):3799–806.
3. Red Nacional para la Anemia de Fanconi. Guía Básica para el diagnóstico y seguimiento de pacientes con Anemia de Fanconi. Red Nacional para la Anemia de Fanconi, editor. Madrid; 2012.
4. Red Nacional De Investigación Cooperativa Sobre la Anemia de Fanconi. Red Nacional De Investigación Cooperativa Sobre La Anemia De Fanconi. Madrid. 2007.
5. Fanconi Anemia Research Fund. Inc. Fanconi Anemia: Guidelines for Diagnosis and Management. Eiler ME, Frohnmayer D, Frohnmayer L, Larsen K, Owen J, editors. 4th ed. 2014;421.
6. Bogliolo M, Surralles J. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. *Mol gen*. 2015 Aug 6;33:32–40.
7. Ceccaldi R, Sarangi P, D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol*. 2016 May 5;17(6):337–49.
8. Meetei AR, Levitus M, Xue Y, Medhurst AL, Zwaan M, Ling C, et al. X-linked inheritance of Fanconi anemia complementation group B. *Nat Genet*. 2004 Nov 24;36(11):1219–24.
9. Ameziane N, May P, Haitjema A, van de Vrugt HJ, van Rossum-Fikkert SE, Ristic D, et al. A novel Fanconi anaemia subtype

associated with a dominant-negative mutation in RAD51. *Nat Commun*. 2015 Dec 18;6:8829.

10. Hays LE, Meyer S, van de Vrugt HJ. A molecular, genetic, and diagnostic spotlight on fanconi anemia. *Anemia*. 2012 Oct 2; 650730.

11. Stecklein SR, Jensen R a. Identifying and exploiting defects in the Fanconi anemia/BRCA pathway in oncology. *Transl Res*. 2012 Sep;160(3):178–97.

12. McLaughlin SE, Diener-West M, Indurkhya A, Rubin H, Heckmann R, Boyle MP. Improving transition from pediatric to adult cystic fibrosis care: lessons from a national survey of current practices. *Pediatrics* . 2008 May 1;121(5):e1160–6.

13. McDonagh JE, Kelly DA. Transitioning care of the pediatric recipient to adult caregivers. *Pediatr Clin North Am*. 2003 Dec 12;50(6):1561–83.

14. Reiss JG, Gibson RW, Walker LR. Health care transition: youth, family, and provider perspectives. *Pediatrics*. 2005 Jan 1;115(1):112–20.

15. Scal P, Evans T, Blozis S, Okinow N, Blum R. Trends in transition from pediatric to adult health care services for young adults with chronic conditions. *J Adolesc Health*. 1999 Apr;24(4):259–64.

16. Kutler DI, Wreesmann VB, Goberdhan A, Ben-Porat L, Satagopan J, Ngai I, et al. Human papillomavirus DNA and p53 polymorphisms in squamous cell carcinomas from Fanconi anemia patients. *J Natl Cancer Inst* . 2003 Nov 19;95(22):1718–21.

17. de Araujo MR, Rubira-Bullen IRF, Santos CF, Dionísio TJ, Bonfim CMS, De Marco L, et al. High prevalence of oral human papillomavirus infection in Fanconi's anemia patients. *Oral Dis* . 2011 Sep;17(6):572–6.

18. Alter BP, Giri N, Pan Y, Savage S a., Pinto L a. Antibody response to human papillomavirus vaccine in subjects with inherited bone marrow failure syndromes. *Vaccine*. 2014 Feb 26;32(10):1169–73.

19. Winer R, Huang C, Cherne S, Stern J, Butsch Kovacic M, Mehta P, et al. Detection of human papillomavirus in the oral cavities of persons with Fanconi anemia. *Oral Dis*. 2015 Apr;21(3):349–54.

20. Chomik S, Klincewicz B, Cichy W. Disease specific knowledge about cystic fibrosis, patient education and counselling in Poland. *Ann Agric Environ Med*. 2014 Jan;21(2):420–4.

SUPPLEMENTARY INFORMATION

Supplementary 1. Questionnaires. Used for evaluation of FA knowledge and follow-up opinion in patients and parents. Some questions of the parents' survey (indicated with *) were asked and evaluated twice in those parents with more than one affected child since each progenitor had to answer specifically for each child.

CUESTIONARIO PARA PACIENTES CON AF

Toda la información proporcionada mediante este cuestionario será totalmente confidencial y solo se manejará en el contexto de este estudio

☐ **Presencial** ☐ **Vía telefónica** **Fecha:**
Número de teléfono: **Código:**

CARACTERÍSTICAS DEMOGRÁFICAS

Voy a empezar haciéndole algunas preguntas generales sobre usted:

- ¿Podría decirme su fecha de nacimiento? ……..
- Sexo: ☐ Hombre ☐ Mujer
- ¿Me podría decir qué estudios ha realizado?:
  - ☐ Ninguno
  - ☐ Primaria completa o EGB (12-14años)
  - ☐ Secundaria completa (16 años)
  - ☐ Bachillerato completo (18 años)
  - ☐ Universitaria o FP
  - ☐ Posgrado completo
  - ☐ Otro, ¿Cuál? ……..……..
- ¿Cuál es su profesión actual o pasada? ……..…
- ¿En qué ciudad reside habitualmente?.....……

INFORMACIÓN RECIBIDA SOBRE ANEMIA DE FANCONI

A continuación voy a hacerle unas preguntas referentes a la AF:

- ¿Ha recibido en algún momento información global y completa acerca de la AF, es decir, información acerca de las posibles manifestaciones, la causa, los posibles tratamientos y los controles requeridos?

  ☐ Sí ☐ No ☐ NS/NC

  En caso afirmativo, ¿quién le ha proporcionado esta información?

  - ☐ Médico de familia
  - ☐ Hematólogo
  - ☐ Pediatra

- ☐ Oncólogo
- ☐ Asesor Genético
- ☐ Otros:……..…..

¿Considera que ha comprendido ésta información?

☐ Sí ☐ No ☐ NS/NC

- ¿Le suena familiar el término “asesoramiento genético”?

☐ Sí ☐ No ☐ NS/NC

En caso afirmativo, ¿ha ido alguna vez a una consulta de asesoramiento genético?

☐ Sí ☐ No ☐ NS/NC

EVALUACIÓN DEL CONOCIMENTO DE LA AF

A continuación le voy a hacer una serie de preguntas relacionadas con la AF:

1. Acerca del diagnóstico de la AF, ¿Me podría decir a qué edad los médicos le diagnosticaron la AF después de valorar sus síntomas? (*Diagnóstico clínico)*……

2. En cuanto a la frecuencia de la AF, ¿cómo definiría la AF?:

   ☐ Rara ☐ Frecuente ☐ NS/NC

3. En relación a su conocimiento de la AF, ¿cuáles son las posibles manifestaciones asociadas a la misma? A continuación le voy a decir diferentes opciones y puede indicar todas las que crea conveniente.

   - ☐ Fallo de la medula ósea
   - ☐ Fallo cardíaco
   - ☐ HTA
   - ☐ Leucemia
   - ☐ Cáncer de boca, faringe o laringe
   - ☐ Cáncer de próstata
   - ☐ Cánceres ginecológicos
   - ☐ Problemas motores y esqueléticos
   - ☐ Diabetes
   - ☐ Problemas de visión
   - ☐ Problemas de audición

4. Considerando su respuesta en la pregunta anterior, ¿hay alguna/s manifestación/es que le genere/n ansiedad y/o preocupación?

   ☐ Sí ☐ No ☐ NS/NC

En caso afirmativo

- ¿Podría indicarme cuál/es?.....
- Indique en una escala del 1 al 5 su grado de ansiedad y/o preocupación dónde 1 significa "No siento ninguna ansiedad ni preocupación" y 5 significa "Siento una ansiedad y preocupación que altera mi vida diaria"

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

5. ¿Me podría decir si la causa de la AF es alguna de las siguientes?
   - ☐ Una infección durante el embarazo.
   - ☐ El consumo de tabaco y/o alcohol durante el embarazo.
   - ☐ Una mala alimentación durante los primeros años de vida.
   - ☐ La causa es genética, es decir, se debe a la mutación de un gen
   - ☐ No se conoce la causa.
   - ☐ NS/NC

6. La causa de la AF es genética, ¿sabría decirme si en su caso se ha encontrado la causa genética concreta que ha originado la AF? (*Diagnóstico Molecular*)

   ☐ Sí ☐ No ☐ NS/NC

   En caso afirmativo:
   - ¿Me podría decir cuál es?.....
   - El hecho de haber confirmado la base genética de la AF ¿qué tipo de sentimiento le ha provocado? *STAI-E6 versión reducida de "Estado del State-Trait Anxiety Inventory (STAI)"*

	Nada	Algo	Bastante	Mucho
1. Me siento cómodo (estoy a gusto)	0	1	2	3
2. Me siento angustiado	0	1	2	3
3. Me siento confortable	0	1	2	3
4. Me siento nervioso	0	1	2	3
5. Estoy preocupado	0	1	2	3
6. En este momento me siento bien	0	1	2	3

   En caso negativo, el hecho de no haber encontrado la causa genética de la AF ¿qué tipo de sentimiento le ha provocado? *STAI-E6 versión reducida de "Estado del State-Trait Anxiety Inventory (STAI)"*

	Nada	Algo	Bastante	Mucho
1. Me siento cómodo (estoy a gusto)	0	1	2	3
2. Me siento angustiado	0	1	2	3
3. Me siento confortable	0	1	2	3
4. Me siento nervioso	0	1	2	3
5. Estoy preocupado	0	1	2	3
6. En este momento me siento bien	0	1	2	3

7. Referente a la transmisión de la AF, cuando una pareja tiene un hijo con AF, ¿qué probabilidad hay de que otro hijo también esté afectado de AF?

   ☐ 100% ☐ 50% ☐ 25% ☐ 0% ☐ NS/NC

8. ¿Usted realiza un seguimiento de la AF?

   ☐ Sí ☐ No ☐ NS/NC

   En caso afirmativo, ¿me podría indicar el centro/hospital?.....

9. Ahora le voy a decir algunos ejemplos de médicos especialistas, ¿me podría indicar a cuál/les de ellos acude para realizar el seguimiento de la AF? (puede indicar más de una opción)

   ☐ Otorrinolaringólogo
   ☐ Médico Maxillofacial
   ☐ Hematólogo
   ☐ Traumatólogo
   ☐ Dermatólogo
   ☐ Neurólogo
   ☐ Ginecólogo
   ☐ Unidad de Consejo Genético
   ☐ Otros.....
   ☐ NS/NC

   Ahora repasaremos cada uno de los médicos que me ha indicado y le voy a hacer unas preguntas cortas acerca de ellos:

   - Especialista 1:……
     - ¿Me podría decir con qué frecuencia acude a este médico especialista?
       - ☐ Cada mes
       - ☐ Cada 3 meses
       - ☐ Cada 6 meses
       - ☐ Cada año
       - ☐ Cada 2 años
     - ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……
   - Especialista 2:……
     - ¿Me podría decir con qué frecuencia acude a este médico especialista?
       - ☐ Cada mes
       - ☐ Cada 3 meses

- ☐ Cada 6 meses
- ☐ Cada año
- ☐ Cada 2 años

- ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……

- Especialista 3:……
  - ¿Me podría decir con qué frecuencia acude a este médico especialista?
    - ☐ Cada mes
    - ☐ Cada 3 meses
    - ☐ Cada 6 meses
    - ☐ Cada año
    - ☐ Cada 2 años
  - ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……
- Especialista 4:……
  - ¿Me podría decir con qué frecuencia acude a este médico especialista?
    - ☐ Cada mes
    - ☐ Cada 3 meses
    - ☐ Cada 6 meses
    - ☐ Cada año
    - ☐ Cada 2 años
  - ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……
- Especialista 5:……
  - ¿Me podría decir con qué frecuencia acude a este médico especialista?
    - ☐ Cada mes
    - ☐ Cada 3 meses
    - ☐ Cada 6 meses
    - ☐ Cada año
    - ☐ Cada 2 años
  - ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……
- Especialista 6:……
  - ¿Me podría decir con qué frecuencia acude a este médico especialista?

- ☐ Cada mes
- ☐ Cada 3 meses
- ☐ Cada 6 meses
- ☐ Cada año
- ☐ Cada 2 años

- ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……

- Especialista 7:……
  - ¿Me podría decir con qué frecuencia acude a este médico especialista?
    - ☐ Cada mes
    - ☐ Cada 3 meses
    - ☐ Cada 6 meses
    - ☐ Cada año
    - ☐ Cada 2 años
  - ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……

10. ¿Qué opina acerca del seguimiento de la AF, cree que hay pocos o demasiados controles? Indique un número en una escala del 1 al 5 dónde 1 significa "El seguimiento es totalmente insuficiente" y 5 significa "El seguimiento es totalmente suficiente"

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

11. ¿Considera que usted acude a todos los especialistas que le indica su médico para el correcto seguimiento de la AF?

☐ Sí ☐ No ☐ NS/NC

11.1 En caso afirmativo, ¿me podría indicar si alguna de las siguientes opciones le facilita hacer un seguimiento completo? (puede indicar más de una opción)

- ☐ Tener todos los médicos en el mismo centro/hospital.
- ☐ Que programen varias visitas en un mismo día.
- ☐ Que avisen desde el centro/hospital para recordar cada visita.
- ☐ Vivir cerca del centro/hospital done hago el seguimiento.
- ☐ No hay ninguna razón en concreto.
- ☐ Otras:.....

11.2 En caso negativo, ahora le voy a decir algunos ejemplos de médicos especialistas, ¿me podría indicar a cuál/cuáles debería ir porqué su médico se lo indica pero finalmente no acude a ellos? (puede indicar más de una opción)

- ☐ Otorrinolaringólogo
- ☐ Médico Maxillofacial
- ☐ Hematólogo
- ☐ Traumatólogo
- ☐ Dermatólogo
- ☐ Neurólogo
- ☐ Ginecólogo
- ☐ Unidad de Consejo Genético
- ☐ Otros.....
- ☐ NS/NC

¿Me podría indicar si alguna de las siguientes opciones le dificulta hacer un seguimiento completo? (puede indicar más de una opción)

- ☐ Vivo lejos del centro/hospital donde hago el seguimiento y no puedo pagar tantos desplazamientos.
- ☐ No puedo faltar al trabajo para venir a todos los controles.
- ☐ Me olvido de muchas visitas porqué el centro no me las recuerda.
- ☐ A veces, el centro cambia las citas sin avisarme.
- ☐ Siento mucho estrés y ansiedad con tantas visitas así que prefiero no ir a todas.
- ☐ No hay ninguna razón en concreto.
- ☐ Otras:.....

CUESTIONARIO PARA PADRES DE PACIENTES CON AF

Toda la información proporcionada mediante este cuestionario será totalmente confidencial y solo se manejará en el contexto de este estudio

☐ **Presencial** ☐ **Vía telefónica** **Fecha:**
Número de teléfono:

Usted es: ☐ Padre ☐ Madre **de (código)*:**

CARACTERÍSTICAS DEMOGRÁFICAS

Voy a empezar haciéndole algunas preguntas generales sobre usted:

- ¿Podría decirme su fecha de nacimiento?……
- Sexo: ☐ Hombre ☐ Mujer
- ¿Podría indicarme su estado civil?:
  - ☐ Soltero/a
  - ☐ Casado/a
  - ☐ Separado/a
  - ☐ Viudo/a
  - ☐ Otros (especificar)
- ¿Me podría decir qué estudios ha realizado?:
  - ☐ Ninguno
  - ☐ Primaria completa o EGB (12-14años)
  - ☐ Secundaria completa (16 años)
  - ☐ Bachillerato completo (18 años)
  - ☐ Universitaria o FP
  - ☐ Posgrado completo
  - ☐ Otro, ¿Cuál? ……..…
- ¿Cuál es su profesión actual o pasada? ……..
- ¿En qué ciudad reside habitualmente?…….

INFORMACIÓN RECIBIDA SOBRE ANEMIA DE FANCONI

A continuación voy a hacerle unas preguntas referentes a la AF:

- ¿Ha recibido en algún momento información global y completa acerca de la AF, es decir, información acerca de las posibles manifestaciones, la causa, los posibles tratamientos y los controles requeridos?

  ☐ Sí ☐ No ☐ NS/NC

  En caso afirmativo, ¿quién le ha proporcionado esta información?

  - ☐ Médico de familia
  - ☐ Hematólogo
  - ☐ Pediatra
  - ☐ Oncólogo
  - ☐ Asesor Genético
  - ☐ Otros:……..……

  ¿Considera que ha comprendido ésta información?

  ☐ Sí ☐ No ☐ NS/NC

- ¿Le suena familiar el término “asesoramiento genético”?

  ☐ Sí ☐ No ☐ NS/NC

  En caso afirmativo, ¿ha ido alguna vez a una consulta de asesoramiento genético?

  ☐ Sí ☐ No ☐ NS/NC

EVALUACIÓN DEL CONOCIMIENTO DE LA AF

A continuación le voy a hacer una serie de preguntas relacionadas con la AF:

1. Acerca del diagnóstico de la AF, ¿Me podría decir a qué edad los médicos le diagnosticaron a su hijo la AF después de valorar sus síntomas? (*Diagnóstico clínico)*…….....

2. En cuanto a la frecuencia de la AF, ¿cómo definiría la AF?:

   ☐ Rara ☐ Frecuente ☐ NS/NC

3. En relación a su conocimiento de la AF, ¿cuáles son las posibles manifestaciones asociadas a la misma? A continuación le voy a decir diferentes opciones y puede indicar todas las que crea conveniente.

   ☐ Fallo de la medula ósea
   ☐ Fallo cardíaco
   ☐ HTA
   ☐ Leucemia
   ☐ Cáncer de boca, faringe o laringe
   ☐ Cáncer de próstata
   ☐ Cánceres ginecológicos
   ☐ Problemas motores y esqueléticos
   ☐ Diabetes
   ☐ Problemas de visión
   ☐ Problemas de audición

4. Considerando su respuesta en la pregunta anterior, ¿hay alguna/s manifestación/es que le genere/n ansiedad y/o preocupación?

   ☐ Sí ☐ No ☐ NS/NC

   En caso afirmativo

   - ¿Podría indicarme cuál/es?.....
   - Indique en una escala del 1 al 5 su grado de ansiedad y/o preocupación dónde 1 significa “No siento ninguna ansiedad ni preocupación” y 5 significa “Siento una ansiedad y preocupación que altera mi vida diaria”

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

5. ¿Me podría decir si la causa de la AF es alguna de las siguientes?
   - ☐ Una infección durante el embarazo.
   - ☐ El consumo de tabaco y/o alcohol durante el embarazo.
   - ☐ Una mala alimentación durante los primeros años de vida.
   - ☐ La causa es genética, es decir, se debe a la mutación de un gen
   - ☐ No se conoce la causa.
   - ☐ NS/NC

6. La causa de la AF es genética, ¿sabría decirme si en su caso se ha encontrado la causa genética concreta que ha originado la AF? (*Diagnóstico Molecular*)

   ☐ Sí ☐ No ☐ NS/NC

   En caso afirmativo:
   - ¿Me podría decir cuál es?.....
   - El hecho de haber confirmado la base genética de la AF ¿qué tipo de sentimiento le ha provocado? *STAI-E6 versión reducida de "Estado del State-Trait Anxiety Inventory (STAI)"*

	Nada	Algo	Bastante	Mucho
1. Me siento cómodo (estoy a gusto)	0	1	2	3
2. Me siento angustiado	0	1	2	3
3. Me siento confortable	0	1	2	3
4. Me siento nervioso	0	1	2	3
5. Estoy preocupado	0	1	2	3
6. En este momento me siento bien	0	1	2	3

   En caso negativo, el hecho de no haber encontrado la causa genética de la AF ¿qué tipo de sentimiento le ha provocado? *STAI-E6 versión reducida de "Estado del State-Trait Anxiety Inventory (STAI)"*

	Nada	Algo	Bastante	Mucho
1. Me siento cómodo (estoy a gusto)	0	1	2	3
2. Me siento angustiado	0	1	2	3
3. Me siento confortable	0	1	2	3
4. Me siento nervioso	0	1	2	3
5. Estoy preocupado	0	1	2	3
6. En este momento me siento bien	0	1	2	3

7. Referente a la transmisión de la AF, cuando una pareja tiene un hijo con AF, ¿qué probabilidad hay de que otro hijo también esté afectado de AF?

