Finding genes related to homologous recombination as modifiers of the number of dermal neurofibromas in Neurofibromatosis type 1 patients

Estudi sobre la implicació dels gens de recombinació homòloga com a modificadors del nombre de neurofibromes dèrmics en pacients amb Neurofibromatosi tipus 1

Carles Garcia Linares

UNIVERSITAT DE BARCELONA

FACULTAT DE BIOLOGIA

DEPARTAMENT DE GENÈTICA

PROGRAMA DE DOCTORAT EN GENÈTICA

(Bienni 2006-2008)

Finding genes related to homologous recombination as modifiers of the number of dermal neurofibromas in Neurofibromatosis type 1 patients

Estudi sobre la implicació dels gens de recombinació homòloga com a modificadors del nombre de neurofibromes dèrmics en pacients amb Neurofibromatosi tipus 1

CARLES GARCIA LINARES

September 2012

Finding genes related to homologous recombination as modifiers of the number of dermal neurofibromas in Neurofibromatosis type 1 patients

Estudi sobre la implicació dels gens de recombinació homòloga com a modificadors del nombre de neurofibromes dèrmics en pacients amb Neurofibromatosi tipus 1

Memòria presentada per **Carles Garcia Linares** per optar al grau de **Doctor en Biologia per la Universitat de Barcelona.**

La present tesi ha estat realitzada sota la direcció del Dr. Eduard Serra Arenas al Departament de Genètica de l'Idibell, i a l'Institut de Medicina Predictiva i Personalitzada del Càncer (IMPPC).

Badalona, Setembre 2012

Dr. Eduard Serra Arenas
Director

Dr. Daniel Grinberg
Tutor

Carles Garcia Linares

Als meus pares

TABLE OF CONTENTS

ACRONYMS

5-FoA	5-Fluoroorotic acid
α-aa	α-aminoadipate
AcLi	Lithium acetate
AI	Allelic imbalance
ALT	Alternatively Spliced Exon
APC	Adenomatous Polyposis Coli
BAF	B-allele frequency
BER	Base excision repair
BIR	Break induced replication
BLM	Bloom
BRCA1	Breast cancer 1,
BRCA2	Breast cancer 2
CI	Confidence interval
CNC	Copy-number change
CNV	Copy-number variation
CORE	COunterselectable REporter
CS	Cancer syndrome
CV	Coeficient of variation
D-loop	Displacement loop
DDR	DNA damage response
dHJ	Double holliday junction
DMEM	Dulbecco's Modified Eagle Medium
dNF	Dermal neurofibroma
DSB	Double strand break
ERM	Ezrin, Radixin and Moesin
EV12	Ecotropic Viral integration site
FA	Fluctuation analysis
FAP	Familial adenomatous polyposis
FBS	Fetal bovine serum
Fo	Forskolin
G418	Geneticin
GAP	GTPase activating protein
GDP	Guanosine diphosphate
GEF	Guanine nucleotide exchange factor
GPHMM	Global parameter hidden markov model
GRD	Gap-related domain
GTP	Guanosine-*5′*-triphosphate
GWAS	Genome wide association studies
His	Histidine
HML	Hidden MAT left
HMR	Hidden MAT right
HoReYe	Homologous Recombination Yeast
HR	Homologous recombination
HSC	Hematopoietic stem cells
IBMX	3-iso-butyl-1-methilxantine
ISCS	Intenstinal stem cells
K	Constant of proportionality
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LA	Linkage analysis

LB/A	Lysogeny Broth/with ampicilin
LD	Linkage disequilibrium
Leu	Leucine
LOH	Loss of heterozygosity
Lys	Lysine
MAPK	Mitogen activated protein kinase
MCS	Multi cloning site
MLPA	Multiplex ligation-dependent probe amplification
MMP	Microsatellite multiplex PCR
MMPA	Microsatellite multiplex PCR analysis
MMR	Mismatch repair
MPNST	Malignant peripheral nerve sheath tumor
MRN	Mre11-Rad50-Nbs1
MZ	Monozygotic
NAHR	Non allelic homologous recombination
Nat	Nourseothricin
NCSC	Neural Crest Stem Cell
NER	Nucleotide excision repair
NF1	Neurofibromatosis type-1
NF2	Neurofibromatosis type-2
NGS	Next-generation sequencing
NHEJ	Non homologous end joining
NIH	National Institutes of Health
NMD	Non-sense mediated decay
OMGP	Oligodendrocyte-Myelin Gycoprotein
OOR	Out of range
PBS	Phosphate buffered saline
PH	Pleckstrin homology
PRA	Paralog ratio analysis
PRS	Paralogous recombination site
PTEN	Phosphatase and tensin homolog
QTL	Quantitative trait loci
QTN	Quantitative trait nucleotide
RB	Retinoblastoma
RDTB	Risk of Dermal Tumor Burden
REP	Repetitive sequence
SC	Schwann cell
SCM	Schwann cell medium
SDSA	Synthesis dependent strand annealing
SGA	Segregation analysis
SKP	Skin-derived precursor
SMARCB1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin
SNP	Single Nucleotide Polymorphisms
SSA	Single-strand annealing
SSB	Single strand break
SUZ12	Suppressor of zeste 12 homolog
TP53	Tumor protein p53
Trp	Tryptophan
TSC	Tuberous Sclerosis

TSG	Tumor suppressor gene
Ura	Uracil
UV	Ultraviolet light
X-QTL	Extreme Quantitative trait loci
YPD/A	Yeast peptone dextrose/with adenine

INTRODUCTION

Cancer

A cell can be seen as the simplest unit of life, but still it remains a complex biological entity. Most cellular processes like DNA transcription and translation, cellular division, etc. are regulated throughout numerous systems and pathways that preserve cellular integrity and homeostasis. Cancer defines a group of complex diseases caused by deregulations in these biological processes that preserve the homeostasis of a cell, leading to an uncontrolled cell division, invasion and metastasis. Disruption of this cellular homeostasis is the result of alterations in the regulation of multiple genes, which can increase, decrease or loss their functions by different means, and transform a normal cell into a cancerous one [1,2]. The complexity of the cell makes it difficult to summarize genes and pathways necessary for tumor development. Therefore Hanahan and Weinberg described some capabilities a normal cell should acquire to become tumorigenic [3], revisited in [2], which include: sustain of proliferative signaling; evasion of growth suppression; avoidance of immune destruction; deregulation of cellular energetics; alteration of the repair machinery; alteration of the inflammatory response; resistance to cell death; activation of invasion, metastasis and angiogenesis; and an immortal replicability.

Cells present mechanisms to protect themselves from becoming tumorigenic by, for instance, repairing DNA alterations or forcing the cell to suicide (apoptosis). Therefore, most times, genes involved in tumorigenesis are the ones responsible of these cellular processes. These genes can be classified in three different types: oncogenes, which promote cell survival and proliferation; tumor suppressor genes (TSGs), which control the cellular proliferation and promote cell death; and stability genes, which maintain genome integrity and are related mostly to repair mechanisms [4,5]. Tumorigenesis would arise after a genetic or epigenetic alteration of these genes (increasing, decreasing or losing their biological functions) that would affect pathways related to the cell capabilities previously explained. Tumor formation also depends on the cell microenvironment. Tumors are the result of alterations in specific cells and their interactions with the surrounding stroma, composed by different cell-types like fibroblasts, endothelial cells, pericytes or leukocytes contributing to cancer initiation and progression [2,6].

But, how do genes controlling cellular homeostasis get altered? The genes of a cell

suffer from both internal and external aggressions that can produce DNA damage. To avoid this damage, cells have evolved an effective DNA damage response (DDR). However, sometimes these mechanisms are not enough to impede the accumulation of DNA alterations, which can end up altering gene function and modifying the capabilities of a cell. There is not a unique process by which cells adopt a tumorigenic state. Each cancer is unique, and it progresses through particular stages in its development that might be different in time, length and number, depending on the cancer or tumor type. Even the same tumor types can appear using different pathways. In addition, some will be caused by punctual alterations affecting particular pathways, while others might present alterations causing huge genomic instability. Tumor formation will depend on the particular alterations suffered in genes, for each particular cell-type, contained in a particular tissue. A well-documented process of cancer progression was described for colorectal cancer, which exemplifies a multi-stage process of tumorigenesis [4,7]: first, detectable lesions in crypt foci present lack of *APC* gene expression [people with Familial Adenomatous Polyposis (FAP) inherit a germline loss-of-function mutation in the *APC* gene]. In early stage adenomas methylation plays an important role by inhibiting (hypermethylation) or activating (hypomethylation) CpG sequences of specific genes involved in tumorigenesis. In 50% intermediate and late adenomas, mutations in *KRAS* oncogene are found (in contrast to 10% early adenomas). 50% of late adenomas and carcinomas present loss of heterozygosity (LOH) in 18q, which it is proposed produces the loss of *SMAD4* gene. Finally, colorectal carcinomas, but not adenomas, present a high percentage of *TP53* gene alterations. This multi-stage process has been found to explain ~60% of the cases reported.

One of the main issues arisen in recent years regarding tumorigenesis is the cell-type promoting tumor formation. Tumors are heterogeneous at a cellular, functional and phenotypic level, and recent evidence suggests that only stem cells, or progenitor cells developing stem-like properties, might be the cell-type suffering the original alterations in tumor formation [2,8,9]. This could explain tumor cell properties like self-renewal, longevity or proliferative potential. Examples are mutations in the long-term hematopoietic stem cells (HSCs), responsible for chronic myeloid leukemia [10], alterations in the intestinal stem cells (ISCs), responsible for colorectal cancer [11], or mutations in stem cells from luminal cellular compartments, responsible for prostate

cancer [12]. It has been proposed, in addition, that cancer stem-cells could be responsible for anticancer drug resistance after chemotherapy treatments, as they could be able to differentiate and repopulate the tumor [13,14].

Cancer as a complex trait

A trait is considered complex when it is affected by multiple factors, genetic (involving many genes), environmental, and by the interplay between them [15]. Most cancers can be considered as complex diseases, as they result from alterations in multiple genes together with external environment interactions. Contribution of genes and environment is variable between cancers: some have a strong genetic component (e.g. hereditary cancer syndromes) and in others the main cause might be the environment, like external aggressions (e.g. tumors arising from radiation exposure). Between these two extreme models, cancers can show any combination of both types of factors contributing to tumorigenesis (Figure 1).

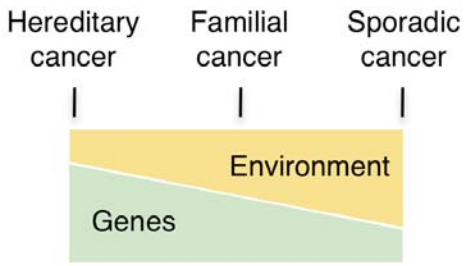


Figure 1. Importance of genes and environment in the different cancer-types.

The genetic predisposition of an individual to develop cancer reflects its risk of tumorigenesis caused by the inherited genetic content. Changes in certain genes can make an individual to be more susceptible to develop cancer. For instance, the cells of an individual can be more or less efficient in repairing certain DNA aggressions. The combination of its inherited DNA (genetic sequence, genomic structure, DNA modifications...), and environmental aggressions will confer the global risk of an individual to develop cancer. Thus, cancer predisposition could be seen as a gradual combination of both inherited and environmental factors (Figure 1). Considering the appearance of cancer in families, we can identify sporadic, familial or hereditary cancers: sporadic cancer appears as a unique (or almost unique) case in a family.

Familial cancer is considered when there is a clear aggregation of a cancer type, appearing in several individuals of a family. Could be due to genetic factors that have not been identified, normally associated with non-Mendelian patterns of inheritance that might be caused by more than a single gene. They could also be to the common exposure to an environmental factor by all members developing cancer. Finally, hereditary cancer syndromes appear in individuals of a family with a clear Mendelian pattern of inheritance, in which environment would play a minor role (Figure 1).

Hereditary Cancer Syndromes

An hereditary cancer syndrome (CS) is a genetic disease with Mendelian inheritance caused by mutations in specific genes that confer, to individuals carrying mutations, a high predisposition to develop cancer [16]. The genes responsible for CS can be classified as oncogenes, TSGs and stability genes. The pattern of inheritance of CS is normally autosomal dominant for oncogenes and TSGs (there are also some X-linked syndromes), and autosomal recessive (with some exceptions showing dominant inheritance) for stability genes [5].

The majority of CS are caused by TSGs. These CS present a dominant inheritance pattern, with the only requirement of the presence of one mutated allele in order to develop the disease. However, tumors associated to these CS arise, almost always, due to a second mutation in the remaining wild-type allele, showing a recessive nature at a cellular level.

This process was explained by Knudson's *two-hit* hypothesis postulated when studying retinoblastoma patients. Knudson statistically demonstrated that patients with hereditary retinoblastoma inherited a germline mutation in one of the *RB* gene alleles, while the other *RB* allele was mutated somatically in the tumor. By contrast, non-hereditary retinoblastoma required mutations in both *RB* gene alleles somatically [17] (Figure 2).

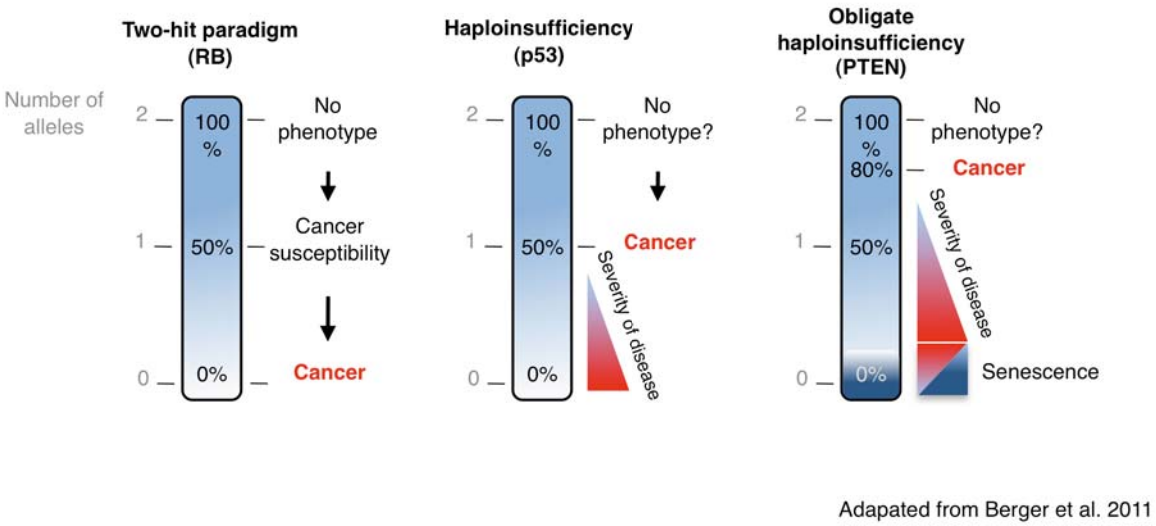


Figure 2. Different models of TSGs function in tumor development.

This idea explained why patients with hereditary cancer syndromes usually showed tumorigenesis earlier than patients with sporadic cancers (in patients with CS only one additional somatic mutation in the normal allele of tumor-initiating cells is sufficient to start the process of tumor formation), and showed that the same gene can be responsible for both sporadic and hereditary cancers in different patients. This model was experimentally confirmed for the first time by Cavenee et al [18], a work that demonstrated the different types of mutational mechanisms that can lead to a somatic inactivation of the Rb gene.

The two-hit hypothesis has evolved, and today it has been demonstrated that the somatic mutation generating loss of the second allele (the wild-type allele) is not always necessary to contribute for tumor development. For instance the *PTEN* gene, a tumor suppressor gene involved in the appearance of multiple cancers, its haploinsufficiency is more tumorigenic in prostate cancer initiation (first steps) than the complete loss of the gene (also called obligate haploinsufficiency) [19], and in advanced tumors its complete loss enhances proliferation (Figure 2). Another example is the tumor suppressor gene *TP53*, whose haploinsufficiency is enough to contribute to tumor formation [20] (Figure 2). The two-hit hypothesis has therefore evolved to a dosage-dependent TSG function, in which each mutated TSG contributes to tumorigenesis depending on how its residual function affects the pathway it is involved [21] (Figure 2). This reasoning applies also for some oncogenes like *RAS*,

whose increased activity could lead to senescence and not to cell proliferation [22].

The study of CS has proven very important in cancer research. Cancer syndromes provide a useful framework to study cancer biology. The high penetrance of these diseases allowed the identification of genes related to tumorigenesis. Genes causing CS are involved in signaling pathways and cellular mechanisms that have facilitated the understanding of the molecular basis of cancer [5]. For instance, the alterations in the gene *BRCA1* were first identified by analyzing patients with familial Breast-ovarian cancer syndrome [23], and were later related to sporadic breast cancer [24]. In addition, the today the genome sequencing of different types of tumors has shown the recurrent identification of mutations in genes responsible for cancer syndromes, like *TP53* or *NF1*.

Neurofibromatoses

The neurofibromatoses, Neurofibromatosis type-1 (NF1), Neurofibromatosis type-2 (NF2) and Schwannomatosis, are a group of inherited diseases that share the development of nervous system tumors as one of the main clinical manifestations, in particular the tumorigenesis of the peripheral nervous system initiated by Schwann cells [reviewed in [25,26]]. However, they differ in their genetic and molecular pathogenesis, as the genes and pathways affected by these diseases are clearly different. NF1 patients bear a mutation in the *NF1* gene, which is mainly involved in the regulation of the RAS/mitogen activated protein kinase (MAPK) pathway. Café-au-lait spots and multiple dermal neurofibromas are the most common clinical manifestations. NF2 patients bear a mutation in the *NF2* gene. The protein product Merlin is related to the group of ERM proteins (Ezrin, Radixin and Moesin), and provides a linkage between membrane proteins and actin cytoskeleton. The appearance of bilateral vestibular Schwannomas is the most distinctive feature of NF2 patients. The gene causing Schwannomatosis is SMARCB1. These patients develop schwannomas, the same tumor-type observed in NF2 patients with the only difference that they are localized at any peripheral nerve, but rarely at the vestibular nerve [25].

Advances in molecular genetics has led to the recognition of a new group of genetic disorders, the RASopathies, which share the alteration in genes related to the Ras/MAPK pathway [27], involved in cell cycle regulation and differentiation.

RASopathies comprise NF1, Noonan syndrome, LEOPARD syndrome and Legius syndrome, among others.

Neurofibromatosis type-1

Neurofibromatosis type-1 (NF1) is a cancer syndrome characterized by the generation of neural crest-derived tumors [28,29]. It is considered a rare disease, with a prevalence in adults of 1/3000 − 1/4000 [30], and an almost virtually complete penetrance at the age of 8 [31]. It is inherited as an autosomal dominant disease, with 50% of cases being *de novo* cases, with no familial history. A hallmark feature of NF1 is its high variability in its clinical manifestations, and its pleiotropy, affecting different tissues [26,32].

Molecular basis of Neurofibromatosis type-1

Gene structure and composition

The gene responsible for Neurofibromatosis type-1 is the *NF1* gene [33-35]. The *NF1* gene is located at 17q11.2, the pericentromeric region of the 17th chromosome [33-35] (Figure 3). It spans ~300Kb, and is composed by 60 exons, three of which are alternatively spliced exons, 9a, 23a and 48a. Intron 27b contains three genes (Figure 3) transcribed in oposite direction with respect to the *NF1* gene, OMGP (Oligodendrocyte-Myelin Gycoprotein), a membrane protein involved in cell adhesion [35], and EV12A and B (Ecotropic Viral integration site), related to murine leukemia [34,36]. The promoter contains a CpG island spanning 10Kbs located upstream the ORF [37], and the 3'UTR spans 3,5Kb. *NF1* gene is surrounded by low-copy repeats termed NF1-REP A, NF1-REP B and NF1-REP C [38-40] (Figure 3). These repeats are responsible for large deletions (also called microdeletions) due to non-allelic homologous recombination (NAHR), identified in NF1 patients as germline mutations. There exist 3 types of microdeletions mediated by REPs (see Germline mutational spectrum section). There are up to 30 *NF1* pseudogenes scattered throughout the genome with different sequence homology with the *NF1* gene [41], some of which are expressed at the mRNA level [42].

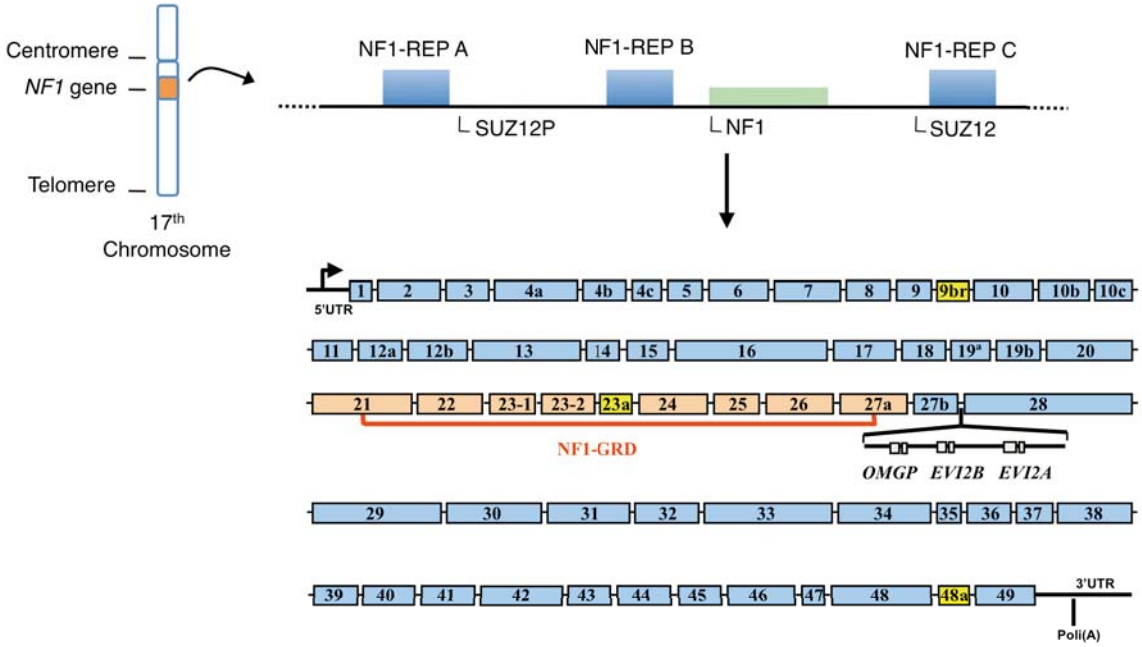


Figure 3. *NF1* gene structure and NF1-REPs localization.

Gene expression

The *NF1* gene is ubiquitously expressed in all tissues, but its expression is especially significant in the brain, spinal cord, and the peripheral nervous system [33,43]. It encodes for an mRNA of 11-13Kb (due to a variable polyA) [44], which by alternative splicing can generate five different isoforms. The main isoforms are type I and type II. Type I isoform contains 57 exons [45]. Type II isoform contains 58 exons and includes the alternative spliced exon 23a [46]. Exon 23a is inserted in the middle of the GRD domain (see below), and neurofibromin produced by this isoform has a lower capacity of negatively regulating Ras than Type I isoform (see below). Type II isoform is the most abundant *NF1* transcript found in adult tissues [47]. Other isoforms are 3'ALT (Alternatively Spliced Exon), which contains 58 exons including the alternative spliced exon 48a [48], 5'ALT2, which contains 58 exons including the alternative exon 9a, and 5'ALT1, that misses exons 21 to 27a and is expressed in the brain (49). *NF1* mRNA can also be regulated by post-transcriptional editing. A cytidine deaminase protein, APOBEC-1, can mediate a C-U change in nucleotide 3916 (CGA-UGA), creating an in frame translation stop codon [49]. This editing process is another way of

inactivating the expression of the *NF1* wild-type allele. However, it doesn't seem to be a mechanism involved in the generation of tumors in NF1 patients, as most somatic inactivation found in NF1 patients are caused by mutations [50].

Neurofibromin

The *NF1* gene encodes a protein called neurofibromin. It has 2818 amino acids and a molecular mass of 250-280 KDa [51,52]. Different domains have been proposed [53], but only 2 have been well characterized at a functional level: the GAP-related domain (GRD), and the Sec14-PH module [composed by Sec14 and PH-like domains]. The GRD domain is composed by 360 Aas and at the DNA sequence corresponds to exons 20−27a. It is conserved among different species, and preserves a high degree of homology and sequence identity with proteins like Ira1p and Ira2p of Saccharomyces cerevisiae (55, 56), Drosophila's GAP1 [54], and also with other mammalian RAS GTPase activating proteins (GAPs) (58). Neurofibromin crystallization helped elucidating the Ras-neurofibromin interaction [55]. GRD domain contains an arginine finger structure that interacts with Ras, stabilizing the transition state of the GTPase reaction. This transition favors the Ras GTPase activity, generating RAS-GDP, and thus inactivating RAS (56, 60). The Sec14-PH module comprises AAs 1560-1816. The Sec14 domain is homologous to a domain in yeast Sec14p protein, which is involved in secretory and lipid-regulated processes [56]. The PH (pleckstrin homology) domains are involved in membrane recruitment processes and it is believed they can promote binding to specific proteins. The PH-like domain seems to modulate the activation of Sec14 domain when interacting with membrane phospholipids.

The NF1 protein is mainly expressed in the nervous system. It has been found mostly in the cytoplasm of neurons, oligodendrocytes and nonmyelinating Schwann cells, and is absent in astrocytes or myelinating Schwann cells [43]. Neurofibromin expression is reduced after the activation of G protein-coupled receptors or receptor tyrosine kinases in growth factor signaling pathways. Its degradation is mediated by the ubiquitin-proteasome pathway. Its regulation follows a cycle of degradation after growth factor signaling, with Ras presenting its maximal peak of activity, followed by a rapid re-expression, promoting Ras inactivation and attenuating the Ras pathway.

Neurofibromin and RAS

RAS, or p21Ras, is a protein anchored to the cell membrane by a farnesyl group [57], and transduces signals from cell surface receptors to intracellular pathways (63, 64) (Figure 4). It is a GTPase [58] that cycles between an active form, bound to GTP, to an inactive form, bound to GDP. This transition is regulated by two different mediators: guanine nucleotide exchange factors (GEFs), which favors Ras activation by accelerating the release of GDP, enabling the insertion of GTP (66, 67), and GTPase activating (GAP) proteins, that stimulates endogenous Ras-GTP hydrolysis [59], hence favoring its inactivation. NF1 is a GAP protein that interacts with RAS (58, 60) as a negative regulator (Figure 4). Ras is involved in different signaling pathways that control several cellular processes. For instance, the Ras/Raf/Mek/Erk pathway, involved in cell proliferation, differentiation, survival and cell death [60], and the Ras/PI3K/Akt/Pkt/mTOR pathway, involved also in cell survival and proliferation (70, 71)(Figure 4).

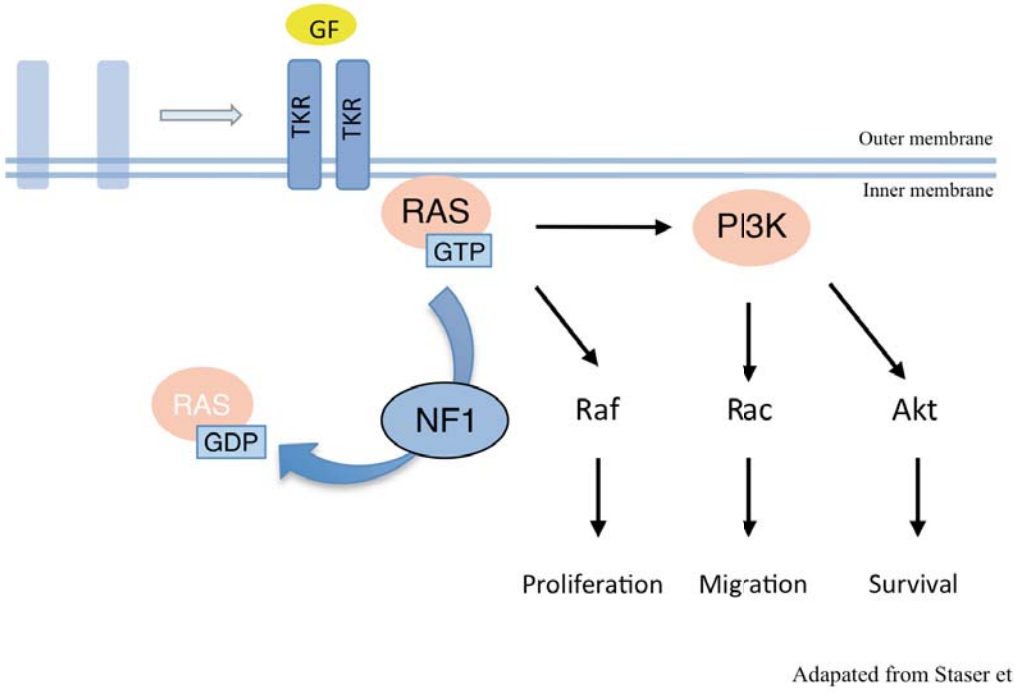


Figure 4. Scheme of *RAS* regulation by the *NF1* gene. GF: growth factor. TKR: Tyrosine kinase receptor.

There are multiple findings that define *NF1* as a TSG. It is germinally altered in all tumors developed. In addition to the germline inactivation, tumors present somatic mutations affecting the wild-type *NF1* gene (72, 73), following the *two-hit* hypothesis. *In vivo* studies demonstrated that mice bearing germline *NF1* mutations generated malignancies associated to NF1 (74). And *in vitro* studies using Schwann cells showed that, in the absence of neurofibromin, cells increased their proliferation and survival capacities (75, 76).

Molecular pathogenesis of Neurofibromatosis type 1

Clinical aspects

Neurofibromatosis type 1 patients exhibit a variable clinical expressivity. The most frequent manifestations are alterations of skin pigmentation, Lisch nodules (iris hamartomas) and multiple benign dermal neurofibromas (29, 77). NF1 patients can develop also severe manifestations like skeletal abnormalities, learning disabilities, vasculopathies, tumors in the central nervous system or Malignant Peripheral Nerve Sheath Tumors (MPNSTs) (29, 77), and present a life expectancy reduced by 15 years [61].

Diagnostic criteria were developed to help identifying and managing NF1 patients [National Institutes of Health (NIH) Consensus Conference in 1987]. NF1 patients should show at least two of the following clinical manifestations: 1) Six or more café-au-lait macules with a diameter >5mm, 2) Two ore more neurofibromas or one plexiform neurofibroma, 3) Freckling in the axillary or inguinal regions, 4) Optic glioma, 5) Two or more Lisch nodules, 6) A distinctive osseous lesion, 7) A first-degree NF1 relative. In NF1 patients these clinical criteria are accomplished with almost full penetrance by the age of 8 [31].

Patients can also develop NF1 in a mosaic form. NF1 mosaic patients suffer the first *NF1* allele mutation during embryonic development in a particular cell. Therefore, these NF1 patients bear two different cell populations: the normal *NF1*(+/+) cells, and the descendants of the *NF1*(+/-) mutated cell. These patients are difficult to identify, as they can appear with a mild NF1 phenotype (mosaic generalized NF1), or with localized NF1 manifestations (mosaic segmental NF1) [62].

Neurofibromas

Neurofibromas are one of the most characteristic features of NF1. They are tumors arising from the peripheral nerve sheath, and can be grossly divided in discrete neurofibromas, which arise from a single peripheral nerve, or plexiform neurofibromas, which involve multiple nerve fascicles [63].

Developing multiple dermal neurofibromas (dNFs) (discrete neurofibromas) is a hallmark of NF1. dNFs are benign tumors affecting almost all NF1 patients. They normally range from 0.1 to some centimeters in size, and can appear in a great number in NF1 patients [28]. Depending on their location they can be painful and cause disfigurement. dNFs normally appear at puberty, and increase their number throughout life. Plexiform neurofibromas affect ~30% of NF1 patients [29]. They are congenital tumors that develop internally along a nerve plexus and can become quite large. Some plexiform neurofibromas can progress to malignancy generating an MPNST. This type of malignancy arises in 8-13% of NF1 patients (81, 82)

Neurofibromas, in contrast to other neoplasias, present a high cellular heterogeneity, being composed by Scwhann cells, perineural cells, fibroblasts, mast cells and endothelial cells (83, 84).

Molecular pathogenesis of neurofibromas

A double inactivation of the *NF1* gene is necessary for neurofibroma development. This was established after a sequence of molecular studies performed in dNFs. First, LOH affecting the *NF1* gene was found in 8/22 dNFs, indicating the possible involvement of somatic *NF1* inactivation in benign tumor development [64]. Then, the germline and somatic *NF1* mutations were identified for one dNF [65]Finally, LOH analysis of 60 dNFs from familial cases showed that an important percentage of somatic inactivations (25%) were generating LOH in the *NF1* region, and established that alterations evidenced by LOH were affecting the allele not transmitted with the disease, confirming the presence of a double *NF1* inactivation in these tumors [66].

To further understand the molecular pathogenesis of neurofibromas mouse models have been widely used. First studies in mice [67] confirmed the rate-limiting role for the loss of the wild-type *NF1* allele for neurofibroma development. In addition, they found that mice bearing germline mutations in the *NF1* and *TP53* genes developed

MPNSTs, which supported the causal role *TP53* had in the formation of NF1 malignancies [67].

Cell-type of neurofibroma origin

Despite its multicellular composition, dermal neurofibromas have a unicellular origin [68]. Deciphering which cell-type was originating this neoplasia was difficult: previous studies had shown the importance of SC in different neoplasias originating in the peripheral nervous system [69], and histological studies established Schwann cells as the most prevalent cell-type in dNFs [40-80% determined by Menon et al. [70]], so this cell-type was the best candidate to bear the *NF1* second hit. With the improvements made for the selective growth of SC over other cell-types (92), it was possible to analyze pure populations of dermal neurofibroma SC. The results obtained for only one neurofibroma [71], were later confirmed in a study of 10 dNFs from 6 unrelated patients [72]: germline and somatic mutations were established for these neurofibromas, and their culture led to the identification of somatic inactivations only in pure SC cultures, in contrast to fibroblast cultures for the same dNFs[72]. In addition, two populations of SC were isolated from dNFs: SC with two hits (-/-), and SC with only the germline mutation (+/-), which demonstrated the coexistence of these SC subtypes[72].