   ☐ 100% ☐ 50% ☐ 25% ☐ 0% ☐ NS/NC

8. ¿Su hijo realiza un seguimiento de la AF?*

   ☐ Sí ☐ No ☐ NS/NC

   En caso afirmativo, ¿me podría indicar el centro/hospital?.....

9. Ahora le voy a decir algunos ejemplos de médicos especialistas, ¿me podría indicar a cuál/les de ellos acude para realizar el seguimiento de la AF? (puede indicar más de una opción)*

   - ☐ Otorrinolaringólogo
   - ☐ Médico Maxillofacial
   - ☐ Hematólogo
   - ☐ Traumatólogo
   - ☐ Dermatólogo
   - ☐ Neurólogo
   - ☐ Ginecólogo
   - ☐ Unidad de Consejo Genético
   - ☐ Otros.....
   - ☐ NS/NC

   Ahora repasaremos cada uno de los médicos que me ha indicado y le voy a hacer unas preguntas cortas acerca de ellos*:

   - Especialista 1:……
     - ¿Me podría decir con qué frecuencia acude a este médico especialista?
       - ☐ Cada mes
       - ☐ Cada 3 meses
       - ☐ Cada 6 meses
       - ☐ Cada año
       - ☐ Cada 2 años
     - ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……
   - Especialista 2:……
     - ¿Me podría decir con qué frecuencia acude a este médico especialista?
       - ☐ Cada mes
       - ☐ Cada 3 meses
       - ☐ Cada 6 meses
       - ☐ Cada año
       - ☐ Cada 2 años

- ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……

- Especialista 3:……
  - ¿Me podría decir con qué frecuencia acude a este médico especialista?
    - ☐ Cada mes
    - ☐ Cada 3 meses
    - ☐ Cada 6 meses
    - ☐ Cada año
    - ☐ Cada 2 años
  - ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……..

- Especialista 4:……
  - ¿Me podría decir con qué frecuencia acude a este médico especialista?
    - ☐ Cada mes
    - ☐ Cada 3 meses
    - ☐ Cada 6 meses
    - ☐ Cada año
    - ☐ Cada 2 años
  - ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……

- Especialista 5:……
  - ¿Me podría decir con qué frecuencia acude a este médico especialista?
    - ☐ Cada mes
    - ☐ Cada 3 meses
    - ☐ Cada 6 meses
    - ☐ Cada año
    - ☐ Cada 2 años
  - ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……

- Especialista 6:……
  - ¿Me podría decir con qué frecuencia acude a este médico especialista?
    - ☐ Cada mes
    - ☐ Cada 3 meses
    - ☐ Cada 6 meses

- ☐ Cada año
- ☐ Cada 2 años

- ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……

- Especialista 7:……
  - ¿Me podría decir con qué frecuencia acude a este médico especialista?
    - ☐ Cada mes
    - ☐ Cada 3 meses
    - ☐ Cada 6 meses
    - ☐ Cada año
    - ☐ Cada 2 años
  - ¿Hasta la fecha, este especialista le han encontrado algún problema? Indíqueme cuál/es por favor……

10. ¿Qué opina acerca del seguimiento de la AF, cree que hay pocos o demasiados controles? Indique un número en una escala del 1 al 5 dónde 1 significa "El seguimiento es totalmente insuficiente" y 5 significa "El seguimiento es totalmente suficiente"

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

11. ¿Considera que su hijo acude a todos los especialistas que le indica su médico para el correcto seguimiento de la AF?*

☐ Sí ☐ No ☐ NS/NC

11.1 En caso afirmativo, ¿me podría indicar si alguna de las siguientes opciones le facilita hacer un seguimiento completo? (puede indicar más de una opción)*

- ☐ Tener todos los médicos en el mismo centro/hospital.
- ☐ Que programen varias visitas en un mismo día.
- ☐ Que avisen desde el centro/hospital para recordar cada visita.
- ☐ Vivir cerca del centro/hospital done hago el seguimiento.
- ☐ No hay ninguna razón en concreto.
- ☐ Otras:.....

11.2 En caso negativo, ahora le voy a decir algunos ejemplos de médicos especialistas, ¿me podría indicar a cuál/cuáles debería ir su hijo porqué su médico se lo indica pero finalmente no acude a ellos? (puede indicar más de una opción)*

- ☐ Otorrinolaringólogo
- ☐ Médico Maxillofacial

- ☐ Hematólogo
- ☐ Traumatólogo
- ☐ Dermatólogo
- ☐ Neurólogo
- ☐ Ginecólogo
- ☐ Unidad de Consejo Genético
- ☐ Otros.....
- ☐ NS/NC

¿Me podría indicar si alguna de las siguientes opciones le dificulta hacer un seguimiento completo? (puede indicar más de una opción)*

- ☐ Vive/vivimos lejos del centro/hospital donde se hace el seguimiento y no se puede pagar tantos desplazamientos.
- ☐ Sólo puedo acompañar a mi hijo yo y no puedo faltar al trabajo para venir a todos los controles.
- ☐ Me olvido/nos olvidamos de muchas visitas porqué el centro no me las recuerda.
- ☐ A veces, el centro cambia las citas sin avisarme.
- ☐ Siento mucho estrés y ansiedad con tantas visitas así que prefiero no ir a todas.
- ☐ Mi hijo no quiere venir tantas veces al médico y yo no quiero forzarlo.
- ☐ No hay ninguna razón en concreto.
- ☐ Otras:.....

Supplementary 2. Informed consent. Voluntary and informed participation of FA patients and their parents to this project was undergone through the filling of an informed consent.

CONOCIMIENTO DE LA ANEMIA DE FANCONI.
HOJA INFORMATIVA Y CONSENTIMIENTO INFORMADO

Proyecto de investigación titulado: *Anemia de Fanconi: evaluación del seguimiento y de la figura del asesor genético*
Investigador principal: *Judith Reina Castillón*
Servicio: *Servicio de Oncología Médica del Hospital Vall d'Hebron*.
Promotor: *Dra. Judith Balmaña y Estela Carrasco*

Objetivos:
Le solicitamos su participación en este proyecto de investigación cuyo objetivo principal es profundizar en el conocimiento de factores que puedan influir en (I) el grado de adherencia al seguimiento médico y (II) la información proporcionada a las familias con un diagnóstico de enfermedad de Anemia de Fanconi.

Beneficios:
Es posible que de su participación en este estudio no se obtenga un beneficio directo. Sin embargo, la identificación de posibles factores relacionados con el grado de adherencia al seguimiento médico y de información proporcionada a las familias podría beneficiar en un futuro a otros pacientes que la sufren y contribuir a un mejor conocimiento y tratamiento de esta enfermedad.

Procedimientos del estudio:
Dicho estudio consiste en un cuestionario de unos 15 minutos de duración aproximadamente que será realizada por Judith Reina Castillón, una estudiante del Máster de Asesoramiento Genético de la Universitat Pompeu Fabra-Idec de Barcelona bajo la supervisión directa de la asesora genética Estela Carrasco y la oncóloga coordinadora de la unidad la Dra. Judith Balmaña.

Protección de datos personales:
De acuerdo con la Ley 15/1999 de Protección de Datos de Carácter Personal, los datos personales que se obtengan serán los necesarios para cubrir los fines del estudio. En ninguno de los informes del estudio aparecerá su nombre, y su identidad no será revelada a persona alguna salvo para cumplir con los fines del estudio, y en el caso de urgencia médica o requerimiento legal. Cualquier información de carácter personal

que pueda ser identificable será conservada por métodos informáticos en condiciones de seguridad por Judith Reina, o por una institución designada por ella. El acceso a dicha información quedará restringido al personal de la unidad de Oncología Médica del Hospital Vall d'Hebron, designado al efecto o a otro personal autorizado que estará obligado a mantener la confidencialidad de la información.

De acuerdo con la ley vigente, tiene usted derecho al acceso de sus datos personales; asimismo, y si está justificado, tiene derecho a su rectificación y cancelación. Si así lo desea, deberá solicitarlo al médico que le atiende en este estudio.

De acuerdo con la legislación vigente, tiene derecho a ser informado de los datos relevantes para su salud que se obtengan en el curso del estudio. Esta información se le comunicará si lo desea; en el caso de que prefiera no ser informado, su decisión se respetará.

Si necesita más información sobre este estudio puede contactar con los investigadores responsables, el/la Dr./a. Judith Balmaña o Estela Carrasco del Servicio de Oncología Médica. Tel.932746000 (Ext: 4717)

Su participación en el estudio es totalmente voluntaria, y si decide no participar recibirá todos los cuidados médicos que necesite y la relación con el equipo médico que le atiende no se verá afectada

Nombre participante: _____

_____	_____	_____	_____
Firma del interesado	Fecha	Firma del responsable	Fecha

Supplementary 3. Demographic data of FA cohort. This table shows demographic data of FA cohort as average age, age and gender distribution, current live status and consanguinity and gypsy ethnicity. Individuals of each category and percentages (in brackets) are shown.

N = 28	
Average age	23.71 yo
Age Distribution	
<16 yo	3 (10.7%)
16-20 yo	6 (21.4%)
21-25 yo	7 (25%)
>25 yo	12 (42.9%)
Gender distribution	
Male	16 (57.1%)
Female	12 (42.9%)
Current live status	
Alive	25 (89.3%)
Death	3 (10.7%)
Consanguinity and gypsy ethnicity	
Yes	3 (10.7%)
No	25 (89.3%)

Supplementary 4. Demographic data of questionnaire responders. On the top of this table, rates of responders obtained for each subgroup are shown. For this, real available cohort size was firstly calculated for each group. In patients' sample set, it consisted of 25 individuals after considering 3 exitus. In fathers group, the real amount of available participants was 21 after excluding 2 deceased fathers, fathers of 3 death patients and counting only once two fathers with more than one affected child in the cohort. In mothers' subgroup, the real sample size was 23 after excluding mothers of 3 death patients and counting only once two mothers with more than one affected child in the cohort. This table also shows general data related to responders that was collected in the demographic data section of the survey as age, gender, level of studies, actual profession, marital status (for parents) and distance between home city and HVH.

	Patients	**Fathers**	**Mothers**
Initial N	28	28	28
Exitus	3	2 (+3)	(+3)
Family relationships to consider	2 pairs of brothers	2 fathers for 4 FA patients	2 mothers for 4 FA patients
Available N	25	21	23
Responders	18 (72%)	13 (61.9%)	19 (82.6%)
Mean age(yo)	24.06	54.85	48.5

Age Distribution			
<16 yo	0 (0%)		
16-20 yo	5 (27.8%)		
21-25 yo	5 (27.8%)		
26-30 yo	6 (33.3%)		
31-35 yo	1 (5.5%)		
36-40 yo	1 (5.5%)	1 (7.69%)	3 (15.78%)
41-45 yo	-	2 (15.38%)	3 (15.78%)
46-50 yo	-	6 (46.15%)	4 (21.05%)
51-55 yo	-	1 (7.69%)	6 (31.57%)
56-60 yo	-	3 (23.07%)	2 (10.52%)
>60 yo	-		1 (5.26%)
Gender distribution			
Male	9 (50%)	-	-
Female	9 (50%)	-	-
Level of studies			
None	0 (0%)	0 (0%)	1 (5%)
Elementary school (6yo)	4 (22%)	4 (31%)	7 (37%)
Secondary school (16yo)	5 (28%)	3 (23%)	3 (16%)
High school (18yo)	2 (11%)	1 (7%)	2 (10%)
University/Professional formation	5 (28%)	4 (31%)	4 (21%)
Postgraduate	1 (5.5%)	1 (8%)	2 (11%)
Others	1 (5.5%)	0 (0%)	0 (0%)
Profession			
Student	7 (39%)	0 (0%)	1 (5%)
Active worker	7 (39%)	11 (84%)	13 (69%)
Temporary unemployed	0 (0%)	1 (8%)	0 (0%)
Never had a job	4 (22%)	0 (0%)	4 (21%)
Retired	0 (0%)	1 (8()	1 (5%)
Marital status			
Unmarried	-	0 (0%)	0 (0%)
Married	-	11 (85%)	11 (58%)
Divorced/Separated	-	2 (15%)	7 (37%)
Widow/er	-	0 (0%)	1 (5%)
Distance between home city and HVH			
Barcelona	4 (22%)	2 (15%)	4 (21%)

1-10Km from Barcelona	1 (6%)	0 (0%)	1 (5%)
10-30Km from Barcelona	3 (17%)	3 (23%)	2 (10.5%)
30-60Km from Barcelona	2 (11%)	2 (15%)	2 (10.5%)
60-100Km from Barcelona	2 (11%)	2 (15%)	3 (16%)
>100Km from Barcelona	6 (33%)	4 (32%)	7 (37%)

Supplementary 5. Clinical medical registry developed for FA patients' follow-up coordination. Parts in blue are the updates performed in this work.

TEST para PACIENTES FANCONI

Derivado por:
Fecha:

Etiqueta del paciente

INFORMACIÓN GENERAL:

Consanguinidad en la familia:

☐ Sí Grado:…..
☐ No

Datos antropométricos

Talla al nacer: Talla actual:
Peso al nacer: Peso actual:

Factores de riesgo al cáncer

Tabaco:

☐ Sí Num cig/dia:
☐ No

Alcohol:

☐ Sí Número de veces/mes
☐ No

Educación sexual recibida:

☐ SÍ
☐ No

MANIFESTACIONES:

Alteraciones de la pigmentación de la piel
- ☐ café au lait spots, núm:...
- ☐ Hipopigmentación, ¿Dónde?.....
- ☐ Hiperpigmentación, ¿Dónde?.....

Malformaciones:
- ☐ Pulgares
- ☐ Antebrazo
- ☐ Sistema esquelético (escoliosis, hemivertebras…)
- ☐ Cuello
- ☐ Ojos (macroftalmia, epicantos…)
- ☐ Orejas (forma anómala…)
- ☐ Riñones y trato urinario
- ☐ Corazón
- ☐ Cavidad oral
- ☐ Sistema gastrointestinal
- ☐ Sistema nervioso central
- ☐ Craneoafacial:

Cabeza: PC:
Forma Cara:……..
Especificar:……...

Otros:
- ☐ Pérdida auditiva
- ☐ Problemas visuales
- ☐ Hipogonadismo Tratamiento:……
- ☐ Retraso en el desarrollo
  - ☐ Menarquia edad:……. Edad cambio a la pubertad:…
  - ☐ Nivel de estudios:…… CI:……..
  - ☐ Ocupación actual:……

Fallo medular: (generalmente entre los 7-8 años)
- ☐ Trombocitopenia. Edad: ……
- ☐ Anemia. Edad: ……
- ☐ Leucopenia. Edad……

TMO:
- ☐ Sí Fecha:../../… de quien? ……
- ☐ No

Anemia aplásica en adultos:
- ☐ Macricitostosis
- ☐ Elevados niveles de F hemoglobina

- ☐ Mielodisplasia
- ☐ Leucemia

Tumores sólidos:

- ☐ Carcinomas de células escamosas de cabeza y cuello
- ☐ Carcinoma de células escamosas de esófago
- ☐ Carcinoma de células escamosas de vulva
- ☐ Cáncer cervical
- ☐ Tumores del hígado (asociados a veces a los ttos con andrógenos)
- ☐ Cáncer de piel
- ☐ Otros……

Fecha diagnóstico: ……
Tratamiento: ……

Toxicidad a quimioterapia o radiación

- ☐ Sí
- ☐ No

TESTING:

Estudios de fragilidad cromosómica:

Sangre Fecha…/…/…

- ☐ Sí ☐mosaico ☐ no determinante
- ☐ No

Fibroblastos Fecha…/…/…

- ☐ Sí ☐ mosaico ☐ no determinante
- ☐ No

Parada del ciclo celular Fecha…/…/…

- ☐ Sí ☐ no determinante
- ☐ No

Monobiquitinación de la proteína FANCD2 Fecha…/…/…

- ☐ Sí ☐ No determinante
- ☐ No

Grupo de complementación: ……..…….. Fecha…/…/…
DNA: exone sequencing ☐Sí ☐No

SEGUIMIENTO:

ESPECIALISTA	FECHA	HALLAZGO	FECHA	HALLAZGO
OTORRINO				
MAXILOFACIAL				
DENTISTA				
HEMATOLOGIA				
TRAUMA				
DERMATO				
NEURO				
GINE				

Vacunación HPV

- ☐ Si Fecha: ../../…
- ☐ No
- ☐ Test VPH:
  - ☐ Positivo → CIN… Fecha: ../../…
  - ☐ Negativo

Aspirados Medula Ósea: ☐Si ☐No

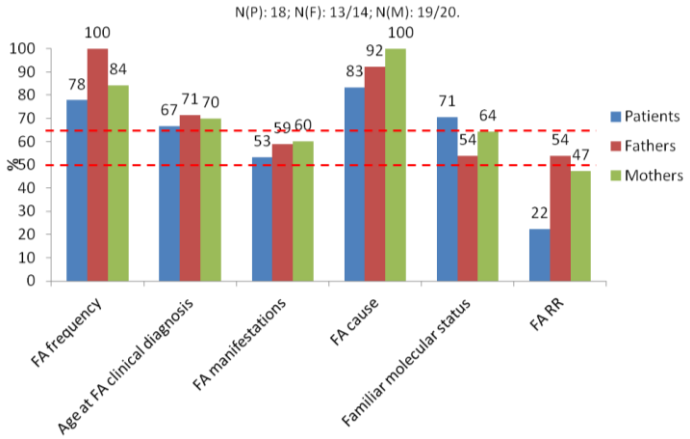
FECHA	HALLAZGO	FECHA	HALLAZGO

Supplementary 6. American FA adult follow-up guidelines. This table provides information about, from left to right, specialists needed for FA adult care (column 1), the kind of controls that each one should do (column 2), what they should evaluate (column 3), frequency of visits (column 4) and age at which each control should start (column 5) (5). CBC: complete blood count; CFC: colony forming cells; yo: years old; m: months; HSCT: human stem cells transplantation.

Specialty	Biological sample	What is evaluated?	Frequency	Beginning age
Hematology	Peripheral blood	CBC	3-4 months (1-2m if abnormality)	Since diagnosis
	Bone marrow aspiration	Cytogenetics, morphology	1 year (1-6m if abnormality)	>2yo
Otorhinolaryngology	Vigilance	Buccal cavity, nasopharynx, oropharynx, hypopharynx and larynx	6 months (2-3months + biopsy in case of lesion)	>10 yo o since HSCT

Maxillofacial medicine	Vigilance	Buccal cavity, nasopharynx, oropharynx, hypopharynx and larynx	6 months (2-3months + biopsy in case of lesion)	>10 yo o since HSCT
Dentist	Vigilance	Buccal cavity	6 months	>1-1.5yo
Gynecology	Vigilance	Papanicolau test	1 year	Beginning sexual activity or >18yo (if sexual inactive)
	Vigilance	Genital examination	1 year	>13yo
	Vigilance	HPV test	In parallel to Papanicolau test	
	Vigilance	Breast	1 year	>20yo
Dermatology	Vigilance	Suspicious nevi or other skin abnormalities	1 year	>18 yo o since HSCT

Supplementary 7. Summary of success rate in knowledge questions of survey. High knowledge (>66.6% of individuals got the right answer or the same information reported in medical record) was detected in FA's frequency, age at FA clinical diagnosis and FA genetic origin in all groups. Intermediate knowledge (between 50-66.6% of individuals got the right answer) was obtained in concepts as FA manifestations and familiar molecular status. Poor knowledge (<50% of individuals got the right answer) was observed in recurrence risk of FA (fathers would have an intermediate knowledge (54%)).



DISCUSSION

Traditionally, human genetics searching has focused on uncovering genetic variation responsible for monogenic disorders or, more recently, complex disorders that are transmitted through the germline. Most studies performed to address this have been designed under the assumption that the vast majority of cells in the human body are genetically identical, that is to say, that genome of somatic cells is stable along the human lifespan. However, during last decades, it has become clear that cells can acquire post-zygotic genetic changes along the time generating what is named somatic mosaicism. This is defined as the coexistence of genetically distinct cell linages in an organism derived from a single zygote (3). Historically, genetic mosaicism has not been considered to be the major factor behind diseases but the idea of that it can modulate, and even cause, certain disorders is increasing in parallel with the development of high-throughput techniques with higher resolution that let to detect more mutations in a lower percentage of mosaicism (7,26).

Until known, genetic mosaicism phenomenon has been observed in the context of certain health conditions as aging. Some publications have reported an increased prevalence of detectable clonal mosaicism related to elderly (5,25,30–32,70,76,257), with specific studies focused on mosaic loss of sex chromosomes (38,39,72,73). A large list of diseases has also been associated somehow with genetic mosaicism. This would include several monogenic disorders as Neurofibromatosis type 1, in which somatic mosaicism attenuates the disease phenotype (26), or McCune Albright (40) or Proteus Syndrome (41), whose causative mutations are found in mosaicism to avoid lethality. Aneuploidies causing diseases as trisomy 21 in Down syndrome (44), might be only present in a proportion of cells whereas some specific examples, as Turner syndrome (45), are believed to be lethal and causing miscarriages when the aneuploidy is constitutional or affects a too elevated fraction of cells. Finally, another typical form of mosaicism is the heteroplasmy phenomenon in the context of mitochondrial diseases.

One of the more studied and clinically relevant forms of mosaicism is cancer, where somatic events promote cells' malignization when accumulated or arise secondary to the neoplastic process itself. This thesis is focused on mosaicism and its role in cancer development and prognosis. We describe for the first time the possible dual role of mosaicism as a promoting or protecting cancer factor in the context of three different hematologic disorders or malignancies. Thus, this work contributes to expand the knowledge of mosaicism effects on cancer, especially in its most unknown role: mosaicism rescue function.

MOSAICISM AS A BIOMARKER OF CANCER RISK AND PROGNOSIS

The need of less invasive cancer screening techniques in Fanconi anemia

Several studies have reported an increased prevalence of detectable chromosomal mosaic events (CMEs) among patients with hematologic cancer (OR:22-30) or, with slighter evidence, certain solid tumors (OR: 4) studying blood and/or saliva DNA samples by *Single Nucleotide Polymorphism* array (30–32,34,70). This data suggests that CMEs detection in easy to obtain samples, as blood or saliva, through a non-invasive technique could be used as a biomarker for early detection of cancer. The development of effective and less invasive cancer screening approaches is particularly important in the context of chronic diseases characterized by an increased baseline cancer risk that require periodic invasive tests, as bone marrow aspirates (BMAs), for early hematologic malignancies detection. BMAs require the use of a large gauge needle to penetrate the muscle and bone and finally reach the bone marrow. This invasive process is painful, complex and generates patients' anxiety; moreover, it may lead to complications as bleeding or infections; thus, these will be obstacles for good patients' care and vigilance.

On this direction, a recent publication described cell free DNA sequencing for detection of mutations related to multiple myeloma could be an alternative to the traditional tumor genetic profiling through BMAs, since the new and less invasive approach reached a sensitivity and specificity over 96%. Longitudinal genetic monitoring of myeloma tumors through BMAs is necessary for multiple myeloma patients who participate in clinical trials that are developing targeted therapies tumor genotype-specific. However, the requirement of periodic BMAs is an obstacle for enrollment and retention of patients in these studies and providing an alternative non-invasive test as cell free DNA sequencing could positively impact in the success of these clinical assays (258).

Fanconi anemia (FA) could be another example of disease that could take some benefit from less invasive screening tests. This is a chromosome instability syndrome where patients have a basal

increased cancer risk secondary to the incapacity of repairing interstrand DNA crosslinking (179,259) and are highly recommended to undergo a BMA every 1-1.5y as part of their hematologic medical vigilance (169,205). With the evaluation of the accomplishment of FA follow-up recommendations given by the *Guía Básica para el Diagnóstico y Seguimiento de Pacientes con Anemia de Fanconi (2012)* from the *Red Nacional para la Anemia de Fanconi* (205) in 21 FA adult patients, we detected a clear underfollowing through BMAs. We observed that half of evaluated patients had never performed any BMA in all their life. Additionally, among those who had performed at least one BMA, 60% only had done between 1-20% of the expected BMAs according to their actual age, their age at diagnosis and the recommended BMAs' frequency. Importantly, we detected that BMAs underscheduling by doctors was probably the reason of the low BMAs rate since we detected a general good patients' attendance to BMAs appointments.

Higher CMEs prevalence associated to increased cancer risk with poorer prognosis in FA

Considering all these points, we wanted to study the prevalence and evolution of detectable clonal mosaicism in blood and/or saliva samples from FA patients and its relationship with cancer and survival. The main goal of this was to determine whether CMEs detection in easy to obtain samples by SNP array could be a biomarker of increased cancer risk and potentially used as screening test in this disease. With the study of 167 FA patients, we detected a 170-180 fold increased risk of detectable clonal mosaicism among FA patients compared to age-matched controls. Considering the basal high cancer risk in FA, our data was in agreement with the increased prevalence of CMEs reported in cancer patients by different publications (30–32,34,70). In FA CMEs carriers, who harbored mosaic events of 0.5-2Mb or more in size, we observed a global 5.6 fold increased risk of prevalent and incident cancer. We also detected a 4.5 fold augmented risk of developing cancer after mosaicism detection with 4 fold shorter cancer free time and a 3 times higher death risk secondary to malignancy compared to those FA patients without detectable clonal mosaicism. When considering hematologic and solid incident cancer separately, we curiously observed a little bit higher risk of solid incident cancer (OR=5.9) than hematologic (OR=3.6) after mosaicism detection but with poor statistical significance due to low sample size. Despite of this, our data would be close to the range

of the 10-5.4 fold increased risk of incident hematologic malignancy among carriers of CMEs (>0.05Mb) reported by other groups (31,34). Importantly, the detection of larger CMEs (>2Mb) is reported to be associated with an even higher risk of incident hematologic cancer (35-19.2 fold) (30,34). Giving more strength to our message and similarly to previous data (31), the fold increase of global cancer risk would go from 5.6 to 40 when individuals harbor more than one CME suggesting that the amount, besides the length, of CMEs can modulate the final individuals' cancer risk.

The study of serial samples during a follow-up period could reinforce the idea of using CMEs as an early cancer biomarker since this study would allow monitoring mosaicism evolution and detecting new CMEs and new cancer diagnosis secondary to them. For this, we longitudinally studied a total amount of 28 out of 167 individuals from whom we had available one to three additional samples to the initially analyzed but all of them from the same tissue (blood or saliva). Individuals were studied for a 5 years follow-up period on average with a maximum lapse of time between two samples of 15 years. This study did not reveal any new CME but showed some fluctuations of mosaicism percentage of CMEs previously detected (in four individuals) along the time. Thus, we did not observed any new possible cancer case secondary to CMEs acquired during the vigilance. In fact, any of the 28 serially studied patients did not have any cancer diagnosis until their last recorded medical control.

Altogether, our results and previous published data in the field report a significant high risk of cancer associated with detection of mosaic anomalies in blood or saliva but very few have been published about the specific mechanisms underlying this association. Despite of this, the exacerbated risk of CMEs among FA patients could be explained by the dysfunction of the DNA reparation system "FA/BRCA pathway" that they present. Impaired FA/BRCA function is known to be responsible of the intrinsic increased cancer risk of these patients due to the lack of DNA damage reparation and the chromosome instability generated. Here, we propose that this deficit also could generate genomic heterogeneity in somatic cells increasing in this way the probability of detectable clonal mosaicism after the clonal expansion of a cell harboring a CME. Hence, a hypothetic FA patient with a notably impaired FA/BRCA function would have greater risk of both developing cancer and detectable CMEs than other patient with more preserved functionality of the pathway. Thus, cancer and

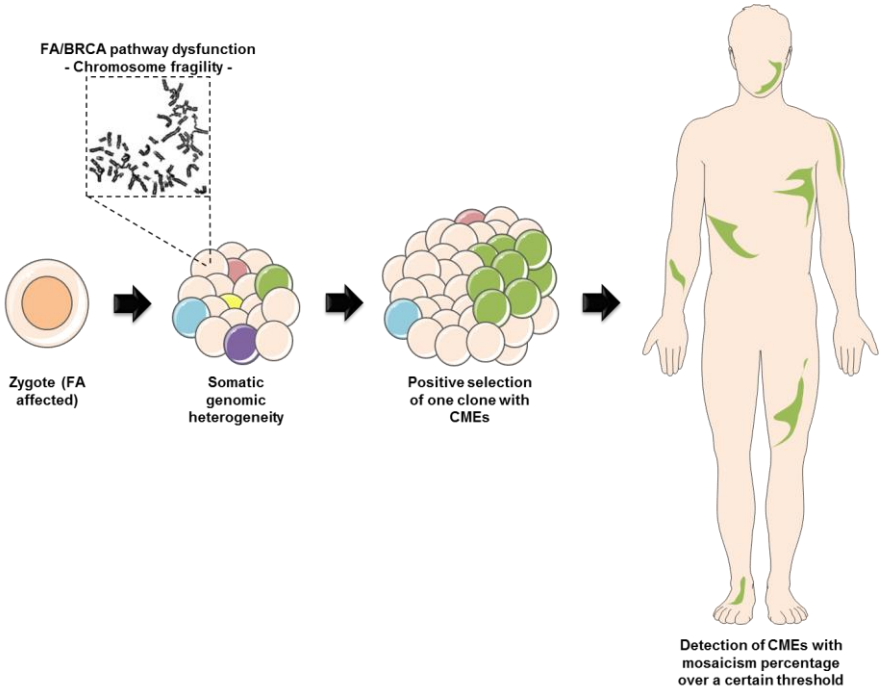


Figure 1. Possible mechanism by which CMEs could arise in FA patients. FA/BRCA pathway dysfunction, besides promoting cancer in these patients, could facilitate the emergence of clones harboring different chromosomal mosaic events (CMEs) (cells painted with different colors represent genomic heterogeneity). The positive selection and proliferation of a specific clone until a certain threshold could lead to its detection in the adulthood. From *Servier Medical Art.*

detectable CMEs could appear by a common mechanism in FA patients: impaired FA/BRCA function (**figure 1**).

In contrast, individuals with preserved FA/BRCA function would be expected to have lower CMEs prevalence and cancer risk. Reverse mosaicism is a phenomenon described in up to a 20-25% of FA patients by which one of the mutated alleles is reverted to a WT form (169). This would lead to a reduction of chromosome fragility and a hematologic clinical stabilization of the patient. In other words, reverse mosaicism in FA is more and more considered as a kind of "natural therapy" for hematologic phenotype in these patients and has inspired genetic therapy studies started until now (231,245). Among 167 studied individuals, we detected 35 (20.9%) individuals with reverse mosaicism which is defined as the detection of less than 50% of cells with chromosomal aberrations after diepoxybutane treatment. Three of them carried one CME in blood (and saliva in one case) and one of these three individuals was diagnosed of

cancer (vulvar SCC). Importantly, only one out of the 35 patients with reverse mosaicism was diagnosed of hematologic malignancy, reinforcing the idea that reversion of a FA germline mutation in hematopoietic cells protects against genomic instability and, consequently, CMEs and hematopoietic cancer emergence. In fact, the reversion of a FA mutation could become visible following the same clonal process as any CME: a preexisting reverted WT clone from very early in the development initially contributed to the somatic genomic heterogeneity but, because of its selective advantage, could undergo progressive monoclonal expansion until repopulating an important fraction (not all) of the bone marrow. This process would be very similar than the one in which a clone with a chromosomal event is positively selected and leads to detectable mosaicism.

Regarding cancer in general and hematologic malignancies in particular, mechanisms behind their relationship with chromosomal mosaicism are also poorly understood but some possibilities would be feasible. Similarly to FA, alterations in tumor suppressor genes that control proper centrosome formation, mitotic checkpoints and movement of duplicated chromosomes during cell division may be dually involved in leading both CMEs and cancer (34). Focusing on checkpoints, when they are weakened, cell cycle can progress to mitosis before chromosomes are fully replicated, aligned to the mitotic spindle or cleared of DNA damage for example. If the impaired checkpoint is the spindle assembly checkpoint (SAC) that monitors chromosome segregation during mitosis, an anaphase lag will appear. This is a process defined by the initiation of cytokinesis before the complete chromosome alignment and attachment to the mitotic spindle leading to the unequal chromosome distribution to daughter cells and generating aneuploidies and general chromosome instability. Defects in centrosome, mitotic spindle and cohesins can also alter the correct chromosome segregation in cell division and generate CMEs. Thus, alteration of proteins that regulate the cell cycle could lead to (I) CMEs detectable after its clonal expansion whenever they affect a proportion of cells enough for its detection and, of course, the tissue that we are analyzing and (II) the uncontrolled cell division typically preceding a tumorigenic process (260). Another possibility would be one of the most well-known sources of both mosaicism and cancer. In the 70s, Alfred G. Knudson developed the “two-hit hypothesis” in which he postulated that some tumors need two mutational events inactivating a tumor suppressor gene or overactivating an oncogene to occur. A “first hit” would be a dominant inherited and constitutional mutation or a somatic change acquired during the

development whereas the “second hit” would be always a genetic change acquired postzygotically in somatic cells. Either way, both inherited and non-inherited forms could generate both mosaicism and cancer disease (57). Finally, environmental factors should be considered in the acquirement of mutations in some tissues generating somatic mosaicism and, sometimes, disease (34). Although further studies are needed to identify mechanisms underlying CMEs-cancer relationship, its association is a reality.

Is there any cause-effect relationship between CMEs and cancer in FA?

There are some mechanisms that, theoretically, could generate both neoplasia and CMEs, but, could CMEs lead to certain cancer subtypes depending on the genome region affected? Studying genomic features of breakpoint intervals of CMEs can guide to which mechanisms are underlying these rearrangements as well as their possible phenotypic effect on the patient. We detected that FA breakpoint intervals were enriched of coding genes going in favor with the previously reported possible link between transcriptional activity and DNA repair, which is impaired in FA patients (261,262).

Mosaic rearrangements most frequently detected in our FA sample set were concentrated at 1q (8 gains and 1 uniparental disomy (UPD)), 3p (5 UPDs), 3q (8 gains), 6p (2 UPDs, 1 gain and 3 losses) and chromosome 7 (3 monosomies, one 7p loss and two 7q losses). Focusing on these genome regions, we observed several interesting points that could give some light in the potential relationship between CMEs and cancer development. Before commenting each specific case, just to mention that, the surprisingly high prevalence of UPD 3p will be deeply discussed later since this is the starting point of one of the main messages of this thesis.

The detection of 6 rearrangements with one breakpoint overlapping with Major Histocompatibility Complex (MHC) region in 5 FA patients suggested that this region is a hotspot for somatic rearrangement in FA and we would have described it for the first time. Structural variation at the MHC locus is a potential substrate for somatic rearrangements in immune system cells. Considering the generalized diminished DNA reparation ability in FA patients, the also reduced reparation of MHC rearrangements may facilitate their clonal proliferation and detection. Interestingly, UPD leading to loss of heterozygosity involving the MHC locus has been reported

in acquired aplastic anemia (263) and primary nervous system lymphoma (264). Loss of the mismatched HLA haplotype by UPD has also been reported in patients with leukemia after haplo-identical hematopoietic stem cell transplantation and infusion of donor T cells, leading to relapse through the escape of leukemic cells from the donor's antileukemic T cells (265–267). Considering this, 4 out of 5 FA patients carrying 6p CME developed cancer (2 solid tumors and 2 acute myeloid leukemias (AML)) at the same time or after mosaicism detection suggesting a possible link between cancer (but not a specific subtype) and CMEs in 6p.

In addition to the new possible 6p hotspot for somatic events in FA disease, we also observed mosaic events frequently in already known leukemia-associated hotspots: 1q and 3q regions and chromosome 7 (223). A fact that would provide additional support for the association between clonal mosaicism and cancer development would be that, in our sample set, we observed that 90% of FA CME carriers with myelodysplastic syndrome (MDS) or AML harbored, at least, one leukemia-related CME (1q gain, 3q gain and/or chromosome 7 monosomy) in blood and/or saliva. Our observations go in the line of previous publications (30,31,34). For example, Laurie et al. detected deletions concentrated in 2p (*DNMT3A*), 4q (*TET2*), 13q (*DLEU* genes, miRNA-15a/16-1, *RB*), 20q (*L3MBTL1*), 22q (*PRAME*) chromosome regions which overlapped with genes associated with hematologic cancer (some examples are specified between brackets) (31). Schick et al. almost obtained identical results since they most frequently detected mosaic deletions in regions overlapping with leukemia-related genes as 2p (*DNMT3A*), 4q (*TET2*), 13q (*DLEU* genes, miRNA-15a/16-1, *RB*), 17q (*NF1*) and 20q (no candidate genes mentioned) (34). Finally, Jacobs et al. also detected 13q and 20q as the most common mosaic deleted regions. Importantly, they observed that 2/4 CMEs carriers with diagnosis of chronic lymphocytic leukemia (CLL) after mosaicism detection harbored the typical 13q14.3 deletion CLL-related affecting some *DLEU* genes and miRNA-15a/16-1 in mosaicism in blood obtained as many as 14 years before diagnosis. With this, they suggested that detectable clonal mosaicism could be a marker for early hematological cancer detection or even its precursor form, such as CLL and monoclonal B-cell lymphocytosis (MBL) respectively (30,75).

Altogether, these results are the first step to define a possible cause-effect relationship between CMEs affecting certain genome regions and the development of hematologic malignancies speci-

fically. Regarding solid tumors, less is known but there are some slight insights about a possible causative effect of CMEs on solid cancers. Previous published data showed that four unrelated bladder cancer patients harbored similar proportion of cells carrying the same CMEs both in the blood and bladder mucosa demonstrating an early embryonic origin of events and opening the door to a possible cause-effect relationship between CMEs and bladder cancer (32). In agreement with these data, in our sample set, we detected one subject (FA13) who harbored two CMEs in both blood and tumor (anal squamous cell carcinoma (SCC)) samples and at different mosaicism percentage suggesting a possible link between clonal mosaicism and tumor development. Moreover, we identified another subject (FA825) carrying five CMEs in saliva who was diagnosed of esophageal cancer in parallel with sample collection. Given the proximity between buccal cavity (from where saliva was taken) and esophagus under a physical and embryologic origin point of view, this data suggested that saliva CMEs could be related somehow with the diagnosed esophageal neoplasia. However, analyzing tumor DNA for detecting the same CMEs observed in saliva would be crucial to stablish whatever potential relationship between CMEs and tumor development.

Although data above explained hints some mechanistic relationship between CMEs and cancer, we actually are far from describing specific CMEs responsible of certain malignancies. In fact, we are still unable to determine whether CMEs that we are detecting are cause or consequence of malignancies. However, all these data would fix with both ideas and, consequently, with the fact that CMEs can be a cancer biomarker.

CMEs detected in FA patients have an early embryonic origin

Studying two tissues derived from different germ layers, as blood (from mesoderm) and saliva (from ectoderm), would be useful to determine a possible early and common embryonic origin of CMEs (32). From 6 CMEs carriers, we had available saliva DNA besides blood DNA where mosaicism was initially detected. We observed that all 6 mosaic carriers harbored their CMEs both in blood and saliva revealing an early embryonic origin of these events. Importantly, mosaicism percentage of events in saliva were similar or higher than the observed in blood in all 6 cases, demonstrating that finding those CMEs also in saliva was not secondary to sample

contamination with blood cells carrying CMEs but by the presence of those rearrangements also in cells from buccal cavity. In other words, since up to 74 % of the DNA in saliva is estimated to come from white blood cells (268), in case of observing CMEs in saliva due to contamination with blood CME carrying cells, we would expect to detect those events at lower mosaicism percentage, below the fraction observed in blood, and this was not our case.

There were 2 out of these 6 CMEs carriers where saliva was obtained 3 and 7 years after blood sample collection respectively giving some kind of evolutionary information but indirectly. In both cases, we observed an increase of mosaicism percentage in saliva. However, since we had information from two different tissues we could not conclude whether mosaicism level was globally increasing or specifically only in saliva.