Studies in mice helped also to confirm the identity of the mutated cell-type causing neurofibroma initiation. The CreLox system was used to ablate *NF1* function specifically in SC, in mice with *NF1* (+/-) genetic background [73]. Mice losing the wild-type *NF1* gene in SC generated neurofibromas, demonstrating that *NF1* mutations in SCs are sufficient for tumor development. In addition, mice losing *NF1* function in SC, but with an *NF1* (+/+) genetic background didn't generate neurofibromas, which highlighted the importance of the cellular background in tumor formation. In particular, authors discovered the importance that mast cells have in neurofibroma initiation, as mice with SC *NF1*(-/-) could only generate dNFs with mast cells *NF1*(+/-), and not with mast cells (+/+) [73]. This was later confirmed by studies disrupting the tyrosine kinase receptor c-kit, implicated in mast cell migration and proliferation, that end up reducing neurofibroma size in mice [74]. In addition, some studies tried to elucidate the stage of Schwann cell development providing tumorigenic properties by *NF1* deficiency. It was shown that differentiated glial cells,

and not Neural Crest Stem Cells (NCSC), were responsible for tumor development when the wilt-type *NF1* gene was lost [75]. This inactivation, however, should occur at a particular time in embryonic development (in mice it happened at E12.5 days) [76]. Posterior studies proposed skin-derived precursors (SKPs), a neural crest-like stem cells present in both human and mouse dermis, as the possible neurofibroma cell of origin [77].

Dermal versus plexiform neurofibroma

Dermal and plexiform neurofibromas are the most common types of neurofibroma found in NF1 patients [63]. They are both composed by the same cell-types, and need the inactivation of the *NF1* gene in SC to develop [32]. Both tumor types present a very similar expression profiling [78]. They differ, however, in the time at which they appear, as dermal neurofibromas appear at puberty or during pregnancy, and increase in number with age, while plexiform neurofibromas are though to be congenital. They also differ in the nerve-type where they develop, the nerve plexus in the case of plexiform neurofibromas and single peripheral nerves for dNFs. Finally, they differ in their capacity to progress to malignancy, a feature only found in plexiform neurofibromas [32]. The genetic or molecular basis for these differences has not been established yet, and no recurrent mutations have been determined for both dermal and plexiform neurofibromas in studies analyzing genomic integrity.

Germline mutational spectrum

The *NF1* gene has one of the highest mutation rates found in human genes [79]The most comprehensive study regarding *NF1* mutations in the germline used 1770 NF1 patients [80]. Mutations included: 27% splice-site mutations, 26% frameshift, 21% nonsense, 18% missense, 5% deletions, 2% intragenic copy-number changes (CNC) and 1% of other mutations. The mechanisms found generating these alterations were DNApolymerase slippage, methyl-cytosine deamination, tandem-repeat mediated deletions by NAHR mechanism, and spontaneous adenine deamination. Recently, also insertions of Long interspersed (L1) and ALU elements have been determined to cause splice defects [81].

Three types of deletions, accounting for 5% of germline mutations, have been described: Type-1 deletions (38-40) span 1.4Mb, are the most frequent and involve 14 genes within the lost region. The breakpoints are two recombination hotspots, the paralogous recombination sites 1 and 2 (PRS1 and PRS2), located inside the low-copy repeats NF1-REP A and NF1-REP C [82]. Type-2 deletions comprise 1.2Mb of DNA, with breakpoints within the SUZ12 gene and its pseudogene SUZ12p, located close to NF1-REPs A and C (105, 106). Type-3 deletions span 1Mb, with breakpoints within NF1-REP B and NF1-REP C (107). These three types of deletions are generated by the mechanism of NAHR (104, 108, 109). Atypical deletions have been described too, which are not generated by the NAHR mechanism and their breakpoints are variable [62].

Somatic mutational spectrum

Somatic *NF1* inactivation is a key event in the initiation of dNFs. Some studies have analyzed and compiled the spectrum of somatic alterations affecting the *NF1* gene in neurofibromas and other NF1 neoplasms [126 dNFs from 32 patients in Serra et al. [83], 38 cultured SC derived from neurofibromas from 9 NF1 patients in Maertens et al. [50], 89 dNFs from 3 patients in Thomas et al. [84], and 109 dNFs from 46 patients in Thomas et al. [85]]. The main conclusions that can be extracted from these studies are that most, if not all, somatic inactivations are generated by mutations. Point mutations, including nonsense, missense, splice site, microdeletions/microinsertions, indels and larger deletions/insertions, are the most frequent generating somatic *NF1* inactivations. And LOH accounts for 20-25% of somatic inactivations, but it is not detected in microdeletion patients.

Genetic basis of disease expression variability

A high variability in clinical expressivity is a characteristic of NF1 disease. Most traits developed by NF1 patients are complex. To elucidate how much of the phenotypical variability was explained by an inherited component, and how much could be affected by environmental factors, different studies were performed (113-115). These studies analyzed a large number of patients from different families [175 individuals from 48 families in Easton et al. [86], 904 individuals from 373 families in

Szudek et al. [87], and 750 individuals from 275 families in Sabbbagh et al. [88]] for quantitative (quantifiable by number or by a continuous measurement) and binary (if the trait was present or not) features. All works found, despite some little divergence, a major correlation between closer relatives than in distant ones for the different traits analyzed. This was indicative of a strong genetic component responsible for NF1 expression variability. In addition, the study performed in monozygotic (MZ) twins, which showed stronger correlations than siblings, minimized the weight environmental factors could have in the traits studied (113, 115). These studies also showed that almost none of the features analyzed had any correlation to the constitutional *NF1* mutation, as only Sabbagh et al. [88] found a possible correlation between the *NF1* germline mutation and the presence of cutaneous neurofibromas. All together these studies demonstrated the importance of modifier genes in the phenotypical variability of NF1 patients. Regarding the effect of the germline mutation in the clinical manifestations of NF1 patients, so far, only two genotype-phenotype correlations have been established in NF1: a 3bp in-frame deletion in exon 17, associated with an absence of dermal neurofibromas [89], and patients with a microdeletion (type-1, type-2) that show a significant severe phenotype [90].

Human somatic mutation rate and DNA damage response

Human somatic mutation rate

The somatic mutation rate of any organism is the rate at which mutations get fixed in the genome of the cells that compose its body. This rate is difficult to measure [91], and it varies between organisms, but also in different cell-types of the same organism [92], and even in different DNA regions [93].

Germline mutations generate genetic variation, which is necessary for organism evolution, but are also responsible for the appearance of inherited diseases. Somatic mutations can contribute to the appearance of cancer, among other disease conditions [94]. Human somatic mutation rate is the result of an equilibrium between the generation of mutations, caused by exogenous or endogenous factors, and their repair, organized in a DNA damage response (DDR). Therefore, this DDR is one of the most important biological factors influencing the somatic mutation rate of an individual organism.

DNA damage response

Genomic instability is one of the causes of tumorigenesis [95,96]. This instability can be caused by exogenous aggressions, like ultraviolet light (UV), cigarette smoke or industrial chemicals, or by endogenous factors, like reactive oxygen species created by cellular metabolism or errors in DNA replication (124-126).

To maintain genomic integrity cells have evolved to contain a complex mechanism of DNA repair, the DNA damage response (DDR), which counteract DNA lesions. The process of repair consists in the detection of DNA damage, the accumulation of factors needed to repair it, and the physical repair of the lesion [97]. The main repair mechanisms are: 1) Base excision repair (BER), which repairs subtle modifications of DNA produced by base oxidation, alkylation, deamination or single strand breaks (SSBs) [98]; 2) Nucleotide excision repair (NER), that repairs helix-distorting base lesions, such as those generated by UV light or chemicals. It also repairs transcription-coupled alterations that block DNA transcription [99]; 3) Mismatch repair (MMR), that repairs nucleotide misincorporations, insertions or deletions that occur during DNA replication [100]; 4) Homologous recombination (HR) [96] non-homologous end joining (NHEJ) [101]and single-strand annealing (SSA) [102], employed to repair double strand breaks (DSBs), and intra/inter-strand crosslinks generated in the DNA sequence.

Repairing DSBs

HR, together with NHEJ and SSA, are the major mechanisms responsible for repairing DSBs (Figure 5).

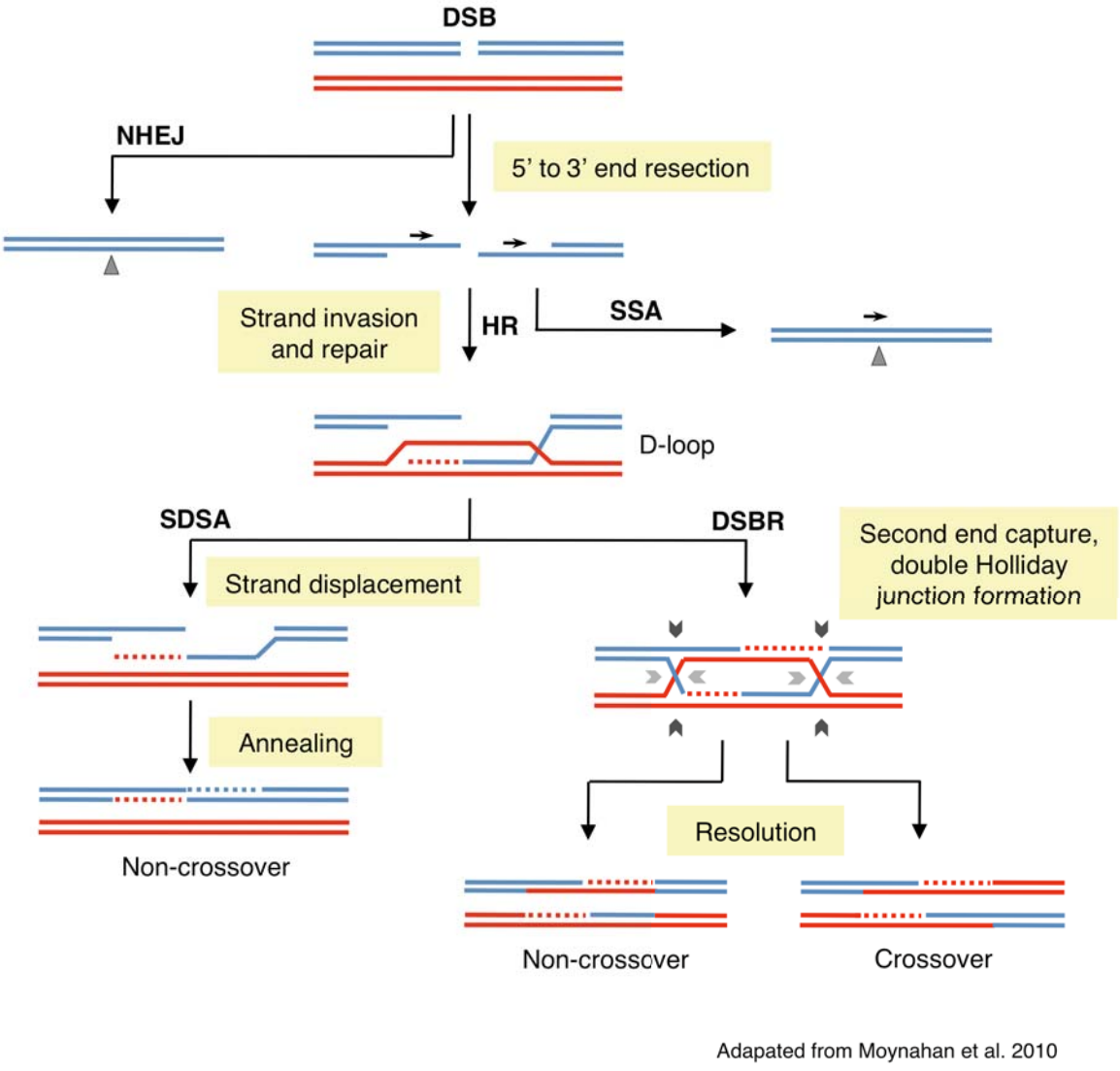


Figure 5. HR, NHEJ and SSA mechanisms repairing a DSB.

DSBs are one of the most cytotoxic forms of DNA damage [103], being able to cause physical dissociation of the chromosome between the two sites around the break. These DNA breaks can be generated by ionizing radiation, chemicals, mechanical stress on chromosomes or by alterations in DNA replication [104]. HR, NHEJ and SSA compete to repair the same type of alteration, but at different cell-cycle phases and

with different results [96,105], being NHEJ and SSA error-prone repair mechanisms while HR is not. NHEJ functions throughout the cell cycle but not in the late S and G2 phases, when homologous sequences of DNA are available. This mechanism involves the ligation of DNA ends with minimal processing of the DNA surrounding the DSB, resulting in a loss of DNA material [101]. The repair of DSBs during V(D)J recombination is also done exclusively by NHEJ [106]. SSA is activated when there are homologous sequences surrounding the DSB. This mechanism recombines these homologous sequences, resulting in the loss of the interstitial DNA and one of the repeated sequences [102]. HR is an error-free repair mechanism. It acts mainly in the late S and G2 phases, when replicated DNA is present, and uses homologous sequences as templates for DNA repair [96]. The resolution of crossovers in meiosis is also exclusively done by the HR mechanism [107].

Homologous recombination mechanism

When a DSB is generated either in the late S or G2 phases, then repair mediated by the HR mechanism takes place (123, 133, 136) (Figure 5):

The first step is the *resection of the 5'-ends* of the DSB, resulting in 3' single-stranded tails (Figure 5). Involved are four nucleases like the MRN (Mre11-Rad50-Nbs1) complex, Exo1, Dna2 and CtIP, and the helicase Blm.

The second step is a *filament formation* in the 3'-single stranded tails. First, RPA (replication protein A) binds to the 3'-ends, eliminating possible secondary structures. Then, the assembly of Rad51 to form filaments is needed, but mediator proteins are required to allow Rad51 filament formation. These are not very well understood, being three the classes of mediators described so far: the Rad51 paralogs Rad51B, Rad51C, Rad51D, Xrcc2 and Xrcc3, the Rad52, only described in *S. Cerevisiae*, and BRCA2, which seems to target Rad51 filament nucleation.

The third step is the *DNA strand invasion* (Figure 5). Once Rad51 is assembled, it performs DNA-homology search and strand invasion, generating a displacement loop (D-loop). Then, this invaded strand is used as a template to perform the DNA synthesis. Proteins like Rad54 are used in this step. It has also been described that MMR machinery is involved in this step by helping in the discrimination between homologous and homeologous (non 100% homologous) sequences [108].

The forth and final step is the *D-loop resolution* (Figure 5), which can be resolved by three different processes: BIR (Break-induced replication), SDSA (Synthesis-dependent strand annealing) or dHJ (double Holliday junction). BIR is activated if one of the ends of the DSB is displaced, resulting in the restoration of the rest of the chromosome. If the template used is the homologous chromosome, it can produce LOH to all the region restored. SDSA anneals the newly synthesized strand to its original position, allowing for the synthesis of the rest of the sequence. It is the preferred pathway, as it avoids the resolution of crossovers that can generate genomic rearrangements. dHJ subpathway is used if two crossovers are generated. This resolution, typically in meiosis, has only recently been described in somatic cells {Bzymek, 2010 #229}. dHJ can suffer a crossover or a non-crossover resolution. Crossover resolution is usually avoided (Rad51 blocks dHJ formation to favor SDSA) as it has the risk to generate deletions, inversions, or translocations if non-allelic sequences are involved. Up to now, it is still not clear which are the most important proteins related to dHJ resolution.

Homologous recombination and cancer

HR is a crucial repair mechanism in mammalian cells, and alterations in any of the HR genes can cause genome instability (123, 138, 139). These alterations could affect HR mechanism in two ways: by negatively affecting the HR mechanism performance, or promoting it over other repair mechanisms. Mutations in HR genes could inhibit the HR mechanism, promoting other error-prone repair mechanisms like NHEJ or SSA, which could generate DNA losses. In addition, gene alterations could affect HR efficiency in any of the steps required for DNA repair, which could generate other alterations besides the one being repaired. Mutations affecting HR genes could also increase the mechanism involvement in DNA repair. These alterations could promote the HR mechanism over other more adequate repair pathways, causing, for instance, large extensions of LOH in the genome.

The biological importance of the genes playing key roles in HR mechanism is supported by the fact that some of them are tumor suppressor genes, like BRCA1, BRCA2, NBS1 or Bloom, among others, which are responsible for the hereditary cancer syndromes of Familial Breast and Ovarian Cancer (BRCA1 and BRCA2

genes), Nijmegen Breakage syndrome (NBS1 gene) and Bloom syndrome (BLM) (139, 140). Homologous recombination is also a mechanism responsible for the generation of copy-neutral losses of heterozygosity of large genomic regions in tumor samples like in dNFs in patients with NF1 (141, 142), and can also generate DNA deletions by intra or inter-chromosomal recombination, or by intervening in replication slippage events [109].

Yeast as a model system for dissecting complex traits associated to cancer

The study of complex traits in humans

Most biological traits are complex. A complex trait results from the interactions between the DNA products of multiple loci, and their interplay with the environment [15]. The study of a complex trait can be divided in three parts: identifying genes and their variation (in relation to expression regulation and function) at a population level, affecting the variation of the trait; understanding the interactions between them; and elucidating how the environment modulates the phenotype.

Mainly two different strategies have been used to identify Quantitative Trait Loci (QTL) involved in complex traits, linkage analysis (LA) and association studies. Linkage analysis identified chromosomal segments shared by affected individuals in a family more often than random Mendelian segregation would predict. These genetic loci would be physically close ("linked") to the genetic determinants responsible for, at least, part of the phenotype of interest. This method was able to identify genes responsible for Mendelian disorders like Cystic fibrosis (143, 144) or Huntington disease [110]. However, for most complex trait studies, LA didn't provide robust results. LA wasn't adequate to study most complex traits, as it was able to identify only few loci with strong effect affecting the phenotype of interest, and most complex traits might be explained also by the involvement of multiple genetic loci with weak effect in the phenotype. Association studies provided more power to detect weak genetic loci responsible for the trait of interest [111]. Association studies identified markers in linkage disequilibrium, which were statistically more frequent in affected individuals

than in the general population. The most employed genetic markers in Genome Wide Association Studies (GWAS) were Single Nucleotide Polymorphisms (SNPs). This technique has led to the identification of numerous loci implicated in diseases like type 1 and type 2 diabetes (147), prostate cancer (148) or breast cancer (149). However, GWAS have been able to explain only a little fraction of the heritability of the traits studied (understanding heritability of a trait as the proportion of the total variance that is explained by genetic variation). This technique is based on the common variant-common disease hypothesis, and it commonly analyzed variants present at least in 2-3% of the population [as described in HapMap Project [112]]. It has been suggested that rare variants (non-detectable by GWAs) could be responsible for the unexplained missing heritability in complex traits [113]. In addition, GWAS provides a statistical measure, not a genotype-phenotype relation, and genetic loci identified in one population frequently are not applicable to other populations.

Researchers have raised many explanations for the missing heritability not explained by current techniques like the involvement of rare variants, the involvement of structural variation, or heritable epigenetic effects, among others [114]. Recently, next-generation sequencing (NGS) technology has appeared as a new methodology to map loci related to human diseases [115]. NGS obtains information of all the genetic variants, including both common and rare. Therefore, NGS methods can improve the degree of heritability explained by current techniques like GWAS for a particular trait [116]. Today, exome-NGS is one of the best options to study complex traits. It analyzes the sequences of the entire DNA protein-coding regions described in the literature and is cheaper than whole-genome sequencing. Exome-NGS has succesfully been used to identify genes responsible for different Mendelian disorders [117], and it is now being applied to study complex diseases [118]. However, NGS still requires large populations (10.000 individuals) to obtain enough statistical power, and it doesn't solve the problem of identifying genotype-phenotype correlations of the variants found.

Model organisms

Finding QTLs require three main features: an accurate description of the phenotype analyzed, the management of high-density complete genotypes, and the identification of functional genotype-phenotype implications of the variants found. Today, these

aspects are difficult to comprise when performing analyzes regarding humans. In addition, the study of biological processes in humans can be challenging due to ethical and technical issues. That is why model organisms are a good option when trying to study complex traits, as they can provide an adequate framework in which to study some of these processes. They can be used to overcome complicate aspects like the understanding of genetic heterogeneity, the study of processes under controlled environmental variables, the deciphering of the basic genetic architecture regarding complex traits, the identification of genes together with functional genotype-phenotype validation, etc… The success of employing model organisms is sustained by their ease of genetic manipulation and genome resources, breeding characteristics and the capacity of *in vivo* phenotyping [119]. Examples of different model organisms used to study complex traits are *Saccharomyces cerevisiae* [120], *Caenorhabditis elegans* [121], *Drosophila melanogaster* [122] or *Mus musculus* [123].

Yeast as a model organism

Yeast is a good model organism because: 1) It has the complexity of a eukaryotic organism, with the ease of manipulation of a unicellular organism, 2) It is genetically stable, and presents both haploid and diploid forms, 3) Its genome is not large, and multiple techniques are available to manipulate it, 4) It maintains numerous processes conserved with higher organisms, like transcription, cell cycle regulation, metabolism, DNA repair, homologous recombination or apoptosis, among others, 5) it is considered as non-pathogenic, so no special precautions are needed for its manipulation, 6) it presents a quick cell-growth in cheap medium (Joaquin Ariño, Treballs de la SCB, 2011).

Yeast model has hugely contributed to the understanding of important aspects related to eukaryotic cell biology, thanks to the ease of introducing mutations in the yeast genome and studying its effects on particular phenotypes, which provided information on gene-protein functions. It was the first eukaryotic organism fully sequenced, which contributed decisively to the establishment of functional genomics and systems biology fields. It led to the identification of the biological role for most protein-coding genes detected (85%), to the ascertainment of the important genomic conservation between species [1000 yeast genes are members of orthologous gene families related

to human diseases [124]], or to the analysis of yeast genomic variation by high-throughput studies [125].

Yeast features

Genome

Saccharomyces Cerevisiae is a unicellular eukaryotic organism. 16 chromosomes, with a total genomic size of 12Mb, compose its genome, in addition to mitochondrial DNA of 85Kb. Chromosome sizes range from 200Kbs the smallest (Chr. I) to 2.2Mb the biggest (Chr. IV). Yeast cells can exist in haploid (16 chromosomes) or diploid (16 pairs of chromosomes) forms. ~6800 functional genes have been characterized, from which ~5800 proteins have been described. It has historically been used to ferment sugars to produce alcoholic beverages and in the baking industry.

Yeast life cycle

Yeast presents a particular life cycle (Figure 6). Haploid yeast cells have two sexual forms, or mating-types, a and α. Haploid a and α cells bear 16 different and unique chromosomes. Haploid cells can change their mating-type, from a to α and vice versa. When a and α cells are close together, they secrete pheromones: a cells secrete a-factor, which signals the presence of a cells to surrounding α cells. α cells respond by secreting α-factor. Both a and α-factors make the opposite cell-type to grow creating a shmoo, a cellular projection, in the direction where pheromones are detected. This process makes the cells to elongate and, once they find each other, they mate (only with cells of the opposite mating-type). a and α cells mate creating a diploid a/ α cell, with 16 pairs of chromosomes. Haploid and diploid cells can reproduce asexually by mitosis, usually by an asymmetric division called budding. When the environmental conditions become stringent, that is, they present low carbon or nitrogen resources, diploid yeast cells can reproduce sexually by going through meiosis and sporulating. In this process, the diploid cell suffers two meiotic divisions, generating 4 progeny cells, or ascospores, inside an ascus. From these 4 ascospores, 2 are a cells and 2 α cells. When the ascus breaks, the spores get spread. Spores remain in this biological form until the availability of the surrounding energetic resources improves. Once carbon or nitrogen levels rise to normal conditions, haploid spores retake the yeast life

cycle, and secrete pheromones to mate with opposite mating cell-types. The study of the 4 individual spores has been vital to understand, for instance, mechanisms involving structural DNA changes like those caused by DNA repair mechanisms, or by the process of meiosis, as all the resulting genetic DNA variants can be collected and studied.

The activation and repression of the genes responsible for an a or α cells is regulated by the MAT locus. MAT is a locus located in the 3rd chromosome, and surrounded by two loci, HML (hidden MAT left) on the left and HMR (hidden MAT right) on the right, which contain silenced MATα and MATa genes, respectively. The mating-type of the yeast cell is determined by which genes are expressed at the MAT locus, either MATα or MATa. Haploid switch is regulated by the HO gene, which encodes for a DNA-endonuclease. When a mating-type change is needed, HO cleaves DNA at the MAT locus, generating a DSB. This DSB is repaired by the mechanism of HR, which uses any of the two silenced MAT loci (located at HML and HMR) as a template. This process replaces the MAT locus by either MATa or MATα genes, generating a MATa or a MATα cell, respectively. MATa and MATα cells express different genes, which regulate the expression of different pheromones. When two haploid cells mate to generate a diploid, the resulting a/α cell will bear two MAT loci. The expression of a and α genes will activate the genes responsible for diploidy, and inhibit those responsible for haploid processes, like the HO gene. At wild, the diploid state of *Saccharomyces Cerevisiae* is the most stable form.

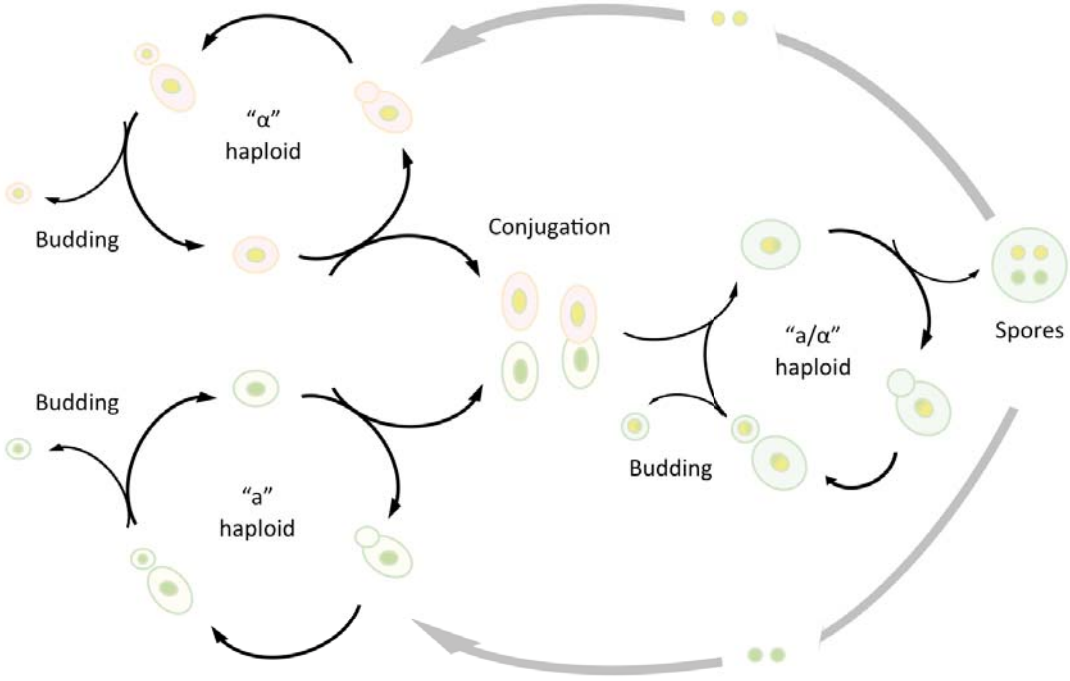


Figure 6. Yeast life cycle

Yeast selection markers

In order to use *Saccharomyces Cerevisiae* in the laboratory, and perform yeast genetics, the employment of selection markers in yeast recombinant DNA technology is essential. Many yeast markers are genes encoding enzymes for the biosynthesis of essential amino acids or nucleotides (auxotrophic markers), which can be selected positively (cells are able to grow when the selection marker is present), negatively (cells grow in the absence of the selection marker) or both (depending on the growing conditions cells are able or not to grow). Auxotrophic genetic markers complement the auxotrophy of the yeast strain when it is unable to synthesize a particular nucleotide or AA. Examples are the selection marker LEU2, which complements yeast cells unable to synthesize the AA of leucine, or URA3, complementing cells that cannot synthesize the nucleotide uracil. Other markers are exogenous genes used to resist the action of certain antibiotics. One example is the selection marker NAT, which confers resistance to the antibiotic nourseothricine.

Yeast-human biological conservation

Yeast has been used to decipher genes and variants involved in human biological processes thanks to the high degree of biological conservation between both species at unicellular level: pathways related to metabolism, cell cycle regulation, DNA repair, aging or apoptosis, among others, are highly conserved between both species (164, 165).

A clear example for the elucidation of human processes using yeast as a model is, for instance, the mechanism of homologous recombination [126]: the proteins involved in the repair mechanism (167), the model of the mechanism itself (168, 169) the generation of Holliday junctions and their resolution (170, 171), or the gene conversion as a consequence of HR [127]. All these biological aspects were first determined in yeast, and identified later in other eukaryotes like humans [105]. Hence, yeast could be considered an adequate model organism to study any process related to HR in humans. Even genes related to human diseases have been identified using yeast, like the identification of the genes PML1, MLH1 and MSH2, involved in Hereditary colorectal cancer, which were first described as candidate genes for the disease after the analysis of the DNA mismatch repair in yeasts [128]. Or the restoration of yeast RAS activity, involved in numerous cancers, by the insertion of mammalian RAS in yeast cells [129], demonstrating the appropriateness of this model organism to study certain aspects related to human cancer.

Studying complex traits in yeast

In the last years yeast model has also been used to successfully identify QTLs affecting complex traits. These studies were based on the analysis of yeast progeny obtained by mating yeast strains with extreme phenotypes. For instance, if the trait of interest was the rate of sporulation, by mating a yeast strain with high rate of sporulation and a yeast strain with low rate of sporulation. In the first assays DNA of the progeny showing the phenotype of interest was analyzed by oligonucleotide arrays, and genetic loci showing low probability of random segregation (being selected for the trait analyzed) were identified [130]. Association studies were performed in yeasts bearing these genetic loci, and no significance was obtained. However, the ease of yeast DNA manipulation allowed researchers to perform Reciprocal-

hemizygosity analysis, a tool for analyzing each allele contained in the selected loci in a hybrid genetic background. This strategy provided a direct genotype-phenotype correlation between the allelic variants obtained and the phenotype studied. Three quantitative trait genes were identified for the high-temperature growth phenotype using this methodology [130]. These assays were later improved by using yeast backcrosses: the progeny showing the trait of interest (following the original example, high rate of sporulation) was backcrossed with the parental strain not showing the phenotype of interest (low rate of sporulation), and, again, another selection for the trait of interest was performed. This backcrossing step, repeated numerous times, maintained the variants responsible for a high rate of sporulation, while enriching the entire genome of the strain with a low rate of sporulation. Then, high-density oligonucleotide arrays analysis at first [131], or direct DNA sequencing later [132], allowed investigators to identify allelic variants coming from the yeast with with the trait of interest contributing to the trait analyzed. The combination of genetic mapping and functional verification provided an experimental and analytical framework to successfully identify not only QTLs, but even Quantitative Trait Nucleotides (QTNs) responsible for the trait studied. Recently, the Extreme QTL-mapping (X-QTL) technique has been developed [120]. This technique doesn't require the backcrossing strategy, as it is able to analyze millions of individuals for the trait of interest and obtain allelic variants comparing yeasts showing the studied phenotype (grown under selection pressure) with control yeasts (grown without selection). X-QTL technique is based in three key steps:

1) Segregation population of very large size: up to 10^7 individuals descending from the parental cells showing extreme phenotypes are grown

2) Selection-based phenotyping of these populations: segregation progeny is divided in two different populations. One is grown under selection for the trait of interest, while the other is grown without selection (control individuals).

3) Quantitative measurements of pooled allele frequencies across the genome by micro-array based genotyping or Next Generation Sequencing (NGS): Tens or hundreds of thousands of individuals from both populations are analyzed. Gene alleles over-represented in the population grown under selection pressure when compared with the control population will be those involved in the trait studied.

With this technique multiple loci affecting complex traits have been identified, in

some cases explaining up to 70% of the complex trait analyzed (158, 178).

OBJECTIVES

Neurofibromatosis type 1 patients present a high variability in their clinical expressivity. The most common manifestation is the appearance of dermal neurofibromas, benign tumors that arise in the peripheral nervous system. They appear at puberty and increase their number throughout life, with patients showing a great variation in their number, ranging from tens to thousands.

The main objective of this thesis consisted in the **identification of genes and variants influencing the number of dermal neurofibromas developed in NF1 patients**. By studying the genetic basis affecting the number of neurofibromas developed the thesis was meant to contribute to:

- Providing biomedical insight for the management and disease prognosis of NF1 patients

- Exploring new approximations to the study of the genetic architecture of complex traits

- Studying the relationship between the human somatic mutation rate and the number of neurofibromas developed

PREMISES AND EXPERIMENTAL DESIGN

Premises

A complex trait with a high heritability

The trait "dermal neurofibroma number" is a complex trait, which means that both genetic and environmental factors influence and modulate it. In this particular case, however, environmental factors are thought to play a minor role, as the peripheral nervous system is a tissue that is not highly exposed to external aggressions (like for instance the colon, the lungs or the skin). Therefore, we hypothesized that a great part of the variation in neurofibroma number among NF1 patients will be caused by variation in genetic factors, meaning a high heritability of the trait. Genetic factors account for both the inherited NF1 mutation and also other modifier genes. Different studies have tried to elucidate the involvement of the *NF1* germline mutation in the number of neurofibromas developed by patients, and only two genotype-phenotype correlations have been determined so far: a 3bp in-frame deletion in exon 17, associated with the absence of dermal neurofibromas, and the presence of a microdeletion, involving the *NF1* gene and surrounding genes adjacent to it, associated with a severe phenotype. However, these correlations account only for a low number of NF1 patients. Moreover, three studies have been performed to elucidate how much of the phenotypical variability was explained by an inherited component (113-115). They concluded that certain quantitative traits associated with the disease, like the number of dermal neurofibromas developed, have a strong genetic component, suggesting the presence of multiple modifier genes involved in this trait, and giving a minor role (as a whole) to germline mutations.

Decomposing a phenotype

Dermal neurofibromas are composed by different cell types (perineural cells, fibroblasts, Schwann cells, macrophages). Neurofibromas develop due to a double inactivation of the *NF1* gene, but only Schwann cells bear it. This inactivation is necessary for neurofibroma development, and it has been demonstrated that it is always caused by mutation [50]. Since neurofibroma development is highly dependent on somatic *NF1* inactivation, we reasoned that genes influencing the somatic mutation

rate of the individual could be good candidates as NF1 modifier genes. Among repair mechanisms, HR has been found to be responsible for a high percentage of somatic *NF1* inactivations in neurofibromas. Homologous recombination is a mechanism that shows interindividual variation [133], and it has been suggested that genes controlling this mechanism *in vivo* might substantially influence the somatic mutation rate of NF1 patients [83]. Thus, from the whole complexity (at a cellular and molecular level) that represents the development of a neurofibroma, and the number of neurofibromas developed, we decided to center our studies only to the SC component and the *NF1* gene inactivation mechanisms. From all possible mutational mechanisms that can cause this inactivation, we focused in HR as a way to decompose the phenotype and facilitate the dissection of the genetic variance [134]. By doing this we were aware that the study was only going to capture part of the heritability of the trait, but at the same time the project was more amenable to a certain extent.