Similarly, we detected a FA patient that in 2010 did not carry any CME in blood but, after 6 year follow-up, he presented an interstitial mosaic deletion 20q in saliva coinciding with the diagnosis of a MDS. This data could suggest an individual with acquired clonal mosaicism along the time with possible consequences on cancer development. However, because of the time lapse between blood and saliva collection, we could not obtain a firm conclusion and different options were compatible with our observations. In case of finding the 20q mosaic deletion in a blood sample from the same time that saliva was collected (2016), we would have identified a new CME carrier after 6 years of follow-up considering that blood from 2010 did not harbored any CME; moreover, we would also have verified the early embryonic origin of the mosaic event. On the contrary, in case of not finding the 20q mosaic deletion in a blood sample taken in 2016, the emergence of a CME specifically in saliva along time would be suggested. However, testing the presence of the mosaic rearrangement in saliva from 2010 would be crucial to really confirm this possible new CME carrier. Again, the key point in this case would have been to perform the longitudinal study correctly by analyzing the same tissue along the time or the transversal study properly by studying different tissues but sampled in parallel.

Despite sampling limitations, confirming CMEs' early embryonic origin in all six tested CMEs carriers would be in agreement with the hypothetic mechanism underlying mosaicism in FA patients that suggest that FA/BRCA pathway dysfunction would generate genomic heterogeneity in somatic cells since zygote formation and

affecting all body because FA mutations are constitutional and transmitted though germline. The positive selection of one of these clones carrying a CME for whatever reason, as a selective advantage for example, would lead to its detection due to affecting a proportion of cells over a certain threshold.

Mosaicism-age association also in FA

A part from the increased CMEs prevalence among FA patients and the associated higher risk of cancer with poorer prognosis, we observed a higher mosaicism prevalence among individuals between 19-50 years of age (34%) compared to those below 18 years (7%) in our sample set concordantly with previous data (5,25,30–32,70,76,257). A possible mechanism underlying this association could be that cells with abnormal karyotype could arise easily early in the human development as previously suggested (260) and become detectable with the pass of the time. Specifically, a single stem cell would acquire a somatic event during embryogenesis and could become apparent when cell population diversity decreases with age and the remaining cell population become oligoclonal. Another possible explanation for increased mosaicism prevalence among elderly individuals could be augmented rates of somatic mutations, higher ability to form large clones or diminished genomic stability maintenance such as incremented telomere attrition leading, as a result, the clonal proliferation of somatically altered cells. The existence of a survival bottleneck could also allow observing some cells carrying CMEs when they overpass the threshold for their detection after their clonal expansion due to a positive selection. Finally, the influence of environmental exposures along the life may contribute to the increase of mosaicism with age (30,31). Despite all these possibilities, further work is required to uncover the underlying mechanisms in age-mosaicism association, particularly in respect to how and when altered clones appear, tissue specificity and the expansion timing of cell populations with age.

Somatic UPDs as biomarkers of hematologic malignancies

Although cancer development in FA patients has been traditionally associated with chromosome copy gains and losses, we detected that copy neutral changes could play an important role on MDS/AML and SCC development in FA patients since 17.14% (12/70) of

CMEs detected in 26 FA patients were UPDs. Our observations were in agreement with the generally increased UPD prevalence reported in a wide range of malignancies (both solid and hematologic) (60). The reason why UPD prevalence is high in a vast list of cancers could be that loss of heterozygosity (LOH) through UPD is the major mechanism associated with somatic tumor suppressor genes inactivation (59) and oncogene activation (60) and, consequently, with cancer development. On this direction, several examples of UPDs leading to disease have been reported. Interestingly, UPD affecting chromosome 16 leading to *FANCA* and *FANCP* mutations homozygosis has been recently observed in four FA patients becoming the first published cases where UPD was the cause of the disease. Thus, these families would have a reduced recurrent risk of having another child with FA since only one parent was actually FA carrier (269). UPD promoting MDS/AML through mutations homozygosis in *FLT3* (13q), *WT1* (11q), *NF1* (17q) genes has been widely reported (62,63). Another remarkable example would be UPD13q leading to homozygosis of deletions affecting the typical 13q14.3 CLL-related region (including *DLEU* genes and miRNA-15a/16-1) or *Rb* gene promoting CLL and retinoblastoma development respectively (64,65). Finally, JAKV617F mutation homozygosis through UPD9p is a well-known example responsible of some myeloproliferative neoplasms (MPN) as polycythemia vera (PV), primary myelofibrosis (PMF) or essential thrombocytosis (ET)) depending on the JAK2V617F mutational burden (60). Besides being a source of an important proportion of cancers under the molecular point of view, UPDs can also arise as consequence of the neoplasia itself due to the existence of generalized chromosome instability. Thus, this also would contribute to the increased UPD prevalence in lots of cancer types.

Among eight non-leukemia individuals harboring both mosaic UPD13q and del13q14.3 in blood, we detected that the UPD13q was acting as a second hit mechanism leading to del13q14.3 homozygosis in six of them. With these results, it would be logical to propose that these six subjects could actually be in the starting point of a malignant process in agreement with the classical view of UPD as a source of LOH associated to CLL development when affecting 13q14.3 deleted region. However, despite of the well-known causal relationship between 13q14.3 deficit and CLL, detection of del13q14.3 in homozygosis thought UPD13q in individuals undiagnosed of leukemia does not automatically mean that the disease is or will be present in them. What this data reflects is that monoclonal B-cell sometimes harboring del13q14.3 (in homozygosis through

UPD13q or in heterozygosis) can expand in peripheral blood of asymptomatic persons, especially in aging people. This condition, known as MBL, is the precursor stage of all CLL diseases but does not always progress to cancer (144,145). In fact, a mosaic del13q14.3 frequency of 0.073% has been reported in non-leukemia individuals and also an accumulation of this mosaic event with age (75). Thus, early detection of mosaic UPD13q and/or del13q14.3 in blood could identify an important proportion of future CLL patients despite of some possible "false positive" cases corresponding to MBL cases that will not progress to CLL. However, the possibility of that these rearrangements could appear as a result of the impairment in genomic maintenance capacities associated to aging could not be discarded due to the documented increased del13q14.3 prevalence with aging.

Finally, we found an individual carrying an UPD9p that lead to somatic JAK2V617F mutation homozygosis and, after 3 years, PV disease. Thus, we were able to detect different individuals with UPD leading to deletion/mutation homozygosis that could potentially cause two different hematologic malignancies (CLL or PV). Consequently, we found two mosaic UPDs that could possibly be used as hematologic cancer biomarkers. Moreover, the mechanism underlying these processes would be the same: a somatically acquired genetic alteration (mutation or deletion) associated with an specific disease would reach the homozygosis status due to a somatically acquired UPD affecting the same region. Due to a selective advantage, these homozygous clones would be positively selected and clonally expanded, leading to a certain disease once reaching a mutational burden (**figure 2**).

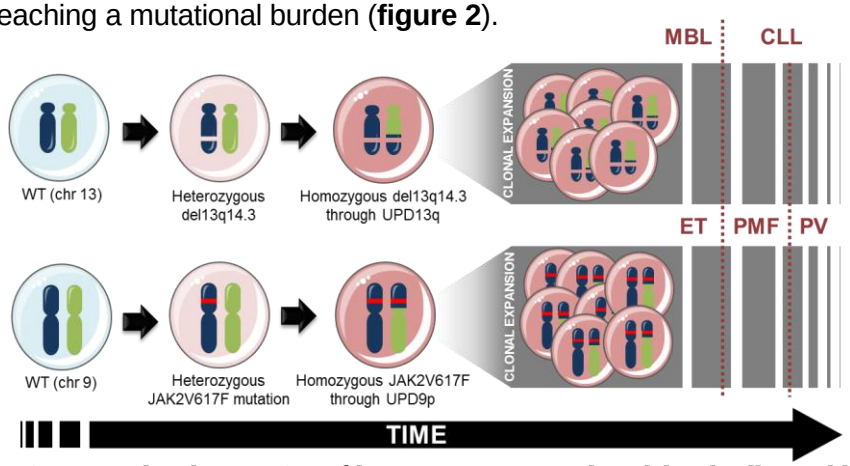


Figure 2. Clonal expansion of homozygous mutated or deleted cells would be the underlying causative mechanism of some hematologic malignancies. From *Servier Medical Art*.

MOSAICISM AS A RESCUE MECHANISM IN HEMATOLOGIC CANCER DEVELOPMENT

The other side of UPDs: a rescue function

Despite the well-known and above explained role of UPDs on causing disorders by leading mutations homozygosis, an opposite function has also been described: UPDs can select the WT allele of a given mutated locus, generate "health clones" in mosaicism and, ultimately, modulate disease development. There is the growing idea that this phenomenon is more frequent than what would be expected having an important influence on disease expression and patients' treatment. Indeed, reverting cells represents a "natural gene therapy" that could be the base of several genetic therapy approaches. This "rescue" function of UPDs has been described in some severe sporadic skin diseases. For example, Ichthyosis with confetti (IWC) is caused by dominant mutations in *keratin 1* or *10 (KRT1, KRT10)* genes and is characterized by the accumulation of thousands "health" skin spots harboring homozygous WT alleles through UPD17q. The observed high frequency of somatic reversion in IWC patients suggests that reverted clones are under strong positive selection and/or the reversion rate is elevated (68,270). Similarly, six patients with dyskeratosis congenital (DC) from four different families were detected to have somatic reversion thought UPD3q of the mutant *TERC* (telomerase RNA component) allele showing that reverting mosaicism is frequent in this disease and can have some implications on DC phenotype variability (69).

But, what is about cancer? Jacobs et al. reported that the four regions concentrating CMEs most recurrently and spanning well-known cancer genes were chromosomes 9p (copy-neutral LOHs), 13q (deletions), 14 (copy-neutral LOHs) and 20q (deletions) (30). Interestingly, they observed that most recurrent mosaic events in these regions were detected in cancer-free individuals as well as across multiple individuals with solid tumors but less frequently in leukemia subjects. Moreover, frequency of these mosaic events was comparable in cancer-free individuals and solid cancer patients for three of the regions whereas chromosome 14 copy-neutral LOH abnormalities were more frequent in non-hematological cancer cases. The fact of finding a higher frequency of UPDs in patients with solid tumor and cancer free individuals compared to leukemia patients could be explained by the existence

of UPDs acting as a second hit mechanism in the first group but by the possibility of UPDs with rescuing functions in the second one.

Altogether, there is consistent evidence that UPD can have a recue role in, at least, some severe sporadic skin diseases. Therefore, uncovering the possibility of rescue UPDs in the context of cancer diseases could have a positive impact on patients' prognosis and developing treatment's strategies.

Somatic UPD3p in the rescue of FA

In the context of mosaicism prevalence study in FA, we identified five patients harboring a mosaic UPD3p encompassing the *FANCD2* mutated gene without any additional CMEs. This mosaic rearrangement was detected both in blood and saliva for four cases demonstrating the early embryonic origin of the event (in one case no saliva sample was available). All these five FA individuals were compound heterozygous for *FANCD2* mutations and the less truncating mutation was the one selected by the UPD3p. These data together with the fact that none of the UPD3p carriers had a reported cancer diagnosis would suggest that UPD3p was acting as a rescue mechanism trying to preserve at maximum FANCD2 function by reducing the prevalence of the most truncating allele. Considering these findings, it would be important to note that mosaic UPD3p is not actually a CMEs biomarker of increased cancer risk but a potential protective factor. Thus, when estimating cancer free time after mosaicism detection in CMEs carriers versus non-carriers by using Kaplan-Meier algorithm, these "protective" rearrangements should be excluded (**figure 3**). With this, although excluding only 5 individuals from the analysis, we even obtained a little bit higher hazard risk (HR) ratios compared to the previous analysis shown in chapter 1 (part 2), reinforcing the idea of shorter cancer-free time in CMEs carriers compared to non-CMEs carriers.

UPD3p found in all five FA patients had a right breakpoint coinciding with a region known to be typically homozygous in Africa, Middle East and Europe populations (271) suggesting a new mechanism of UPD where the two identical DNA strands due to the homozygosity could favor homologous recombination and UPD emergence. UPDs usually include an entire chromosome but they sometimes are limited only to a subchromosomal region generating what is known as segmental UPD. Ohtsuka et al. identified four consensus motifs for mitotic recombination frequently appearing within breakpoint

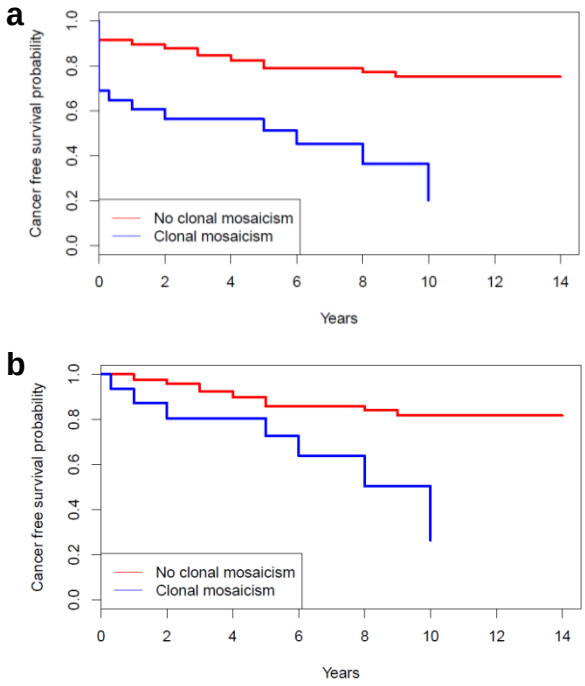


Figure 3. Kaplan-Meier (KM) plots show even slightly shortened cancer free time in CMEs carriers compared to non-carriers after excluding protective UPD3p. a) KM performed considering all cancer cases (prevalent and incident cancers after sampling) and doing an age-adjustment (n=156, 6 no informative cases, 42 events of cancer. HR=4.1, CI95%=2.1-8.0, $p=2.0x10^{-5}$). **b)** KM performed considering only incident cancer cases after sampling and doing an age-adjustment (n=135, 6 no informative cases and 21 exclusions due to cancer free time equal to 0, 22 events of cancer. HR=4.3, CI95%=1.6-10.3, $p=2.5x10^{-3}$).

regions of segmental UPD11p in nine patients with Beckwith-Wiedemann syndrome. In addition, seven consensus motifs previously found in meiotic recombination hot-spots were not detected in these UPDs suggesting that mechanisms of mitotic recombination leading to segmental UPDs are different from those of meiotic recombination. Doing similar analysis in UPD3p breakpoint intervals found in FA patients would give some light about possible mechanisms underlying rearrangements emergence in this new hot-spot.

Despite more studies remain to be done to better define mechanisms underlying rescue UPD, what it is clear is that not all mosaic events should be considered as cancer risk biomarkers, especially when one isolated UPD is detected in a clinically stable individual. In these cases, depending on UPD localization versus cancer-related genes, the possibility of a rescue UPD should be considered.

Somatic UPD13q in the rescue of CLL

Although UPD13q leading to 13q14.3 deletions (del13q14.3) homozygosis is a well-known cause of CLL, Rodríguez-Santiago et al. reported a mosaic UPD13q together with a mosaic del13q14.3 in two unrelated bladder cancer patients without any reported hematologic problem suggesting a potential relationship between both events different from the known cancer-inducing effect (32). Following this data, we performed a meta-analysis of 70144 individuals without any diagnosis of leukemia and studied by SNP array for mosaicism detection. We detected a prevalence of UPD13q-del13q14.3 around 20.51% confirming, in this way, the initially suspected selection of this co-occurrence in the general population by Rodríguez-Santiago et al. This co-occurrence was also detected with a prevalence of 5.01% in a group of 722 CLL patients. Finding a higher co-occurrence of both events in the non-leukemia group compared to the leukemia one could be explained somehow by the fact that del13q14.3 found in non-CLL group were smaller than in CLL group having less probability to cause cancer and consequently having more chance to be detected together with UPD13q during more the time.

At this point, we wanted to further investigate the potential mechanistic or functional relationship between these two events that could explain their unexpected high prevalence in non-leukemia subjects. The frequency at which these two rearrangements were detected together in our sample set (20.51%) was not significantly different to the published co-occurrence prevalence of UPD-deletion (46.6%) in other genome region and in the context of other cancer disease as Neurofibromatosis 2 (50). This data together with the fact that UPD13q and del13q14.3 found in our sample set were very different in size in all cases (data not shown) would discard a mechanistic relationship between UPD13q and del13q14.3.

Since our data revealed a high prevalence of UPD13q-del13q14.3 in non-leukemia individuals, we suspected two different options: the possibility of having an important proportion of individuals with MBL condition caused by del13q14.3 homozygosis through UPD13q or being in front of another example of rescue UPD, in this case, trying to avoid MBL/CLL development. To further investigate this, we used mathematic approaches together with a simulator developed by our group to infer the percentage of cells harboring each rearrangement (separately or together) by using SNP array parameters (LRR and BAF) in eight non-leukemia individuals with both mosaic events. In

other words, trying to determine whether SNP array plots obtained in all nine samples (corresponding to eight individuals since one subject had two samples obtained at different times) with both rearrangements were compatible with having an UPD13q acting as a second hit mechanism leading to deletion homozygosis or having a rescue UPD. With this analysis, we detected that 7 out of 9 samples (corresponding to 6 out of 8 individuals) had an UPD13q acting as a second hit mechanism leading del13q14.3 homozygosis (explained in the previous section). Interestingly, three of these samples (2 individuals) presented another extra UPD13q different in size to the one acting as a second hit mechanism. This extra UPD13q was selecting the WT allele (corresponding to the same allele deleted in other cell fraction harboring the del13q14.3) and acting as a rescue mechanism of the global 13q14.3 deficit, surely in 1 sample (1 individual) and possibly in 2 samples (1 individual). To functionally compensate 13q14.3 deficit, we propose that the WT allele selected by the UPD, and consequently the allele deleted in the other cell fraction, should correspond to the unmethylated and functional 13q14.3 allele. Finally, we did not obtained conclusive results for the two remaining individuals but our data would be more in favor of considering the UPD13q as a rescue in one of them. Interestingly, in 2 out of 8 individuals, we found UPDs in other chromosomes different from 13 with a very similar percentage of mosaicism indicating a common origin in time.

Somatic UPD in the rescue of FA and CLL: similarities and differences

Our findings suggest an important evidence of the rescue function that UPDs can play in two different scenarios, FA and CLL, and we would distinguish between two quite different rescue mechanisms. In FA, considering that UPD is a rather common event at embryonic stages but it does not use to show up due to the low proportion of cells with the rearrangement, a preexisting clone harboring an UPD3p selecting the most benign/less truncating *FANCD2* allele would be positively selected and clonally expanded to guarantee the maximum protein function in the patient (**figure 4a**). In CLL context, considering that secretory tumor-suppressive miRNAs as miRNA-15a/16-1 can act as a death signal in a cell competitive process/ paracrine way (132), we hypothesize that del13q14.3 affecting the functional (unmethylated) allele would lead to a microenvironment defective in miR-15a/16-1/*DLEU1*/*DLEU2* function. Consequently, a proliferative and antiapoptotic media suitable for tumor formation

would appear, being exacerbated by the emergence of clones homozygous for del13q14.3 through an UPD13q. Then, we propose that the existence of a microenvironment deficient in 13q14.3 function would promote a positive selection of a putative preexisting clone with UPD13q of the functional allele (different from the previous one acting as a second hit mechanism) in order to functionally rescue the global MDR deficit and going against MBL/CLL development (**figure 4b**). Thus, besides the classical involvement of UPD in tumorigenesis by LOH secondary to second hit mechanism, UPD could also play a remarkable role on going against tumor development. We have identified two UPD rescue mechanisms in two hematologic malignant conditions that, although being different, they have a common characteristic: clonal expansion of the most advantaging clone is the basis of both processes since the ultimate goal is restoring, at the maximum possible extend, the altered protein function.

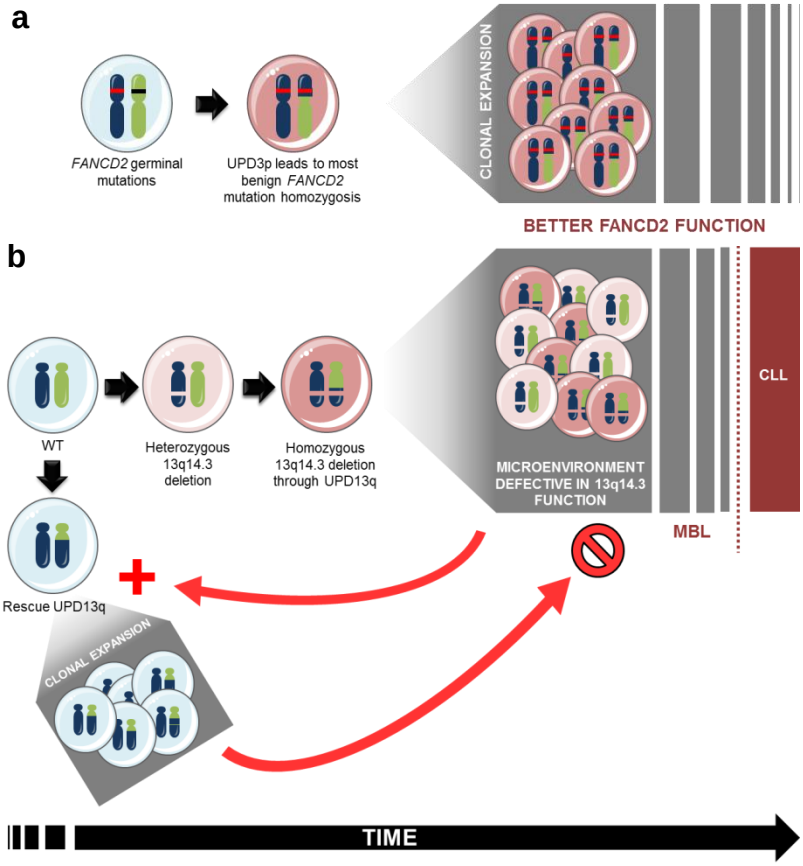


Figure 4. UPD rescue function in FA and CLL. From *Servier Medical*

GENETIC COUNSELLING

The importance of the genetic counselling in FA families

FA has always been considered a pediatric disease due to the low life expectancy in last decades so that FA patients have been traditionally controlled by pediatric hematologists and, occasionally, by other specialists depending on their needs. However, the improvement of treatments' quality and efficacy are more and more enlarging patients' life expectancy and increasing the need of FA patients' transition to adult services to receive a well-orchestrated multidisciplinary follow-up by Otorhinolaryngology, Maxillofacial medicine, Hematology and Gynecology. This transition of health care would have a positive impact in several aspects as avoiding the over accumulation of FA patients in pediatric services, ensuring that FA patients are controlled by adult specialists and helping young adults to assume a personal responsibility for their healthcare (169). The transition from pediatric to adult follow-up system is an important issue to be carefully addressed in chronic and multiorganic illnesses as FA since there is clear evidence that an anticipated and coordinated transition process benefits patients and their families. In fact, specific transition programs have been successfully developed for patients with cystic fibrosis (CF) and other chronic diseases (254). Since some programs have denoted that the most successful transitions are those initiated during the late teenage years and those in which families receive education about future issues (254), FA patients attended in Hospital Vall d'Hebron (Barcelona) tend to do the transition to adult system controls around 16 years of age. Despite of recommendations and probably due to the lack of a specific transition protocol for FA patients, we found that 4 out of 6 pediatric patients of our cohort would have probably already done the transition to adult care considering their age over 16. However, as previously mentioned, the age is not the unique factor to be considered to initiate the transition and other aspects, as information received or patients' maturity could play a key role on this.

One of the aspects that would encourage FA patients' transition to adult follow-up system is to guarantee to them a complete and well-orchestrated multidisciplinary follow-up, definitely, make to feel them sure, comfortable and equally controlled as in pediatric

services. For this purpose, in Hospital Vall d'Hebron, there is the *Unitat d'Alt Risc i Prevenció de Càncer (UARPC)* consisting of a medical oncologist, nurses and genetic counsellors that constitute a FA referral unit. One of the main functions of this unit in FA context is receiving FA families once pediatric stage has finished by doing some initial visits to place patients and their family in the new scenario. During these visits, a complete medical record including patients' personal medical history and family history is performed as well as an explanation of FA basic concepts that families need to know or reinforce. Explaining in detail how the medical follow-up will be as well as the healthy habits that are crucial to follow for FA patients are the main points that will be addressed during not only the first visits but also along patients' supervision in the unit. Moreover, during first visits, a genetic counselling process will be performed in order to identify relatives at risk of having affected offspring or even to introduce the possibility of reproductive strategies to avoid mutation transmission depending on the familiar context. Genetic counselling will have an especial importance in those families with identified mutations in FA genes associated to increased cancer risk in carriers. So, in these cases, an accurate genetic counselling and an extension of genetic studies to other risky relatives should be offered to also promote cancer prevention strategies. In those cases where FA mutations are not identified, professionals can offer to patients the possibility of participating in studies based on the use of new technologies as Whole Exome Sequencing to try to identify the gene mutation/s responsible of the disease in his/her specific case.

Once, patients and relatives are situated in the new scenario and have entered in the new dynamics, genetic counsellors will have as main functions to (I) guarantee a complete and well-orchestrated medical follow-up without losing controls with any specialist and encouraging patients' adherence to all visits, (II) keep patients and relatives informed about the most important issues and concepts related to FA disease, solving doubts when necessary, and (III) provide psychological support to both patients and relatives.

Considering all these points we wanted to determine whether FA patients visited at least once by a genetic counsellor in the UARPC underwent a follow-up according to the Spanish guidelines' recommendations (205) which, in fact, were stablished according to the American guidelines from *Fanconi Anemia Research Fundation, Inc* (169). Firstly, we detected that almost all patients had follow-up, at least for a period, in all specialties required for guidelines (Otorhino-

laryngology, Maxillofacial Medicine, Hematology and Gynecology) with the exception of Dentist from which very poor information was recorded. Importantly we detected only one case of loss of follow-up in last two years in all specialties which was not explained by the exitus of the patient nor care controls in other center showing the efficacy of genetic counsellors on promoting a complete FA follow-up. Secondly, regarding time between visits, we observed underfollowing in Hematology and overfollowing in Gynecology with a tendency to concentrate all visits every six/seven months, probably in order to improve patients' quality life by reducing hospital stays but without compromising their health. Thirdly, we detected a drastic underfollowing through BMAs, as explained and discussed at the beginning of the section, together with also slightly poor following through HPV testing and hormonal controls. Finally, we detected a good HPV vaccination rate among FA female patients and, because of discrepancies regarding the possible role of HPV in HNSCC in both male and female FA patients (238–241), no vaccination among male patients was recorded. Importantly, after this work, male vaccination was started in UARPC in agreement with publications that support HPV contribution to HNSCC in both male and female FA patients. Altogether, our results showed a global accomplishment of recommendations for FA chronic follow-up established by the guidelines, although some aspects, as low BMAs following, should be urgently addressed. These observations would suggest that genetic counsellors are succeeding on coordinating controls in all specialties in terms of patients' adherence and low loss follow-up rate. In this sense, the main reason for adherence to follow-up reported in our sample set were specialists' coordination in the same center, receiving visits reminders and scheduling controls in the same day.

Adherence to chronic and multidisciplinary follow-up is a serious concern especially during adolescence and young adulthood where behaviors that increase the risk of cancer are more prevalent (169). FA patients are highly recommended to avoid cancer agents as tobacco, alcohol, sun exposure without protection and unsafe sexual conducts. Initially, we detected a high rate of habitual smokers in our cohort but half of those habitually visited by genetic counsellors stopped smoking in last two years, showing the important role and effectiveness of these professionals in terms of health education.

In agreement with genetic counsellors' functions above explained, we also wanted to determine whether FA patients visited at least once by a genetic counsellor in the UARPC or by a similar professional in the pediatric service had a good FA disease

knowledge and basic psychological necessities fulfilled. It has been demonstrated that good patients' and parents' disease knowledge on chronic onset childhood illnesses is important to ensure adherence to treatments and follow-up. Misconceptions, gaps and errors in CF knowledge, for example, have been proven to result in non-adherence to treatment impacting on the progression and outcome of the disease (272). In our sample set, without sex neither remarkable patients-parents differences, we detected poor knowledge on knowing recurrence risk of FA; intermediate knowledge in FA manifestations and being conscious of the correct familiar molecular status; and finally, high knowledge in concepts as FA frequency, age at FA clinical diagnosis and FA genetic origin. Regarding disease psychological impact, mothers showed a higher intensive worrying towards FA manifestations compared to fathers and patients. In fact, we detected a general mothers' worrying towards everything related to the disease, since we observed that mothers had worrying in both situations of knowing or not knowing FA genetic cause in their child compared to fathers, who preferred knowing genetic cause, and patients, who were comfortable in both situations. Considering that molecular analyses results are forgotten overtime and that most of parents that did not know FA mutation showed high worrying and anxiety, genetic counsellors should ensure that all families with identified mutations know correctly this information to avoid at maximum the anxiety linked to the lack of information. Importantly, a general lower worrying towards FA manifestations in patients compared to their parents was detected probably explained by the fact that patients usually do not feel that they are "sick" and they are not really conscious about their real risks. In this sense, genetic counsellor should work on this aspect and ensure good patients' healthy lifestyle and follow-up adhesion.

Thus, we observed a general accomplishment of guidelines recommendations for chronic FA follow-up and good patients' adherence to these controls although some issues as BMAs testing should be urgently addressed. We also detected quiet good FA knowledge in both patients and parents although some effort should be done to solve the identified errors, gaps and misconceptions. Finally, psychologic necessities were found to be somehow unfulfilled, especially in mothers' subgroup. Altogether, our data would suggest that regular contact of FA families with genetic counsellors or other health professional with similar roles have positive effects on promoting FA chronic vigilance and FA knowledge but some emotional necessities are quiet uncovered and more psychological accompaniment should be done.

Implications of CMEs on genetic counselling

According to our data, we propose that detecting CMEs in easy to obtain samples as blood or saliva could be used as a biomarker of cancer risk. Thus, the situation of having to communicate to an individual that is carrying one or several CMEs should be considered together with all questions that both patients and relatives could do. One important aspect that we should be aware of is that when the word “genetics” is heard by an individual without a scientific background, it is often misused as a synonym of “inheritable”. Thus, in a putative situation where communicating CMEs detection in an individual, it is logical to think that one of the questions that he/she could ask could be whether these CMEs can be transmitted to his/her offspring. Here, the key concept that should be clearly explained with adapted language is that all these rearrangement are in mosaicism, so, they have appeared after zygote formation and do not affect all body cells. However, we do not know whether germline cells are affected or not by the rearrangement and, in case of affectation, up to which extent. Despite of this, a hypothetic embryo carrying one CME detected in a progenitor would not harbor this rearrangement in mosaicism but constitutionally. Depending on the event, this pregnancy would derive into a miscarriage due to the unviability of the embryo or could continue but as a risky pregnancy needing further studies to address the real risk of affected offspring.

In the specific case of FA patients, the same explanation should be performed but considering several additional issues. Firstly, individuals with FA are known to present reduced or absent fertility, more evident in male FA patients (169). These reproductive difficulties could be influenced in some cases by the presence of CMEs in the germline and the derived miscarriages probably undetectable most of times. Secondly, reproductive issues are acquiring an increasingly importance for FA patients since they more and more frequently reach reproductive ages. Thus, explaining the real possibilities of reproduction due to FA condition, the implications of having a baby in the context of a family where a progenitor is affected by FA, the risk of FA mutations transmission to the offspring and the possibility to study the other member of the couple for FA carrying condition testing should be clearly explained besides possible CMEs inheritability.

LIMITATIONS OF THE STUDIES

Although the main objectives of this thesis have been reached (this will be further discussed in "Concluding remarks" section), some limitations could have obstructed to obtain firmer conclusions or definitive results in some of the issues treated in this work.

Regarding the first chapter of the thesis, difficulties on having a complete medical record due to the sometimes slow communication with physicians or the loss of patients' follow-up may influenced a little in our study since some individuals had to be discarded for certain analysis due to the lack of certain information. Fortunately, these excluded individuals did not change the main results and conclusions derived from this work. The longitudinal study of CMEs carriers to monitor mosaicism evolution and new cancer diagnosis potentially related to CMEs accumulation as well as the longitudinal study of non-CMEs carriers to detect new CMEs and cancer diagnosis associated to them were a key issue that could definitively demonstrate the use of CMEs detection as cancer biomarker. Due to the small sample size, our longitudinal study did not bring light in this sense. Thus, having a more extensive longitudinal study in terms of number of individuals serially analyzed and the average vigilance period evaluated would be needed. Importantly, we possibly detected a new CME carrier coinciding with MDS diagnosis but this observation was derived from the analysis of blood and saliva samples taken with several years of difference. The collection of samples from different tissues (blood and saliva) was initially thought for testing the possible early embryonic origin of CMEs detected. In order to do it correctly, all blood and saliva samples should have been taken in parallel or with one year of difference at maximum. In the case above mentioned, both samples were not obtained coordinately, as some other few cases, leading to confusion at the time to obtain firm conclusions. In this case, we were not able to distinguish whether the rearrangement detected was acquired along the time or was present in saliva until the beginning. This situation provides an example of how important is to know what we are measuring in every moment and the time at which samples should be collected to perform useful analysis. In case of doing transversal analysis, samples from the same or different tissues should be collected in parallel whereas in case of longitudinal analysis, the same tissue should be studied along the time. Fortunately, this situation was detected in very few cases and only induced to confusion in the example here explained. Finally, although being aware of the

bureaucratic difficulties to access to patients' biopsies or tumor excisions, determining the presence of blood or saliva CMEs also in solid tumors developed after mosaicism detection would be a key strategy to reinforce the idea of mosaicism detection as a biomarker of also solid tumors.

Regarding the second chapter of the thesis, the main limitation that we had was that the rescue function of UPD13q in MBL/CLL development was based on SNP array data and the use of mathematic approaches and simulations. Although some experimental validations were done, they did not confirm our data and designing new experimental assays to really obtain firm conclusions, similarly as in the case of UPD3p rescuing FANCD2 mutations, would have been very useful. However, considering that non-CLL individuals were from EGCUT project and we did not have easy access to their blood DNA in combination with the fact that findings related to UPD13q rescue function were done at the end of the project, no more informative experimental assays could be designed to reinforce our observations. Possible limitations in the study of UPD3p rescue function on FANCD2 patients could be the fact of having some unknown *FANCD2* mutations and some unavailable parental samples. However, these issues did not negatively influence to our final message.

Finally, in the third chapter of the thesis we were conscious of some limitations that could influence in our data although we obtained remarkable results. Regarding the evaluation of FA adult follow-up in Hospital Vall d'Hebron, similarly to the first chapter of the thesis, the main problem that we had was incomplete medical records (especially in those patients with external controls) and this fact could diminish the strength of our results. In the assessment of the importance of a genetic counsellor in a FA unit, the main difficulty was that our initial available cohort was small because of the low frequency of the disease and the fact that we were restricted to patients from one hospital. Consequently, we had to considered pediatric patients no specifically attended by genetic counsellors but by professionals with similar functions in order to increase or sample size. The existence of these "two subgroups" attended by different health professionals could be responsible of the confusion detected in certain FA concepts, may be secondary to differences in the way of transmitting the information. Despite of this, our final sample size was also not enough to do certain statistical analysis as determining main reasons for no adhesion to FA follow-up among patients without follow-up or with incomplete medical care since we only had two

individuals in this group. We had another important limitation which was that we could not perform all questionnaires face to face and most of them were done by phone due to participants' availability. This difference in the method could introduce bias in the final results, however, we read-aloud all questions to participants always in the same way in order to add homogeneity to the process and obtain valuable and comparable results. Finally, low knowledge and confusion detected in some FA concepts could be attributable to the quiet lower intelligence quotient sometimes found in some FA patients.

CONCLUDING REMARKS

Genetic mosaicism is more and more considered as modulator even a possible cause of many disorders. In fact, the possible use of detectable CMEs as early cancer biomarker has been proposed by different groups during the last years. Thanks to the development of new high-throughput strategies with higher resolution as SNP array technology, CMEs at increasingly lower mosaicism percentage can be detected.

This thesis proposes that genetic mosaicism can have a dual role, depending on the type of rearrangement and the genome region affected. The vast majority of CMEs would be considered as cancer biomarkers (to either solid or hematologic malignancies) whereas certain CMEs, specifically segmental UPDs, could act as cancer protective factors in certain contexts. Here, we have detected a high CMEs prevalence in FA patients together with low rate follow-up thought BMAs. Thus, incorporating the regular use of SNP array technology in FA patients' vigilance for early cancer detection would have a positive impact in terms of improving cancer patients' prognosis and survival (due to an earlier cancer diagnosis) and patients' adherence to FA follow-up (due to the low invasiveness of the technique). Specific mosaic UPDs potentially used as biomarkers of CLL and PV have been also reported by us in agreement with previous published data.

In the line with previous knowledge about UPDs' role one mutation rescuing in severe sporadic skin disease, this thesis shows two examples of UPD to the rescue of hematologic disorders: CLL and FA. The detection of these CMEs could be particularly important for the management of these patients.

Finally, genetic counselling has been proven to have a positive impact on FA families in terms of follow-up guidelines accomplishment by the hospital, patients' follow-up adherence, families' knowledge about the disease and psychological support (**figure 5**).

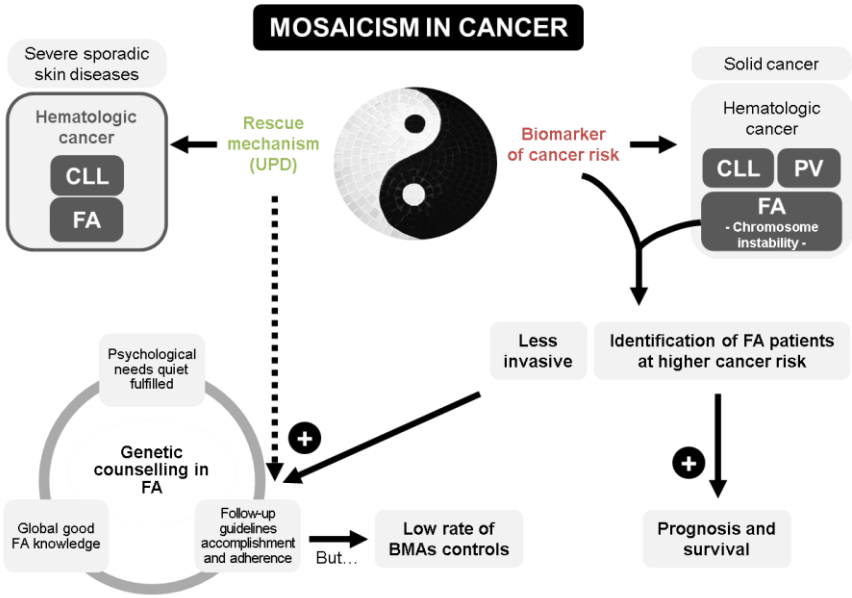


Figure 5. Dual role of genetic mosaicism in cancer. A summary of main messages derived from this thesis us shown in this figure. Note that diseases in which we have described a possible role of detectable CMEs as cancer biomarker and/or rescue mechanism are marked in dark grey.

CONCLUSIONS

CONCLUSIONS

1. FA patients present 170-180 fold increased risk of harboring detectable CMEs in blood and/or saliva in agreement with previous knowledge that report increased genetic mosaicism prevalence in cancer risk populations.

2. Age is an influencing factor of detectable mosaicism prevalence also in FA disease since adult FA patients show higher prevalence of detectable CMEs than younger patients.

3. CMEs carriers present a global 5.6 fold increased risk of cancer leading to shorten cancer free time and 3 fold higher exitus rate. This make evidence that detectable clonal mosaicism could be a biomarker to identify those FA patients at higher risk of cancer.

4. All tested CMEs had an early embryonic origin since they were detected in both blood and saliva, derived from mesoderm and ectoderm germ layer respectively.