A functional approximation to genotype-phenotype correlations

Deciphering the genetic architecture of complex traits in humans is complicated. Since the beginning of XXI century, this has been performed at high resolution by using approximations like genome-wide association studies (GWAS) applying SNP-arrays. GWAS have been used to identify numerous loci related to important diseases like type 1 and type 2 diabetes (147) or prostate cancer (148). However, until now the majority of the identified loci account for a rather small fraction of the variation on the phenotype studied. Often, after identifying statistically significant loci, functional studies fail to identify a relationship between identified loci and the trait under study. In addition, a GWAS approximation requires a large number of individuals to obtain significant results. We new upfront, before starting this study, that it was difficult to obtain the frequency of dermal neurofibromas exhibiting LOH and HR as a second hit for a large number of NF1 patients. Therefore, since homologous recombination mechanism is well conserved between humans and yeast, we decided to identify candidate genes by modeling *in vivo* homologous recombination rates in different yeast strains. *Saccharomyces cerevisiae* present an unprecedented ease of DNA manipulation for an eukaryotic organism that facilitate genomic modifications for setting up new yeast assays, and, what for us was very important, it allows rapid functional analyses to confirm any genotype-phenotype correlation found in the

studies performed.

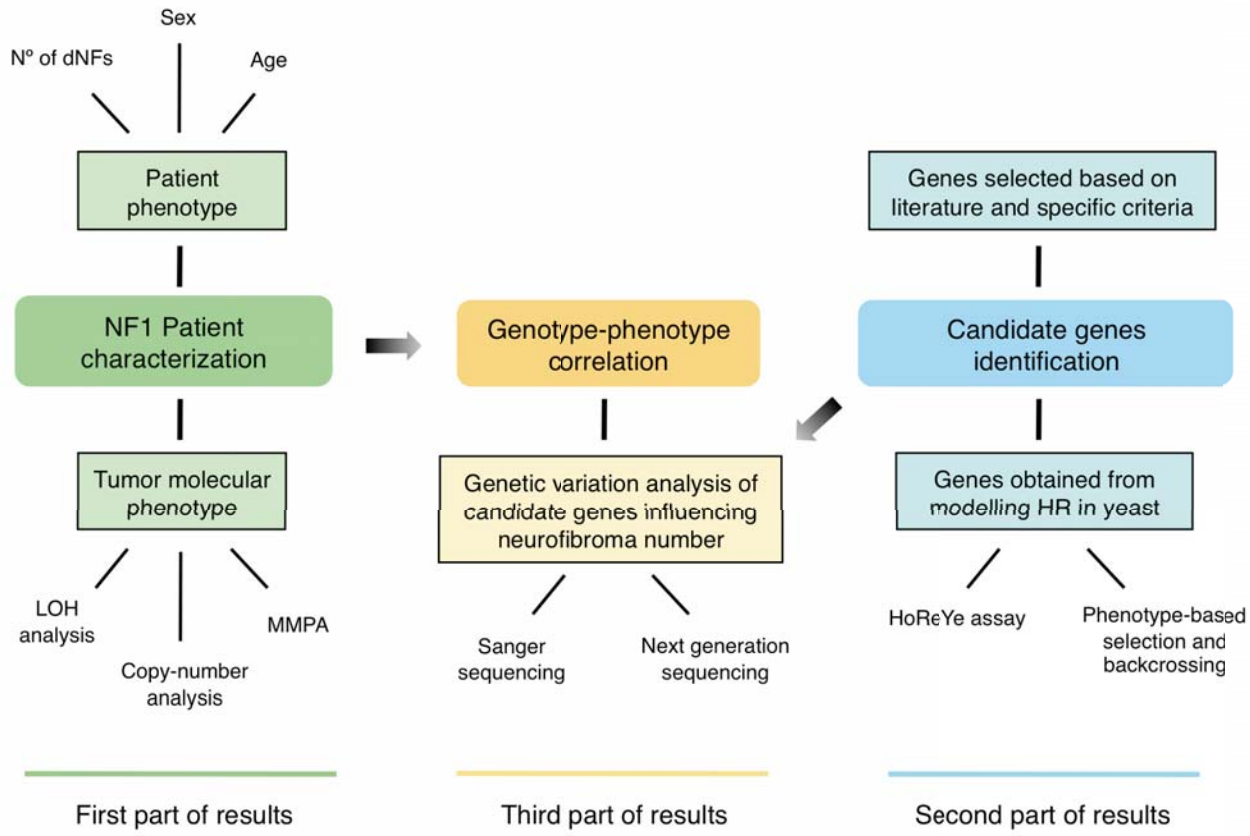
Common and rare variants

Another issue raised in the field of quantitative trait analysis is the type of DNA variants that mainly compose the genetic architecture of complex traits. GWAS studies were based on the "common disease-common variant" hypothesis [135], which proposed that a complex disease is attributable to a moderate number of common variants. The fact that the variants obtained by GWAS only explained a small fraction of the genetic variance of the trait suggested that rare genetic variants could also be responsible for the variation in complex traits [113]. We decided to analyze candidate genes by sequencing (in contrast to analyze specific known polymorphisms) in order to capture all sorts of genetic variants. We first started by Sanger sequencing, but the advent of Next-generation sequencing (NGS) in the middle of the thesis made us to rethink the sequencing technology. We used NGS to sequence targeted exons related to repair mechanisms, avoiding the high cost of analyzing the entire exome of each patient.

Human somatic mutation rate in health and disease

One of the main hypothesis sustaining the present work is that human somatic mutation rate influences the number of dermal neurofibromas developed by NF1 patients. In this regard, if this relationship is strong, dermal neurofibroma number could be seen as an indicator of the *in vivo* somatic mutation rate of NF1 patients, and we would be able to start studying the role of somatic mutation rate in health and disease. By finding genes influencing the number of neurofibromas, we would be finding those genes influencing the somatic mutation rate in humans. There exist other cancer syndromes where patients also develop a variable number of benign tumors, like the number of colon adenomas in Adenomatous Polyposis Coli (APC) patients or the number of hamartomas in Tuberous Sclerosis (TSC) patients. Genes influencing the number of dNFs in NF1 patients could also be influencing the number of benign tumors developed in these other cancer syndromes.

Experimental design scheme



MATERIAL AND METHODS

Patients and samples

NF1 patients

All NF1 patients were diagnosed according to standard diagnostic criteria [31], and gave a written informed consent for the molecular studies performed.

Dermal neurofibromas

Dermal neurofibromas were completely removed after minor surgery, undertaken by either a dermatologist or a surgeon. Surrounding non-neurofibroma tissue was removed and tumors were cut into different pieces. A piece of each tumor was preserved at −80ºC until DNA was extracted. Whenever possible, remaining pieces of each dNF were further cut, immersed in freezing solution [10% DMSO, 90% fetal bovine serum (FBS)], and preserved in liquid nitrogen until they were used to establish cell cultures.

DNA extraction

DNA from skin, neurofibromas, Schwann cells or fibroblast cultures was extracted using the Gentra Puregene Kit (Qiagen, Chatsworth, CA) following manufacturer's instructions. In some cases DNA from venous blood of the patients was also used, and was isolated by the Wizard Genomic DNA purification kit (Promega, Madison, WI) according to the manufacturer's instructions. All DNA samples were stored and preserved at 4ºC and at −80ºC.

DNA quality and quantity assessment

DNA quality and quantity was assessed using gel electrophoresis analysis and a nanodrop spectrophotometer. For an accurate quantification of DNA concentration samples were further quantified using Quant-iT PicoGreen reagent (Invitrogen, Carlsbad, CA), according to manufacturers instructions. Briefly, this method employs an ultrasensitive fluorescent nucleic acid stain that specifically binds dsDNA. ssDNA or other compounds have little effect on Picogreen measures. DNA is quantified by

extrapolating the fluorescence obtained in the sample in a standard curve generated in the same Picogreen assay.

Primers

All primers used are summarized in Table annex A1.

LOH/AI analysis

Microsatellite Multiplex PCR (MMP)

The primers used in the MMP are described in Table 1. Primers were dye-labeled with four different fluorophores, FAM (blue), NED (yellow), VIC (green), and PET (red) (Invitrogen). We designed PCR products of different sizes labeled with different colors in order to differentiate all of them in a single genotyping run. MMP reaction was performed using the Multiplex PCR Kit (Qiagen) in a 25 µl reaction with 80 ng of DNA as follows: an initial cycle of denaturation at 95ºC for 15 min followed by 25 cycles of denaturation at 94ºC, annealing at 56ºC, and extension at 72ºC for 30sec, 3 min and 1.5 min, respectively, and by a final cycle of 60ºC for 30 min. A total of 2 µl of PCR product was mixed with 7.8 µl of highly deionized formamide (Applied Biosystems, Bedford, MA) and 0.2 µl of LIZ500 Size Standard (Applied Biosystems). PCR fragments were separated by capillary electrophoresis on an ABI 3130xl Genetic Analyzer. Peak height values were extracted using Peak Scanner software (Applied Biosystems).

Microsatellite Multiplex PCR Analysis (MMPA)

The primers (Table 1) and conditions used to perform the MMPA are the same as the previous MMP, with the only difference that for the MMPA 22 cycles of PCR amplification were performed: an initial cycle of denaturation at 95ºC for 15 min followed by 22 cycles of denaturation at 94ºC, annealing at 56ºC, and extension at 72ºC for 30sec, 3 min and 1.5 min, respectively, and by a final cycle of 60ºC for 30 min.

Primer name	Primer sequence	Fragment length	Primer final concentration (μM)	Dye
D2S2314	F- GGTGTCAGTGAGACCCTGT	96-118	0,1	FAM
	R- ATTTCTAGCGGCCCTAAAAC			
D17S1879	F- GAAGTTTACGAAAATTGCTGTC	135-173	0,08	FAM
	R- AAGGGTAGTTCTGCGTGC			
D17S1307	F- TAGGAGACCTGCTGCCTTT	209	0,05	FAM
	R- AGGGCAGAGAAACCTAAGGA			
D17S798	F- ACAACATTGTCAAACCCCATC	226-246	0,05	FAM
	R- TGTTCTTGGGAGTGCAG			
D17S1800	F- CTAAACTAGGTTGGGTTGAAATCTC	268-284	0,4	FAM
	R- TCTGGCACAAAGACCTGAG			
D17S250	F- GGAAGAATCAAATAGACAAT	151	0,4	VIC
	R- GCTGGCCATATATATATTTAAACC			
D17S933	F- ACTCACTGGGGTCCTGG	187-206	0,08	VIC
	R- TGTGGTTTCCTTATAGACTGTAGA			
D17S1303	F- CTCTCCAAGGCTCACTCAAA	225-245	0,05	VIC
	R- TGGTCTTTTTCCATTCCAAA			
D17S807	F- TCCACCTGTAGACCTGGTAAA	114-138	0,6	NED
	R- AGTGCTGCGTCTTACAACCT			
D17S789	F- ACTCCAAATCAAGTTTGTACTGAGA	153-170	0,08	NED
	R- CTGCATACGAAGGGTAGGAC			
D17S783	F- GATTTTCACCACTGAAACCA	200-214	0,1	NED
	R- CCTGGGGTAACAGAAAACTT			
D17S1294	F- TGGCATGCAATTGTAGTCTC	248-272	0,08	NED
	R- TTCTTTCCTTACTAAGTTGAGAACG			
IVS38AC53.0	F- CAGAGCAAGACCCTGTCT	170-187	0,3	PET
	R- CTCCTAACATTTATTAACCTTA			
IVS27AC28.4	F- GTTCTCAACTTAAATGTAAGT	207-219	0,3	PET
	R- GAACATTAACAACAAGTACC			
3'NF1	F- CTTCCATGGCTGCTAACATC	233-249	0,08	PET
	R- CCCTGTGGTGTAGTTCAACA			
D17S841	F- AGTTGGACTTTCTTACATGGC	275-295	0,08	PET
	R- AGAATCTCATATTAGGGGAAGG			

Table 1. Primers used in MMP and MMPA

LOH/AI analysis

Loss of heterozygosity (LOH) and allelic imbalance (AI) are used to refer to a loss of one allele (LOH), or the imbalance generated by a loss or gain of one allele (AI), when analyzing microsatellite peak height intensities. To assess the presence of LOH/AI we performed an allelic imbalance analysis, based in the expression Q^{AI} [136]that represents the quotient between the tumor allelic ratio of a given heterozygous microsatellite marker within the MMP reaction and the control sample allelic ratio of the same marker [137]. LOH/AI was always assessed by comparing pairs of control and tumor DNA from the same patient. We considered that a microsatellite marker was showing LOH/AI when, comparing the ratios of the tumor and control alleles (Q^{AI}), there was a difference between ratios equal or greater than 0.2.

Semi-automated analysis of MMPA results

To analyze the data extracted by Peak Scanner software we have developed a tool that automates all the necessary MMPA calculations. This automated analysis has been implemented using Ruby programming language in a script named mmpa.rb. Ruby is a high level interpreted programming language, thus it is very easy to learn and use and can be run on any operating system. The script requires Ruby to be installed on the computer and has to be run from the command line. The program works as follows: mmpa.rb takes as input three plain text files in tabulated format which can be easily generated using any spreadsheet program. The first file must contain the peak heights and sizes of the multiplex PCR as extracted from an electropherogram for no less than a control and a test sample. The second file must contain the denomination, the dyes used, and the expected size range for each microsatellite surveyed by the multiplex PCR. Optionally, chromosome coordinates can be added to allow for an automatic positional ordering of the output. The third required file must indicate the pairings of test and control samples. From these inputs the script batch processes the data and outputs the results of the MMPA in a file; this ouput file consists of a serialized description of each test-control pair analysis performed, as defined by the user in the configuration files. For each paired analysis the script calculates and outputs the Kμ±s of the reaction, the average percentage of normal tissue contamination, and the overall mechanism of AI for the test sample. For each informative microsatellite, calculated Ks, AI determination, observed peak height ratios, expected peak height ratios, % of non-AI cells and mechanisms generating the AIs are also calculated. All output files are written in tabulated format and can be opened with any spreadsheet program for easier inspection of the results.

mmpa.rb is released under the GNU GPL license. Source code and detailed instructions for its installation and use are available as supplementary information and can also be requested to the authors.

NF1 copy-number assessment

Multiplex Ligation-dependent Probe Amplification (MLPA)

We used the Multiplex Ligation-dependent Probe Amplification (MLPA) (MRC-Holland) to analyze the number of copies of the *NF1* gene present in DNA samples. Because the company is improving the assay with time, different kits were used for the characterization of all dNFs: Salsa MLPA Kit P122 versions P122, P122-B1, and P122-C1. This technique is based on a ligation-dependent probe amplification of 12 to 28 probes (depending on the kit used) encompassing the NF1 area together with 11 to 15 additional control probes (also depending on the kit used). Probes differ in size by means of spacers of different length that are probe-specific. Only probes that are properly ligated can be amplified in the subsequent PCR reaction. Once amplified, PCR fragments are separated by capillary electrophoresis and peak intensities are analyzed. For the Salsa MLPA Kit P122, PCR products of different length (ranging from 100 bp to 400 bp, depending on the kit used) were separated on an ABI 3130xl Genetic Analyzer. Peak height values were extracted using Peak Scanner software (Applied Biosystems) and were normalized as described by Wimmer et al. [2006] for peak areas. Briefly, peak heights from individual probes were divided by the sum of all peak heights of a given sample, obtaining a relative value. Relative values of each probe were then divided by the mean of the relative values of that probe obtained only from control samples. After this normalization of peak heights, it was possible to analyze the NF1 copy number. The criteria used to assess the number of copies of the NF1 gene present was the following: a value between 0.8 and 1.2 indicated that two copies of the NF1 gene were present; a value lower than 0.8 implied that only one copy of the NF1 gene was present.

Paralog Ratio Analysis (PRA)

This method was developed for the *NF1* region by De Raedt et al. [138] and determines the *NF1* copy number status of a sample DNA by comparing the peak height values of two PCR products of a single amplification reaction. This PCR reaction coamplifies a fragment of exon 22 present in the NF1 gene and the same fragment bearing an insertion of 4bp from a single NF1 pseudogene located in chromosome 15. PRA

reaction was performed in a 10 µl final reaction with 80ng of DNA as follows: an initial cycle of denaturation at 94ºC for 3 min followed by 26 cycles of denaturation at 94ºC, annealing at 55ºC, and extension at 72ºC for 30sec each step. A final cycle of 72ºC for 7min was performed.

We used a dosage quotient analysis comparing the ratios between gene and pseudogene peaks obtained from a tumor and a control sample from the same patient (in triplicate) and established the presence of a deletion when the quotient of the ratios obtained were below 0.8, that is, when the differences between tumor and control DNAs were greater than 20%. With this technique only deletions covering the region of exon 22 from the NF1 gene can be detected. In addition, it has to be taken into account that the *NF1* pseudogene located in chromosome 15 is placed in a copy number variation (CNV) region, and thus only control and tumor pairs from the same patient can be compared in the analysis, because both DNAs will contain the same number of copies in the CNV region.

SNP-Array

We used the Illumina Infinium technology (Illumina, San Diego, CA). In all cases, pairs of control and neurofibroma DNA were analyzed. Twelve dNFs were analyzed using the Illumina 370-Quad beadchip (~351,000 genotyping SNPs, ~9,300 located in the 17th chromosome), and seven dNFs samples were investigated using the Human660W-Quad beadchip (~592,000 genotyping SNPs, ~15,600 located in the 17th chromosome). DNA used for this type of analysis was of the highest quality. Raw data were analyzed using the Illumina software GenomeStudio v2009.1 with the Genotyping v1.1.9 module. The GenomeStudio program provides two types of measurements. On the one hand, it provides a quantitative measurement of the B-allele frequency (BAF) that reflects the allelic imbalance of each SNP. A four band pattern appears when an AI/LOH is present in the region analyzed. On the other hand, the program generates the logR ratio metric, which reflects the number of DNA copies. The logR ratio is a log-transformed ratio of the measured SNP signal intensity by the expected intensity if two copies of DNA were present. Two different methods were used to calculate the percentage of normal cells present in the tumor sample: the method by Assie et al. [139], that applies the BAF value to different equations depending on the somatic event generating the AI, and the Global Parameter Hidden Markov

Model (GPHMM) algorithm [140] that quantitatively models signal baseline shift due to aneuploidy and also the presence of normal diploid cells.

REP-Mediated Deletion Analysis

After performing Microsatellite Multiplex PCR Analysis and/or MLPA analysis, a few neurofibromas were suspicious of bearing deletions mediated by recombination between duplicated sequences in the NF1 gene region (REP-mediated deletion). In these cases, we performed several diagnostic PCR reactions, as described by Steinmann et al. (186), to identify somatic type-1 or type-2 NF1 deletions with breakpoints either in the NF1-REPs or in the SUZ12 sequences.

Cell culture

Cell culture solutions

Scwhann Cell Medium (SCM): DMEM (Gibco, Grand Island, NY) with 10% FBS, 500U/ml penicillin/streptomycin (Gibco), 0.5 mM 3-iso-butyl-1-methilxantine (IBMX, Sigma), 2.5 μg/ml insulin (Sigma), and 10 nM b1-heregulin (R&D Systems, Minneapolis, MN).

SCM+Fo: SCM with 0.5 μM forskolin (Sigma).

Dissociation solution: SCM with 160U/ml collagenase type 1 (Sigma) and 0.8U/ml dispase (Roche)

Freezing medium: 10% DMSO, 40% FBS and 50% DMEM

Paraformaldehid 4%: Dissolve paraformaldehid in hot water (60ºC), and add NaOH 1M until it gets totally dissolved. Add PBS

Antibody S100: Dilute antibody S100 (Polyclonal rabbit anti-cow S100, Dako) 1:1000 in PBS-FBS 1%. Centrifuge at 4ºC for 5 minutes

Antibody anti-rabbit: Dilute antibody anti-rabbit (Goat anti-rabbit IgG-FITC, Southern Biotech) 1:200 in PBS-FBS 10%

Cell culture experimental procedures

Coating: Filtered Poly-L-Lysine (0.1mg/ml) was added to the well-plate and left for 1h at RT. The well-plate was washed three times with sterile PBS, and filtered laminin (4µg/ml) was then added and left O/N at 4ºC, or 1h at 37ºC. The well-plate was washed three times with sterile PBS and it was ready to use.

Trypsinization: Culture medium was removed from the plate, and the cell monolayer was washed with sterile PBS. Pre-warmed trypsin was added to cover the cell monolayer, and the plate was incubated at 37ºC for 3 minutes. After cells were detached, DMEM was added to inactivate the trypsin. Cell suspension was spun down and cells were ready to be resuspended in the desired medium or solution.

Cell freezing: Cells were spun down and 500µl to 1ml of freezing medium was added. The cell tube was placed in a Mr. Frosty at -80ºC, until putting them in liquid nitrogen.

Neurofibroma-derived Schwann cell culture (SC +/+ or +/-): Neurofibroma pieces preserved in liquid nitrogen were mechanically and enzymatically dissociated (using a Dissociation solution), and cultured in the necessary volume of SCM+Fo. Plates should be previously coated with poly-L-Lysine and laminin, and incubated at 37ºC and 10% CO2. The next day, the medium was replaced by SCM without forskolin for 2–3 additional days. To culture SC (-/-) the culture medium was replaced by SCM+Fo, and the process was repeated in cycles (1 day SCM+Fo, 2-3 days SCM, 1 day SCM+Fo…). To culture SC (+/+) the culture medium was replaced only with SCM. Cells should be passaged when they were confluent, and differentiation should be avoided.

Neurofibroma-derived Schwann cell purity assessment by S100 staining: Cell purity was assessed using antibodies specific for the protein S100, expressed in nervous-system derived cells, like SC. Each washing step was performed for 10 minutes. Cells grown on a cover-slide were first washed with PBS, and fixed with 100 µl p-formaldehid 4% for 15 minutes at RT. After two PBS washes, the cover-slide was treated with 100 µl PBS-triton 0.1% and left for 10 minutes at RT. After two PBS washes, one using PBS-FBS 1% and another with PBS-FBS 10%, 100 µl of antibody S100 was added and the cover-slide was left 1h at RT. Three washing steps with PBS-FBS 1% were performed, and 100 µl antibody anti-rabbit was added

and left for 45 minutes at RT. Three washing steps with PBS-FBS 1% were performed, and the cover-slide was dried on a filter paper. 5 µl of Vectashield with DAPI was placed in a slide, placing the cover-slide on it. The preparation was visualized in an inverted microscope (Leica DMI 6000B).

Candidate gene analysis

Candidate gene experimental procedures:

RT-PCR: Random primers were used to amplify the mRNA. First, two solutions were prepared. Solution 1: Random primers (200ng/µl) 1µl, mix dNTPs (10mM each) 1µl, RNA (1ng-5µg) 5µl, water 6µl. Solution 2: 5x First Strand Buffer 4µl, 0,1M DTT 2µl, SuperScript II RT 1µl. Both solutions were mixed obtaining a total of 20µl, and one PCR cycle of 25ºC for 10 minutes, 42ºC for 50 minutes and 70ºC for 15 minutes was performed, obtaining de cDNA.

cDNA sequencing: The cDNA was purified (e.g with Marligen Kit), and a sequencing-PCR was performed. The PCR mix was composed by DNA (~10ng) 1µl, Reaction Mix 1µl, Primers (3,2pmol) 3,2µl, Buffer 1µl, Water 3,8µl. The PCR conditions were an initial cycle of denaturation at 96ºC for 1 min followed by 25 cycles of denaturation at 96ºC, annealing at 50ºC, and extension at 60ºC for 10sec, 5 seconds and 4 min, respectively. The final PCR product was purified to eliminate remaining ssNTPs (e.g using Sephadex).

Strategies to identify candidate genes

Bacteria solutions:

LB liquid (1L): Add 5gr Bacto yeast extract, 10gr Bacto tryptone, 10gr NaCl, a "GALLETA" NaOH to 1L water and autoclave

LB solid (1L): Add 5gr Bacto yeast extract, 10gr Bacto tryptone, 10gr NaCl, a "GALLETA" NaOH and 12gr Bacto agar to 1L water and autoclave. Cool to 65ºC

and pour plates

LBA (1L): Add 5gr Bacto yeast extract, 10gr Bacto tryptone, 10gr NaCl, a "GALLETA" NaOH to 1L water and autoclave. Cool to 65ºC, and add 100mg/ml of ampiciline

SOB (500ml): Add 10gr tryptone, 2,5gr Bacto yeast extract, 0.25gr NaCl, to 500ml water and autoclave. Then add 5ml 1M MgCl2 and 5ml 1M MgSO4, both sterilized

SOC (100ml): Add 1ml of Glucose 40% to 99ml of SOB

Bacteria experimental procedures:

Bacteria preservation: Bacterial cells were stored by adding 0.5ml bacterial culture to 0.5ml sterile 80% (v/v) glycerol solution

Electrocompetent cell production: Cells were made electrocompetent by a glycerol treatment. All the process was performed at 4ºC. Bacteria glycerinate was inoculated in 10ml LB, and incubated O/N. Cells were transferred to 1L pre-warmed LB medium, and incubated until O.D = 0.6 to 0.7. The culture was transferred to four ice-cold plastic centrifuge bottles and placed to an ice-water bath for 40 minutes to 2h. Cells were pelleted by centrifuging at 4000g for 15 minutes at 4ºC, and resuspended in 10ml ice-cold water. Then, the volume was raised until 250ml (full) x4. Sequences of centrifugation and aspiration were repeated to wash the cells in 250ml (x4), 100ml (x4), finally adding to the pelleted cells 45ml of ice-cold 10% glycerol (20ml per bottle). Bottles were joined, centrifugation and aspiration was repeated, and cells were resuspended in 3ml of 10% glycerol (final volume ~5ml). 10µl (in 990µl H2O, 1/100 dilution) was measured at OD600. It should be around 1. Aliquots of 50µl were transferred to eppendorfs in a bath of dry-ice EtOH 70%, and placed to −80ºC.

Bacterial electroporation: The electroporation process was performed on ice. The elements to be transformed (for instance, a plasmid) were added to the thawed competent cells. All the suspension was placed in an electroporation cuvette, and electroporation was performed (Voltage: 1500v, Resitance: 0125Ohms, Capacity: 50µF). 750 µl of SOC was added, suspension was inserted in an eppendorf and incubated for 45 minutes at 37ºC. Cells were spun down and plated.

Yeast solutions:

YPDA liquid (1L): Add 10gr Bacto yeast extract, 20gr Bactopeptone, 50mg Adenine to 950ml water. After autoclaving, cool to 65ºC and add 50ml of 40% glucose solution.

YPDA solid (1L): Add 10gr Bacto yeast extract, 20gr Bactopeptone, 50mg Adenine, 20gr Bacto agar to 950ml water. After autoclaving, cool to 65ºC, add 50ml of 40% glucose solution and pour plates.

SC-X (Synthetic complete medium lacking amino acid X) liquid (1L): Add 6,7 gr Yeast nitrogen base, 0.77 gr Complete supplement mixture (CSM) minus X (X is the lacking amino acid), and the amino acids needed to complement the media (see below) to 950ml water. After autoclaving, cool to 65ºC and add 50ml of 40% glucose solution.

SC-X solid (1L): Add 6,7 gr Yeast nitrogen base, 0.77 gr Complete supplement mixture (CSM) minus X, the amino acids needed to complement the media (see below) and 20gr Bacto agar to 950ml water. After autoclaving, cool to 65ºC, add 50ml of 40% glucose solution and pour plates. Amino acids added to SC-X (1L): Ade (40mg/L), Arg (20mg/L), His (20mg/L), Leu (100mg/L), Met (20mg/L), Thr (200mg/L), Tryp (40mg/L), Ura (20mg/L).

Nourseothricin or Geneticin selection media solid (1L): For solution 1 add 1,7gr Yeast nitrogen base without (NH4)2SO4, 0.77gr CSM, 1gr Glutamic acid, 20gr Glucose to 150ml water, and filter sterilize. For solution 2 add 20gr Bacto agar to 850ml water and autoclave. Cool both solutions to 65ºC, and mix them. Add the correspondent antibiotic (Nourseothricin: 100mg/L, Geneticin: 200mg/L), mix thoroughly and pour plates.

5-FoA selection media solid (1L): For solution 1 add 7gr Yeast nitrogen base, 0.77gr CSM-URA, 0.08gr Uracil, 20gr Glucose, 1gr 5-FoA to 500ml water, and mix with 65ºC heat. Filter sterilize. For solution 2 add 20gr Bacto agar to 525ml water and autoclave. Cool solution 2 to 65ºC, mix them and pour plates.

Transformation mix: Mix 240 µl PEG 3350 (50% w/v), 36 µl LiAc 1M, 50 µl ssDNA carrier (2ml/ml) and 34 µl of the DNA to transform

Yeast freezing solution: Add 1ml Tris 1M, 4ml MgSO4 1M, 15ml autoclaved water and 20ml glycerol 100%

Breaking buffer (50ml): Add TX100 2% 1ml, SDS 1% 50µl, NaCl 0.1M 1ml, TrisHCl pH8.0 10mM 500 µl and EDTA 1mM 100 µl to 47.350ml of autoclaved water

Delitto Perfetto Solution 1 (50ml): Add 5ml AcLi 1M, 5ml TE 10x to 40 ml water

Delitto Perfetto Solution 2 (10ml): Add 1ml AcLi 1M, 1ml TE 10x to 8 ml PEG 3350 50%

Yeast experimental procedures:

Yeast preservation: Yeast cells are stored by adding 0.5ml yeast culture to 0.5ml yeast freezing solution

High-efficiency yeast transformation protocol: To transform yeast cells we used the method of Gietz. A yeast colony was cultured in 5ml YPDA liquid O/N. OD600 was measured amd the suspension was diluted at 0.15ODs in 50ml YPDA. Cells were incubated at 10ºC until OD = 0.6 (for two cell divisions), and harvested by centrifugation at 2000g at room temperature for 5 minutes. Cells were washed in 25ml of sterile water, resuspended in 1ml of sterile water, cells were transferred to an eppendorf and spun down again. Sterile water was added to a final volume of 1ml. 100 µl of cell suspension are used for each transformation. Cells were spun down and 360µl of transformation mix was added and resuspended to each transformation tube. Cells were incubated in a water bath at 42ºC for 40 minutes. Cells were spun down, and plated in the selection media.

Delitto Perfetto: The Delitto Perfetto method, here used to delete genes, was performed in two steps. First, a CORE (COunterselectable REporter) cassette was integrated into the yeast genome, replacing the gene to be deleted. The CORE used in the present work contained the URA gene (Counterselectable gene) and the KanMX4 gene (Reporter). Once transformed, yeast cells were selected for the presence of KanMX4 by Kanamycin. In the second step, yeast cells were transformed by targeting oligonucleotides, which replaced the CORE cassette. To select for the insertion of the

oligonucleotides, yeast cells were grown in 5-FoA, which selected for the cells that lost the URA3 gene, and were replica-plated in Kanamycin plates, that assured the complete loss of the CORE construct. This method did not leave any track of the deletion performed.

Colony PCR: We used the NaOH method to perform colony-PCR. A yeast colony was diluted in 10µl NaOH 20mM, and heated for 10 minutes at 98ºC. 2 µl of supernatant was used to perform the PCR reaction. PCR mix: Primers (10µM) 2µl, dNTPs (10mM) 0,25µl, Mg2+ (25mM) 1,5µl, Buffer (10x) 2,5µl, Taq 0,15µl, Water 16,6µl. PCR conditions: an initial cycle of denaturation at 94ºC for 3 min followed by 35 cycles of denaturation at 94ºC, annealing at 55ºC, and extension at 72ºC for 30sec, 30 seconds and 2 min, respectively, and by a final cycle of 72ºC for 10 min

Yeast plasmid extraction: The yeast cells containing the plasmid were grown in 2ml SC-X (X is the plasmid resistant gene) O/N. 1.5ml cell suspension were transferred to an eppendorf, spun down, and resupended in 400 µl of Breaking buffer. The solution was mixed by inversion. 0,3gr of glass beads, together with 200µl of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) were added, always on ice. Suspension was vortexed at 4ºC for 2-4 minuts, and centrifuged at 4ºC for 10 minutes. Three different phases appeared at this point, and the supernatant was carefully pipetted and placed to a new in the eppendorf. This solution was purified (for instance using UltraClean 15 DNA purifiaction Kit, MoBio Laboratories) and precipitated by EtOH.

DNA precipitation: We used the EtOH method to precipitate the DNA. 2 volumes of EtOH and 10% volume of NaAc was added to 1 volume of DNA solution. Solution was mix by inversion and left O/N at −80ºC. Solution was centrifuged 10 minutes at 4ºC, the supernatant was pellet was resuspended in TE or water.

Cassette construction: The construction of genes (CGL105) was created by a sequential insertion of each gene in the pRS415 plasmid. The method used was the high-efficiency yeast transformation protocol. Each gene insertion process was as follows: the pRS415 was cut (linearized) and purified in an agarose gel. In parallel, the selection marker gene was amplified and purified also in an agarose gel. Both the linearized plasmid and the marker gene were transformed together in a yeast cell. The marker gene contained in its extremes 50bps homologous to the flanking regions of the plasmid target locus. These flanking regions were used by the yeast machinery to

repair the DSB present in the plasmid pRS415 (as shown in Figure 7) After this process, both plasmid and marker gene were perfectly matched together. After transformation, yeast cells were grown for the presence of the plasmid using the auxotrophic gene LEU2. Then, the new plasmid was extracted from the yeast cells by the Phenol:Chloroform:Isoamyl Alcohol method, introduced in bacteria by electroporation, amplified, and extracted using a yeast plasmid extraction method. After each transformation, the resulting recombinant plasmid was sequenced to ensure the correct insertion of the marker gene. Then, the next gene marker was introduced following the same steps.