5. Although longitudinal studies are good strategies to detect new CME carriers and secondary malignancies along time, we did not detect any new mosaic case during an average 5 years period of vigilance of a small group of FA patients. However, a possible new CME carrier was identified when comparing blood and saliva samples obtained with a time lapse of 6 years.

6. Around 90% of FA CMEs carriers with MDS/AML diagnosis harbored leukemia-associated events (1q gain, 3q gain and 7 monosomy). MHC loci (6p region) could be a new hotspot for CMEs in FA leading to some events affecting genome regions previously associated with hematologic malignancies. Thus, our data would be on the direction with a possible causative relation-ship between CMEs and hematologic cancer development.

7. The detection of the same CMEs in both blood and anal SCC sample in one FA individual would be in favor of the potential use of genetic mosaicism detection as solid cancer biomarker.

8. Somatic UPD can act as second hit mechanism by leading mutation/deletion homozygosis and causing hematologic malignancies as CLL (associated with 13q14.3 deletions), PV (associated with JAK2 V617F mutation) or, more recently

reported by other groups, FA. These mosaic UPDs could serve as early hematologic cancer biomarker for these diseases.

9. However, we observed that mosaic UPDs also can be selected to rescue two different hematologic disorders or malignancies as CLL and FA-D2. UPD would select the most functional allele in order to increase, at the maximum possible extent, the functionality of the truncated protein/s. Hence, these events would not be cancer risk biomarkers but cancer protecting factors.

10. A high prevalence of UPD13q and del13q14.3 co-occurrence in non-CLL individuals was validated which suggested the possibility of having an important proportion of individuals with MBL condition by del13q14.3 homozygosis through UPD13q or being in front of another example of rescue UPD, in this case, trying to avoid MBL/CLL development

11. Adult FA patients regularly visited by genetic counsellors perform a chronic follow-up globally according to Spanish guidelines' recommendations. However, there is the tendency to concentrate all visits twice a year leading to an underfollowing in Hematology and an overfollowing in Gynecology.

12. Very poor hematologic follow-up through BMAs was detected in FA patients. This was due to a reduced test scheduling by doctors probably because of test invasiveness. This observation denotes the potential great positive impact of setting up CMEs detection by SNP array in easy to obtain samples as part of the chronic follow-up in FA, at least, to identify those patients that should not avoid BMAs due to their increased cancer risk.

13. A high adherence to chronic follow-up was detected among FA patients motivated by specialists' coordination in the same center, receiving visits reminders and scheduling controls in the same day increase adherence to follow-up.

14. The importance of genetic counsellor functions in FA families was demonstrated since we detected a global satisfactory FA knowledge, quite fulfillment of psychologic needs and a decreasing rate of habitual smokers in FA families periodically attended by a genetic counsellor or by a professional with similar functions.

REFERENCES

REFERENCES

1. Zlotnikoff M. A Human Mosaic. *J Hered*. 1945;36:163–8.

2. Fernández LC, Torres M, Real FX. Somatic mosaicism: on the road to cancer. *Nat Rev Cancer*. 2016;16(1):43–55.

3. Forsberg LA, Gisselsson D, Dumanski JP. Mosaicism in health and disease — clones picking up speed. *Nat Rev Genet*. 2016 Dec 12;18(2):128–42.

4. Machiela MJ, Chanock SJ. Detectable clonal mosaicism in the human genome. *Semin Hematol*. 2013 Oct;50(4):348–59.

5. Machiela MJ, Chanock SJ. The ageing genome, clonal mosaicism and chronic disease. *Curr Opin Genet Dev*. 2017 Feb 6;42:8–13.

6. Nussbaum, Robert L; McInnes, Roderick R; Willard HF. Capítulo 5: Principios de citogenética clínica. In: Elsevier Masson, editor. Thompson & Thompson Genética en Medicina. 7a ed. Barcelona; 2008. p. 68–75.

7. Biesecker, Leslie G.; Sppiner NB. A genomic view of mosaicism and human disease. *Nat Rev*. 2013;14(5):307–20.

8. De S. Somatic mosaicism in healthy human tissues. *Trends Genet*. 2011 Jun;27(6):217–23.

9. Hancks DC, Kazazian HH, Jr. Roles for retrotransposon insertions in human disease. *Mob DNA*. 2016;7:9.

10. van den Hurk JAJM, Meij IC, del Carmen Seleme M, Kano H, Nikopoulos K, Hoefsloot LH, et al. L1 retrotransposition can occur early in human embryonic development. *Hum Mol Genet*. 2007 Apr 9;16(13):1587–92.

11. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, et al. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev*. 2009 Jun 1;23(11):1303–12.

12. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, et al. L1 retrotransposition in human neural progenitor cells. *Nature*. 2009 Aug 27;460(7259):1127–31.

13. Kim S, Cho C-S, Han K, Lee J. Structural Variation of *Alu* Element and Human Disease. *Genomics Inform*. 2016 Sep;14(3):70.

14. Lisch D. How important are transposons for plant evolution? *Nat Rev Genet*. 2012 Dec 18;14(1):49–61.

15. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nat Rev Genet*. 2009 Aug;10(8):551–64.

16. Gonitel R, Moffitt H, Sathasivam K, Woodman B, Detloff PJ, Faull RLM, et al. DNA instability in postmitotic neurons. *Proc Natl Acad Sci U S A*. 2008 Mar 4;105(9):3467–72.

17. Messer A, Manley K, Shirley TL, Flaherty L. Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat Genet*. 1999 Dec 1;23(4):471–3.

18. Jorda J, Kajava A V. Protein Homorepeats: Sequences, Structures, Evolution, and Functions. In: Advances in Protein Chemistry and Structural Biology. 2010. p. 59–88.

19. Sung P, Klein H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol*. 2006 Oct 23;7(10):739–50.

20. Stephens K, Weaver M, Leppig KA, Maruyama K, Emanuel PD, Le Beau MM, et al. Interstitial uniparental isodisomy at clustered breakpoint intervals is a frequent mechanism of NF1 inactivation in myeloid malignancies. *Blood*. 2006 Sep 1;108(5):1684–9.

21. McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet*. 2008 Nov;24(11):529–38.

22. Nagaoka SI, Hassold TJ, Hunt PA. Human aneuploidy: mechanisms and new insights into an age-old problem. *Nat Rev Genet*. 2012 Jun 18;13(7):493–504.

23. Hook EB, Warburton D. Turner syndrome revisited: review of new data supports the hypothesis that all viable 45,X cases are cryptic mosaics with a rescue cell line, implying an origin by mitotic loss. *Hum Genet*. 2014 Apr 30;133(4):417–24.

24. Youssoufian H, Pyeritz RE. Mechanisms and consequences of somatic mosaicism in humans. *Nat Rev Genet*. 2002 Oct;3(10):748–58.

25. Forsberg LA, Rasi C, Razzaghian HR, Pakalapati G, Waite L, Thilbeault KS, et al. Age-related somatic structural changes in the nuclear genome of human blood cells. *Am J Hum Genet*. 2012 Feb 10;90(2):217–28.

26. Forsberg LA, Absher D, Dumanski JP. Non-heritable genetics of human disease: spotlight on post-zygotic genetic variation acquired during lifetime. *Postgrad Med J*. 2013 Jul 18;89(1053):417–26.

27. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet*. 1998 Oct 1;20(2):207–11.

28. Weimer J, Heidemann S, von Kaisenberg CS, Grote W, Arnold N, Bens S, et al. Isolated trisomy 7q21.2-31.31 resulting from a complex familial rearrangement involving chromosomes 7, 9 and 10. *Mol Cytogenet*. 2011 Dec 5;4:28.

29. Frigerio M, Passeri E, de Filippis T, Rusconi D, Valaperta R, Carminati M, et al. SNPs and real-time quantitative PCR method for constitutional allelic copy number determination, the VPREB1 marker case. *BMC Med Genet*. 2011 May 5;12:61.

30. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet*. 2012;44(6):651–8.

31. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet*. 2012 Jun;44(6):642–50.

32. Rodríguez-Santiago B, Malats N, Rothman N, Armengol L, Garcia-Closas M, Kogevinas M, et al. Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *Am J Hum Genet*. 2010 Jul 9;87(1):129–38.

33. González JR, Rodríguez-Santiago B, Cáceres A, Pique-Regi R, Rothman N, Chanock SJ, et al. A fast and accurate method to detect allelic genomic imbalances underlying mosaic rearrangements using SNP array data. *BMC Bioinformatics*. 2011 Jan;12(1):166.

34. Schick UM, McDavid A, Crane PK, Weston N, Ehrlich K, Newton KM, et al. Confirmation of the reported association of clonal chromosomal mosaicism with an increased risk of incident hematologic cancer. *PLoS One*. 2013;8(3):e59823.

35. Pique-Regi R, Cáceres A, González JR. R-Gada: a fast and flexible pipeline for copy number analysis in association studies. *BMC Bioinformatics*. 2010 Jul 16;11(1):380.

36. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al.

PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007 Nov 1;17(11):1665–74.

37. Wang K, Bucan M. Copy Number Variation Detection via High-Density SNP Genotyping. *CSH Protoc*. 2008 Jun 1;2008.

38. Dumanski JP, Lambert J-C, Rasi C, Giedraitis V, Davies H, Grenier-Boley B, et al. Mosaic Loss of Chromosome Y in Blood Is Associated with Alzheimer Disease. *Am J Hum Genet*. 2016 Jun 2;98(6):1208–19.

39. Forsberg LA, Rasi C, Malmqvist N, Davies H, Pasupulati S, Pakalapati G, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat Genet*. 2014 Jun 28;46(6):624–8.

40. Happle R. The McCune-Albright syndrome: a lethal gene surviving by mosaicism. *Clin Genet*. 1986 Apr;29(4):321–4.

41. Lindhurst MJ, Sapp JC, Teer JK, Johnston JJ, Finn EM, Peters K, et al. A Mosaic Activating Mutation in *AKT1* Associated with the Proteus Syndrome. *N Engl J Med*. 2011 Aug 18;365(7):611–9.

42. Aypar E, Yildirim MS, Sert A, Ciftci I, Odabas D. A girl with metopic synostosis and trisomy 13 mosaicism: Case report and review of the literature. *Am J Med Genet Part A*. 2011 Mar;155(3):638–41.

43. Tucker ME, Garringer HJ, Weaver DD. Phenotypic spectrum of mosaic trisomy 18: Two new patients, a literature review, and counseling issues. *Am J Med Genet Part A*. 2007 Mar 1;143A(5):505–17.

44. Papavassiliou P, York TP, Gursoy N, Hill G, Nicely LV, Sundaram U, et al. The phenotype of persons having mosaicism for trisomy 21/Down syndrome reflects the percentage of trisomic cells present in different tissues. *Am J Med Genet A*. 2009 Feb 15;149A(4):573–83.

45. Levitsky LL, O'Donnell Luria AH, Hayes FJ, Lin AE. Turner syndrome. *Curr Opin Endocrinol Diabetes Obes*. 2015 Feb;22(1):65–72.

46. Freriks K, Timmers HJLM, Netea-Maier RT, Beerendonk CCM, Otten BJ, van Alfen-van der Velden JAEM, et al. Buccal cell FISH and blood PCR-Y detect high rates of X chromosomal mosaicism and Y chromosomal derivatives in patients with Turner syndrome. *Eur J Med Genet*. 2013 Sep;56(9):497–501.

47. Denes A-M, Landin-Wilhelmsen K, Wettergren Y, Bryman I, Hanson C. The Proportion of Diploid 46,XX Cells Increases with Time in Women with Turner Syndrome—A 10-Year Follow-Up Study. *Genet Test Mol Biomarkers*. 2015 Feb;19(2):82–7.

48. Izumi K, Krantz ID. Pallister-Killian syndrome. *Am J Med Genet Part C Semin Med Genet*. 2014 Dec;166(4):406–13.

49. Mertzanidou A, Wilton L, Cheng J, Spits C, Vanneste E, Moreau Y, et al. Microarray analysis reveals abnormal chromosomal complements in over 70% of 14 normally developing human embryos. *Hum Reprod*. 2013 Jan 1;28(1):256–64.

50. Vanneste E, Voet T, Le Caignec C, Ampe M, Konings P, Melotte C, et al. Chromosome instability is common in human cleavage-stage embryos. *Nat Med*. 2009 May 26;15(5):577–83.

51. Spinner NB, Conlin LK. Mosaicism and clinical genetics. *Am J Med Genet Part C Semin Med Genet*. 2014 Dec;166(4):397–405.

52. Wallis GA, Starman BJ, Zinn AB, Byers PH. Variable expression of osteogenesis imperfecta in a nuclear family is explained by somatic mosaicism for a lethal point mutation in the alpha 1(I) gene (COL1A1) of type I collagen in a parent. *Am J Hum Genet*. 1990 Jun;46(6):1034–40.

53. Bermúdez-López C, Teresa BG, Angel AG, Alcántara-Ortigoza MA. Germinal Mosaicism in a Sample of Families with Duchenne/Becker Muscular Dystrophy with Partial Deletions in the *DMD* Gene. *Genet Test Mol Biomarkers*. 2014 Feb;18(2):93–7.

54. Helderman-van den Enden ATJM, de Jong R, den Dunnen JT, Houwing-Duistermaat JJ, Kneppers ALJ, Ginjaar HB, et al. Recurrence risk due to germ line mosaicism: Duchenne and Becker muscular dystrophy. *Clin Genet*. 2009 May;75(5):465–72.

55. McFarland R, Chinnery PF, Blakely EL, Schaefer AM, Morris AAM, Foster SM, et al. Homoplasmy, heteroplasmy, and mitochondrial dystonia. *Neurology*. 2007 Aug 28;69(9):911–6.

56. Taylor RW, Giordano C, Davidson MM, d'Amati G, Bain H, Hayes CM, et al. A homoplasmic mitochondrial transfer ribonucleic acid mutation as a cause of maternally inherited hypertrophic cardiomyopathy. *J Am Coll Cardiol*. 2003 May 21;41(10):1786–96.

57. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971 Apr;68(4):820–3.

58. Hino O, Kobayashi T. Mourning Dr. Alfred G. Knudson: the two-hit hypothesis, tumor suppressor genes, and the tuberous sclerosis complex. *Cancer Sci*. 2017 Jan;108(1):5–11.

59. Tuna M, Knuutila S, Mills GB. Uniparental disomy in cancer. *Trends Mol Med*. 2009 Mar;15(3):120–8.

60. Makishima H, Maciejewski JP. Pathogenesis and consequences of uniparental disomy in cancer. *Clin Cancer Res*. 2011 Jun 15;17(12):3913–23.

61. Lapunzina P, Monk D. The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer. *Biol Cell*. 2011 Jul;103(7):303–17.

62. Bullinger L, Krönke J, Schön C, Radtke I, Urlbauer K, Botzenhardt U, et al. Identification of acquired copy number alterations and uniparental disomies in cytogenetically normal acute myeloid leukemia using high-resolution single-nucleotide polymorphism analysis. *Leukemia*. 2010 Feb 17;24(2):438–49.

63. Flotho C, Steinemann D, Mullighan CG, Neale G, Mayer K, Kratz CP, et al. Genome-wide single-nucleotide polymorphism analysis in juvenile myelomonocytic leukemia identifies uniparental disomy surrounding the NF1 locus in cases associated with neurofibromatosis but not in cases with mutant RAS or PTPN11. *Oncogene*. 2007 Aug 23;26(39):5816–21.

64. Lehmann S, Ogawa S, Raynaud SD, Sanada M, Nannya Y, Ticchioni M, et al. Molecular allelokaryotyping of early-stage, untreated chronic lymphocytic leukemia. *Cancer*. 2008 Mar 15;112(6):1296–305.

65. Mol BM, Massink MPG, van der Hout AH, Dommering CJ, Zaman JMA, Bosscha MI, et al. High resolution SNP array profiling identifies variability in retinoblastoma genome stability. *Genes, Chromosom Cancer*. 2014 Jan;53(1):1–14.

66. Kralovics R, Passamonti F, Buser AS, Teo S-S, Tiedt R, Passweg JR, et al. A Gain-of-Function Mutation of *JAK2* in Myeloproliferative Disorders. *N Engl J Med*. 2005 Apr 28;352(17):1779–90.

67. Middeldorp A, van Puijenbroek M, Nielsen M, Corver WE, Jordanova ES, ter Haar N, et al. High frequency of copy-neutral LOH in MUTYH-associated polyposis carcinomas. *J Pathol*. 2008 Sep;216(1):25–31.

68. Choate KA, Lu Y, Zhou J, Elias PM, Zaidi S, Paller AS, et al.

Frequent somatic reversion of KRT1 mutations in ichthyosis with confetti. *J Clin Invest*. 2015 Apr 1;125(4):1703–7.

69. Jongmans MCJ, Verwiel ETP, Heijdra Y, Vulliamy T, Kamping EJ, Hehir-Kwa JY, et al. Revertant Somatic Mosaicism by Mitotic Recombination in Dyskeratosis Congenita. *Am J Hum Genet*. 2012 Mar 9;90(3):426–33.

70. Machiela MJ, Zhou W, Sampson JN, Dean MC, Jacobs KB, Black A, et al. Characterization of Large Structural Genetic Mosaicism in Human Autosomes. *Am J Hum Genet*. 2015 Mar 5;96(3):487–97.

71. Pierre R V, Hoagland HC. Age-associated aneuploidy: loss of Y chromosome from human bone marrow cells with aging. *Cancer*. 1972 Oct;30(4):889–94.

72. Dumanski JP, Rasi C, Lönn M, Davies H, Ingelsson M, Giedraitis V, et al. Mutagenesis. Smoking is associated with mosaic loss of chromosome Y. *Science*. 2015 Jan 2;347(6217):81–3.

73. Zhou W, Machiela MJ, Freedman ND, Rothman N, Malats N, Dagnall C, et al. Mosaic loss of chromosome Y is associated with common variation near TCL1A. *Nat Genet*. 2016 Apr 11;48(5):563–8.

74. Machiela MJ, Zhou W, Karlins E, Sampson JN, Freedman ND, Yang Q, et al. Female chromosome X mosaicism is age-related and preferentially affects the inactivated X chromosome. *Nat Commun*. 2016 Jun 13;7:11843.

75. Machiela MJ, Zhou W, Caporaso N, Dean M, Gapstur SM, Goldin L, et al. Mosaic 13q14 deletions in peripheral leukocytes of non-hematologic cancer cases and healthy controls. *J Hum Genet*. 2016 May;61(5):411–8.

76. Machiela MJ, Zhou W, Caporaso N, Dean M, Gapstur SM, Goldin L, et al. Mosaic chromosome 20q deletions are more frequent in the aging population. *Blood Adv*. 2017;1(6).

77. Bonnefond A, Skrobek B, Lobbens S, Eury E, Thuillier D, Cauchi S, et al. Association between large detectable clonal mosaicism and type 2 diabetes with vascular complications. *Nat Genet*. 2013 Sep;45(9):1040–3.

78. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman P V., Mar BG, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N Engl J Med*. 2014 Dec 25;371(26):2488–98.

79. Noveski P, Madjunkova S, Sukarova Stefanovska E, Matevska Geshkovska N, Kuzmanovska M, Dimovski A, et al. Loss of Y Chromosome in Peripheral Blood of Colorectal and Prostate Cancer Patients. Laconi E, editor. *PLoS One*. 2016 Jan 8;11(1):e0146264.

80. Wong AK, Fang B, Zhang L, Guo X, Lee S, Schreck R. Loss of the Y chromosome: an age-related or clonal phenomenon in acute myelogenous leukemia/myelodysplastic syndrome? *Arch Pathol Lab Med*. 2008 Aug;132(8):1329–32.

81. Ganster C, Kämpfe D, Jung K, Braulke F, Shirneshan K, Machherndl-Spandl S, et al. New data shed light on Y-loss-related pathogenesis in myelodysplastic syndromes. *Genes, Chromosom Cancer*. 2015 Dec;54(12):717–24.

82. Persani L, Bonomi M, Lleo A, Pasini S, Civardi F, Bianchi I, et al. Increased loss of the Y chromosome in peripheral blood cells in male patients with autoimmune thyroiditis. *J Autoimmun*. 2012 May;38(2–3):J193–6.

83. Baird DM, Britt-Compton B, Rowson J, Amso NN, Gregory L, Kipling D. Telomere instability in the male germline. *Hum Mol Genet*. 2005 Nov 30;15(1):45–51.

84. Müller AC, Jakobsen MA, Barington T, Vaag AA, Grunnet LG, Olsen SF, et al. Microchimerism of male origin in a cohort of Danish girls. *Chimerism*. 2015 Oct 2;6(4):65–71.

85. Celton-Morizur S, Desdouets C. Polyploidization of liver cells. *Adv Exp Med Biol*. 2010;676:123–35.

86. McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, Cowing-Zitron C, et al. Mosaic copy number variation in human neurons. *Science*. 2013 Nov 1;342(6158):632–7.

87. Piotrowski A, Bruder CEG, Andersson R, de Ståhl TD, Menzel U, Sandgren J, et al. Somatic mosaicism for copy number variation in differentiated human tissues. *Hum Mutat*. 2008 Sep;29(9):1118–24.

88. Rodriguez-Abreu D, Bordoni A, Zucca E. Epidemiology of hematological malignancies. *Ann Oncol*. 2007 Feb 1;18(Supplement 1):i3–8.

89. Lichtman MA. Battling the hematological malignancies: the 200 years' war. *Oncologist*. 2008 Feb;13(2):126–38.

90. Klausner R. Leukemia, Lymphoma, and Multiple Myeloma: Toward a New Understanding. Washington, DC: NIH; 2001.

91. Mugnaini EN, Ghosh N. Lymphoma. *Prim Care Clin Off Pract*. 2016 Dec;43(4):661–75.

92. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016 May 19;127(20):2375–90.

93. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20).

94. Pokharel M. Leukemia : A Review Article. *IJARPB*. 2012;2(3):397–407.

95. World Cancer Research Fund International. Worldwide data | World Cancer Research Fund International. Last reviewed: 16 January 2015.

96. Cancer Research UK. Leukaemia (all subtypes combined) incidence statistics | Cancer Research UK. Last reviewed: 11 June 2014.

97. Hallek M. Chronic lymphocytic leukemia: 2015 Update on diagnosis, risk stratification, and treatment. *Am J Hematol*. 2015 May;90(5):446–60.

98. Kipps TJ, Stevenson FK, Wu CJ, Croce CM, Packham G, Wierda WG, et al. Chronic lymphocytic leukaemia. *Nat Rev Dis Prim*. 2017 Jan 19;3:16096.

99. Santos FPS, O'Brien S. Small Lymphocytic Lymphoma and Chronic Lymphocytic Leukemia. *Cancer J*. 2012;18(5):396–403.

100. National Cancer Institute. National Cancer Institute. Surveillancce, Epidemiology and End Results Program (SEER). Cancer Stat Facts. Last reviewed: 9 December 2016.

101. Nabhan C, Aschebrook-Kilfoy B, Chiu BC-H, Smith SM, Shanafelt TD, Evens AM, et al. The impact of race, ethnicity, age and sex on clinical outcome in chronic lymphocytic leukemia: a comprehensive Surveillance, Epidemiology, and End Results analysis in the modern era. *Leuk Lymphoma*. 2014 Dec 3;55(12):2778–84.

102. Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*. 1999 Sep 15;94(6):1840–7.

103. Brown JR. Inherited predisposition to chronic lymphocytic leukemia. *Expert Rev Hematol*. 2008 Oct;1(1):51–61.

104. Pan JWY, Cook LS, Schwartz SM, Weis NS. Incidence of leukemia in Asian migrants to the United States and their descendants. *Cancer Causes Control*. 2002 Nov;13(9):791–5.

105. Lichtenstein P, Holm N V., Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000 Jul 13;343(2):78–85.

106. Videbaek A. Hereditary in Human Leukaemia and Its Relation to Cancer: A Genetic and Clinical Study of 209 Probands. Copenhagen, Denmark, Nyt Nordisk Forlag, Arnold Busck. 1947;5.

107. Jønsson V, Houlston RS, Catovsky D, Yuille MR, Hilden J, Olsen JH, et al. CLL family “Pedigree 14” revisited: 1947–2004. *Leukemia*. 2005 Jun;19(6):1025–8.

108. Goldin LR, Pfeiffer RM, Li X, Hemminki K. Familial risk of lymphoproliferative tumors in families of patients with chronic lymphocytic leukemia: results from the Swedish Family-Cancer Database. *Blood*. 2004 Sep 15;104(6):1850–4.

109. Sellick G, Catovshy D, Houlston R. Familial Chronic Lymphocytic Leukemia. *Semin Oncol*. 2006 Apr;33(2):195–201.

110. Goldin LR, Sgambati M, Marti GE, Fontaine L, Ishibe N, Caporaso N. Anticipation in Familial Chronic Lymphocytic Leukemia. *Am J Hum Genet*. 1999 Jul;65(1):265–8.

111. Daugherty SE, Pfeiffer RM, Mellemkjaer L, Hemminki K, Goldin LR. No evidence for anticipation in lymphoproliferative tumors in population-based samples. *Cancer Epidemiol Biomarkers Prev*. 2005 May 1;14(5):1245–50.

112. Berndt SI, Camp NJ, Skibola CF, Vijai J, Wang Z, Gu J, et al. Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nat Commun*. 2016 Mar 9;7:10933.

113. Speedy HE, Di Bernardo MC, Sava GP, Dyer MJS, Holroyd A, Wang Y, et al. A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. *Nat Genet*. 2013 Dec 1;46(1):56–60.

114. Baumann Kreuziger LM, Tarchand G, Morrison VA. The impact of Agent Orange exposure on presentation and prognosis of patients with chronic lymphocytic leukemia. *Leuk Lymphoma*. 2014 Jan 15;55(1):63–6.

115. Schinasi LH, De Roos AJ, Ray RM, Edlefsen KL, Parks CG, Howard B V., et al. Insecticide exposure and farm history in relation to risk of lymphomas and leukemias in the Women's Health Initiative observational study cohort. *Ann Epidemiol*. 2015 Nov;25(11):803–810.

116. Hsu W-L, Preston DL, Soda M, Sugiyama H, Funamoto S, Kodama K, et al. The Incidence of Leukemia, Lymphoma and Multiple Myeloma among Atomic Bomb Survivors: 1950–2001. *Radiat Res*. 2013 Mar;179(3):361–82.

117. Hjalgrim H, Rostgaard K, Vasan SK, Ullum H, Erikstrup C, Pedersen OB V., et al. No evidence of transmission of chronic lymphocytic leukemia through blood transfusion. *Blood*. 2015 Oct 22;126(17):2059–61.

118. Delgado J, Villamor N, López-Guillermo A, Campo E. Genetic evolution in chronic lymphocytic leukaemia. *Best Pract Res Clin Haematol*. 2016 Mar;29(1):67–78.

119. Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011 Jun 5;475(7354):101–5.

120. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature*. 2015 Oct 14;526(7574):525–30.

121. Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015 Jul 22;526(7574):519–24.

122. Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T, et al. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell*. 2010 Jan 19;17(1):28–40.

123. Döhner H, Stilgenbauer S, Benner A, Leupolt E, Kröber A, Bullinger L, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*. 2000 Dec 28;343(26):1910–6.

124. Klein U, Dalla-Favera R. New insights into the pathogenesis of chronic lymphocytic leukemia. *Semin Cancer Biol*. 2010 Dec;20(6):377–83.

125. Kalachikov S, Migliazza A, Cayanis E, Fracchiolla NS, Bonaldo MF, Lawton L, et al. Cloning and Gene Mapping of the Chromosome 13q14 Region Deleted in Chronic Lymphocytic Leukemia. *Genomics*. 1997 Jun 15;42(3):369–77.

126. Mertens D, Wolf S, Tschuch C, Mund C, Kienle D, Ohl S, et al. Allelic silencing at the tumor-suppressor locus 13q14.3 suggests an epigenetic tumor-suppressor mechanism. *Proc Natl Acad Sci U S A*. 2006 May 16;103(20):7741–6.

127. Ouillette P, Collins R, Shakhan S, Li J, Li C, Shedden K, et al. The Prognostic Significance of Various 13q14 Deletions in Chronic Lymphocytic Leukemia. *Clin Cancer Res*. 2011 Nov 1;17(21):6778–90.

128. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 2002 Nov 26;99(24):15524–9.

129. Puiggros A, Venturas M, Salido M, Blanco G, Fernandez-Rodriguez C, Collado R, et al. Interstitial 13q14 deletions detected in the karyotype and translocations with concomitant deletion at 13q14 in chronic lymphocytic leukemia: Different genetic mechanisms but equivalent poorer clinical outcome. *Genes, Chromosom Cancer*. 2014 Sep;53(9):788–97.

130. Aqeilan RI, Calin G a, Croce CM. miR-15a and miR-16-1 in cancer: discovery, function and future perspectives. *Cell Death Differ*. 2010 Feb;17(2):215–20.

131. Cui B, Chen L, Zhang S, Mraz M, Fecteau J-F, Yu J, et al. MicroRNA-155 influences B-cell receptor signaling and associates with aggressive disease in chronic lymphocytic leukemia. *Blood*. 2014 Jul 24;124(4):546–54.

132. Kosaka N, Iguchi H, Yoshioka Y, Hagiwara K, Takeshita F, Ochiya T. Competitive interactions of cancer cells and normal cells via secretory microRNAs. *J Biol Chem*. 2012 Jan 6;287(2):1397–405.

133. Weinhold B. Epigenetics: the science of change. *Environ Health Perspect*. 2006 Mar;114(3):A160-7.

134. Herman JG, Baylin SB. Gene Silencing in Cancer in Association with Promoter Hypermethylation. *N Engl J Med*. 2003 Nov 20;349(21):2042–54.

135. Kulis M, Heath S, Bibikova M, Queirós AC, Navarro A, Clot G, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet*. 2012 Oct 14;44(11):1236–42.

136. Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, et al. Locally Disordered Methylation Forms the Basis of Intratumor Methylome Variation in Chronic Lymphocytic Leukemia. *Cancer Cell*. 2014 Dec 8;26(6):813–25.

137. Hanada M, Delia D, Aiello A, Stadtmauer E, Reed JC. bcl-2 gene hypomethylation and high-level expression in B-cell chronic lymphocytic leukemia. *Blood*. 1993 Sep 15;82(6):1820–8.

138. Cahill N, Bergh A-C, Kanduri M, Göransson-Kultima H, Mansouri L, Isaksson A, et al. 450K-array analysis of chronic lymphocytic leukemia cells reveals global DNA methylation to be relatively stable over time and similar in resting and proliferative compartments. *Leukemia*. 2013 Jan 27;27(1):150–8.

139. Baer C, Claus R, Frenzel LP, Zucknick M, Park YJ, Gu L, et al. Extensive promoter DNA hypermethylation and hypomethylation is associated with aberrant microRNA expression in chronic lymphocytic leukemia. *Cancer Res*. 2012 Aug 1;72(15):3775–85.

140. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner H, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood*. 2008 Jun 15;111(12):5446–56.

141. Lanasa MC, Weinberg JB. Immunologic aspects of monoclonal B-cell lymphocytosis. *Immunol Res*. 2011 Apr;49(1–3):269–80.

142. Rawstron AC, Green MJ, Kuzmicki A, Kennedy B, Fenton JAL, Evans PAS, et al. Monoclonal B lymphocytes with the characteristics of “indolen” chronic lymphocytic leukemia are present in 3.5% of adults with normal blood counts. *Blood*. 2002 Jul 15;100(2):635–9.

143. Strati P, Shanafelt TD. Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. *Blood*. 2015 Jul 23;126(4):454–62.

144. Rawstron AC, Bennett FL, O'Connor SJM, Kwok M, Fenton JAL, Plummer M, et al. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *N Engl J Med*. 2008 Aug 7;359(6):575–83.

145. Landgren O, Albitar M, Ma W, Abbasi F, Hayes RB, Ghia P, et al. B-cell clones as early markers for chronic lymphocytic leukemia. *N Engl J Med*. 2009 Feb 12;360(7):659–67.

146. Kalpadakis C, Pangalis GA, Sachanas S, Vassilakopoulos TP, Kyriakaki S, Korkolopoulou P, et al. New insights into monoclonal B-cell lymphocytosis. *Biomed Res Int*. 2014;2014:258917.

147. Rawstron AC, Shanafelt T, Lanasa MC, Landgren O, Hanson C, Orfao A, et al. Different biology and clinical outcome according to the absolute numbers of clonal B-cells in monoclonal B-cell lymphocytosis (MBL). *Cytom Part B Clin Cytom*. 2010;78B(S1):S19–23.

148. Jabbour E, Kantarjian H. Chronic myeloid leukemia: 2016 update on diagnosis, therapy, and monitoring. *Am J Hematol*. 2016 Feb;91(2):252–65.

149. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*. 1973 Jun 1;243(5405):290–3.

150. Chiaretti S, Zini G, Bassan R. Diagnosis and subclassification of acute limphoblastic leukemia. *Mediterr J Hematol Infect Dis*. 2014 Oct 24;6(1):2014073.

151. Inaba H, Greaves M, Mullighan CG. Acute lymphoblastic leukaemia. *Lancet*. 2013 Jun 1;381(9881):1943–55.

152. Hunger SP, Lu X, Devidas M, Camitta BM, Gaynon PS, Winick NJ, et al. Improved Survival for Children and Adolescents With Acute Lymphoblastic Leukemia Between 1990 and 2005: A Report From the Children's Oncology Group. *J Clin Oncol*. 2012 May 10;30(14):1663–9.

153. Pui C-H, Robison LL, Look AT. Acute lymphoblastic leukaemia. *Lancet*. 2008 Mar 22;371(9617):1030–43.

154. Saultz J, Garzon R. Acute Myeloid Leukemia: A Concise Review. *J Clin Med*. 2016 Mar 5;5(3):33.

155. De Kouchkovsky I, Abdul-Hay M. “Acute myeloid leukemia: a comprehensive review and 2016 update.” *Blood Cancer J*. 2016 Jul 1;6(7):e441.

156. Taylor AM, Metcalfe JA, Thick J, Mak YF. Leukemia and lymphoma in ataxia telangiectasia. *Blood*. 1996 Jan 15;87(2):423–38.

157. Stiller CA, Chessells JM, Fitchett M. Neurofibromatosis and childhood leukaemia/lymphoma: a population-based UKCCSG study. *Br J Cancer*. 1994 Nov;70(5):969–72.

158. Hasle H, Clemmensen IH, Mikkelsen M. Risks of leukaemia and solid tumours in individuals with Down’s syndrome. *Lancet*. 2000 Jan 15;355(9199):165–9.

159. Cunniff C, Bassetti JA, Ellis NA. Bloom’s Syndrome: Clinical Spectrum, Molecular Pathogenesis, and Cancer Predisposition. *Mol Syndromol*. 2017 Jan;8(1):4–23.

160. Fanconi G. Familial infantile pernicious-like anemia. *Yearb Pediatr Phys Educ*. 1927;117:257–80.

161. Lobitz S, Velleuer E. Guido Fanconi (1892–1979): a jack of all trades. *Nat Rev Cancer*. 2006 Nov 12;6(11):893–8.

162. Red Nacional De Investigación Cooperativa Sobre la Anemia de Fanconi. Red Nacional De Investigación Cooperativa Sobre La Anemia De Fanconi. Madrid. 2007.

163. Auerbach A. Orphanet: Fanconi anemia. Last reviewed: November 2011.

164. Alter BP KG. Gene Reviews: Fanconi Anemia. University of Washington, Seattle; 1993-2017. Last reviewed: 23 February 2017.

165. Huck K, Hanenberg H, Gudowius S, Fenk R, Kalb R, Neveling K, et al. Delayed diagnosis and complications of Fanconi anaemia at advanced age - a paradigm. *Br J Haematol*. 2006 Apr;133(2):188–97.

166. Shimamura A, Alter BP. Pathophysiology and management of inherited bone marrow failure syndromes. *Blood Rev*. 2010 May;24(3):101–22.

167. Rose SR, Myers KC, Rutter MM, Mueller R, Khoury JC, Mehta PA, et al. Endocrine phenotype of children and adults with Fanconi anemia. *Pediatr Blood Cancer*. 2012 Oct;59(4):690–6.

168. Nathan D, Orkin S, Ginsburg D. Hematology of infancy and childhood. ed 6, vol. Saunders W, editor. Philadelphia; 2003. 285 p.

169. Fanconi Anemia Research Fund. Inc. Fanconi Anemia: Guidelines for Diagnosis and Management. Eiler ME, Frohnmayer D, Frohnmayer L, Larsen K, Owen J, editors. 4th ed. 2014;421.

170. Kutler DI, Singh B, Satagopan J, Batish SD, Berwick M, Giampietro PF, et al. A 20-year perspective on the International Fanconi Anemia Registry (IFAR). *Blood*. 2003 Feb 15;101(4):1249–56.

171. Alter BP, Giri N, Savage SA, Peters JA, Loud JT, Leathwood L, et al. Malignancies and survival patterns in the National Cancer Institute inherited bone marrow failure syndromes cohort study. *Br J Haematol*. 2010 Apr;150(2):179-88.

172. Tamary H, Nishri D, Yacobovich J, Zilber R, Dgany O, Krasnov T, et al. Frequency and natural history of inherited bone marrow failure syndromes: the Israeli Inherited Bone Marrow Failure Registry. *Haematologica*. 2010 Aug 1;95(8):1300–7.

173. Rosenberg PS, Alter BP, Ebell W. Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry. *Haematologica*. 2008 Apr 1;93(4):511–7.

174. Alter BP, Rosenberg PS, Brody LC. Clinical and molecular features associated with biallelic mutations in FANCD1/BRCA2. *J Med Genet*. 2006 Jul 6;44(1):1–9.

175. Myers K, Davies SM, Harris RE, Spunt SL, Smolarek T, Zimmerman S, et al. The clinical phenotype of children with Fanconi anemia caused by biallelic FANCD1/BRCA2 mutations. *Pediatr Blood Cancer*. 2012 Mar;58(3):462–5.

176. Kutler DI, Auerbach AD, Satagopan J, Giampietro PF, Batish SD, Huvos AG, et al. High incidence of head and neck squamous cell carcinoma in patients with Fanconi anemia. *Arch Otolaryngol Head Neck Surg*. 2003 Jan;129(1):106–12.

177. Graham J, Grenier D, Ponnampalam A. Metastatic Breast Cancer in a Patient with Fanconi Anemia. *Blood*. 2014;124(21).

178. Judkins T, Hendrickson BC, Deffenbaugh AM, Eliason K, Leclair B, Norton MJ, et al. Application of embryonic lethal or other obvious phenotypes to characterize the clinical significance of genetic variants found in trans with known deleterious mutations. *Cancer Res*. 2005 Nov 1;65(21):10096–103.

179. Bogliolo M, Surralles J. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics Article. *Mol gen*. 2015;33:32–40.

180. Ceccaldi R, Sarangi P, D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol*. 2016 May 5;17(6):337–49.

181. Meetei AR, Medhurst AL, Ling C, Xue Y, Singh TR, Bier P, et al. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat Genet*. 2005 Sep 21;37(9):958–63.

182. Singh TR, Bakker ST, Agarwal S, Jansen M, Grassman E, Godthelp BC, et al. Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M. *Blood*. 2009 Jul 2;114(1):174–80.

183. Lim ET, Würtz P, Havulinna AS, Palta P, Tukiainen T, Rehnström K, et al. Distribution and medical impact of loss-of-function variants in the Finnish founder population. Cutler D, editor. *PLoS Genet*. 2014 Jul 31;10(7):e1004494.

184. Park J-Y, Virts EL, Jankowska A, Wiek C, Othman M, Chakraborty SC, et al. Complementation of hypersensitivity to DNA interstrand crosslinking agents demonstrates that *XRCC2* is a Fanconi anaemia gene. *J Med Genet*. 2016 Oct;53(10):672–80.

185. Bluteau D, Masliah-Planchon J, Clairmont C, Rousseau A, Ceccaldi R, Dubois d'Enghien C, et al. Biallelic inactivation of REV7 is associated with Fanconi anemia. *J Clin Invest*. 2016 Aug 8;126(9):3580–4.