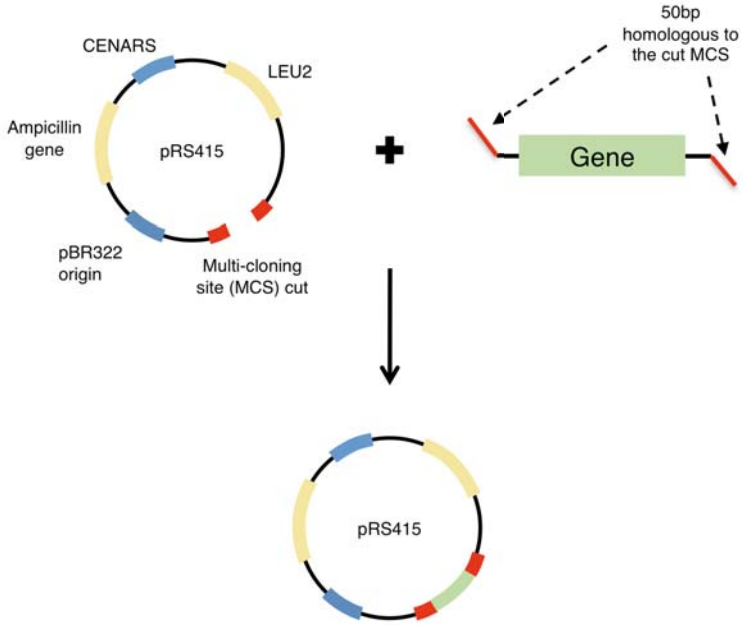


Figure 7. DNA recombination by yeast transformation.

RESULTS

Clinical and molecular phenotype characterization of NF1 patients

This part gathers the clinical and molecular characterization of NF1 patients and dNFs developed. At a clinical level, features like the number of dermal neurofibromas, the age and the sex of the patients were collected (together with information regarding the *NF1* germline mutation). At molecular level, we performed studies to determine the LOH frequencies exhibited in patient neurofibromas, together with the mechanisms generating them. In addition, a new methodological approximation was developed to facilitate the molecular characterization of a large number of neurofibromas.

PART 1

1. Clinical and molecular phenotype characterization of NF1 patients

1.1 Characterization of dermal neurofibroma number in NF1 patients

A correct clinical characterization of the NF1 patients was necessary to perform the present study. Dermal neurofibromas were obtained from 117 NF1 patients, for which clinical data regarding age, sex, the number of dermal neurofibromas developed, and the presence of other acute NF1 manifestations like plexiform neurofibromas or MPNSTs, was intended to be collected.

Clinical data considered:

- Age: Neurofibromas usually appear at puberty, and increase in number throughout the years. To compare the degree at which neurofibromas appear in NF1 patients it is important to normalize by age, as two patients with the same rate of appearance of dNFs can have a different number of neurofibromas if they significantly differ on their ages.

- Number of dermal neurofibromas: Determining the exact number of dermal neurofibromas was key in the characterization of patient phenotype, but at the same time it was complicated, as NF1 patients can develop up to thousands of them. Most of the data presented here regarding the number of dermal neurofibromas was obtained by either direct counting of tumors, or, when they were uncountable, by estimating the total number. However, during the development of this thesis a standardized method to estimate the number of neurofibromas developed by a patient was intended to be implemented. This method consisted in counting the number of dNFs present in 6 defined areas of the body (both arms, both legs, chest and back), and estimate the number of dNFs developed in the whole body. In order to validate the estimation of the number of dNFs developed, the total number of dNFs would be counted for at least 30 patients. With this data, a correlation between the total number of dNFs and the dNFs counted in the different proposed body areas could be established. This correlation could be applied to the NF1 patients to extrapolate the total number of dNFs by only counting neurofibromas in the defined areas. The

validation of the estimation method is still in progress.

- Presence of plexiform neurofibromas or MPNSTs: Information regarding these NF1 manifestations was collected because finding any correlation between the presence of these tumors and the other parameters studied would contribute to a better prognostic of the disease.

1.1.1 Characterization of NF1 patients phenotype concerning age, sex and number of dNFs developed

In order to have a clear picture on the number of dNFs developed by our group of NF1 patients, we analyzed how patients were distributed according to the number of dNFs developed and their age. In this work we analyzed dNFs from 117 NF1 patients. dNF molecular data was collected from most of these individuals, as they came to the hospital mainly to remove developed dermal neurofibromas. We are still in the process of collecting clinical data from all of them to create a database for our cohort of NF1 patients. In addition, in conjunction with the Asociación de Afectados de Neurofibromatosis, the ICO-IMPPC Neurofibromatosis group is collecting clinical data from the Spanish NF1 Community, to have a global and clear picture of the prevalence of the different traits studied.

Therefore, to obtain information regarding the distribution of NF1 patients taking into account the dNFs developed in relation to age, we collected clinical data from previous european studies, as we believe this distribution shouldn't differ a lot from the spanish NF1 population. Easton et al. [86] collected data from the United Kingdom and Southern Wales NF1 population (Figure 8a and Table annex A2), and Sabbagh et al. [88] from french NF1 population (Figure 8b and Table annex A3). Both NF1 populations show similar distributions, with a clear increase in the number of dNFs as the age of NF1 patients was also increasing. Note that both studies differ in the higher range of dNF counting (+500 for Easton et al. study and >100 for Sabbagh et al. study).

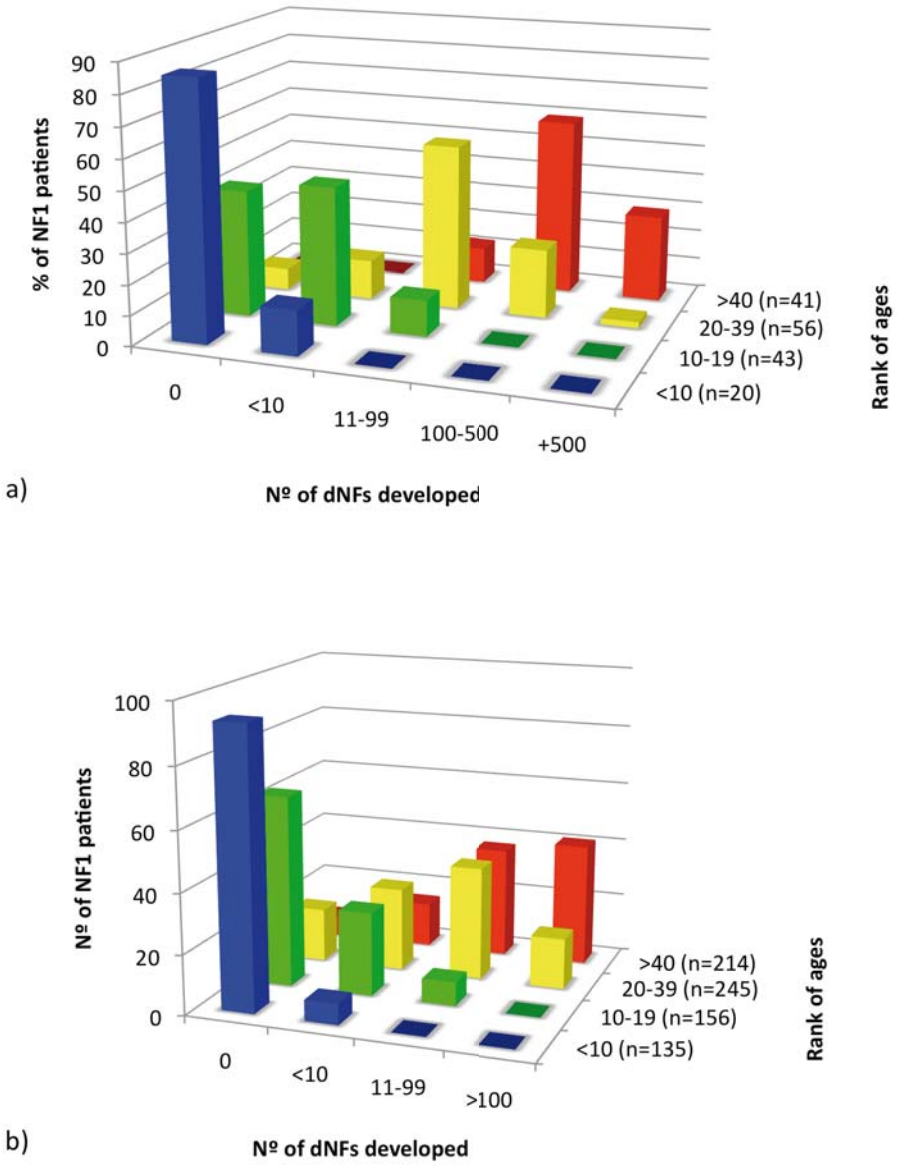


Figure 8. Distribution of dNFs developed in relation to the age of the NF1 patients for UK and Southern Sales (a) and French (b) NF1 populations. The number of patients contained in each age rank is indicated in brackets.

To obtain a clearer picture on the relation between the nº of dNFs developed and the age of the patients, we plotted the increase in the number of dNFs considering the different ages of the NF1 patients (Figure 9). We attributed different values to each rank of dNFs developed (see legend in Figure 9), and obtained qualitative values representing the global number of dNFs developed by all the NF1 patients for each

specific rank of ages. This analysis shows clearly the positive correlation between both variables, the nº of dNFs developed and the age of NF1 patients.

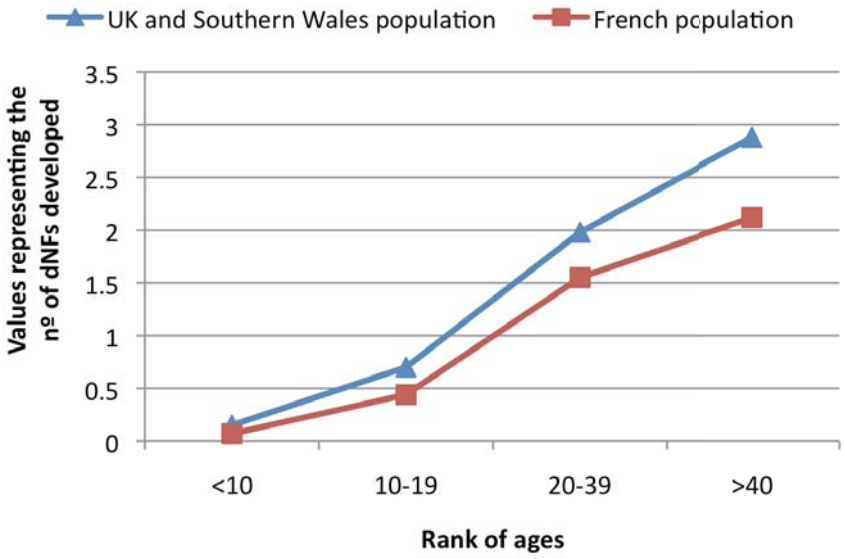


Figure 9. Obtained values representing the nº of dNFs developed for each rank of ages. These values were obtained for each rank of ages by, first, multiplying each percentage of patients bearing 0, 10, 11-99 or >100 dNFs by an arbitrary number (values used for 0 dNFs = 0, >10 dNFs = 1, 11-99 dNFs = 2, >100 = 3), and then, by summing them.

1.1.2 Normalization of the number of dNFs developed by age

Due to the correlation between the number of dermal neurofibromas developed and the age of NF1 patients, to study any relationship between the dNFs developed and data regarding the frequencies of LOH or HR studied in our cohort, it was necessary to correct the number of dNFs presented by the age of the patient. To do so we performed a custom age rank-based correction to obtain a Risk of Dermal Tumor Burden (RDTB) value. RDTB values should be representative of the degree of dNFs developed (independently of patient age) for a given patient. To obtain this value we used data from the study performed by Huson et al. (30, 187). Huson et al. collected data from the Southern Wales population, which was later added to the UK population by Easton et al. We considered the original data from Huson et al. to be more adequate for our study, as the ranks of ages considered were in agreement with

the ones obtained for our NF1 patients, with increases of 10 years, from 10 to 80 years old (in contrast to Easton et al. and Sabbagh et al., which ranked ages with increases of 10 years until 40 years old). Figure 10 shows the distribution of NF1 patients from Huson et al. data. The RDTB value for each NF1 patient was calculated using the collected NF1 general data (represented in Figure 10 and Table annex A4) as follows: first, the patient was positioned in the corresponding column of the NF1 general population regarding its nº of dNFs developed and the age. The median of the percentage of patients for this column, M, was obtained. Then, all the percentages of patients bearing lower nº of dNFs contained in the same rank of ages were summed, obtaining L. Finally, RDTB = M + L (an example of RDTB calculation is shown below).

Using this correction, for a rank of ages, a patient bearing a higher number of dNFs would have a higher RDTB value compared to a patient bearing a lower dNF number, for the same rank of ages.

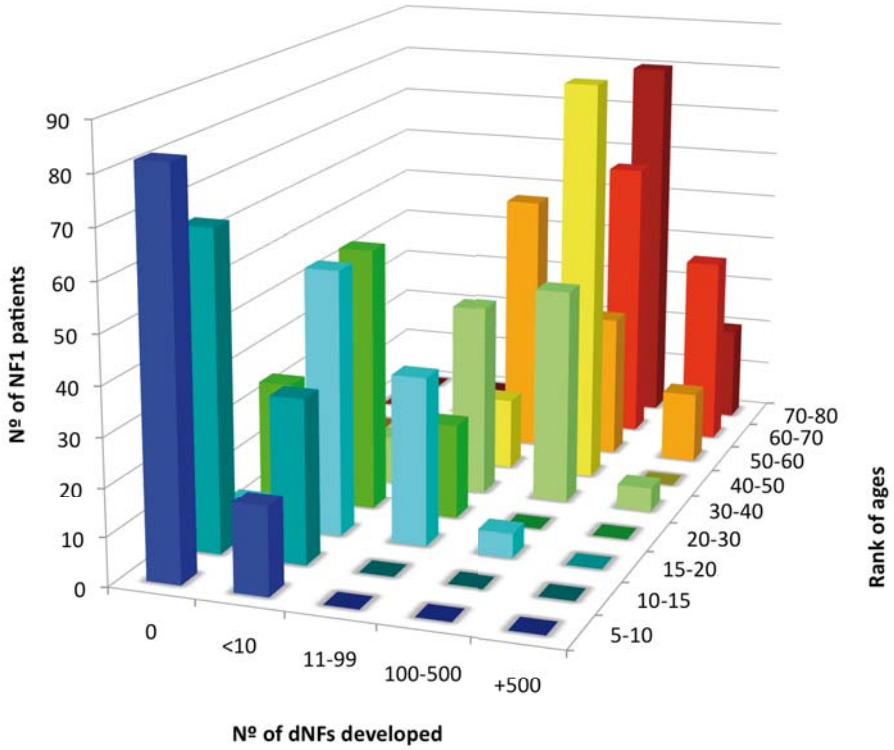


Figure 10. Distribution of dNFs developed in relation to the age of NF1 patients for Southern Wales NF1 population.

Figure 11 shows an example of RDTB calculation for patient P047, which developed 100-500 dNFs at the age of 47. In this case, all the calculations were performed using the columns representing the percentages of patients within the rank of 40-50 years old.

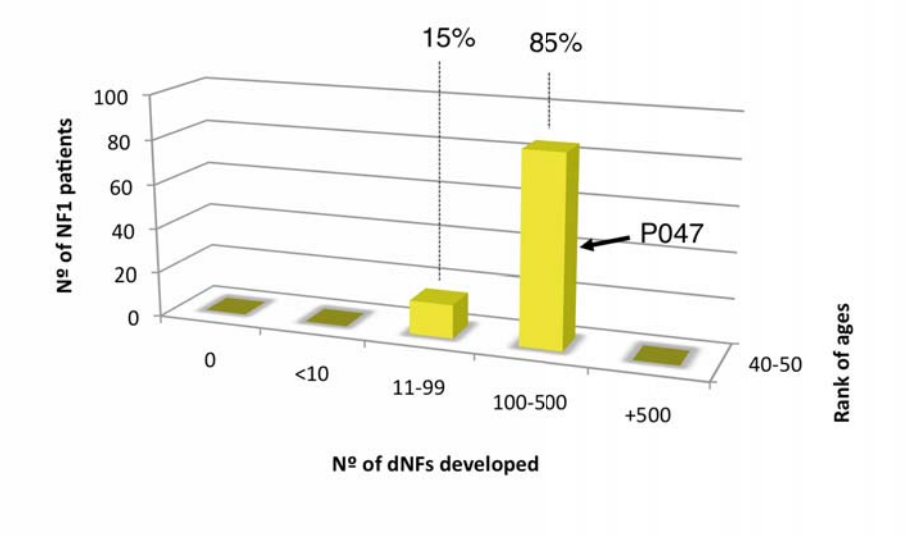


Figure 11. Example of dNF number normalization for patient P047. Patient P047 has developed 100-500 dNFs at the age of 47. In the general NF1 population, 15% of patients have developed 11-99 dNFs, while 85% have developed 100-500 dNFs for the rank of 40-50 years old. P047 is positioned in the 85% of patients developing 100-500 dNFs. The median of this percentage, M, is calculated: M = 0.85/2 = 0.425. Then, the percentages of patients showing lower nº of dNFs developed, L, are also calculated: L = 0.15. RDTB = M + L = 0.425 + 0.15 = 0.575.

1.2 Characterization of tumor molecular phenotype in NF1 patients

The molecular phenotypic characterization of dermal neurofibromas was based on estimating, for each NF1 patient, the LOH frequencies (in this work prevalence of LOH refers to the general percentage of LOH obtained in the NF1 population, while frequency of LOH refers to the individual percentages of LOH obtained for each NF1 patient) as an evidence of second hits implicated in the generation of dNFs, together with the identification of the mechanisms generating these LOHs. To do so, we determined, for every dNF analyzed for all NF1 patients, the presence/absence of LOH, its extension, and the locus copy-number of the region analyzed.

We developed a microsatellite multiplex PCR (MMP) to analyze the whole 17th chromosome (the *NF1* gene is located at 17q11.2) for the presence/absence of LOH by amplifying 16 microsatellites in one PCR reaction (15 located at Chr. 17 and 1 located at Chr. 2) (Figure 12).

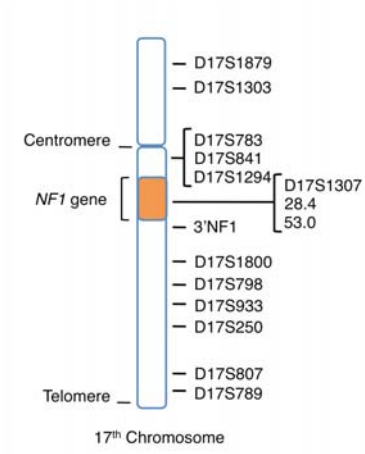


Figure 12. Location of MMP microsatellites

The determination of LOH was based on the Q^{LOH} expression (see M&M section). We analyzed a total of 606 dNFs from 117 NF1 patients (Table annex A5). We found LOH in 144 dNFs, which accounts for the 23.7% of dermal neurofibromas (Table 2). After the LOH analysis, we determined the *NF1* locus copy-number by either Multiplex Ligation-dependent Probe Amplification (MLPA) or Paralog Ratio Analysis (PRA) techniques. Knowing the LOH extension and the locus copy-number of the regions analyzed let us infer the mutation mechanism responsible for each LOH. Moreover, we analyzed 19 dNFs with the SNP-array technique to obtain

information of LOH extent, *NF1* copy-number, mechanism generating the LOH, percentage of cells bearing the LOH, and to study the presence of alterations in other genomic regions.

Nº of patients	117
dNFs analyzed	606
dNFs with LOH	144 (53 deletions / 91 HRs)
% LOH tumors / Total tumors	23.7
% LOH-deletion / Total tumors	8.7
% LOH-recombination / Total tumors	15
% LOH-deletion / LOH-tumors	37
% LOH-recombination / LOH-tumors	63

Table 2. Global view of LOH in neurofibromas

1.2.1 Tumor heterogeneity and locus copy-number determination

dNFs have a particular mixed cellular composition that can hamper the *NF1* locus copy-number analysis: these tumors are composed by an admixture of cells, but only a group of Schwann cells harbor the somatic *NF1* mutation. Hence, high percentages of non-LOH cells within neurofibromas can really hamper the detection of copy-number alterations by MLPA or PRA techniques. Q^{LOH} value is directly related to the percentage of cells not bearing the second hit. Therefore, we performed an assay to, first, study the relation between Q^{LOH} and the percentage of non-LOH cells, and then use it to avoid possible false negatives regarding the *NF1* copy-number detection by MLPA and PRA techniques: we performed serial admixtures of two DNAs, one obtained from a patient carrying a constitutive deletion in the *NF1* gene (*NF1*+/-), and the other from its progenitor that did not transmitted the deletion (*NF1*+/+). These samples allowed us to reproduce hypothetical cellular composition of neurofibromas showing LOH due to the presence of a deletion in the *NF1* gene and carrying different proportions of non-LOH normal cells. We created 11 samples with NF1+/+ cells ranging from 0% (100% of *NF1*+/- cells) up to 100% (0% of *NF1*+/- cells) in intervals incrementing non-LOH *NF1*+/+ cells in 10%. With the same DNA admixtures we performed MMP, MLPA and PRA techniques in parallel, using conditions as described in Material and Methods section (Figure 13). All reactions

were performed in triplicate. We found a good correlation between the Q^{LOH} value and the percentage of non-LOH normal cells within a context of cells bearing a deletion in the *NF1* gene (Figure 13a). By comparing the same admixtures of DNA using MLPA and PRA we were able to establish the Q^{LOH} values that marked the limit of detection for MLPA and PRA techniques. In the case of MLPA, we found a Q^{LOH} upper limit value of 0.58 (Figure 13b). In the case of PRA, the Q^{LOH} upper limit was 0.76 (Figure 13c).

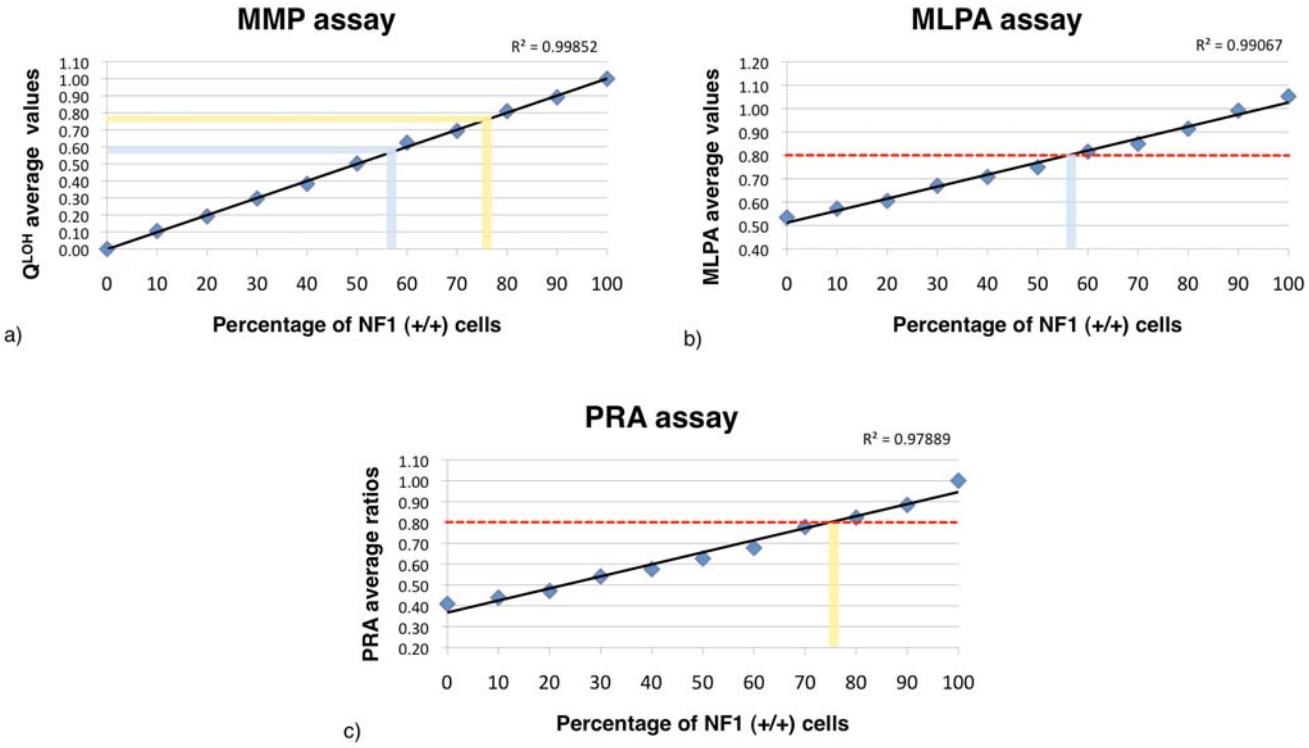


Figure 13. MMP, MLPA and PRA copy-loss limit of detection comparison. Plotted in the X axes there are represented the different percentages of NF1+/+ cells. Plotted in the Y axes there are represented: a) for MMP analysis, the average of Q^{LOH} values obtained from at least 2 informative microsatellites located within the NF1 gene; b) for MLPA technique, the average normalized values obtained from NF1 probes; c) for the PRA analysis, the average normalized ratios obtained from the analysis of NF1 gene/pseudogene. Dashed red line indicates the threshold value for considering the presence of an NF1 deletion in the case of both techniques. Values below the red line (a normalized ratio of 0.8) indicate a copy loss in the NF1 gene. Blue and yellow solid lines indicate the relationship between the limit of detection of an NF1 deletion by MLPA and PRA techniques (respectively) and the Q^{LOH} value obtained in the MMP analysis. R^2 (Coefficient of determination).

1.2.2 Mutational mechanisms leading to LOH in dNFs

After performing LOH and *NF1* copy-number analysis to all dNFs analyzed (606), we found that ~37% of LOHs were generated by the mechanism of deletion, and that ~63% were due to homologous recombination (Table 2). In 15 dNFs showing LOH we weren't able to determine the copy-number status of the *NF1* region due to a lack of enough DNA material (Table annex A5). For these samples we considered that, as determined for the rest of the tumors (see below), in dNFs showing LOH for the entire 17q arm, LOH was caused by HR, and that for the dNFs with LOH localized in the *NF1* and regions surrounding it, loss of heterozygosity was caused by a deletion.

1.2.2.1 LOH affecting copy-number

We found LOH with copy-loss in 53 dNFs (~37% of LOHs) (Table 2, Table annex A5 and see example in Figure 14). In all cases LOH was localized to the *NF1* gene and the regions surrounding it, and one copy of the regions analyzed was present. As shown in Figure 15a, the deletions obtained spanned from ~80Kbs to ~8Mbs. The large distance between microsatellites made it impossible to accurately determine the breakpoints for these deletions. However, for 10 dNFs showing copy-loss, it was possible to narrow down the breakpoints by SNP-array analysis, and non of the LOHs involved more than 8Mbs (Figure 15b). Interstitial deletions are presumed to be caused by double-strand breaks (DSBs) and subsequent illegitimate joining of DNA ends mainly by homologous or nonhomologous end joining repair (138, 188). The structure of the breakpoints can help identifying the repair mechanisms generating the alteration, but characterizing breakpoints and understanding the exact mechanisms generating deletions was beyond the scope of this work.

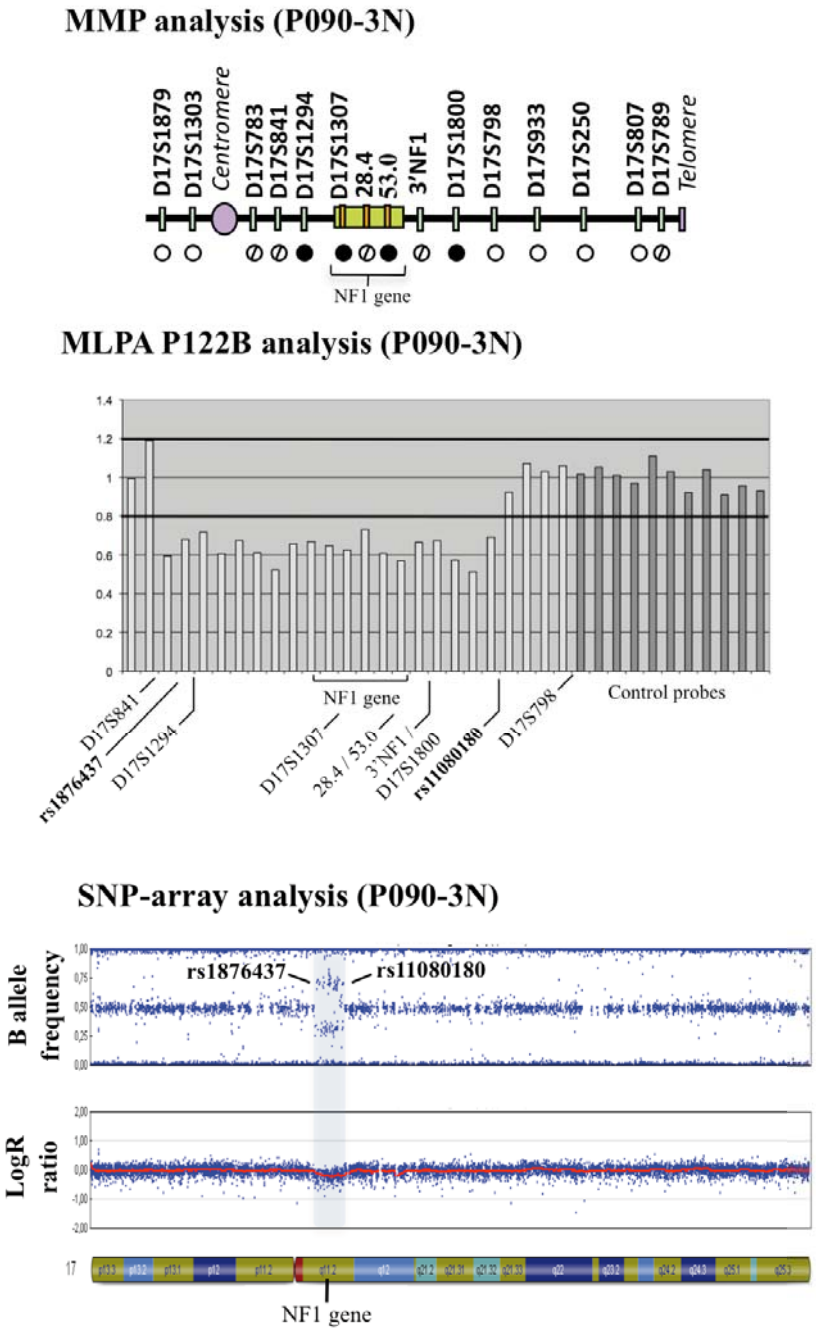


Figure 14. Mechanisms generating deletions. In all cases Microsatellite Multiplex PCR (MMP) evidenced LOH in the NF1 gene and normally also in 5' and 3' regions adjacent to it. MLPA detected only one copy of the NF1 gene. SNP-array analysis detected LOH (four-band pattern in the B allele frequency plot) and one copy of the NF1 gene (LogR ratio <0). MMP (circle = no LOH; dashed circle = noninformative, black circle = LOH). MLPA (Values between 0.8 and 1.2 = two copies. Values <0.8 = one copy). SNP-array: B allele frequency plot (0.5 = heterozygote; 0 or 1 = homozygote, other intermediate values = allelic imbalance); LogR ratio (0 = two copies; values <0 = one copy).

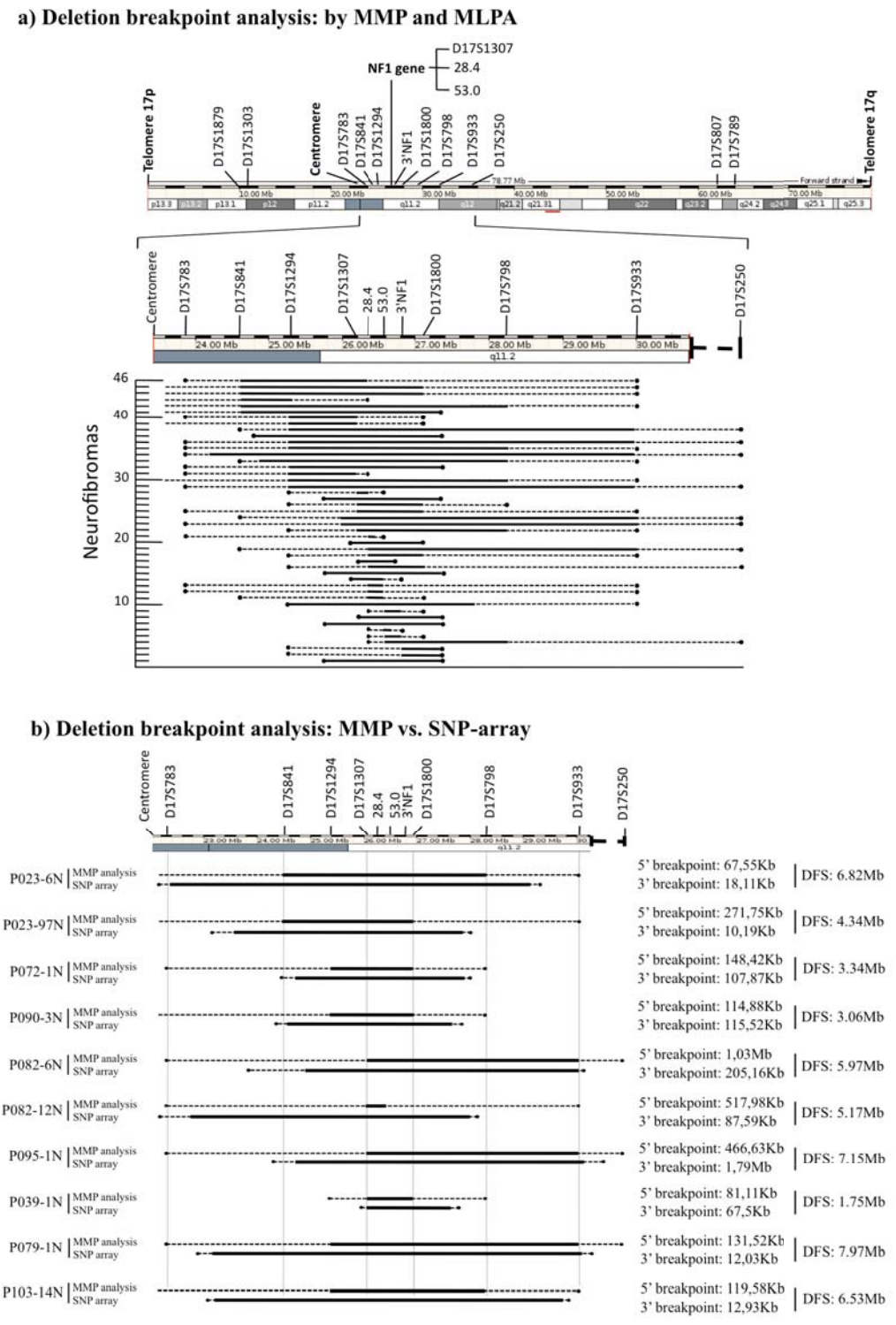


Figure 15. Deletion breakpoint analysis. a) Breakpoint analysis by Multiplex Microsatellite PCR (MMP) and by MLPA analysis. b) Deletion breakpoint mapping refinement by SNP-array. Solid black bar: deletion mapped by either MMP (A, B), by MLPA (A), or SNP-array (B); dashed line: uncertain region by either MMP (A, B) or SNP-array (B). Black circle: no deletion by either MMP (A, B), by MLPA (A) or SNP-array (B). DFS, Deletion fragment size.

In 13 dNFs we detected LOH affecting only the *NF1* gene and flanking regions (Figure 16a) suggesting the presence of an interstitial deletion, but when applying MLPA the same tumors exhibited two copies of the *NF1* gene (Figure 16b).

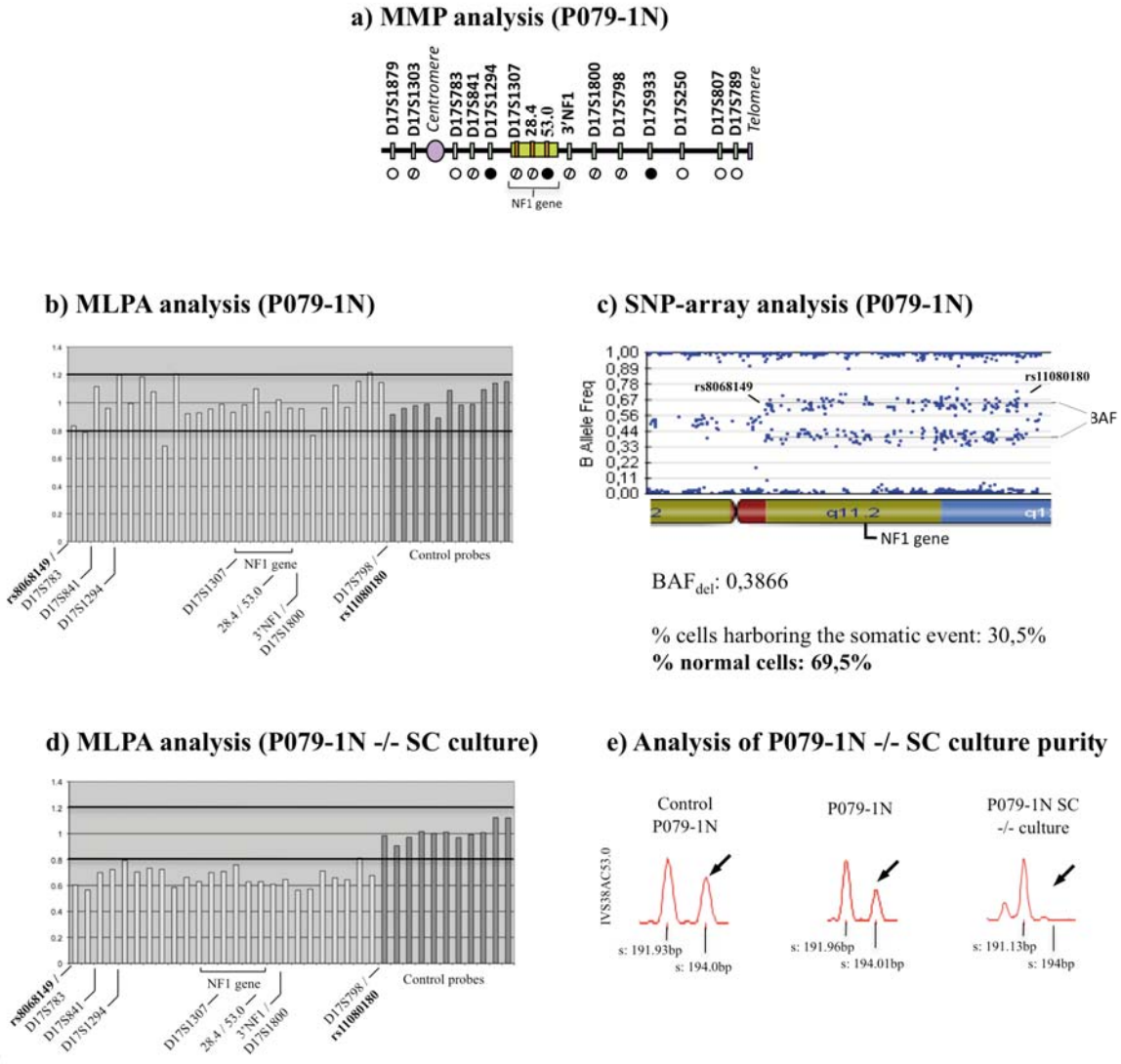
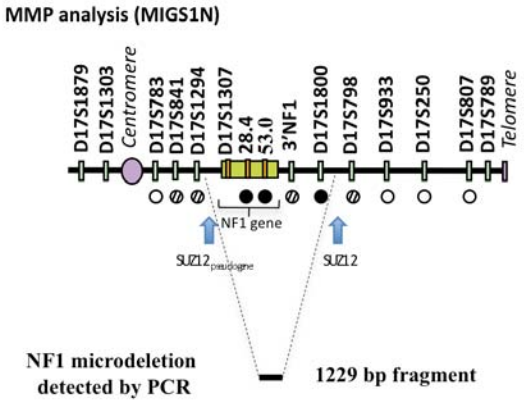


Figure 16. The presence of normal cells in neurofibromas affects the NF1-copy number detection by MLPA. a) MMP analysis of tumor P079-1N (see legend of Figure 14 for nomenclature); b) MLPA of P079-1N detected two copies of the NF1 gene; c) SNP-array analysis evidenced a high proportion (~70%) of normal cells within tumor P079-1N; d) MLPA of P079-1N NF1(-/-) Schwann cell culture detected one copy of the NF1 gene. e) Purity of the culture was evidenced by MMP analysis, comparing control, tumor, and SC culture. A total loss of one allele was detected in the SC culture. MLPA (values between 0.8 and 1.2 = two copies. Values <0.8 = one copy).

These results could indicate a new mechanism generating LOH in dNFs, a homologous recombination with two crosssovers, one at 5' and the other at 3' of the *NF1* gene. However, SNP-array analyses in some of them showed high percentages of non-LOH cells in the tumors (Figure 16c), an indication that was also supported by the high Q^{LOH} values obtained (e.g., Q^{LOH} = 0.62 for P079-1N Table annex A5). MLPA cannot detect deletions in samples with a Q^{LOH} higher than 0.58 (Figure 13). Accordingly, we decided to perform additional analyses. For two neurofibromas we were able to culture pure populations of dNF derived *NF1*(-/-) Schwann cells to determine if the presence of non-LOH cells was masking the copy-loss detection (Figure 16e): in both cases (P079-1N and P062-4N) the re-analysis of these cultured cells by MLPA showed a copy-loss in the LOH region (Figure 16d). Thus, LOH in these tumors was indeed caused by deletions. In 7 of the remaining 11 dNFs copy-loss was assessed by either PRA or SNP-array techniques. In the remaining 4 neurofibromas we were not able to determine the *NF1* copy-number due to the lack of DNA material (Table annex A5).

In two dNFs analyzed (P004-7N, P082-1N), the deletion breakpoints were located close the low-copy repeats *NF1*-REPs and SUZ12, flanking the *NF1* gene region. In both patients the *NF1* germline mutation was already identified. We investigated, in collaboration with the laboratory of Dr. Kehrer-Sawatzki (University of Ulm, Germany), whether these deletions were generated by nonallelic homologous recombination (NAHR) between these low-copy repeats. We performed PCR reactions to detect the breakpoints of previously identified *NF1* microdeletions in patients with constitutional type-1 or mosaic type-2 deletions (142, 186), and identified one dNF, P082-1N, that carried a type-2 deletion (Figure 17). This type of *NF1* deletion is caused by NAHR between the SUZ12 gene and its pseudogene, which are located adjacent to the *NF1*-REPs. First-hit type-2 *NF1* microdeletions are identified in mosaic NF1 patients, and thus they are assumed to occur during early fetal development [62]. This is the first reported dermal neurofibroma with a somatic type-2 deletion causing LOH.



Primer pair	hks4033jtü8r
PCR product designation	br4
Size of PCR product (bp)	1229
Position on Chr. 17	26108479-26108504 27314021-27314046
Type of NF1 microdeletion detected by PCR (breakpoint position)	type-2 (SUZ12 intron 4)

Figure 17. Nonallelic homologous recombination causing deletion. MMP detected LOH involving the NF1 gene and adjacent regions, apparently not going further 5' or 3' of the NF1-REPs. A table summarizing the PCR conditions used for detecting the deletion breakpoint is depicted. Breakpoint localized in SUZ12 intron 4.

1.2.2.2 Copy-neutral LOH

We identified copy-neutral LOH in 91 dNFs (~63% of LOH) (Table 2, Table annex A5 and see example in Figure 18). The LOH began somewhere between the centromere and the *NF1* gene, and extended until the end of the 17q telomere. In all cases, two copies of the 17q arms were present in the tumors, as well as the two copies of the *NF1* gene carrying the constitutional mutation [141]. The mechanism generating LOH in these tumors was homologous recombination (HR), with a unique crossover between the centromere and the *NF1* gene, generating isodisomy from the crossover region up to the 17q telomere (as evidenced by SNP-array analysis) (Figure 18).

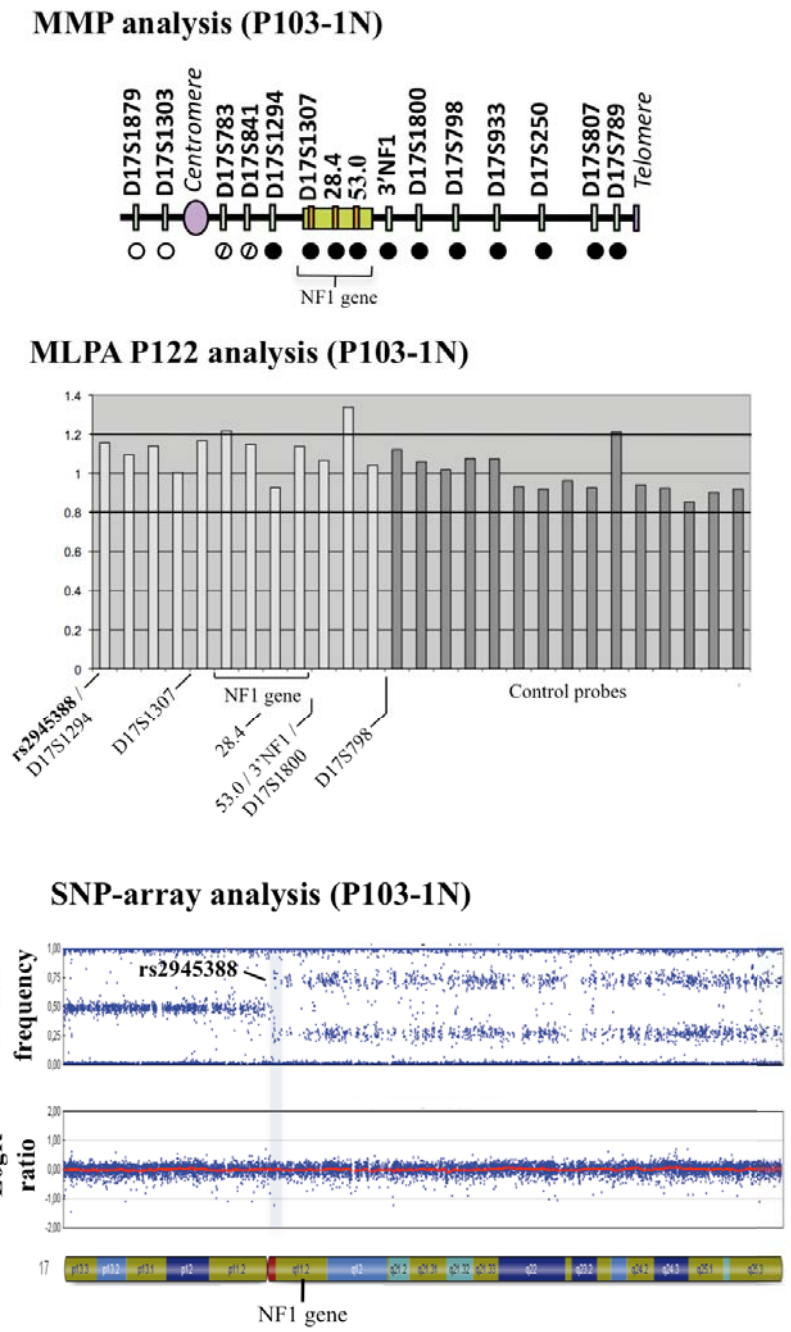


Figure 18. Mechanism of homologous recombination. MMP evidenced LOH of almost all 17q chromosome. MLPA detected two copies of the NF1 gene and the entire region analyzed. SNP-array detected LOH (four-band pattern in the B allele frequency plot) from the centromere to the end of the chromosome (17q telomere) and the presence of two copies (LogR ratio = 0) of the entire chromosome 17q. MMP (circle = no LOH; dashed circles = noninformative, black circles = LOH). MLPA (Values between 0.8 and 1.2 = two copies. Values <0.8 = one copy). SNP-array: B allele frequency plot (0.5 = heterozygote; 0 or 1 = homozygote, other intermediate values = allelic imbalance); LogR ratio (0 = two copies; values <0 = one copy).

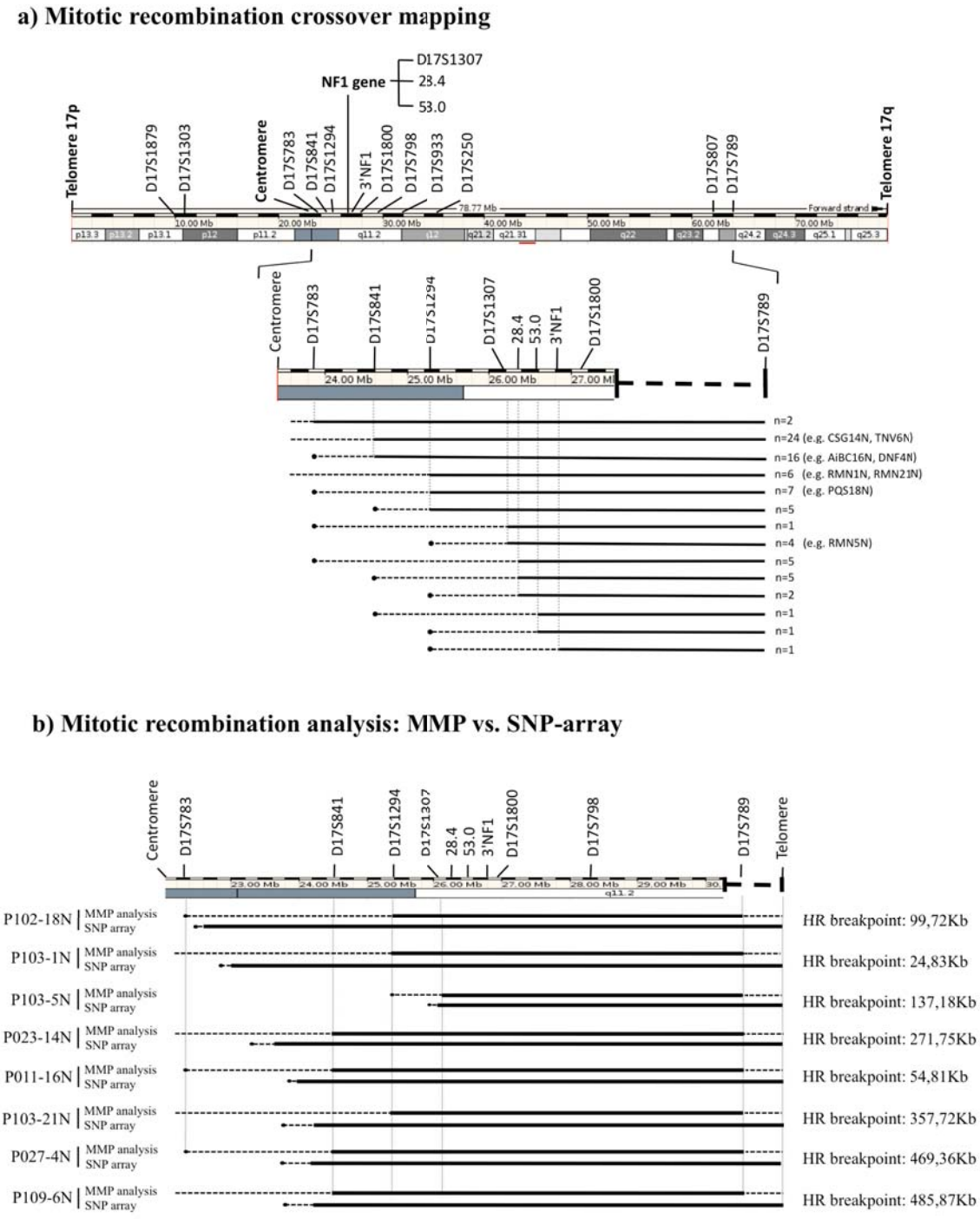


Figure 19. Mitotic recombination crossover mapping. a) Crossover analysis by Multiplex Microsatellite PCR (MMP). b) Mapping refinement by SNP-array compared to MMP. Solid black bar: uniparental isodisomy (B); dashed line: uncertain region; black circle: no uniparental isodisomy. HR breakpoint: distance between mapped informative SNPs.

Again, microsatellite analysis did not facilitate a precise location of recombination breakpoints (Figure 19a). Nevertheless, in eight dNFs we were able to locate the crossover region at a higher resolution using SNP-array analysis (Figure 19b). We found that in 2 dNFs the crossover was centromeric to the D17S783 marker, clearly located closer than 1Mb from the centromere. However, in addition, there were 29 dNFs homozygous (non-informative) for marker D17S783, for which the crossover could be also located near the centromere. For tumors P102-18N and P103-1N this was the case (Figure 19b), as shown by SNP-array analysis. For five dNFs there were no informative 17p microsatellite markers (three from the same patient), for which we were not able to rule out the possibility of a loss of the complete 17th chromosome with an endoreduplication of the remaining one. A significant number of dNFs showed a crossover located closer than 1Mb from the *NF1* gene; in eight the breakpoint was located between marker D17S1294 and the *NF1* gene, and SNP-array showed that in tumor P103-5N the crossover was located in a region 260-400Kbs close to the *NF1* transcription start site. However, in many dNFs the crossover seemed to occur in a region equidistant from the centromere and the *NF1* gene (Figure 19b).

1.2.2.3 Other genomic alterations

19 dNFs were analyzed by SNP-array to obtain more resolution in the LOH analysis and to search for the presence of other genomic alterations. 17 dNFs exhibited no copy number changes or LOH elsewhere in the genome except for the 17q alteration involving the *NF1* gene. We identified two dNFs containing large deletions: one located in chromosome 2q24.2-q31.1 (P090-3N) of about ~15.3Mb and the other in chromosome 3q11.2-q22.1 (P072-1N) of about ~39.8Mb (Figure 20). None of these deletions were present in the matching control DNAs, evidencing somatic events. In tumor P090-3N chromosome 2 and 17q-*NF1* deletions were present in the same proportion of cells (~49%), suggesting that both alterations affected the same cells. In contrast, in tumor P072-1N the 17q-*NF1* deletion was present in ~49% of the cells while the chromosome 3q deletion was only present in ~37% of the cells. This finding could reflect that either the 3q deletion occurred later during tumor

development, or that both deletions together do not confer a proliferation advantage to the cell.

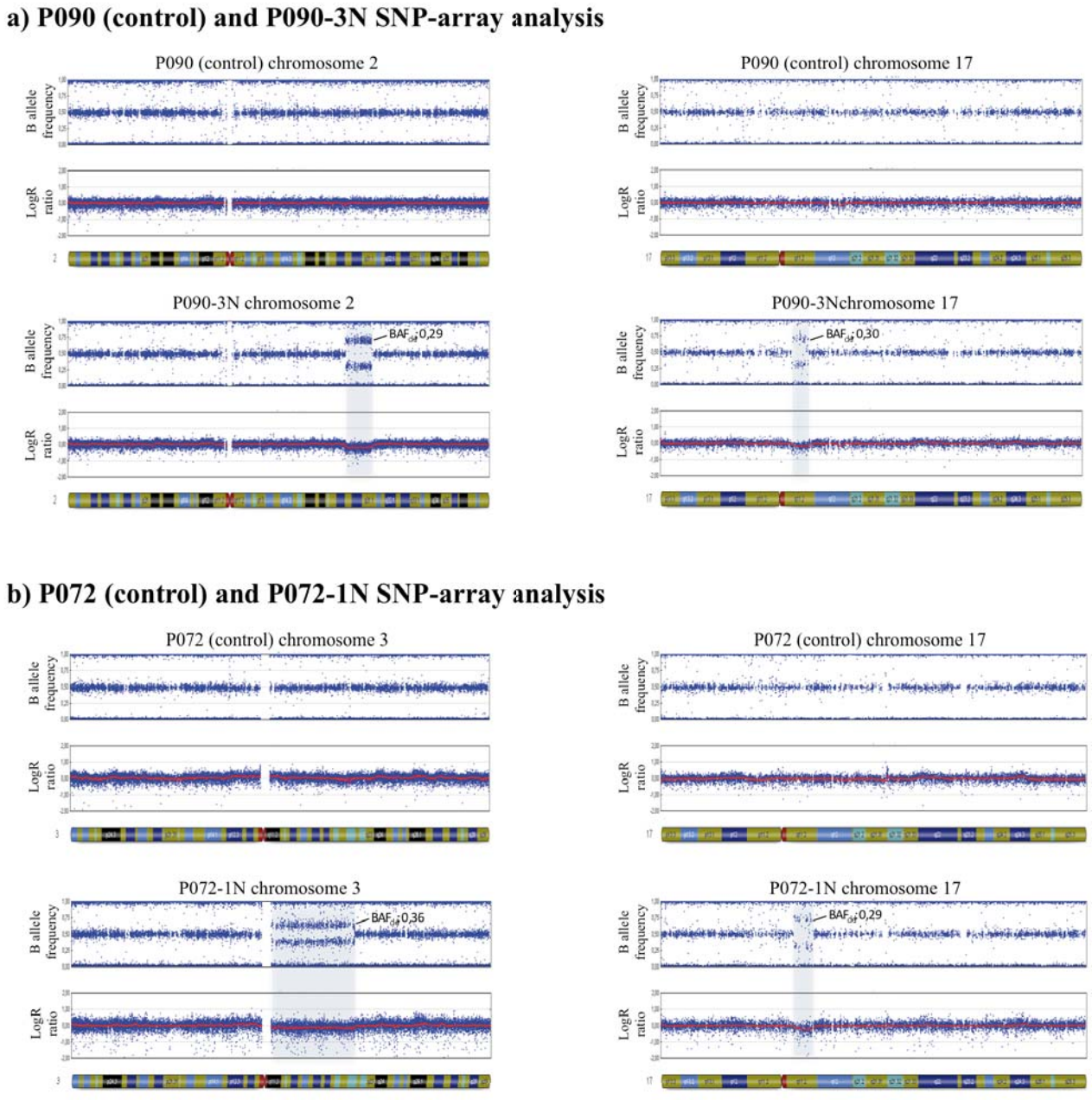


Figure 20. SNP-array analysis of samples P090-3N and P072-1N compared to their respective controls. a) Detection of deletions in 2q and the NF1 region in tumor P090-3N; b) Detection of deletions in 3q and the NF1 region in tumor P072-1N. Deletions were evidenced by a four-band pattern in the B allele frequency plot and a LogR ratio<0.

1.3 LOH frequencies in dNFs from different NF1 patients

One of the hypotheses of this thesis is that variation in the number of dermal neurofibromas developed by NF1 patients is influenced by the variation in genes related to the somatic mutation rate of patients. We focused our studies on those modifying the *in vivo* homologous recombination rate. We characterized NF1 patients for different phenotypic traits and variables like the number of dNFs developed, age, sex, and also for the molecular mechanisms generating second hits in their neurofibromas, like HR and deletion. Then, we studied clinical and molecular data to find possible correlations.

1.3.1 Determination of LOH frequencies in dNFs from NF1 patients

In the present work we collected and analyzed dNFs from 117 NF1 patients. In order to have a correct characterization of the individual LOH frequencies in dNFs of each patient, we calculated the number of dNFs needed to reliably determine these frequencies. Taking data published by different groups in the literature (see, i.e., [138] we assumed a general LOH prevalence in neurofibromas from NF1 patients of 20%. Because there is no information about the variation among NF1 patients regarding this prevalence, we assumed a constant prevalence of 20% in order to assess the minimal number of dNFs to be tested for each NF1 patient to estimate the individual LOH frequencies. As sampling fraction is large, the binomial approximation is no longer appropriate, and therefore we used a finite population correction term to adjust the variance of the normal distribution to the more closely model of the hypergeometric density function [142]. Figure 21 present different scenarios of LOH prevalence and nº of dNFs analyzed. Considering a LOH prevalence of 20%, we needed to analyze at least 10 dNFs to estimate the hypothesized frequencies with a precision of 20% (Figures 21b and 21d). Increasing the number of neurofibromas analyzed over 10 would improve precision, but it should be considered the feasibility of analyzing a large number of neurofibromas from a single patient, both for the patient and for the research team. In our study from 117 NF1 patients, we were able to analyze at least 10 dNFs in 17 of them (Table 3).

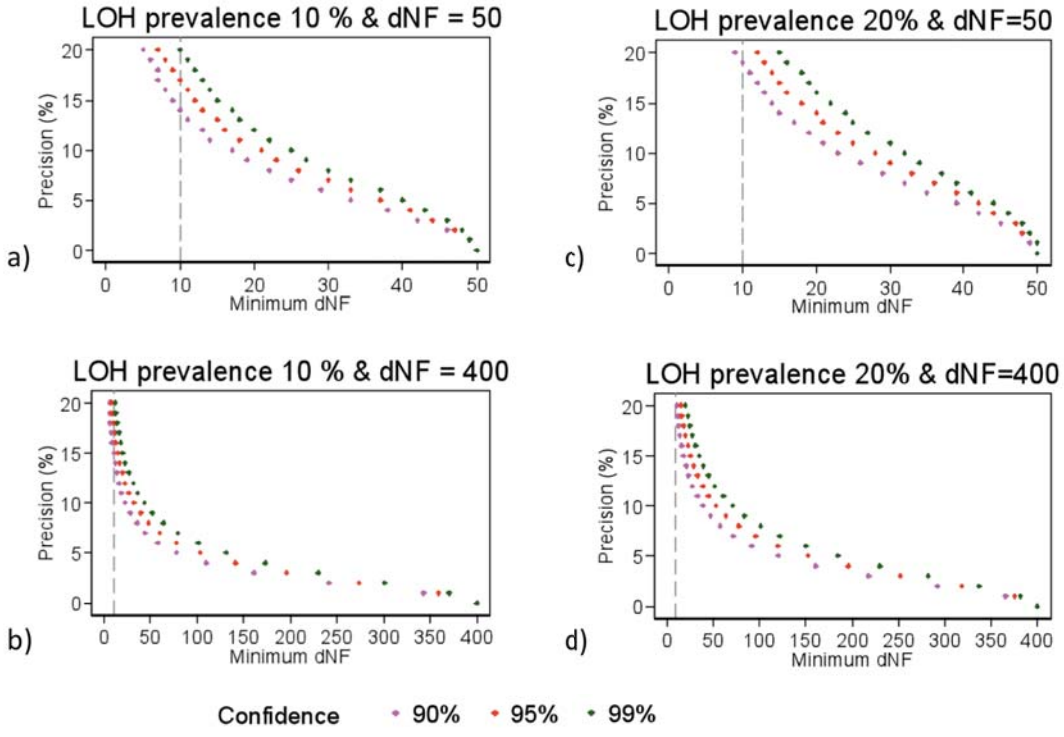


Figure 21. Precision expected in relation to the number of neurofibromas analyzed per patient and the confidence level, under different scenarios of prevalence of dermal neurofibroma (dNF)-LOH and number of dNFs developed by a patient. Dashed line indicates the precision when analyzing 10 dNFs per patient.

NF1 patient	Sex	Age	Nº of dNFs developed	Nº of dNFs analyzed
P011	Female	36	10-50	37
P020	Female	42	10-50	13
P021	Female	54	100-500	14
P022	Male	80	>500	20
P023	Male	58	>500	87
P024	Female	31	50-100	11
P027	Male	26	100-500	14
P047	Female	47	100-500	14
P052	Male	59	50-100	11
P055	Male	51	100-500	11
P062	Female	40	100-500	16
P078	Male	55	100-500	48
P084	Female	63	100-500	14
P102	Female	56	>500	28
P103	Female	51	>500	24
P104	Female	38	100-500	11
P114	Female	41	50-100	18

Table 3. Clinical data from NF1 patients with at least 10 dNFs analyzed

1.3.2 Variation in neurofibroma-LOH frequencies in NF1 patients

We collected data from 17 NF1 patients with at least 10 dNFs analyzed (Table 4). We also added to Table 4 patient P082, with 9 dNFs analyzed (see below).

Patient	RDTB	N° of dNFs analyzed	N° of dNFs with LOH	% of LOH	% of LOH - Deletion	% of LOH - HR	Germline mutation	
							DNA level	Protein level
P020	0.0375	13	2	15.3	0	15.3	c.2041C>T	p.Arg681X
P082	0.0375	9	5	55.5	55.5	0	c.910C>T	[p.Arg304X; p.Lys297_Lys354del]
P114	0.1125	18	4	22.2	0	22.2	c.1466A>G	p.Tyr489CysX
P011	0.2	37	11	29.7	13.51	16.21	c.3826C>T	p.Arg1276X
P084	0.3	14	3	21.4	0	21.4	c.4572C>G	p.Tyr1524X
P024	0.4	11	1	9.1	0	9.1	c.3525_3526delAA	p.Arg1176fsX18
P052	0.4125	11	1	9.1	9.1	0	c.6791_6792dupA	[p.Tyr2264X; p.Ala2253_Lys2286del]
P047	0.575	14	4	28.5	0	28.5	-	-
P062	0.575	16	6	37.5	12.5	25	c.910C>T	[p.Arg304X; p.Lys297_Lys354del]
P021	0.7	14	0	0	0	0	c.6792C>A	p.Tyr2264X
P055	0.7	11	1	9.1	0	9.1	c.5710G>T	p.Glu1904X
P078	0.7	48	6	12.5	0	12.5	c.4537C>T	p.Arg1513X
P104	0.725	11	3	27.3	0	27.3	c.4308G>T	p.Glu1436X
P022	0.9	20	9	45	10	35	c.1756_1759delACTA	p.Thr586fsX18
P023	0.925	87	14	16	5.7	10.3	c.3525_3526delAA	p.Arg1176fsX18
P102	0.925	28	5	17.8	7.1	10.7	c.5242C>T	p.Arg1748X
P103	0.925	24	11	45.8	16.6	29.1	c.2338A>C	p.Thr780X
P027	1	14	1	7.1	0	7.1	c.6226delG	p.Ala2076fsX14

Table 4. Neurofibroma-LOH frequencies of NF1 patients. RDTB: Risk of Dermal Tumor Burden. HR: Homologous recombination.

The frequencies of LOH in dNFs in the group of 18 NF1 patients ranges from less than 10% to more than 50%, (Table 4). There is even one patient, P021, with no LOH in the dNFs analyzed. Most patients had dNFs with LOH caused by either deletions or homologous recombination leading to copy-neutral LOH. However, there were also patients with a clear dominant LOH mechanism. For instance, dNFs of patient P082 exhibited mainly LOH caused by deletions, whereas dNFs of patients P022 and P078 had LOH primarily due to homologous recombination (Table 4), regardless of LOH frequencies in dNFs. There are two pairs of patients bearing the same germline mutation. It is remarkable that patients P082 and P062, bearing the same germline mutation (c.910C4T), exhibited quite different frequencies of LOH in dNFs and a different preference of LOH-generating mechanism. This would suggest that the germline mutation might not play an important role in relation to the mechanisms generating LOH (this cannot be applied to patients P023 and P024, as they are closely related).

1.3.3 Studying the distribution of the dNF number in relation to age in our cohort of NF1 patients

We studied the distribution of dNFs developed in relation to age in the group of 18 NF1 patients for which clinical and molecular phenotypic data was collected (Figure 22). The distribution obtained is similar to those obtained for UK and Southern Wales or French populations, with older patients bearing a higher number of dNFs (see Figure 8 in page 66). However, our population shows some particularities, mainly due to the low number of NF1 patients analyzed per rank of ages. For instance, in the rank of 30-40 years old patients, there are more individuals bearing 100-500 dNFs than in the rank of 40-50 years old patients, as only 4 NF1 patients have been analyzed per rank. In addition, 50-60 years old patients show the highest percentage of individuals with the highest number of dNFs, as the number of patients analyzed for this age rank is the highest.

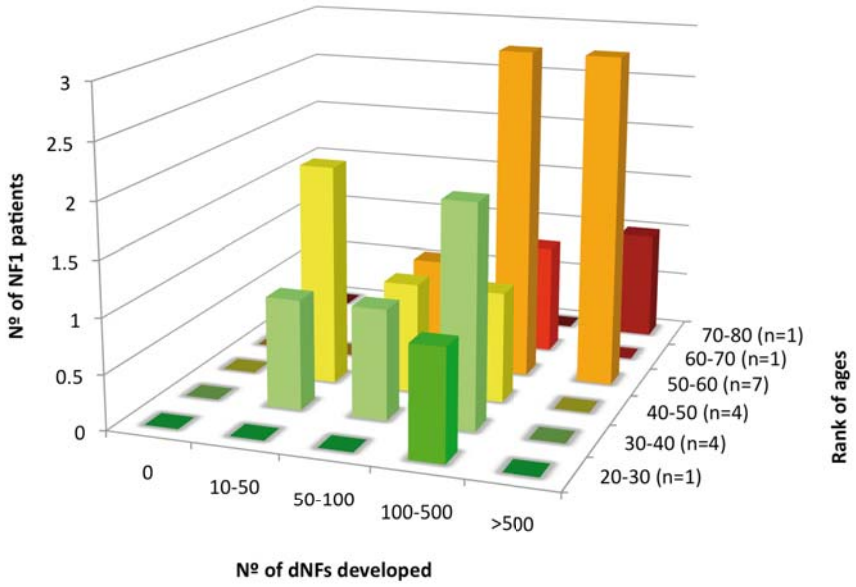


Figure 22. Distribution of dNFs developed in relation to age in our cohort of NF1 patients

1.3.4 Correlation studies between LOH/HR frequencies and the number of dNFs developed

We also considered studying a possible correlation between dNF LOH and HR frequencies and the number of dNFs developed. For this part of the study we used the 18 patients shown in Table 4 except patients P027 and P082. Patient P027, at the time of the study, was 26 years old, with 100-500 dNFs developed (patient represented in dark green in Figure 22). P027 presented an abnormal severe burden of dNFs, with a particular familial history of many different tumor types developed. We excluded him from the study as we considered it was a special case that differed a lot from the NF1 general population. On the other hand, patient P082 had 9 dNFs analyzed, not reaching the desirable 10 dNFs needed to obtain an acceptable statistical significance when determining LOH frequency. In addition, this patient was unique in presenting the highest prevalence of LOH generated solely by deletion. We reasoned that, considering the low number of individuals analyzed, the inclusion of these two NF1 patients could bias our results. Therefore, the correlation study was performed with 16 NF1 patients. This is still a low number of NF1 patients analyzed to obtain significant results, and more NF1 patients will need to be studied in the near future.