186. Meetei AR, Levitus M, Xue Y, Medhurst AL, Zwaan M, Ling C, et al. X-linked inheritance of Fanconi anemia complementation group B. *Nat Genet*. 2004 Nov 24;36(11):1219–24.

187. Ameziane N, May P, Haitjema A, van de Vrugt HJ, van Rossum-Fikkert SE, Ristic D, et al. A novel Fanconi anaemia subtype associated with a dominant-negative mutation in RAD51. *Nat Commun*. 2015 Dec 18;6:8829.

188. Rosenberg PS, Tamary H, Alter BP. How high are carrier frequencies of rare recessive syndromes? Contemporary estimates for Fanconi Anemia in the United States and Israel. *Am J Med Genet Part A*. 2011 Aug;155(8):1877–83.

189. Castella M, Pujol R, Callén E, Trujillo JP, Casado JA, Gille H, et al. Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood*. 2011 Apr 7;117(14):3759–69.

190. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science*. 1994 Sep 30;265(5181):2088–90.

191. Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet*. 2006 Nov 8;38(11):1239–41.

192. Rafnar T, Gudbjartsson DF, Sulem P, Jonasdottir A, Sigurdsson A, Jonasdottir A, et al. Mutations in BRIP1 confer high risk of ovarian cancer. *Nat Genet*. 2011 Oct 2;43(11):1104–7.

193. Kiiski JI, Pelttari LM, Khan S, Freysteinsdottir ES, Reynisdottir I, Hart SN, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proc Natl Acad Sci*. 2014 Oct 21;111(42):15172–7.

194. Peterlongo P, Catucci I, Colombo M, Caleca L, Mucaki E, Bogliolo M, et al. *FANCM* c.5791C>T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. *Hum Mol Genet*. 2015 Sep 15;24(18):5345–55.

195. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet*. 2007 Feb;39(2):165–7.

196. Osorio A, Endt D, Fernandez F, Eirich K, de la Hoya M, Schmutzler R, et al. Predominance of pathogenic missense variants in the RAD51C gene occurring in breast and ovarian cancer families. *Hum Mol Genet*. 2012 Jul 1;21(13):2889–98.

197. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*. 1994 Oct 7;266(5182):66–71.

198. Gresner P, Gromadzinska J, Twardowska E, Rydzynski K, Wasowicz W. Rad51C: A novel suppressor gene modulates the risk of head and neck cancer. *Mutat Res Mol Mech Mutagen*. 2014 Apr;762:47–54.

199. Scheckenbach K, Baldus SE, Balz V, Freund M, Pakropa P, Sproll C, et al. RAD51C--a new human cancer susceptibility gene for sporadic squamous cell carcinoma of the head and neck (HNSCC). *Oral Oncol*. 2014 Mar;50(3):196–9.

200. Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, et al. Exomic Sequencing Identifies PALB2 as a Pancreatic Cancer Susceptibility Gene. *Science*. 2009 Apr 10;324(5924):217–217.

201. Kashiyama K, Nakazawa Y, Pilz DT, Guo C, Shimada M, Sasaki K, et al. Malfunction of Nuclease ERCC1-XPF Results in Diverse Clinical Manifestations and Causes Cockayne Syndrome, Xeroderma Pigmentosum, and Fanconi Anemia. *Am J Hum Genet*. 2013 May 2;92(5):807–19.

202. Niedernhofer LJ, Garinis GA, Raams A, Lalai AS, Robinson AR, Appeldoorn E, et al. A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature*. 2006 Dec 21;444(7122):1038–43.

203. Kelland L. The resurgence of platinum-based cancer chemotherapy. *Nat Rev Cancer*. 2007 Aug 12;7(8):573–84.

204. Stone MP, Cho Y-J, Huang H, Kim H-Y, Kozekov ID, Kozekova A, et al. Interstrand DNA Cross-Links Induced by α,β-Unsaturated Aldehydes Derived from Lipid Peroxidation and Environmental Sources. *Acc Chem Res*. 2008 Jul;41(7):793–804.

205. Red Nacional para la Anemia de Fanconi. Guía Básica para el diagnóstico y seguimiento de pacientes con Anemia de Fanconi. Red Nacional para la Anemia de Fanconi, editor. Madrid; 2012.

206. Hays LE, Meyer S, van de Vrugt HJ. A molecular, genetic, and diagnostic spotlight on fanconi anemia. *Anemia*. 2012 Jan;

207. Stecklein SR, Jensen R a. Identifying and exploiting defects in the Fanconi anemia/BRCA pathway in oncology. *Transl Res*. 2012 Sep;160(3):178–97.

208. Auerbach AD. Fanconi anemia diagnosis and the diepoxybutane (DEB) test. *Exp Hematol*. 1993 Jun;21(6):731–3.

209. Zen PRG, Moraes FN de, Rosa RFM, Graziadio C, Paskulin GA. Características clínicas de pacientes com anemia de Fanconi. *Rev Paul Pediatr*. 2011 Sep;29(3):392–9.

210. Chrzanowska KH, Gregorek H, Dembowska-Bagińska B, Kalina MA, Digweed M. Nijmegen breakage syndrome (NBS). *Orphanet J Rare Dis*. 2012 Feb 28;7(1):13.

211. Andreassen PR, D'Andrea AD, Taniguchi T. ATR couples FANCD2 monoubiquitination to the DNA-damage response. *Genes Dev*. 2004 Aug 15;18(16):1958–63.

212. Seyschab H, Friedl R, Sun Y, Schindler D, Hoehn H, Hentze S, et al. Comparative evaluation of diepoxybutane sensitivity and cell cycle blockage in the diagnosis of Fanconi anemia. *Blood*. 1995 Apr 15;85(8):2233–7.

213. Hanenberg H, Batish SD, Pollok KE, Vieten L, Verlander PC, Leurs C, et al. Phenotypic correction of primary Fanconi anemia T cells with retroviral vectors as a diagnostic tool. *Exp Hematol*. 2002 May;30(5):410–20.

214. Shimamura A, Montes de Oca R, Svenson JL, Haining N, Moreau LA, Nathan DG, et al. A novel diagnostic screen for defects in the Fanconi anemia pathway. *Blood*. 2002 Dec 15;100(13):4649–54.

215. Ameziane N, Sie D, Dentro S, Ariyurek Y, Kerkhoven L, Joenje H, et al. Diagnosis of Fanconi Anemia: Mutation Analysis by Next-Generation Sequencing. *Anemia*. 2012;2012:1–7.

216. Knies K, Schuster B, Ameziane N, Rooimans M, Bettecken T, de Winter J, et al. Genotyping of Fanconi Anemia Patients by Whole Exome Sequencing: Advantages and Challenges. Schuelke M, editor. *PLoS One*. 2012 Dec 20;7(12):e52648.

217. Chandrasekharappa SC, Lach FP, Kimble DC, Kamat A, Teer JK, Donovan FX, et al. Massively parallel sequencing, aCGH, and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. *Blood*. 2013 May 30;121(22):e138–48.

218. Gille JJP, Floor K, Kerkhoven L, Ameziane N, Joenje H, de Winter JP. Diagnosis of Fanconi Anemia: Mutation Analysis by Multiplex Ligation-Dependent Probe Amplification and PCR-Based Sanger Sequencing. *Anemia*. 2012;2012:603253.

219. Cioc AM, Wagner JE, MacMillan ML, DeFor T, Hirsch B. Diagnosis of myelodysplastic syndrome among a cohort of 119 patients with

fanconi anemia: morphologic and cytogenetic characteristics. *Am J Clin Pathol*. 2010 Jan 1;133(1):92–100.

220. Meyer S, Neitzel H, Tönnies H. Chromosomal aberrations associated with clonal evolution and leukemic transformation in fanconi anemia: Clinical and biological implications. *Anemia*. 2012;2012.

221. Stivrins TJ, Davis RB, Sanger W, Fritz J, Purtilo DT. Transformation of Fanconi's anemia to acute nonlymphocytic leukemia associated with emergence of monosomy 7. *Blood*. 1984 Jul;64(1):173–6.

222. Tönnies H, Huber S, Kühl JS, Gerlach A, Ebell W, Neitzel H. Clonal chromosomal aberrations in bone marrow cells of Fanconi anemia patients: Gains of the chromosomal segment 3q26q29 as an adverse risk factor. *Blood*. 2003;101(10):3872–4.

223. Mehta PA, Harris RE, Davies SM, Kim M-O, Mueller R, Lampkin B, et al. Numerical chromosomal changes and risk of development of myelodysplastic syndrome--acute myeloid leukemia in patients with Fanconi anemia. *Cancer Genet Cytogenet*. 2010 Dec;203(2):180–6.

224. Sarma NJ, Takeda A, Yaseen NR. Colony Forming Cell (CFC) Assay for Human Hematopoietic Cells. *J Vis Exp*. 2010 Dec 18;(46):pii: 2195.

225. Kanter-Lewensohn L, Hellström-Lindberg E, Kock Y, Elmhorn-Rosenborg A, Ost A. Analysis of CD34-positive cells in bone marrow from patients with myelodysplastic syndromes and acute myeloid leukemia and in normal individuals: a comparison between FACS analysis and immunohistochemistry. *Eur J Haematol*. 1996 Mar;56(3):124–9.

226. Pomerantz SR. Myelography. In: Handbook of clinical neurology. 2016. p. 193–208.

227. Rosenberg PS, Socié G, Alter BP, Gluckman E. Risk of head and neck squamous cell cancer and death in patients with Fanconi anemia who did and did not receive transplants. *Blood*. 2005;105:67–73.

228. Velleuer E, Dietrich R. Fanconi anemia: young patients at high risk for squamous cell carcinoma. *Mol Cell Pediatr*. 2014 Nov 1;1(1):9.

229. Berrington de Gonzalez A, Berg CD, Visvanathan K, Robson M. Estimated Risk of Radiation-Induced Breast Cancer From Mammographic Screening for Young BRCA Mutation Carriers. *JNCI J Natl Cancer Inst*. 2009 Feb 4;101(3):205–9.

230. Kahraman S, Beyazyurek C, Yesilipek MA, Ozturk G, Ertem M, Anak S, et al. Successful haematopoietic stem cell transplantation in 44 children from healthy siblings conceived after preimplantation HLA matching. *Reprod Biomed Online*. 2014 Sep;29(3):340–51.

231. Tolar J, Becker PS, Clapp DW, Hanenberg H, de Heredia CD, Kiem H-P, et al. Gene Therapy for Fanconi Anemia: One Step Closer to the Clinic. *Hum Gene Ther*. 2012 Feb;23(2):141–4.

232. Freedman ND, Abnet CC, Leitzmann MF, Hollenbeck AR, Schatzkin A. Prospective investigation of the cigarette smoking–head and neck cancer association by sex. *Cancer*. 2007 Oct 1;110(7):1593–601.

233. Freedman ND, Schatzkin A, Leitzmann MF, Hollenbeck AR, Abnet CC. Alcohol and head and neck cancer risk in a prospective study. *Br J Cancer*. 2007 May 7;96(9):1469–74.

234. Marklund L, Hammarstedt L. Impact of HPV in Oropharyngeal Cancer. *J Oncol*. 2011;2011:509036.

235. Golusinski P. Risk Factors for Oral Infection with Human Papillomavirus. In: Recent results in cancer research Fortschritte der Krebsforschung Progres dans les recherches sur le cancer. 2017. p. 73–85.

236. Harper DM. Currently approved prophylactic HPV vaccines. *Expert Rev Vaccines*. 2009 Dec 9;8(12):1663–79.

237. Mollers M, Vossen JM, Scherpenisse M, van der Klis FRM, Meijer CJLM, de Melker HE. Review: Current knowledge on the role of HPV antibodies after natural infection and vaccination: Implications for monitoring an HPV vaccination programme. *J Med Virol*. 2013 Aug;85(8):1379–85.

238. Kutler DI, Wreesmann VB, Goberdhan A, Ben-Porat L, Satagopan J, Ngai I, et al. Human papillomavirus DNA and p53 polymorphisms in squamous cell carcinomas from Fanconi anemia patients. *J Natl Cancer Inst*. 2003 Nov 19;95(22):1718–21.

239. de Araujo MR, Rubira-Bullen IRF, Santos CF, Dionísio TJ, Bonfim CMS, De Marco L, et al. High prevalence of oral human papillomavirus infection in Fanconi's anemia patients. *Oral Dis*. 2011 Sep;17(6):572–6.

240. Alter BP, Giri N, Pan Y, Savage S a., Pinto L a. Antibody response to human papillomavirus vaccine in subjects with inherited bone marrow failure syndromes. *Vaccine*. 2014;32(10):1169–73.

241. Winer R, Huang C, Cherne S, Stern J, Butsch Kovacic M, Mehta P, et al. Detection of human papillomavirus in the oral cavities of persons with Fanconi anemia. *Oral Dis*. 2015;21(3):349–54.

242. Kwee ML, Poll EHA, van de Kamp JJP, de Koning H, Eriksson AW, Joenje H. Unusual response to bifunctional alkylating agents in a case of Fanconi anaemia. *Hum Genet*. 1983 Oct;64(4):384–7.

243. Ellis NA, Ciocci S, German J. Back mutation can produce phenotype reversion in Bloom syndrome somatic cells. *Hum Genet*. 2001 Feb;108(2):167–73.

244. Joenje H, Waisfisz Q, Morgan N V., Savino M, de Winter JP, van Berkel CGM, et al. Spontaneous functional correction of homozygous fanconi anaemia alleles reveals novel mechanistic basis for reverse mosaicism. *Nat Genet*. 1999 Aug 1;22(4):379–83.

245. Gross M, Hanenberg H, Lobitz S, Friedl R, Herterich S, Dietrich R, et al. Reverse mosaicism in Fanconi anemia: natural gene therapy via molecular self-correction. *Cytogenet Genome Res*. 2002;98(2–3):126–35.

246. Resta R, Biesecker BB, Bennett RL, Blum S, Estabrooks Hahn S, Strecker MN, et al. A New Definition of Genetic Counseling: National Society of Genetic Counselors' Task Force Report. *J Genet Couns*. 2006 Apr 19;15(2):77–83.

247. Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, et al. Timing, rates and spectra of human germline mutation. *Nat Genet*. 2015 Dec 14;48(2):126–33.

248. Pyott SM, Pepin MG, Schwarze U, Yang K, Smith G, Byers PH. Recurrence of perinatal lethal osteogenesis imperfecta in sibships: Parsing the risk between parental mosaicism for dominant mutations and autosomal recessive inheritance. *Genet Med*. 2011 Feb 13;13(2):125–30.

249. Rose VM, Au K-S, Pollom G, Roach ES, Prashner HR, Northrup H. Germ-Line Mosaicism in Tuberous Sclerosis: How Common? *Am J Hum Genet*. 1999 Apr;64(4):986–92.

250. Mody R, Li S, Dover DC, Sallan S, Leisenring W, Oeffinger KC, et al. Twenty-five-year follow-up among survivors of childhood acute lymphoblastic leukemia: a report from the Childhood Cancer Survivor Study. *Blood*. 2008 Jun 15;111(12):5515–23.

251. Kearney JA, Hay JL, Halpern L, Boulad F. Peritransplant

Psychiatric Evaluation of Patients With Fanconi Anemia. *J Pediatr Hematol / Oncol*. 2012 Apr;34(3):163–8.

252. Hutson SP, Alter BP. Experiences of siblings of patients with Fanconi anemia. *Pediatr Blood Cancer*. 2007 Jan;48(1):72–9.

253. Zierhut HA, Bartels DM. Waiting for the next shoe to drop: the experience of parents of children with fanconi anemia. *J Genet Couns*. 2012 Feb;21(1):45–58.

254. McLaughlin SE, Diener-West M, Indurkhya A, Rubin H, Heckmann R, Boyle MP. Improving transition from pediatric to adult cystic fibrosis care: lessons from a national survey of current practices. *Pediatrics*. 2008 May 1;121(5):e1160-6.

255. McDonagh JE, Kelly DA. Transitioning care of the pediatric recipient to adult caregivers. *Pediatr Clin North Am*. 2003 Dec 12;50(6):1561–83.

256. Reiss JG, Gibson RW, Walker LR. Health care transition: youth, family, and provider perspectives. *Pediatrics*. 2005 Jan 1;115(1):112–20.

257. Score J, Chase A, Forsberg LA, Feng L, Waghorn K, Jones A V, et al. Detection of leukemia-associated mutations in peripheral blood DNA of hematologically normal elderly individuals. *Leukemia*. 2015 Jul 28;29(7):1600–2.

258. Kis O, Kaedbey R, Chow S, Danesh A, Dowar M, Li T, et al. Circulating tumour DNA sequence analysis as an alternative to multiple myeloma bone marrow aspirates. *Nat Commun*. 2017 May 11;8:15086.

259. Wang AT, Smogorzewska A. SnapShot: Fanconi anemia and associated proteins. *Cell*. 2015 Jan 15;160(1–2):354–354.

260. McCoy RC. Mosaicism in Preimplantation Human Embryos: When Chromosomal Abnormalities Are the Norm. *Trends Genet*. 2017 Jul; 33(7):448-463.

261. García-Rubio ML, Pérez-Calero C, Barroso SI, Tumini E, Herrera-Moyano E, Rosado I V, et al. The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genet*. 2015 Nov;11(11):e1005674.

262. Schwab RA, Nieminuszczy J, Shah F, Langton J, Lopez Martinez D, Liang C-C, et al. The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Mol Cell*. 2015 Nov 5;60(3):351–61.

263. Ogawa S. Clonal hematopoiesis in acquired aplastic anemia. *Blood*. 2016 Jul 21;128(3):337–47.

264. Schwindt H, Vater I, Kreuz M, Montesinos-Rongen M, Brunn A, Richter J, et al. Chromosomal imbalances and partial uniparental disomies in primary central nervous system lymphoma. *Leukemia*. 2009 Oct;23(10):1875–84.

265. Vago L, Perna SK, Zanussi M, Mazzi B, Barlassina C, Stanghellini MTL, et al. Loss of mismatched HLA in leukemia after stem-cell transplantation. *N Engl J Med*. 2009 Jul 30;361(5):478–88.

266. Villalobos IB, Takahashi Y, Akatsuka Y, Muramatsu H, Nishio N, Hama A, et al. Relapse of leukemia with loss of mismatched HLA resulting from uniparental disomy after haploidentical hematopoietic stem cell transplantation. *Blood*. 2010 Apr 15;115(15):3158–61.

267. Dubois V, Sloan-Béna F, Cesbron A, Hepkema BG, Gagne K, Gimelli S, et al. Pretransplant HLA mistyping in diagnostic samples of acute myeloid leukemia patients due to acquired uniparental disomy. *Leukemia*. 2012 Sep;26(9):2079–85.

268. Gajecka M. Unrevealed mosaicism in the next-generation sequencing era. *Mol Genet Genomics*. 2016 Apr;291(2):513–30.

269. Donovan FX, Kimble DC, Kim Y, Lach FP, Harper U, Kamat A, et al. Paternal or Maternal Uniparental Disomy of Chromosome 16 Resulting in Homozygosity of a Mutant Allele Causes Fanconi Anemia. *Hum Mutat*. 2016 May;37(5):465–8.

270. Choate KA, Lu Y, Zhou J, Choi M, Elias PM, Farhi A, et al. Mitotic Recombination in Patients with Ichthyosis Causes Reversion of Dominant Mutations in KRT10. *Science*. 2010 Oct 1;330(6000):94–7.

271. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet*. 2012 Aug 10;91(2):275–92.

272. Chomik S, Klincewicz B, Cichy W. Disease specific knowledge about cystic fibrosis, patient education and counselling in Poland. *Ann Agric Environ Med*. 2014 Jan;21(2):420–4.

273. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, et al. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet*. 2005 Jul;77(1):78–88.

274. Chen J-M, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet*. 2007 Oct 11;8(10):762–75.

275. Molica S, Alberti A. Prognostic value of the lymphocyte doubling time in chronic lymphocytic leukemia. *Cancer*. 1987 Dec 1;60(11):2712–6.

276. Jacobsohn DA, Vogelsang GB. Acute graft versus host disease. *Orphanet J Rare Dis*. 2007 Sep 4;2:35.