We first studied how dNF LOH, HR and deletion frequencies detected in the different patients analyzed were distributed in relation to the Risk of Dermal Tumor Burden (RDTB) values (see results section 1.1.2), that represents the number of dNFs developed corrected by age. We grouped NF1 patients in two different categories, low and high RDTB, and plotted the percentages of the different molecular parameters obtained analyzing dNFs (LOH, HR and Del) (Figure 23). We used an arbitrary cut off value for RDTB of <0.5 (low RDTB) and >0.5 (high RDTB). Figure 23a shows the % LOH obtained for high and low RDTB groups. The median showed that there were no clear differences between both groups, although the group with high RDTB contained more patients with higher LOH percentages than the low RDTB group. The same happened in the analysis of the HR percentages, in which the differences between both groups were even higher (Figure 23b). There were no differences between patients with high and low RDTB for the mechanism of deletion (Figure 23c).

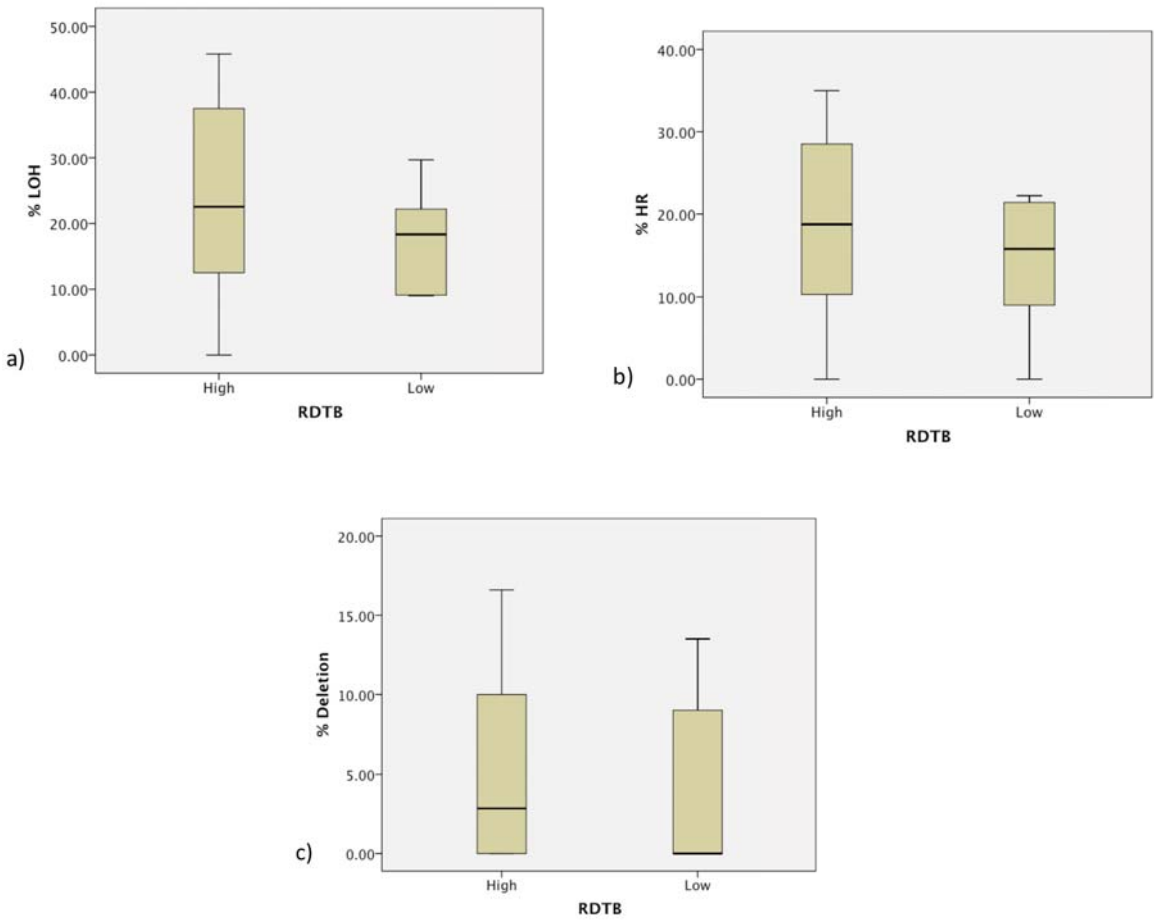


Figure 23. Box plots showing the % LOH (a), % of HR (b) and % of deletion (c) for the two groups of NF1 patients with high and low RDTB.

We further studied whether or not patients showing a higher frequency of LOH and HR in their dNFs were those also developing a higher number of dNFs (higher RDTB). We used a Chi-square statistic test to compare both RDTB groups (high and low values) with the LOH and HR frequencies. The threshold used to consider high and low frequencies of LOH and HR was taken arbitrarily as 25% (the median LOH prevalence found in our NF1 population). Tables 5 and 6 show the results obtained for LOH and HR, respectively. Results regarding the relationship between LOH frequency and RDTB show no significance (p = 0.215), while results regarding HR frequency and RDTB show a value close to significance (p = 0.058). This could suggest that NF1 patients showing higher percentages of *NF1* second hits generated by HR in dNFs, could tend to develop higher number of dermal neurofibromas (higher RDTB values). Again, 16 is a low number of NF1 patients to be analyzed, so these results should be taken as preliminary.

		% LOH		
		.00	1.00	Total
RDTB	.00	5	1	6
	1.00	5	5	10
Total		10	6	16

Fisher's exact test (unilateral exact significance) = 0.215

Table 5. Chi-square calculations for LOH (RDTB 0 and 1 values are low and high RDTB, respectively. % of LOH 0 and 1 values are low and high LOH, respectively).

		% HR		
		.00	1.00	Total
RDTB	.00	6	0	6
	1.00	5	5	10
Total		11	5	16

Fisher's exact test (unilateral exact significance) = 0.058

Table 6. Chi-square calculations for HR (RDTB 0 and 1 values are low and high RDTB, respectively. % of LOH 0 and 1 values are low and high LOH, respectively).

The use of RDTB value eliminates any bias age could produce to the estimation of the number of dNFs developed. However, to rule out any correlation between HR and the number of dNFs developed due to age, we checked whether or not patients showing the highest HR frequencies were also the oldest ones. Figure 24 shows the distribution of patients considering age and HR frequencies and, as shown in this figure, patients with higher HR rates (red dots) were distributed in different ranks of ages. We can conclude that the results obtained were not a bias by the age of the patients.

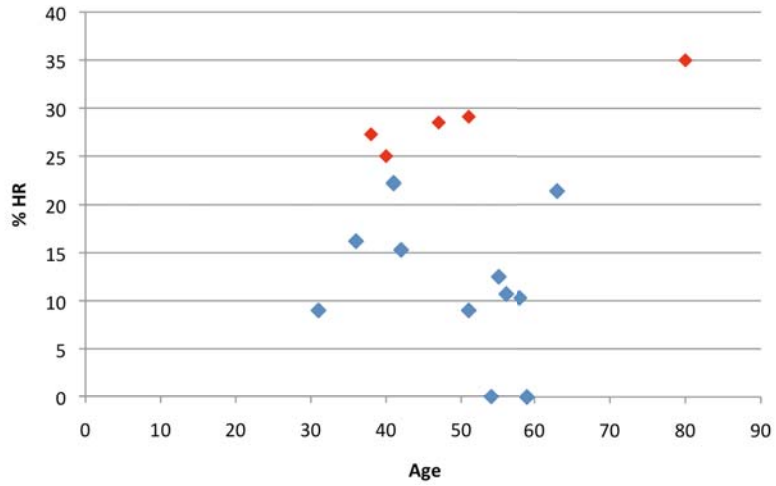


Figure 24. % of HR values in relation to the age of the patients. Red dots highlight patients showing high HR frequencies (with values above 25%).

1.4 Microsatellite Multiplex PCR Analysis (MMPA)

The LOH analysis evidenced the need to study 10 or more dermal neurofibromas from many more NF1 patients (ideally from our 117 patient cohort) in order to obtain significant values in the studies performed regarding any possible correlation between a clinical and a molecular phenotype, or any correlation between these phenotypes and a genetic analysis of candidate genes. Therefore, we decided to improve the methodology employed to analyze neurofibromas, to decrease its cost and speed up its analysis. While performing the LOH analysis of neurofibromas we realized that a microsatellite multiplex PCR could give us more information than just the allelic imbalance status. Understanding the origin of allele peak height intensities for the different mechanisms causing allelic imbalance (AI), together with the knowledge of locus copy-number calculations obtained by the MLPA technique, provided us the necessary information to set up new PCR conditions and create a calculation procedure that, when applied to a microsatellite multiplex PCR of control-tumor DNA pairs, let us determine the presence/absence and extension of the AI, the percentage of cells bearing AI in the tumor, the *NF1* locus copy-number, and the mechanism generating AI. The setting up, the quality controls used and the proposed calculations are described below.

1.4.1 MMPA set up

1.4.1.1 Proportionality between co-amplified amplicons

The MMPA calculations rely on comparing co-amplified microsatellites of control-tumor pairs. Proportionality between co-amplified amplicons must be maintained. In the MMPA set up shown here 22 cycles of PCR amplification ensured an adequate yield of each PCR amplicon while still maintaining an exponential reaction phase, guaranteeing proportionality among the different co-amplified amplicons for each independent multiplex PCR reaction (Figure 25).

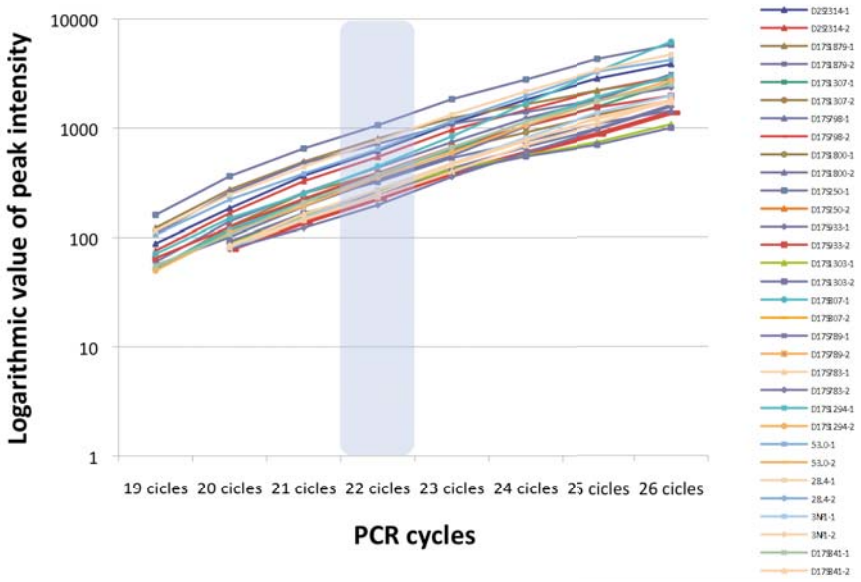


Figure 25. All co-amplified microsatellites in the MMPA reaction need to be analyzed when still in exponential phase of amplification. To determine the optimal number of cycles of the MMPA reaction, microsatellite amplification was monitored from 19 to 26 cycles. Logarithmic values of peak height intensities for the two alleles of each microsatellite are plotted. This analysis was performed in triplicate for sample P114-15N (only one reaction is shown).

1.4.1.2 Constant of proportionality (Kμ) among sample pairs and variation analysis

Control-tumor independent multiplex PCR reactions can be compared and a constant of proportionality (K) can be calculated for each heterozygous microsatellite using peak height values (Figure 26). Each K should be approximately the same for the different co-amplified amplicons within the MMPA reaction. The mean of the different microsatellite Ks (Kμ) reflects the constant of proportionality between the control-tumor compared reactions and can be calculated using either known pre-established control microsatellite markers within the MMPA reaction or by only taking into account those microsatellite markers that show no AI in the tumor sample (see below).

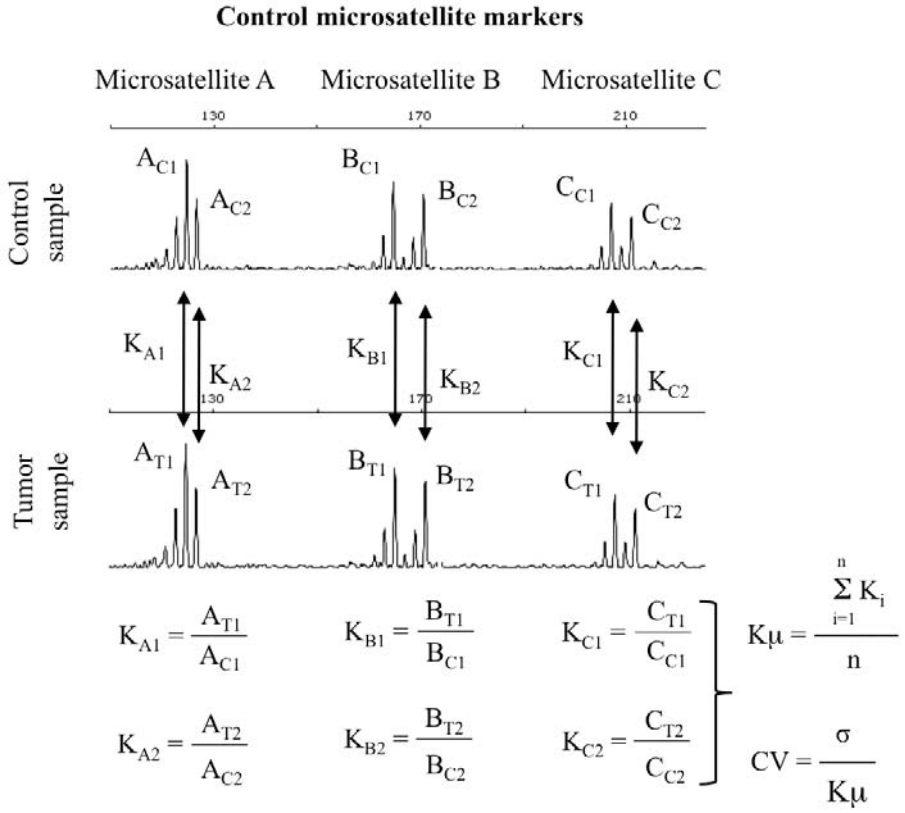


Figure 26. Constant of proportionality calculation. Example illustrating how the average of the proportionality constant (Kμ) between two microsatellite multiplex PCR reactions that compare control and tumor paired samples is calculated. Microsatellite electropherograms are shown. Each microsatellite peak has a constant of proportionality (K). Kμ is the average of all individual Ks from control microsatellites. The coefficient of variation (CV) from Kμ is calculated as a quality control parameter. Black line, observed peak heights (obs); dashed line, expected peak heights (exp).

Variation of Kμ affects the accuracy and sensitivity of proposed MMPA calculations and the coefficient of variation (CV) of Kμ is used as a quality control (Figure 26). To determine the maximum allowed Kμ variation in an MMPA, we studied a set of control samples from four different patients: 6 samples from patient P114, 4 samples from patient P027, 8 samples from patient P024 and 8 samples from patient P013 (Figure 27). CV (sigma/Kμ) values were obtained for each pair of samples, using 1 sample per individual as a control reference. CV did not differ more than 0.1 in all but 2 cases. MMPA should be performed in PCR reactions with CV values ≤ 0.15. Higher CV values could indicate loss of proportionality between the amplicons amplified. Moreover, we studied the behavior of the different co-amplified microsatellite

markers in the MMPA reaction using the same control samples (Figure 28). If proportionality between co-amplified amplicons is maintained, K values obtained from each microsatellite allele should be similar to the Kμ of the reaction. Hence, we calculated [1-(K/Kμ)] values, which gives us information on this variation (Figure 28) and, as before, we used 1 sample of each individual as a control reference. For most microsatellite markers, K values didn't differ more than 12% from the Kμ of the reaction. We established a cut off value on Kμ dispersion (Φ) of 0.12 in our MMPA reaction set up (Figure 28).

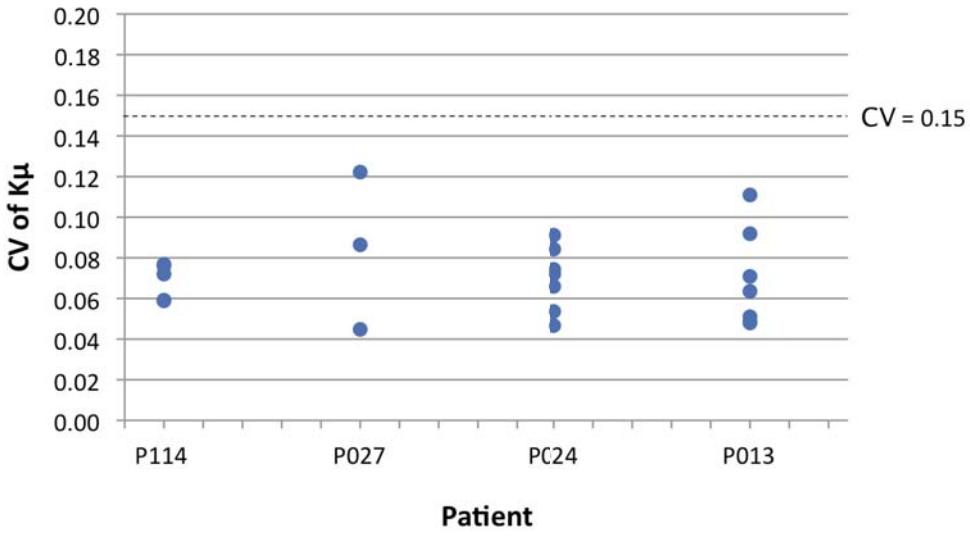


Figure 27. Limit CV value ($\sigma/K\mu$) for the MMPA reaction exemplified. Each blue dot represents the CV of $K\mu$ calculated for each pair of control samples (using one sample per individual as a reference). Only heterozygous microsatellites were used to calculate these ratios.

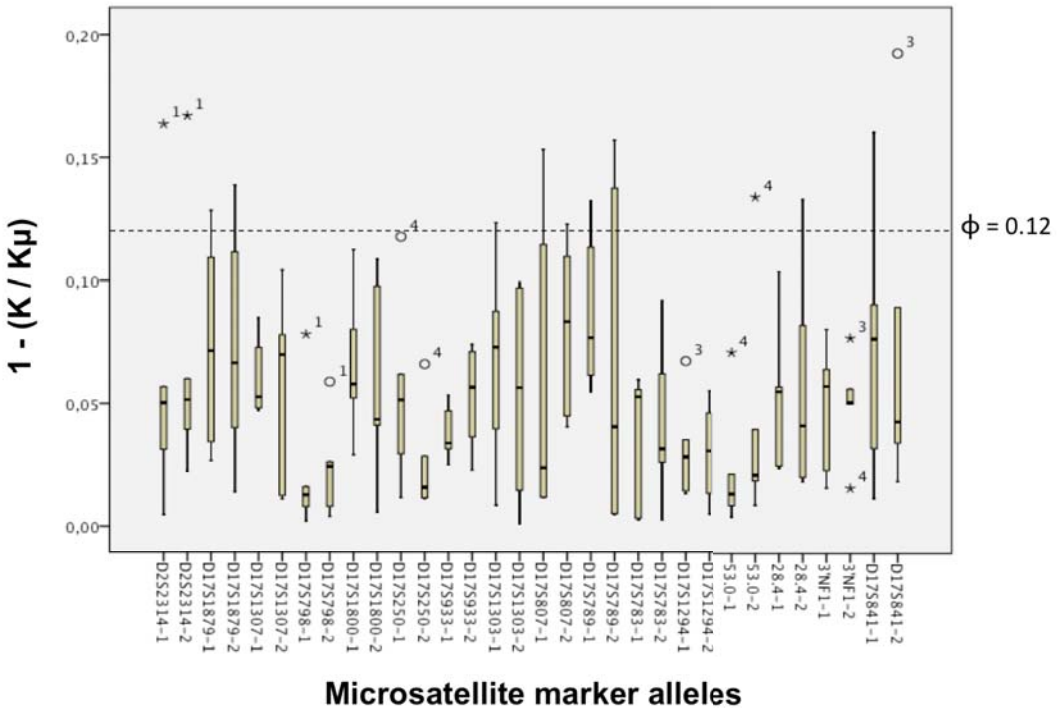


Figure 28. Box-plots of 1-(K/K*μ*), which measures the deviation of K with respect to K*μ* for each co-amplified microsatellite marker allele. Only heterozygous microsatellite markers were used to calculate

1.4.1.3 Detection of AI

AI analysis was based in the expression Q^{AI} (see material and methods). For the MMP we considered AI when there was a difference between ratios equal or greater than 0.2. For the MMPA we studied if this threshold could be more restrictive by studying Q^{AI} variation caused by methodological errors (Figure 29). We used the same control samples from four different patients as before, and one sample per individual was used as control reference. We calculated 1-(Q^{AI}) to measure the degree of variation of each microsatellite allelic ratio (Figure 29). There was an average variation of 5% when comparing pairs of allelic ratios of microsatellite markers not showing AI, and no microsatellite had a variation higher than 10% when comparing pairs of control samples. We considered 0.2 an adequate threshold to minimize false positive results.

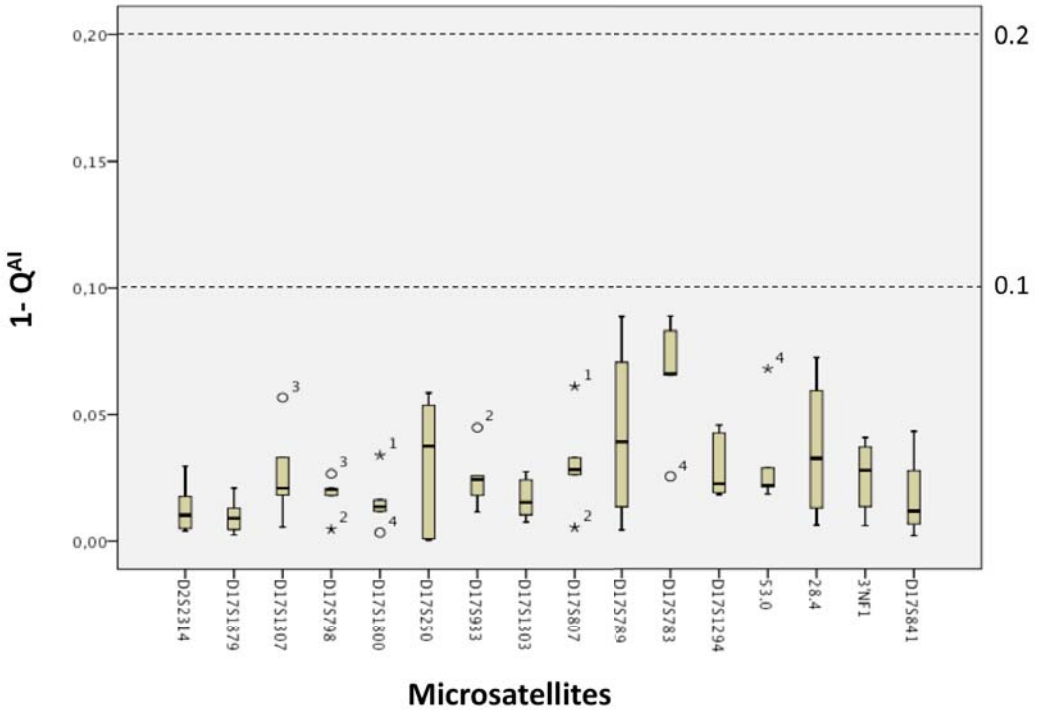


Figure 29. Boxplots of the 1-(Q^{AI}) values of all the ratios obtained from heterozygous microsatellites used in the MMPA reaction.

1.4.2 Expected allele peak height values and comparison of observed vs. expected allele peak heights

MMPA calculations are based in the comparison between the observed peak height values of a heterozygous microsatellite marker ($peak_{obs}$) with the hypothetical situation (expected peak height, $peak_{exp}$) in which there is no AI for that microsatellite marker. $Peak_{exp}$ in the tumor is calculated by multiplying Kμ by the allele peak height value of the relative microsatellite in the control sample ($peak_C$) (Figure 30):

$$Peak_{exp} = peak_C \cdot K\mu$$

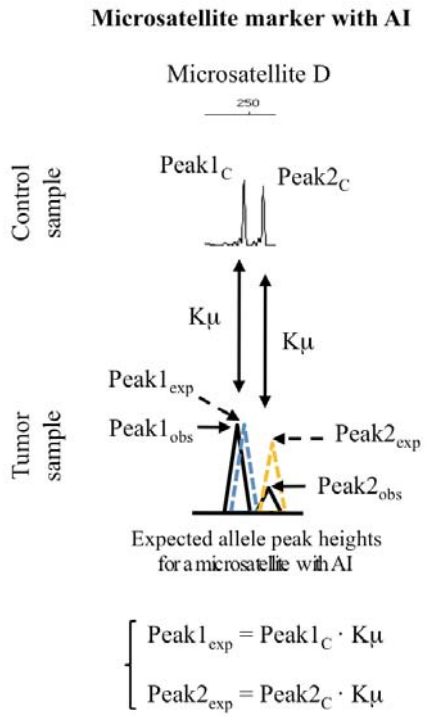


Figure 30. Expected peak height estimation. Example of the calculation of expected peak heights (in the case 100% of the cells were non-AI) of a query microsatellite D with hypothetical allelic imbalance (AI) by using Kμ between control and tumor paired samples. Black line, observed peak heights (obs); dashed line, expected peak heights (exp).

Allele peak heights of a microsatellite with AI will depend on the number of copies of that allele in AI-cells and the percentage of AI-cells within the tumor. Depending on the mechanism generating AI (Figure 31) $peak_{obs}$ values will increase, decrease or will be equal to $peak_{exp}$ values. By comparing $peak_{obs}$ and $peak_{exp}$ values of a given tumor microsatellite with AI, it can be ascertained which allele has lost a locus dosage (observed peak lower than expected), has an equal locus dosage (observed peak equal to the expected) or has a higher locus dosage (observed peak higher than expected) compared to the allelic status of the control sample pair of that microsatellite marker (Figure 31).

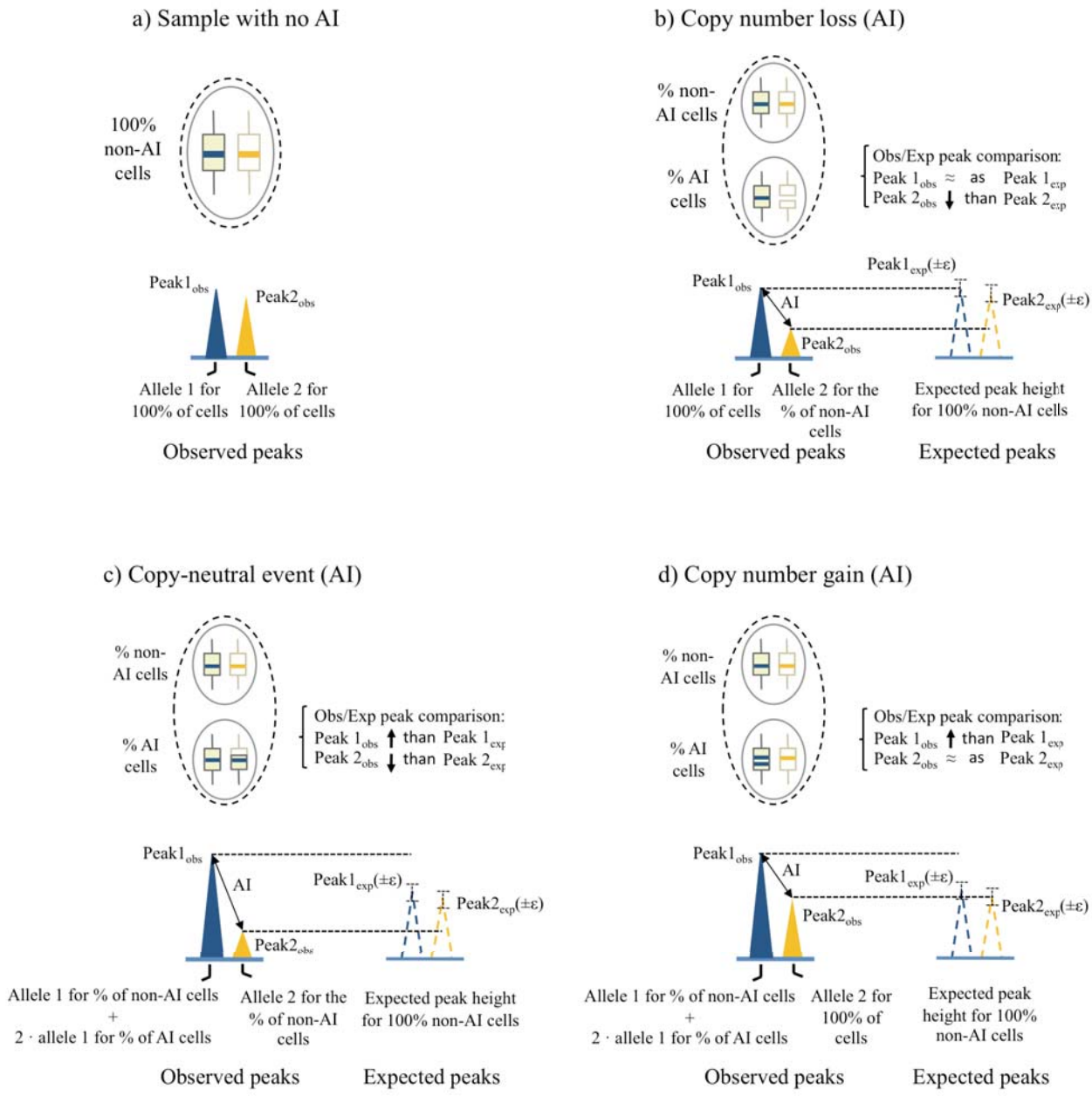


Figure 31. $Peak_{obs}$ vs. $peak_{exp}$ comparison for 1 heterozygous microsatellite in the analysis of AIs generated by different mutational mechanisms. Dashed line circles represent sample tissues (normal or tumor) and solid line circles represent types of genetically distinct cells within tissues. Blue and orange solid lines within chromosome-representing boxes indicate alleles of a heterozygous microsatellite marker. Solid color peaks represent allele peak heights obtained from microsatellite electropherograms after a theoretical MMPA. Dashed color peaks indicate expected peak heights in the case that 100% of the cells were non-AI, with the interval of peak height values indicated. Underneath allele peaks there is a short explanation of the source of peak height intensities regarding the number of amplified alleles according to the percentage of AI and non-AI cells. The Obs/Exp peak comparisons are also

shown for the different mechanisms: a) Sample with no AI; b) Tumor sample with AI due to a copy-number loss; c) Tumor sample with AI due to a copy-neutral event; d) Tumor sample with AI due to a copy-number gain. Note that identical AI values for a given microsatellite in examples b and c will reflect distinct percentages of AI-cells within tumor samples.

To perform this comparison, an interval of $peak_{exp}$ values is calculated by taking into account an empirically determined variation value (ε) that is applied to Kμ:

$Peak_{exp}$ [±ε] = $peak_C \cdot (K\mu \pm \varepsilon)$

ε [ε=Kμ·Φ] depends and is intrinsic to each MMPA reaction set up and takes into account the upper limit threshold (Φ) previously established by analyzing individual deviations of K vs. Kμ for every microsatellite marker used in the MMPA reaction (see Figure 28).

1.4.3 Calculating the percentage of cells exhibiting allelic imbalance from AI-markers caused by copy-loss or copy-neutral events

MMPA can be used to estimate the percentage of cells within tumors that carry an AI in an interrogated heterozygous marker locus. This is possible only in the cases where AI is produced by either copy-loss or copy-neutral events, and assuming a simple model in which tumors are mainly composed of two populations of cells: normal 2n cells and AI-cells. Once AI is identified, a comparison of $peak_{obs}$ and $peak_{exp}$ values is performed. In the case of AI due to copy-loss and copy-neutral events, one of the observed alleles will show a reduced peak height signal compared to the expected peak height (Peak 2_{exp} in Figures 30b and 30c and in Figure 32), and only this allele will be used to calculate the percentage of non-AI cells present in the tumor sample (Figure 32).

The difference between the expected and observed peak heights is directly proportional to the percentage of cells bearing AI for this allele, since the observed peak height signal is only produced by the fraction of non-AI cells (heterozygous 2n cells) present in the tumor:

% non-AI cells = $peak_{obs} / peak_{exp} \cdot 100$

For simplicity, the calculation of non-AI cells present in a tumor will be performed using a single Kμ value, rather than an interval of values.

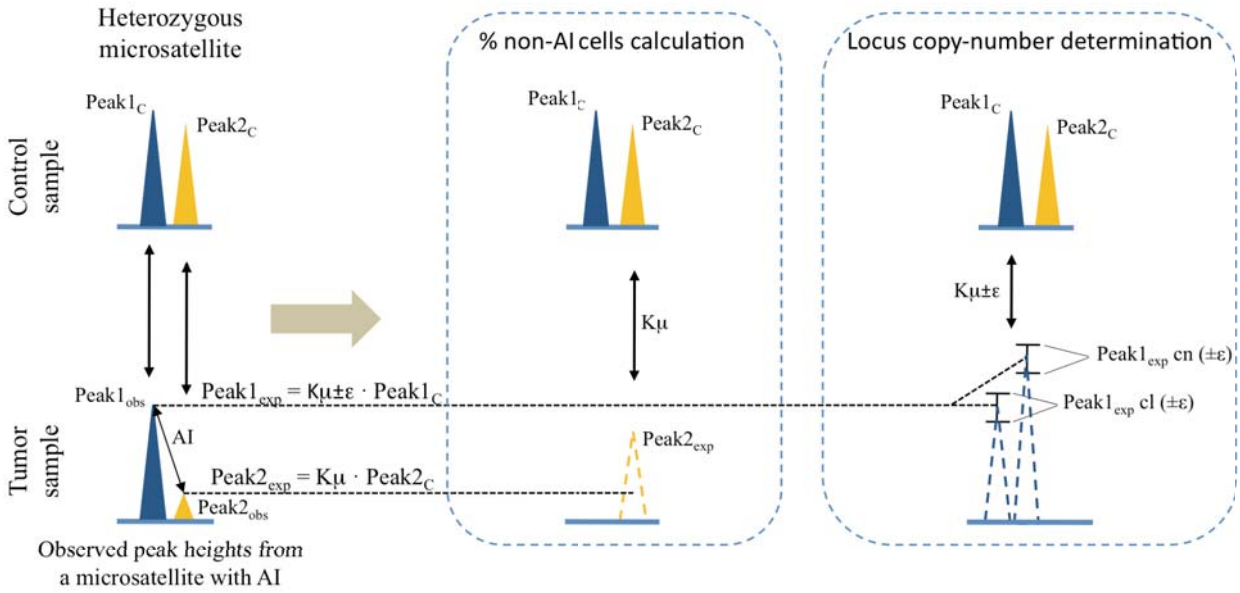


Figure 32. Calculating the percentage of non-AI cells and the locus copy-number of AI-cells from markers with allelic imbalance caused by copy-loss or copy-neutral events. Schematic view of the different use of both microsatellite alleles comparing observed vs. expected allele peak heights, concerning the calculation of the percentage of non-AI cells and the locus copy-number determination. Solid color peaks represent allele peak heights obtained from microsatellite electropherograms after a theoretical MMPA. Dashed color peaks indicate expected peak heights in the case 100% of the cells were non-AI. cl, copy-loss; cn, copy-neutral.

1.4.4 Calculating locus copy-number in cells showing allelic imbalance from AI-markers caused by copy-loss or copy-neutral events

MMPA analysis can provide information on whether each allele of an AI-locus has one, two or more than two copies. Considering the same AI-marker, the allele not used for calculating the percentage of non-AI cells is used to determine both the number of copies of that allele and the mechanism generating AI (peak 1_{obs} in Figures 30b and 30c, and in Figure 32). Using the percentage of AI/non-AI cells previously calculated and taking into account gene dosages that will result from the different mechanisms generating AI, expected peak height ratios are calculated. Epsilon will be

used to generate interval values, considering two scenarios:

Copy-loss (cl)

$Peak_{exp}$ cl (±ε) = [$peak_{exp}$ (±ε) · % non-AI cells] + [$peak_{exp}$ (±ε) · % AI cells]

Copy-neutral (cn) change

$Peak_{exp}$ cn (±ε) = [$peak_{exp}$ (±ε) · % non-AI cells] + [$peak_{exp}$ (±ε) · 2 · % AI cells]

$Peak_{obs}$ values will fit into one of the two intervals of $peak_{exp}$, indicating the locus copy-number status and the mechanism generating AI. In the case $peak_{obs}$ value does not fit any of the two expected intervals, the value should be considered out of range (OOR) and the corresponding microsatellite cannot be used for MMPA calculations. In the case that the two copies of a given locus were lost (e.g. nullosomy) or gained (e.g. tetrasomy) AI would not be detected. However, in these cases affected microsatellites would show K values substantially different from K values of other non-AI microsatellites. A low degree of variation in Kμ for each MMPA reaction is critical for a correct assessment of both the percentage of AI/non-AI cells in tumors and allele copy-number (see Figures annex A1 and annex A2).

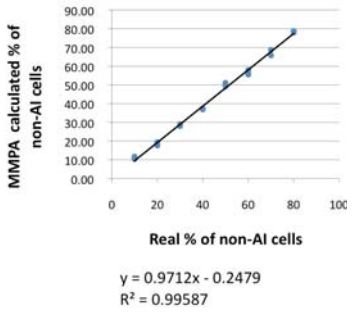
We performed an assay to determine the limit on the percentage of 2n cells present within tumor samples at which MMPA calculations are still reliable for the analysis of control/tumor pairs. We could perform this assay only for an AI caused by a copy-loss. We used the same DNA admixtures from the tumor heterogeneity and locus copy-number determination assay explained in Figure 13. For all microsatellites located within the deleted region the son was missing the maternal allele and only the remaining paternal allele was present. Alleles from heterozygous markers, outside the deleted region and shared by both father and son, were used as controls since, independently of the DNA admixture, they were always present in a 100% proportion (Figure 33c). To reproduce the presence of normal cells within tumors, heterozygous paternal alleles not present in the son and mapping within the deleted region were quantified (see boxed alleles in Figure 33c). A control sample with 100% father's DNA was independently paired with each of the different DNA admixtures. Kμ and

the different calculations were generated in triplicate for each pair of samples. The real percentages of DNA admixtures were compared to the calculated percentages obtained by applying MMPA (Figure 33a and 33b). Copy number analysis and the mechanism generating AI was also determined. MMPA technique was able to calculate the percentage of non-AI cells present in the DNA admixtures up to 70-80% while still detecting the presence of a deletion.

a)

	Calculated % of non-AI cells by MMPA			Mutational mechanism determined by MMPA
Real % of non-AI cells	Replica 1	Replica 2	Replica 3	Replica 1, 2 and 3
10%	11.69 ±2.25	10.38 ±1.46	10.84 ±1.57	Deletion
20%	17.59 ±2.23	19.42 ±2.51	18.13 ±2.63	Deletion
30%	29.06 ±3.17	28.36 ±2.70	27.92 ±1.36	Deletion
40%	37.07 ±2.40	37.41 ±1.96	36.84 ±2.70	Deletion
50%	51.08 ±2.50	50.05 ±2.07	49.20 ±2.99	Deletion
60%	57.91 ±2.70	55.48 ±2.72	56.26 ±1.44	Deletion
70%	68.65 ±2.66	68.65 ±2.44	65.74 ±3.75	Deletion
80%	Non detectable	Non detectable	Non detectable	Non detectable

b)



c)

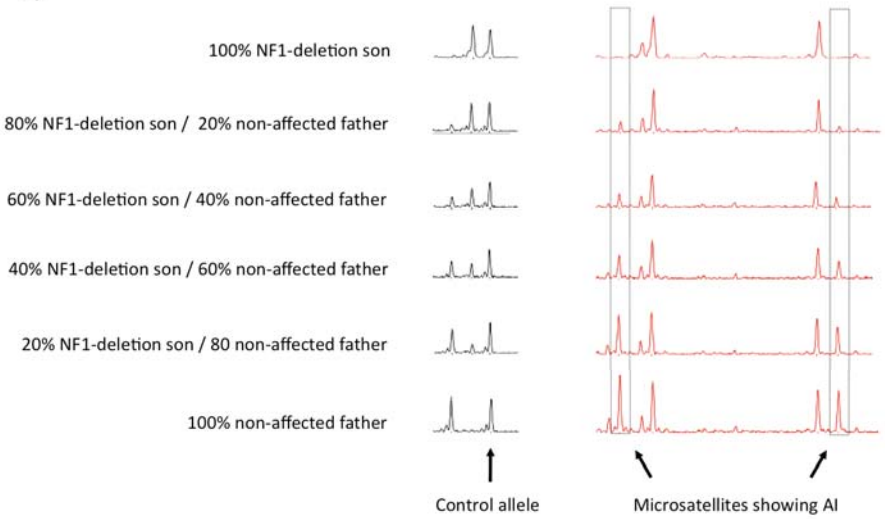
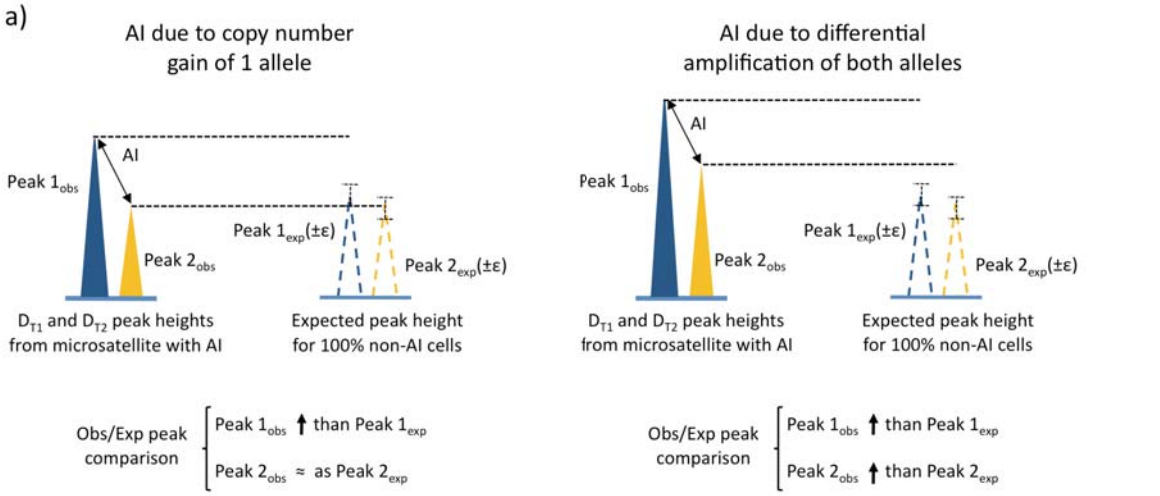


Figure 33. a) Results (in triplicate) of the % of non-AI cells and the mechanism generating AI obtained by applying an MMPA to the different DNA admixtures. b) Regression analysis indicating a good correlation between the real percentages of non-AI cells present in the DNA admixtures and the percentages calculated by applying MMPA. Blue dots represent the percentage of non-AI cells calculated using MMPA of each replica vs. the real percentage of non-AI cells. c) Examples of the electropherograms obtained for control and AI microsatellites for the different DNA admixtures.

1.4.5 Analysis of other mechanisms generating AI in microsatellites

AI can also be caused by copy-number gains of one allele or by the gain of a different number of copies of each allele (see Figure 34 as an example). In fact, equations described to obtain "$peak_{exp}$ cl" or "$peak_{exp}$ cn" could be adapted to calculate any gain in locus copy-number generating AI (Figure 34).



b)

When $Peak_{obs}$ is higher than $Peak_{exp}$:

$Peak_{exp}$ cg (±ε) = ($Peak_{exp}$ · % non-AI cells*) + ($Peak_{exp}$ · X · % AI cells*)

where X is the nº of copies of the microsatellite allele locus, being a whole number greater than 1 that meets:

$Peak_{exp}$ cg (-ε) < $Peak_{obs}$ < $Peak_{exp}$ cg (+ε)

* The % of AI/non-AI cells should be obtained from another microsatellite (see text)
cg: copy-gain

Figure 34. a) Schematic view of AIs generated by either single allele locus amplification (copy-number gain), or by differential amplification of both alleles (multiple amplification). Solid color peaks represent allele peak heights obtained from microsatellite electropherograms after a theoretical MMPA. Dashed color peaks indicate expected peak heights (in the case 100% of the cells were non-AI). b) Determination of allele locus copy-number for the two scenarios described. Only allele peak heights higher than expected will be used (Peak 1_{obs} in copy-gain and Peak 1_{obs} or Peak 2_{obs} for multiple amplification). In these cases, observed peak height values will fall inside the interval of expected peak height values [$Peak_{exp}$ cg (±ε)] obtained by using the actual number of copies of the locus, represented by the variable X. The percentage of AI/non-AI cells used for these calculations will need to

be obtained from a different microsatellite marker showing AI generated by either copy-loss or copy-neutral events.

A microsatellite marker with an AI generated by copy-number gain will show one of the observed allele peak heights higher than expected and the other one similar to its respective expected peak height. A microsatellite marker with an AI generated by multiple differential amplifications will show both observed allele peak heights higher than their expected peak heights. In both scenarios there is no observed peak height lower than expected, and that's why no information regarding the percentage of non-AI cells will be produced. However, in the context of a specific tumor and under the assumption that all genetic alterations causing AI are present in the same tumor cell population, information on the percentage of normal cells present in that tumor could be obtained from other AI-markers if AI is caused by copy-loss or copy-neutral events. In the case that one or more extra copies of one allele is gained, $peak_{obs}$ of that allele will be greater than the $peak_{exp,}$ while the remaining allele will be equal. In the case both alleles of an AI-marker gain extra copies but in different numbers, both $peak_{obs}$ will be greater than the respective $peak_{exp}$ (Figure 34).

1.4.6 Applying MMPA analysis to determine the *NF1* locus status in neurofibromas

We validated the MMPA by re-analyzing a set of 29 dermal neurofibromas with the newly established conditions, and comparing the obtained data regarding the percentage of non-AI cells and the *NF1* locus copy-number to the previously obtained by MLPA, PRA and SNP-array techniques (Tables 6 and 7).

In 26 out of 29 neurofibromas MMPA calculations correctly determined locus copy-number and AI-mechanism, considering all informative microsatellites (Table 7). In addition, AI-mechanism was correctly determined in all 29 neurofibromas, that had at least 3 informative AI-microsatellites, taking into account a correct result when at least 60% of all informative microsatellites were indicating the correct mechanism (Table 7). Overall 144 microsatellites showing AI were identified in these 29 neurofibromas: AI-mechanism was correctly determined after MMPA in 116 (80.5%); 25 (17.4%) were out of range (OOR) meaning that the $peak_{obs}$ value was not

fitting within any of the expected intervals; and only 3 (2.1%) were incorrectly determined (Table 7). We included samples P009-1N and P102-4N (marked with an * in Table 7) in the analysis. These tumor samples contained high percentages of non-AI cells, making copy-number analysis impossible by MLPA (see Table annex A2). However, they were considered to bear a deletion in the *NF1* locus, based in the similarity with other tumor samples bearing the same localized AI (only affecting the *NF1* gene and surrounding regions). MMPA calculations in these samples determined a deletion, supporting that this might be the actual mechanism generating AIs in these tumors. Copy-number ascertainment by MMPA calculations was possible due to the higher sensitivity of this technique when dealing with tumors containing high percentages of non-AI cells (Figure 33).

Regarding the presence of normal cells within tumors, we calculated the percentage of AI/non-AI cells for 9 neurofibromas using the SNP-array data by applying the GPHMM algorithm. We compared these data with the percentages obtained for the same tumors by applying MMPA calculations (Table 8). The results showed a good agreement between both techniques since, for all tumor samples but one, the calculated percentages of non-AI cells present in neurofibromas did not differ more than 8%.

	MMPA calculations (present work)			Garcia-Linares et al. 2011		
Tumor samples	MMPA copy-number determination	Correct microsatellite determination / Total microsatellites with AI	Non determined microsatellites	(Copy-number assessment)		
				MLPA	PRA	SNP-array
P001-1N	Two copies	6/7	1 OOR	Two copies		
P009-1N	One copy	4/4		*		
P011-16N	Two copies	4/5	1 OOR	Two copies		Two copies
P022-19N	Two copies	3/5	2 OOR	Two copies		
P022-21N	Two copies	1/6	5 OOR	Two copies		
P023-14N	Two copies	6/8	2 OOR		Two copies	Two copies
P023-6N	One copy	3/4	1 RH	One copy	One copy	One copy
P023-97N	One copy	3/5	2 OOR	One copy	One copy	One copy
P030-2N	Two copies	4/5	1 OOR	Two copies		
P039-1N	One copy	4/4		One copy		One copy
P047-1N	Two copies	4/5	1 OOR	Two copies		
P054-1N	Two copies	6/6			Two copies	
P062-11N	One copy	2/2		One copy		
P079-1N	One copy	3/3		One copy		One copy
P081-1N	One copy	5/5		One copy		
P082-6N	One copy	5/5			One copy	One copy
P090-3N	One copy	3/3		One copy		One copy
P095-1N	One copy	2/3	1 OOR	One copy		One copy
P102-3N	One copy	4/5	1 OOR	One copy		
P102-4N	One copy	1/2	1 OOR	*		
P102-5N	Two copies	8/9	1 OOR	Two copies		
P102-18N	Two copies	9/9		Two copies	Two copies	Two copies
P103-5N	Two copies	5/5		Two copies	Two copies	Two copies
P103-21N	Two copies	4/6	2 OOR	Two copies		Two copies
P109-1N	Two copies	4/5	1 Del	Two copies		
P109-5N	Two copies	3/5	1 OOR, 1 Del	Two copies		
P109-6N	Two copies	4/4				Two copies
P109-7N	Two copies	3/4	1 OOR	Two copies		
P112-4N	Two copies	3/5	2 OOR	Two copies		

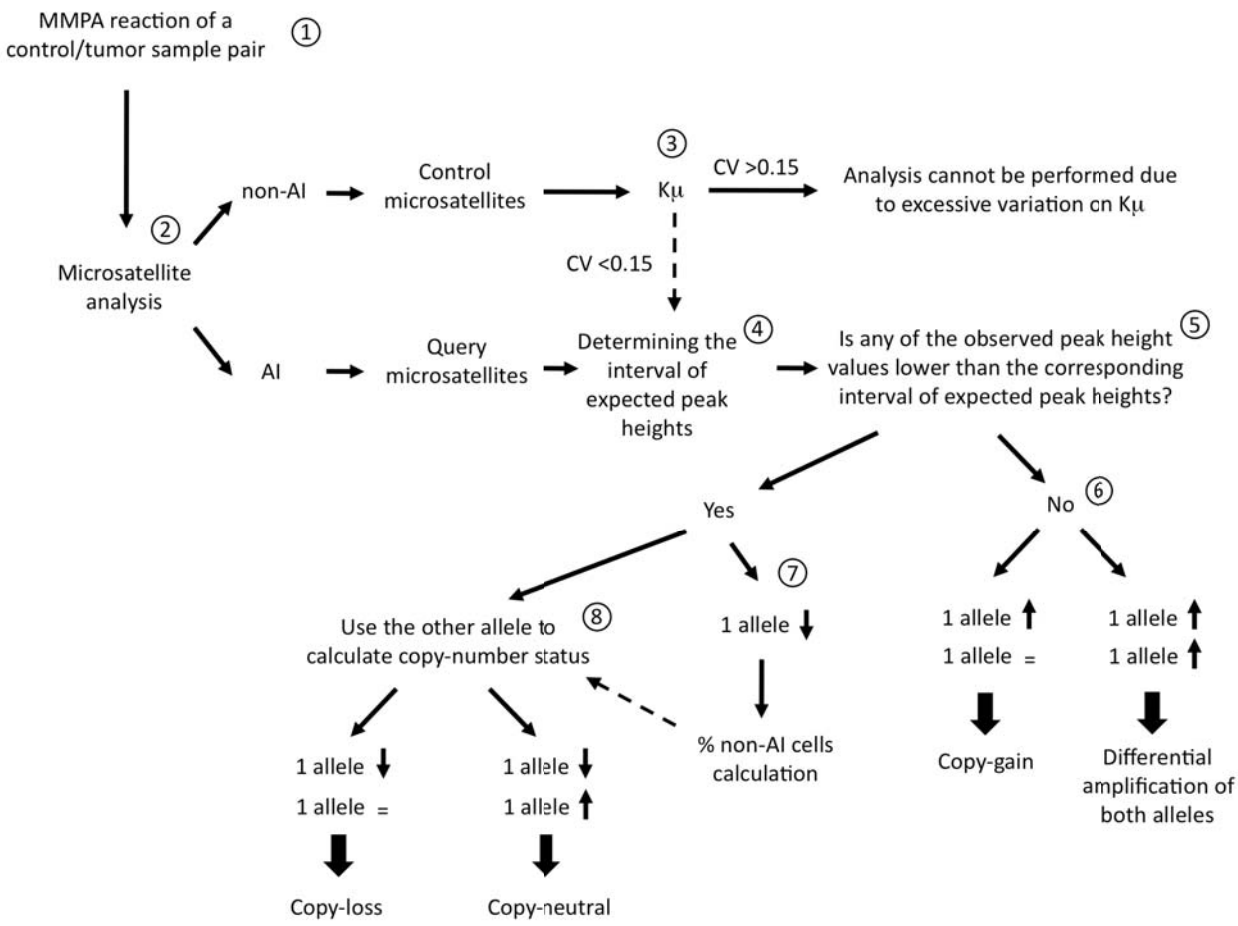
*see text for details

Table 7. MMPA validation of the NF1 locus copy-number. Analysis of the copy number status of the *NF1* locus in 29 neurofibromas using newly developed MMPA calculations and comparison with previous data obtained for the same tumors using MLPA, PRA and SNP-array techniques.

Tumor samples	MMPA (% non-AI cells)	SNP-array (% non-AI cells)	Difference between techniques (%)
P023-6N (Del)	40	42.5	2,5
P079-1N (Del)	66.1	66	0.1
P082-6N (Del)	57	65	8
P011-16N (HR)	67	74.9	7,9
P023-14N (HR)	45	59.5	14,5
P102-18N (HR)	43	47	4
P103-5N (HR)	44	50.6	6,6
P103-21N (HR)	56.1	55.7	0,4
P109-6N (HR)	46.2	53.6	7,4

Table 8. MMPA validation for the % of non-AI cells. Analysis of the percentage of non-AI cells present in 9 neurofibromas using newly developed MMPA calculations and comparison with data obtained applying GPHMM algorithm to previously generated SNP-array data for the same tumor.

In order to facilitate the calculations of the MMPA assay, we developed an automated analysis script that outputs the different parameters of an MMPA reaction together with the different calculations explained above. This script is freely downloadable and can be easily customized to any MMPA design. Figure 35 shows a summary of the MMPA workflow.



① MMPA reaction of control/tumor sample pairs

② Microsatellite analysis by Q^{AI}: designing control and query microsatellites

③ Calculation of Kμ from Ks of control microsatellites and of Kμ's CV

④ Calculation of expected allele peak height values in microsatellites with AI using ±ε, obtaining an interval of values for every microsatellite allele

⑤ Comparison between observed and expected allele peak height values. Independent analysis for each microsatellite with AI.

⑥ Neither observed allele is lower than expected for a given AI-microsatellite. Mechanisms like copy-gain of one allele or multiple differential copy-gains of the two alleles could cause the AI

⑦ For a given AI-microsatellite an observed allele peak height is lower than the interval of its expected peak height. This peak allele is used to calculate the percentage of non-AI (2n) cells present in the tumor

⑧ For the same AI-microsatellite, the other allele is used to calculate the locus copy-number (taking into account the percentage of non-AI cells), and discriminate between copy-loss and copy-neutral events.

Figure 35. MMPA workflow.

Selection of candidate genes influencing the number of dermal neurofibromas

This part compiles different strategies used to select candidate genes. Genes were selected using two different approximations. In a first approximation, genes were selected based on literature and specific criteria. In a second approximation, candidate genes were meant to be selected after modeling the homologous recombination mechanism in different yeast strains, and applying new association strategies and functional characterization to identify genes influencing the homologous recombination rate *in vivo*.

PART 2

2. Selection of candidate genes influencing neurofibroma number

The aim of the present work was to find genes influencing the number of dNFs in NF1 patients. Among the biological processes that could affect this trait, we narrowed our research to the processes related to the somatic mutation rate of NF1 patients, in particular to repair mechanisms. A high percentage of somatic inactivations in neurofibromas are generated by homologous recombination (HR) (141, 142) and, considering that this mechanism presents interindividual variation [133] we thought genes affecting HR rate could be good candidates as modifiers of the number of dNFs. This repair mechanism has recently gained importance in relation to tumorigenesis [96,109], as its abnormal regulation can lead to genome instability, causing genomic alterations that contribute to the appearance of cancer [143]. In addition, numerous cancer syndromes are caused by mutations in genes related to HR, like the *NBS1* gene in Nijmegen Breakage syndrome, or *BLM* gene in Bloom syndrome [144].

HR is a complex mechanism regulated by multiple genes [96,145]. In order to identify candidate genes related to HR and influencing the number of dNFs developed, we defined two different strategies: the analysis of candidate genes based on literature, and the study of candidate genes obtained from analyses using yeast as a model system.

2.1 Literature-based gene selection

Among genes involved in DNA repair, we decided to narrow the analysis to genes that accomplished the following criteria:

1) Focus the analyses in genes involved in the homologous recombination process.

2) Favor genes responsible for cancer syndromes (CS). Most CS caused by mutations in repair genes are hereditary recessive disorders, and thus allowing for the presence of carriers of hypomorphic (or hypermorphic) alleles. In addition, mutations in these genes increase significantly the risk for developing tumors.

3) Favor the selection of human genes with a corresponding ortholog in yeast. This

would allow us to study the candidate genes in this model system, and even perform functional genotype-phenotype analyses to clarify the functional role of the variants identified.

Taking into account these criteria, we started selecting three different candidate genes:

- *BLM*: Mutations in the *BLM* gene cause the inherited Bloom Syndrome. Patients with Bloom syndrome show genomic instability with an increased predisposition to develop cancer [146], together with other manifestations like dwarfism, skin lesions, or immunodeficicency. Blm is a RecQ helicase-like protein implicated in the 3'-end resection in the mechanism of HR. Its alteration generates a high frequency of sister-chromatid exchanges, hyperrecombination and chromosomal breakages [146]. Studies in mice showed that mutations in *BLM* were responsible for the appearance of multiple polyps in the intestines of the animals [147], the majority evidencing APC somatic inactivation by LOH mediated by HR. This gene has an ortholog in yeast, the *SGS1* gene, a nucleolar DNA helicase involved in genome integrity maintenance [148].

- *NBS1*: Mutations in the *NBS1* gene cause Nijmegen breakage syndrome. This syndrome generates an acute chromosomal instability and predisposition to cancer [143]. Patients with this syndrome can show microcephaly, growth retardation or immunodeficiency [149]. The Nibrin (Nbn) protein is a nuclease, member of the MRN complex, which is involved, like Blm protein, in the process of the DSBs 3'-end resection in the HR mechanism. It has an ortholog in yeast called *XRS2* gene [150].

- *RAD51A*: This gene is a member of the RAD51 protein family, and it is involved in the mechanism of homologous recombination by interacting with proteins PRA and RAD52, and by favoring filament formation in the 3'-end strands. It also interacts with BRCA2, which alteration can lead to genomic instability. This gene is altered in numerous cancers, like breast and prostate, and has been associated with poor survival in head and neck cancers, non-small cell lung cancers and in soft tissue sarcomas [143]. The ortholog *RAD51A* has been described in yeast [105].

Genetic variation in candidate genes was first planed to be scanned by direct sequencing, since we wanted to capture all variation regardless of population frequencies. While the work of this thesis was being developed, the new advent of next-generation sequencing (NGS) technologies made us rethink our first

approximation. A NGS-dependent hereditary cancer syndrome target-exome strategy was designed and started to be set up by the genetic testing unit of our laboratory (Castellanos, Gel, Serra et al. unpublished results), using Agilent custom target-exome capture, and the Ion Torrent PGM platform. A targeted exome (sub-exome) approximation was adopted in contrast to analyzing the whole exome. This strategy provides an exhaustive coverage of all exon and exon-intron boundaries of the genes analyzed and reduces the data analysis complexity and cost.

The hereditary cancer targeted exome strategy was designed to cover a total of 106 genes: 97 involved in cancer syndromes and 9 involved in RASopathies. Among these genes, 39 were involved in DNA repair pathways, and 9 were known to be related to the mechanism of homologous recombination (Table 9). The analysis provided exhaustive and deep coverage, together with high sensitivity and specificity (false positive and false negative rates were very low). We took advantage of this methodological approach to start sequencing the group of 18 NF1 patients for whom clinical and molecular characterization was performed, in order to start capturing genetic variation in a fair number of candidate genes at once (see part 3 of results).

Genes	Gene involved in HR pathway
ATM	
BLM	x
BRCA1	x
BRCA2	x
BRIP1	
CHEK2	x
CSA/ERCC8	
DDB2	
ERCC2/XPD	
ERCC3/XPB	
ERCC4/XPF	
ERCC5/XPG	
FANCA	
FANCB	
FANCC	
FANCD2	
FANCE	
FANCF	
FANCG	
FANCL	
MLH1	
MLH3	
MSH2	
MSH3	
MSH6	
MUTYH	
NBS1 /NBN	x
PALB2	x
PMS1	
PMS2	
POLH/rad30	
RAD50	x
RAD51	x
RECQL4	
TP53	
WRN	x
XPA	
XPC	
XPE/DDB1	

Table 9. Summary of the genes involved in repair mechanisms analyzed by NGS. (x) indicates genes known to be involved in the mechanism of HR.

2.2 Candidate gene selection by modeling *in vivo* homologous recombination in yeast

Homologous recombination has been widely studied in yeast [151,152]. Much is known about the genes involved in this mechanism, but not about the genes influencing the *in vivo* homologous recombination rates. We developed a molecular assay in yeast, the Homologous Recombination Yeast assay, or HoReYe assay, to determine the *in vivo* HR rate in different yeast strains. The HR event measured mimicked homologous recombination events produced in neurofibromas. We planned to obtain several different yeast strains exhibiting different rates of HR. The variants responsible for high or low HR rate in these strains were planned to be identified by an adaptation of the X-QTL assay. This strategy is based on the analysis of yeast haploid segregants obtained from the mating of two yeast strains with extreme opposite phenotypes to obtain allelic variants influencing variation in the trait studied in the two strains. If X-QTL wasn't adequate to analyze this trait, and alternative assay based on a phenotype-based selection and backcrossing strategy could be applied (see below). After the identification of candidate gene, a testing for a functional relationship between genotype and phenotype in yeast could be performed, and human orthologs of candidate genes obtained by this methodology would be studied in our group of well-characterized NF1 patients. The variants found by this approximation would be correlated with the molecular and clinical phenotype of the patient. In addition, human allelic variants could be analyzed in yeast to establish functional genotype-phenotype correlations regarding the homologous recombination rate of the patient.

2.2.1 HoReYe (Homologous Recombination in Yeast) assay

We created the HoReYe assay to measure the HR rate in different yeast strains. Systems for determining the HR rate in yeast had already been developed [153], but HoReYe was set up to measure only HR events identical to those generating LOHs in dermal neurofibromas (Figure 36).

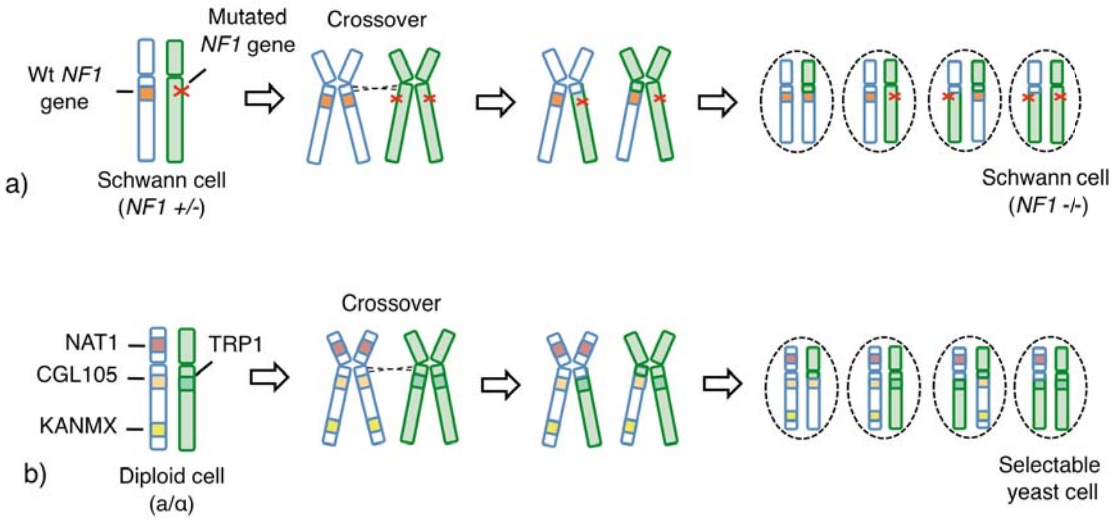


Figure 36. a) Schematic representation of a somatic *NF1* mutation by the homologous recombination mechanism. b) Schematic representation of the construction used in the HoReYe assay to select for cells suffering from a homologous recombination mechanism.

In dNFs HR events are due to a crossover between the centromere and the *NF1* gene, that results in the generation of LOH extending from the breakpoint until the telomere of the long arm of 17th chromosome. In 50% of the cases, and depending on the chromosomal disposition on the metaphase plate, cells with no functional copy of the *NF1* gene (*NF1* -/-) will segregate, and will end up generating the neurofibroma (Figure 36a). In these SC, both 17q arms will be exactly the same with the *NF1* germline mutation and any other variation or polymorphism reduced to homozygosity. The HoReYe assay was meant to mimic the same HR event, and select for the same kind of cells, the ones bearing homozygosity along the chromosome long arm (Figure 36b). Any mutation rate assessment should consider the gain of mutations in the initial population of cells previous to performing the assay. Mutation rates are determined by quantifying the number of individuals bearing a specific mutation compared to the initial population. However, mutations that occur early in a cell culture can inflate their proportion over recent mutations, and thus will be over-represented in the final analysis [154]. To avoid this issue, a fluctuation analysis (FA) should be performed. FA consists in basically performing culture replicates of the *in*

vivo mutation rate assay [155].

To perform the HoReYe assay yeast cells had to be genetically modified. Several selection marker genes were inserted in different regions of chromosome XV of haploid yeast cells. Markers were inserted close to telomeres of both L and R chromosome arms in α cells. In addition, a selection cassette construct was inserted in the R arm close to the centromere. In the a cells gene markers were inserted in the R arm near the centromere (Figure 37).

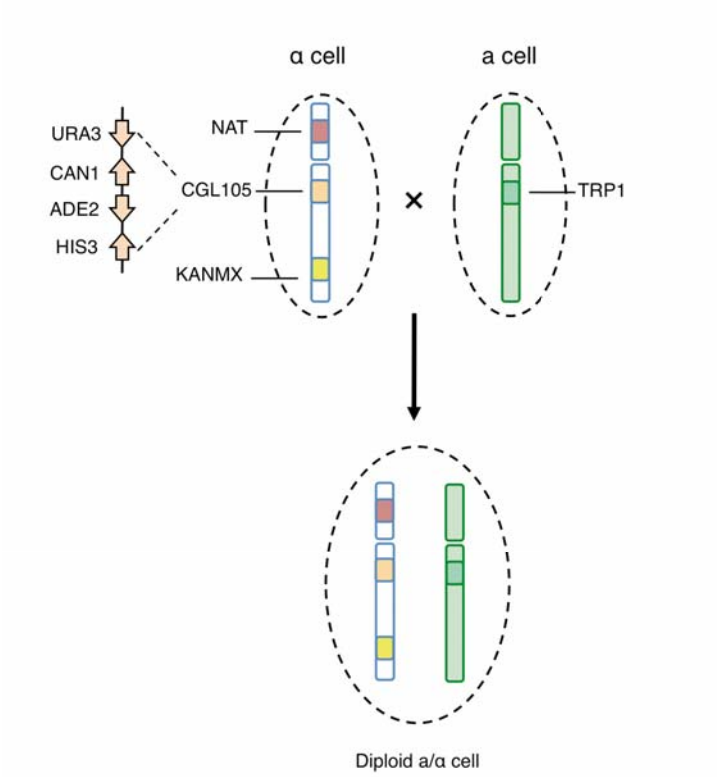


Figure 37. Schematic representation of the selection marker genes inserted in the chromosome XV of the yeast α and a cells. The genes forming the CGL105 construct are orientated in opposite directions.

2.2.1.1 Cassette construction

A cassette construct of four selection marker genes, called CGL105, was designed to be inserted in the R arm of the chromosome XV near the centromere, in an analogous location to the position of the *NF1* gene in chromosome 17. It was created by a serial of insertions of 4 different marker genes in the pRS415 plasmid (Figure 38) by a high-efficiency yeast transformation and recombination method. Each gene marker was

orientated in an opposite direction to avoid possible expression inhibition due to the proximity of gene promoters and terminators (Figure 38).

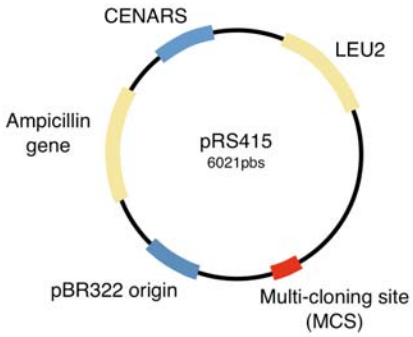


Figure 38. Plasmid pRS415. CENARS and pBR322 origin are the origin of replication of yeast and bacterial cells, respectively. LEU2 is the auxotrophic gene for leucine. The ampicillin gene confers resistance to the ampicillin antibiotic. MCS contains multiple restriction sites.

After each transformation, selection markers were tested functionally: yeasts lacking the marker genes to be tested were transformed with the entire plasmid and put under selective growing conditions for the presence of the marker. Once all genes were tested, we inserted the entire construct in the yeast genome, near the centromere, by the same transformation method.

2.2.1.2 Selection markers in cassette construct

The genes composing the CGL105 construct were:

1) *URA3*: It is an auxotrophic gene responsible for the biosynthesis of pyrimidines, and it can be used as a positive or a negative selector. Yeast strains need this gene to survive in the absence of uracil in the medium (positive selection), and in the presence of 5-Fluoroorotic Acid (5-FoA) it generates 5-fluorouracil, which is toxic to the cell (negative selection). In the HoReYe assay it was used as a negative selector.

2) *CAN1*: It codifies for an arginine permease, and it is used as a negative selector. This gene lets the entrance of the compound of canavanine, which is toxic to the cell. So its mutation, the loss of the gene, confers canavanine resistance. We used two

negative selection markers in tandem to minimize false positive rate (yeast colonies growing in selection media). Each of the genes has a high mutation rate (~10^5) higher than transformation efficiency. The use of a double selection makes the mutation rate lower than the transformation efficiency, practically avoiding any false positive.

3) *ADE2*: It is a carboxylase the mutation of which affects the color of the yeast cells. Cells with this gene mutated, and without a source of adenine in the medium, generate an accumulation of a red pigment. Its loss confers a color selection.

4) *HIS3*: It is an auxotrophic gene involved in histidine biosynthesis. A mutant for the *HIS3* gene cannot survive in the absence of histidine in the medium, so it is used as a positive selector.

2.2.1.3 Other selection markers

NAT1: Inserted in the L arm of the yeast chromosome XV in an α cell. It is a gene present in *Streptomyces noursei*, and encodes for the protein nourseothricin N-acetyl-transferase. This protein confers resistance to the antibiotic Nourseothricin (clonNat), and it is used as a positive selector. *KANMX*: Inserted near the telomere in the R arm of the yeast chromosome XV in an α cell. It is a hybrid module construct that contains the kanR open reading-frame of a E. Coli transposon. It confers resistance to the transformant yeast cells grown against geneticin (G418). *TRP1*: Inserted near the centromere in the R arm of the yeast chromosome XV in an a cell. It is an auxotrophic gene involved in tryptophan synthesis. A mutant in *TRP1* won't be able to grow in the absence of tryptophan in the medium, and it is used as a positive selector.

2.2.1.4 Yeast strain preparation

The HoReYe assay was based on the selection of markers inserted in the yeast genome. As some of these are auxotrophic genetic markers (markers that complement the auxotrophy of the yeast strain, that commonly is unable to synthesize either an AA or a nucleotide), they were, in some strains, already present in other genomic locations (Table 10). In these cases, it was necessary first to delete these genes to create auxotrophy in these strains from their original location before introducing them

in the chromosome XV. The method used to delete these genes was the "Delitto Perfetto" [156]. This is a site-directed mutagenesis method that uses the yeast recombinational machinery to accurately perform chromosomal modifications. Its name comes from the fact that this method involves a complete elimination of any sequence used in the deletion process (in opposition for instance to a Cre-loxP strategy). The method used to insert the selection marker genes into the yeast cells was the same employed in the CGL105 construction, the high-efficiency yeast transformation protocol: this time the marker genes contained in its extremes 50bps homologous to the flanking regions of the genomic target locus. They were transformed alone, and selected for their correct expression. Both laboratory and wild-type strains of *Saccharomyces cerevisiae* were chosen to perform the assay to assure the presence of different genetic backgrounds in our study.

Yeast strain	Isogenic	Yeast type	CAN1Δ	URA3Δ	ADE2Δ	HIS3Δ	TRP1Δ	NAT	KANMX	CGL105
BMA64-1A	W303	Laboratory								
BMA64-1B	W303	Laboratory								
SY991	S288C	Laboratory								
SY992	S288C	Laboratory								
CEN-PK2-1C	CENPK	Laboratory								
CEN-PK2-1D	CENPK	Laboratory								
YJM145		Wild-type								
YJM320		Wild-type								
YJM326		Wild-type								

Table 10. Table with the yeast strains that have been, or are planned to be, genetically modified. Δ = deletion. Orange squares: genes already mutated. Red squares: genes deleted by the Delitto Perfetto method. Green squares: genes inserted by transformation and recombination.

2.2.1.5 HoReYe assay

Figure 39 shows a summary of the Homologous Recombination Yeast assay. The starting point of the assay was the generation of haploid a and α cells with the respective selection markers and the cassette construct. To analyze the HR rate diploid a/α cells were generated (by mating the haploid ones). Diploid cells were grown in media lacking histidine and triptophane to select for yeasts containing HIS and TRP markers, assuring for the presence of both modified chromosomes XV.

When cells reached a desired density, they were transferred and diluted in a medium with no selection pressure (YPDA). Cells were then grown just for one doubling time (~90-120min). It was in this period of time when double strand breaks and repair mechanisms took place, some of them involving the XV chromosome. Then, cell density was measured and 10^6 cells were plated in canavanine plates (at the right density for being able to count and analyze colonies). After three days under this selection, only cells that had suffered mutations in the CAN1 gene were able to grow. Once colonies reached a desired size replica-platting was performed in the selection media conditions: SD+5-FoA (negative selection of *URA3*), SD+G418 (positive selection of KANMX), SD+clonNat (positive selection of *NAT1*), SD-His/Trp (negative selection of *HIS* and *TRP*) and YPD (color selection) (Figure 39).

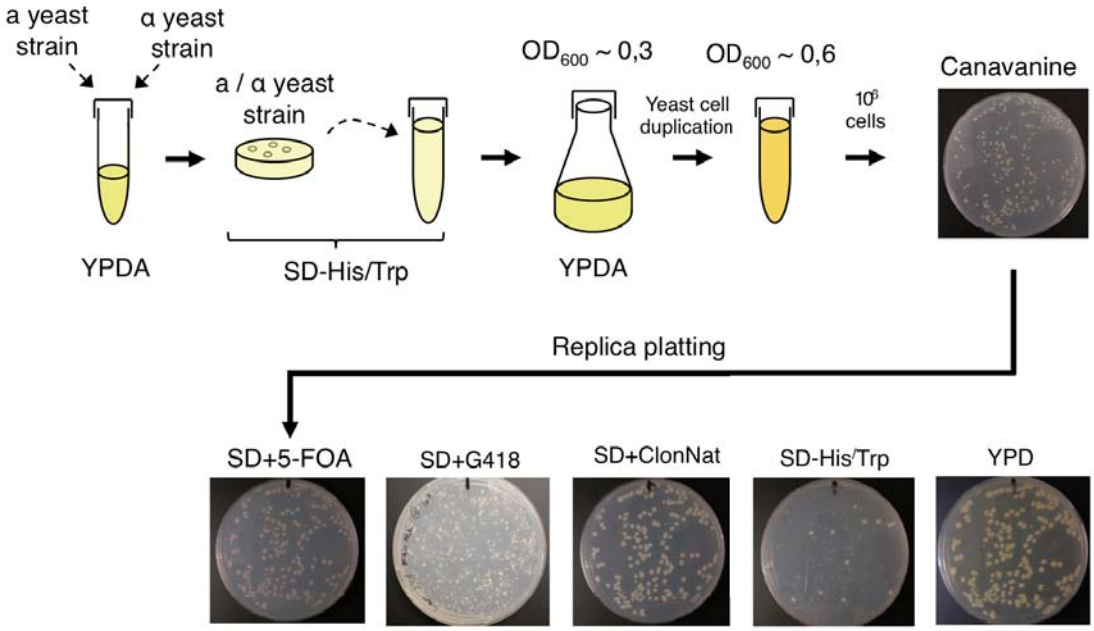


Figure 39. Summary of the HoReYe assay.

Each colony grew in different plates depending on mutational mechanisms involved in the loss of CAN1 gene function. For each mechanism a different combination of colony growth/non-growth in the different selection media was obtained. We considered four mutational mechanisms involved in the loss of CAN1, as described in Table 11, and explained in Figure 40:

1) Point mutation (Figure 40a): A point mutation inactivating CAN1. In this case colonies grew white in color (presence of ADE2) in all plates but in SD+5-FoA, since growth was inhibited by the presence of URA3. As the first step in the yeast assay was the selection against canavanine, only point mutations in this gene could produce this combination of growth/non-growth pattern. If the first platting were against 5-FoA (instead of canavanine), then point mutations in the URA3 gene could be analyzed.

2) Deletion (Figure 40b): Deletion involving the whole cassette of selection markers CAN1-URA3-ADE2-HIS3. In this case, red colonies (loss of ADE2) grew in all plates (loss of CAN1 and URA3, and presence of NAT an KANMX), but in SD-His/Trp (loss of the HIS3 gene).

3) Homologous recombination (Figure 40c): Mechanism of homologous recombination with a crossover between the centromere and the cassette, involving the entire R arm. Red colonies (loss of ADE2) grew in all plates (loss of CAN1 and URA3, and presence of NAT), but SD-His/Trp (loss of HIS3) and SD+G418 (loss of KANMX). The selection marker located in the telomere (KANMX) was lost, but the one located in the L arm (NAT) was maintained.

4) Loss of a chromosome and endoreduplication (Figure 40d): A total loss of the chromosome containing the cassette would cause growth of red colonies (loss of ADE2) in the plates SD+canavanine and SD+5-FoA (loss of CAN1 and URA3), but not in SD+ClonNat (loss of NAT), SD+G418 (loss of KANMX) and SD-His/Trp (loss of HIS3).

One mechanism non identifiable by the HoReYe assay, but that would generate the same pattern of growth/non-growth as a HR event, would be a physical loss of the R arm, maintaining the L arm. We believe losing an entire fragment of a chromosome shouldn't be a usual endogenous mutation event, as it would hamper the integrity if the yeast cells. However, the number of copies of the R arm could be analyzed by a simple quantitative PCR reaction of the chromosome XV, and the presence of two copies of the R arm in yeast cells suffering a HR event could be confirmed.

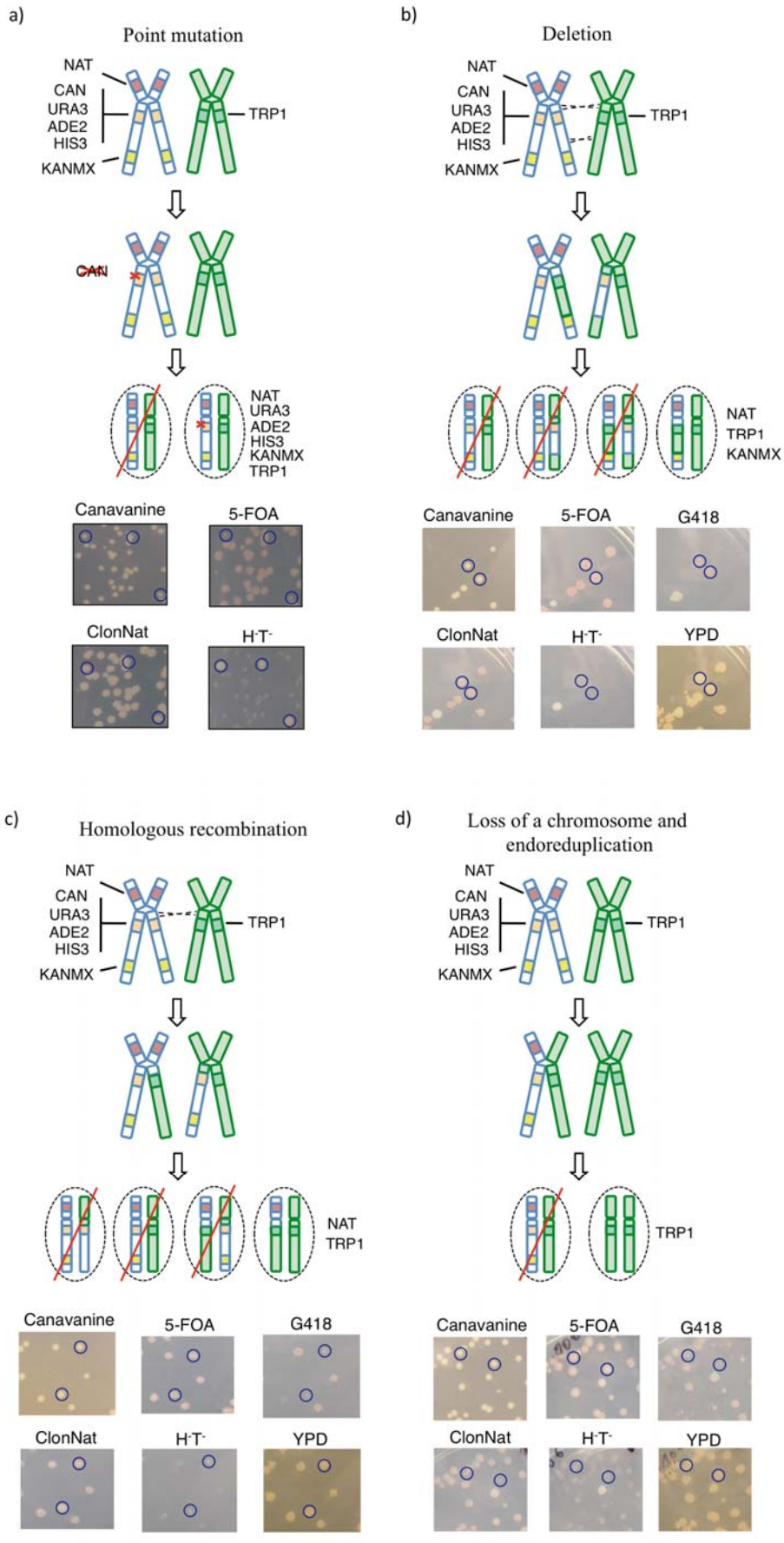


Figure 40. Mutational mechanisms considered in the HoReYe assay

Considering the different number of yeast colonies suffering different mutational mechanisms, and taking into account the number of yeast cells plated, it was possible to calculate the rates for each mechanism, and in particular the homologous recombination rate.

Mutational mechanism	SD+canavanine	SD+5-FoA	SD+ClonNat	G418	SD-His/Trp	YPD
Point mutation	X		X	X	X	White
Deletion	X	X	X	X	X	Red
Homologous recombination	X	X	X			Red
Chromosome loss and endorreduplication	X	X				Red

Table 11. Description of the selection plates where yeast cell grow (marked with x) after each mutational mechanism analyzed.

2.2.1.6 HoReYe assay protocol:

1) Generation of diploid a/α cells:

- One haploid colony a and one haploid colony α, both with the corresponding selection marker genes, are grown together in 5ml YPDA overnight at 30ºC.

- The next day a sterile toothpick is used to streak yeast in SD-His/Trp plated to isolate single colonies. Cells are grown for 2 days at 30ºC (diploid cells are bigger than haploid cells).

2) Mutational event:

- A single isolated colony is inoculated in 5ml SD-His/Trp media and grown O/N at 30ºC (To perform fluctuation analysis inoculate, for each replicate culture, a single colony, and proceed with the protocol for all the cultures in parallel).

- The next day, diploid cell culture O.D. is measured, and the culture is diluted again in SD-His/Trp to obtain an O.D. read of 0.15.

- The culture is grown at 30ºC for one doubling time, until reaching 0.3 O.Ds. At this point the culture is centrifuged and cells resuspended in the same volume of media (to

maintain the O.D.) that is switched to YPDA to get rid of the selection conditions for maintaining both XV chromosomes.

- The culture is left growing at 30ºC for one doubling time, until reaching 0.6 O.Ds. (in this doubling time the different mutational mechanisms that will be measured in the assay take place). When the 0.6 O.Ds. are reached cell growth is stopped, and cells are centrifuged and resuspended in sterile H_2O (the quantity of H_2O will depend on the O.D. reached and on the number of cells that need to be plated).

3) Yeast selection:

- 10^6 and 10^7 cells are plated in two different canavanine plates (taking into account the frequency of the mutational event to be screened, since platting should produce a density of 50-150 colonies per plate for an early counting).

- Once colonies are grown in canavanine plates, a replica-platting is performed using different selective plates: SD+5-FoA, SD+ClonNat, SD+α-aa, SD-His/Trp and YPD, always in the same order.

4- Yeast growth analysis:

- The growth of the colonies is analyzed in the different selection plates. Each mutational mechanism is indicated by the pattern of growth/non-growth under different selections.

2.2.2 *In vivo* homologous recombination rate for BMA64 strain

We performed the HoReYe assay for the yeast strain BMA64. Two replicates of the assay were performed, so a definitive HR rate estimate for this strain would require additional assays (Table 12). However, the result of the two assays were in agreement. The most frequent event generating the loss of the CAN1 gene was found to be the homologous recombination mechanism, followed by the mechanism of deletion. The two replicates differed on the third most frequent mutational mechanism, being loss of a chromosome and endoreduplication in the first one, and point mutation in the second replicate. Increasing the replicates would clarify which of the two mechanisms was more frequent. We should take into account, however, that in the present work we were not interested in the particular frequencies of the

different mutational mechanisms, but in the homologous recombination rate. In the first and second assays we obtained 69 and 52 yeast colonies bearing CAN1 inactivation due to a HR mechanism. As shown in Figure 40c, only 50% of yeast cells bearing a HR event are able to grow (due to the random disposition of chromosomes in the metaphase plate), so the total number of yeast colonies bearing a HR event should be multiplied per 2: 69·2 (138) and 52·2 (104) for assay 1 and assay 2, respectively. Considering we plated 10^6 yeast cells in both assays, the average HR rate obtained from these two assays was of $1.21 \cdot 10^{-4}$. The same assay should be performed using other yeast strains, and considering several replicates to elucidate the individual homologous recombination rates for each different strain.

	Assay 1		Assay 2	
Mutational mechanism	**Nº of colonies**	**% mechanism**	**Nº of colonies**	**% mechanism**
Point mutation	7	5.6	3	4
Deletion	43	33.8	20	26.6
Homologous recombination	69	54.4	52	69.3
Loss of a chromosome and endoreduplication	8	6.2	0	0
Total number of colonies	127		75	

Table 12. Yeast mutational mechanism assay results for the BMA64 strain.

2.2.3 Genetic dissection of homologous recombination rate in yeast

Although we only calculated the HR rate of one yeast strain, we planned on how to identify candidate genes affecting the HR rate *in vivo* by crossing two strains exhibiting the two opposite extremes of the HR rate variation identifiable. We planned to perform an X-QTL assay. The X-QTL assay [120]was designed to obtain information on the most important quantitative trait loci by performing association analysis with large pools of yeasts. X-QTL represents a powerful extension of bulk segregation analysis (SGA) [157], which was a first attempt to use large populations to dissect the genetic basis of complex traits. We planned to use this method to find allelic variants involved in the *in vivo* homologous recombination rate. This technique is based in three steps:

1- Generation of a large number of segregating populations: Two haploid parental strains with opposite phenotypes are mated and sporulated. Meiosis will create a

haploid progeny of millions of descendants containing a shuffle of parental gene combinations.

2- Selection-based phenotyping: up to 10^7 MATa individuals are plated on selection medium, and a large number of progeny is recovered (~10^3-10^4 yeast cells). This number of individuals provides the posterior association analysis with a strong statistical power.

3- Quantitative measurement of pooled allele frequencies: DNA arrays or next generation sequencing are used to obtain allele frequencies present in the cells selected by phenotype. These frequencies are compared with allele frequencies from control cells, grown in the same conditions but without any selection. The significant differences between allelic frequencies will indicate the regions containing genes involved in the trait studied.

In the original work describing X-QTL analysis, the array was performed in haploid cells to facilitate the quantification of allele frequencies. MATa cells were selected over α or a/α cells by an ingenious method (Figure 41): a construction composed by the MFA1 promoter (a promoter specific and only expressed in a cells), linked to the marker LEU2 (MFA1pr-LEU2) was inserted in the CAN1 locus of the parental MATα cell (can1::MF1pr-LEU2). MATa and α cells were mated, and diploids were sporulated and grown in medium lacking LEU2. Only MATa cells bearing MFA1pr-LEU2, transferred from the MATα cell by a recombination event in the meiosis, were able to grow. MATα, MATa/α or even original haploid cells that hadn't diploidized, weren't able to grow (MFA1pr is repressed in α or a/α cells). One complication following this approach would be if a mitotic recombination event occurred between the MAT locus and the chromosome III of both yeast strains, generating MATa/a diploids that would grow in the conditions lacking leucine, and would hamper the posterior analysis. To avoid this complication, two recessive markers were introduced in the original haploid MATα cells, can1Δ and lyp1Δ. The selection by canavanine has been explained previously, and LYP1 is a lysine permease that, when active, allows for the entrance of thialysine, a toxic that kills the cell. Using these marker genes as negative selectors would avoid the growing of MATa/a cells, as only those diploid cells suffering three independent recombination events (in MF1pr-LEU2, CAN1 and LYP1 genes), a very unlikely combination, would be able to grow. In a

medium lacking leucine, with canavanine and thialysine, only MATa cells would be able to grow.

To perform the dissection analysis of HR rate we would use the same yeast strains studied by the HoReYe assay already modified for the MATa, CAN1, and LYP1 loci (α cells in Figure 41) available upon request (Boone Lab).

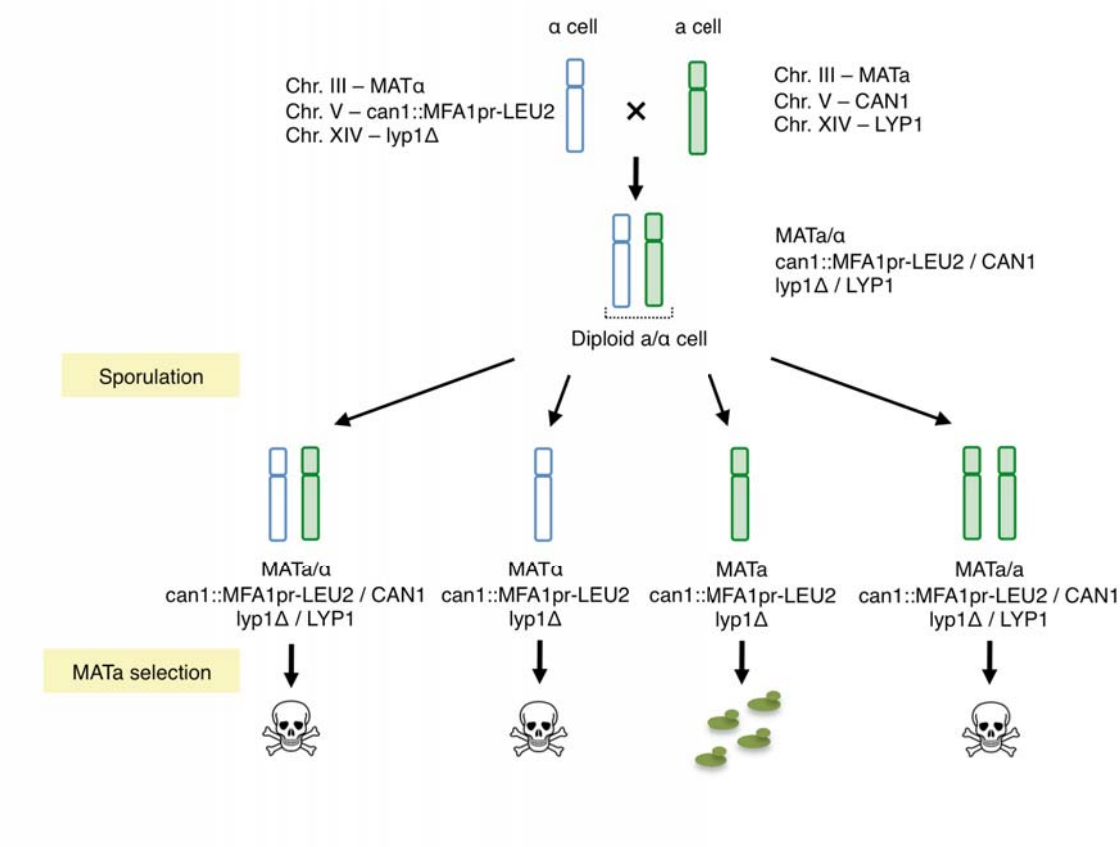


Figure 41. Selection of MATa cells over MATα, MATa/α and MATa/a yeast cells.

2.2.3.1 Adapting X-QTL to HR rate based selection

In order to perform an X-QTL assay to dissect the genetic architecture behind the *in vivo* homologous recombination rate between two yeast strains with a clear distinct phenotype, we had to adapt the original assay to the particular characteristics of the HoReYe methodology developed in this thesis. In addition, some aspects concerning the selection marker genes used in the HoReYe assay should be also modified (see below).

The first steps of the X-QTL assay would be the same: mating of two yeast strains with extreme phenotypes (one with high HR rate, and the other with low HR rate),

sporulation of the generated diploid MATa/α cells, and recovery of a MATa segregating pool (Figure 42).

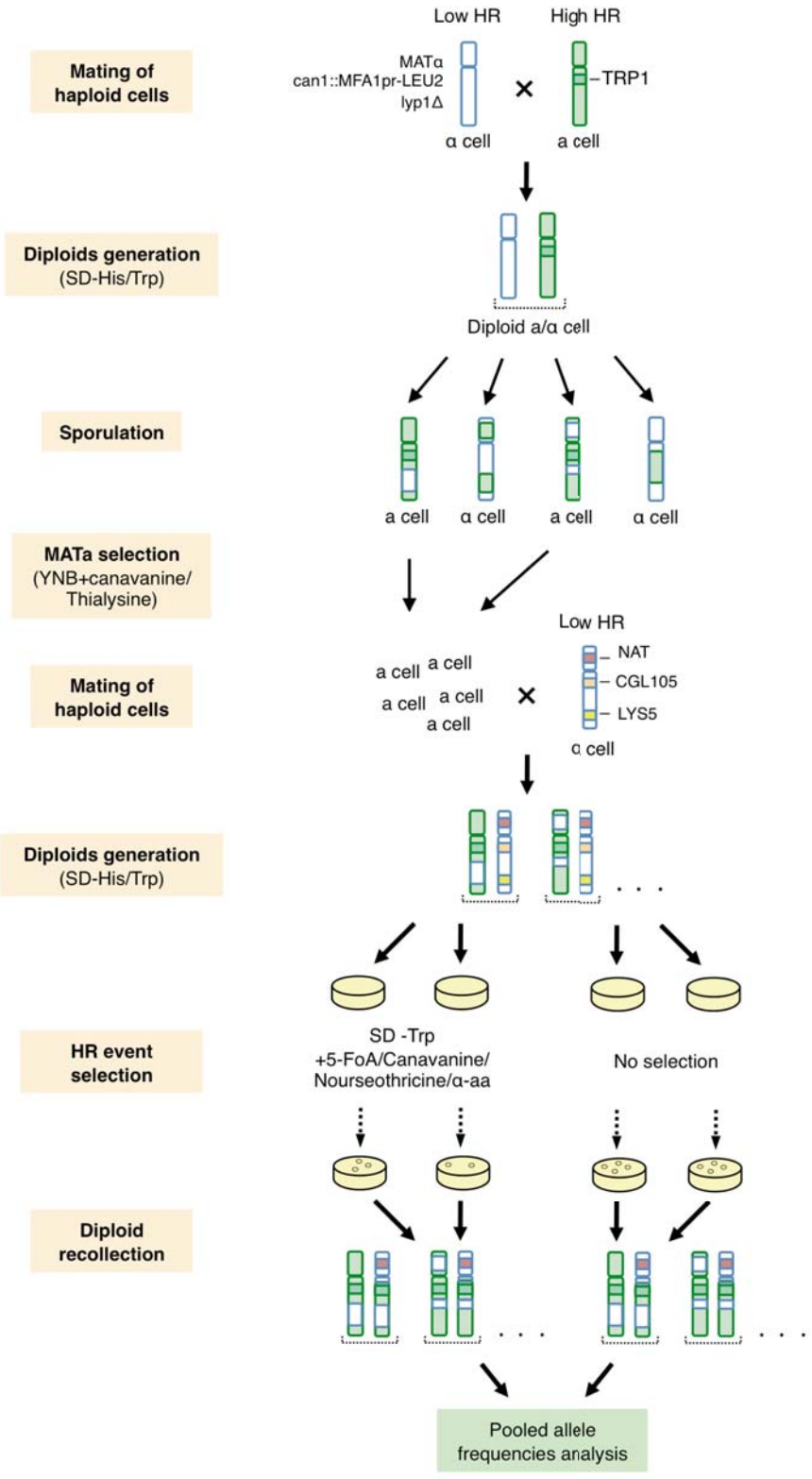


Figure 42. Modified X-QTL assay

These MATa cells contain millions of combinations of both parental genomes. In the original X-QTL assay these MATa cells were plated under selection media, and the cells grown were recovered to analyze allele frequencies responsible for the trait studied. In the modified X-QTL assay we introduced a mating step, in which the pool of MATa cells would be mated to a pool of MATα cells containing the selection marker genes used in the HoReYe assay, with some modifications (see below) (Figure 42).

The diploids obtained would be grown for one duplication-time (in which mutational mechanisms in the chromosome XV would take place, as in the HoReYe assay), and cells would be plated to select for the HR event in the XV chromosome. In addition, the same cells would be also plated without selection (Figure 42). Once cells were recovered, allele frequencies from selected and non-selected cells would be obtained by sequencing. The comparison of the allele frequencies from these two populations (selected and non-selected cells) would provide information of alleles significantly over-represented in the selected cells, which would be responsible for the trait studied, the HR rate of the yeast cells.

Unlike the HoReYe assay, here yeast cells should be grown in only one plate of selection media. Cells should be able to grow due to the loss of the selection markers located near the centromere, and at the telomere of the R arm. KANMX, located in the telomere, is a positive selector, and should be replaced by a negative one. We would use the positive/negative selection marker LYS5. LYS5, together with LYS2, are auxotrophic markers involved in lysine biosynthesis by encoding an αAA reductase. It can be used as a positive selector by culturing yeast cells without lysine, and a negative selector by adding α-aminoadipate (α-aa) to the medium. α-aa is a metabolite processed by the αAA reductase, which creates a toxic intermediate that is deleterious for the yeast cell. The mutation of both *LYS5* and *LYS2* genes prevents the formation of this toxic intermediate. In a yeast cell *LYS5* lys2, the loss of *LYS5* would let the survival of the cell in a medium with α-aa, acting as a negative selector.

2.2.3.2 Phenotype-based selection and backcrossing for HR rates

The success for obtaining significant allele frequencies in the quantitative analysis

depends on three features. First, the size of the segregating pool. The higher this pool, the more combinations of both parental genomes will be analyzed in the assay. Second, the number if individuals recovered after the phenotype-based selection, which will determine the capacity to identify significant allele frequencies in the final analysis (the number of individuals recovered depends on the number of segregating cells, together with the frequency of the event analyzed). And third, the genetic architecture of the trait studied. A complex trait affected by lots of genes with low effect will be more difficult to analyze than a trait with few genes responsible for most of the phenotypic expression for the trait.

For those cases in which the number of individuals recovered after the phenotype-based selection was not enough to obtain significant allele frequencies, a backcrossing strategy could be applied. Backcrosses would be performed between the selected offspring (those yeasts with higher HR rates), and the other parental strain expressing the lowest rate in the HoReYe assay [131] (Figure 43). After several generations of backcrosses we would obtain a hybrid population homozygous for most of the genome coming from the parent with a low HR rate according to HoReYe, while loci from the other parental strain, contributing to a high HR rate, would be retained. The enrichment of the alleles responsible for the trait of interest, in a genome mostly composed by the low HR rate strain genome, would decrease the number of individuals needed to obtain significant values in the measurement of allele frequencies [158]. When performing the backcrossing approach, regions coming from the yeast strain with high HR rate would be the candidate regions to contain genes involved in this trait. Therefore, direct sequencing of these cells would elucidate the identification of the HR rate candidate genes.

In this case, gene marker modifications should also be performed from the original HoReYe assay construction. As explained in the modified X-QTL assay, cells should be able to grow due to the loss of the cassette construct and KANMX in the R arm, and the presence of NAT in the L arm. However, the MATa cells generated from the recovered diploids should loss also the NAT gene present in the L arm, as these cells will be mated with α cells containing this gene. Therefore, we should replace NAT gene by a positive/negative selection marker, and KANMX by a negative selector. NAT could be replaced by LYS5, which, as explained before, is a positive/negative selector. KANMX could be replaced by Gal4pΔ. GAL4 is expressed in the presence

of galactose and acts by binding to DNA promoter sites of GAL genes, activating them. Gal4pΔ is a truncated protein of Gal4p that is able to bind the GAL DNA promoter sites, but it is not able to activate them. If Gal4pΔ is over-expressed (placed under the control of a constitutive strong promoter), it competes with the endogenous Gal4p, inhibiting cell growth when galactose is used as a carbon source. Therefore, yeast cells grown in galactose would be able to grow only if they lost the inserted Gal4pΔ gene.

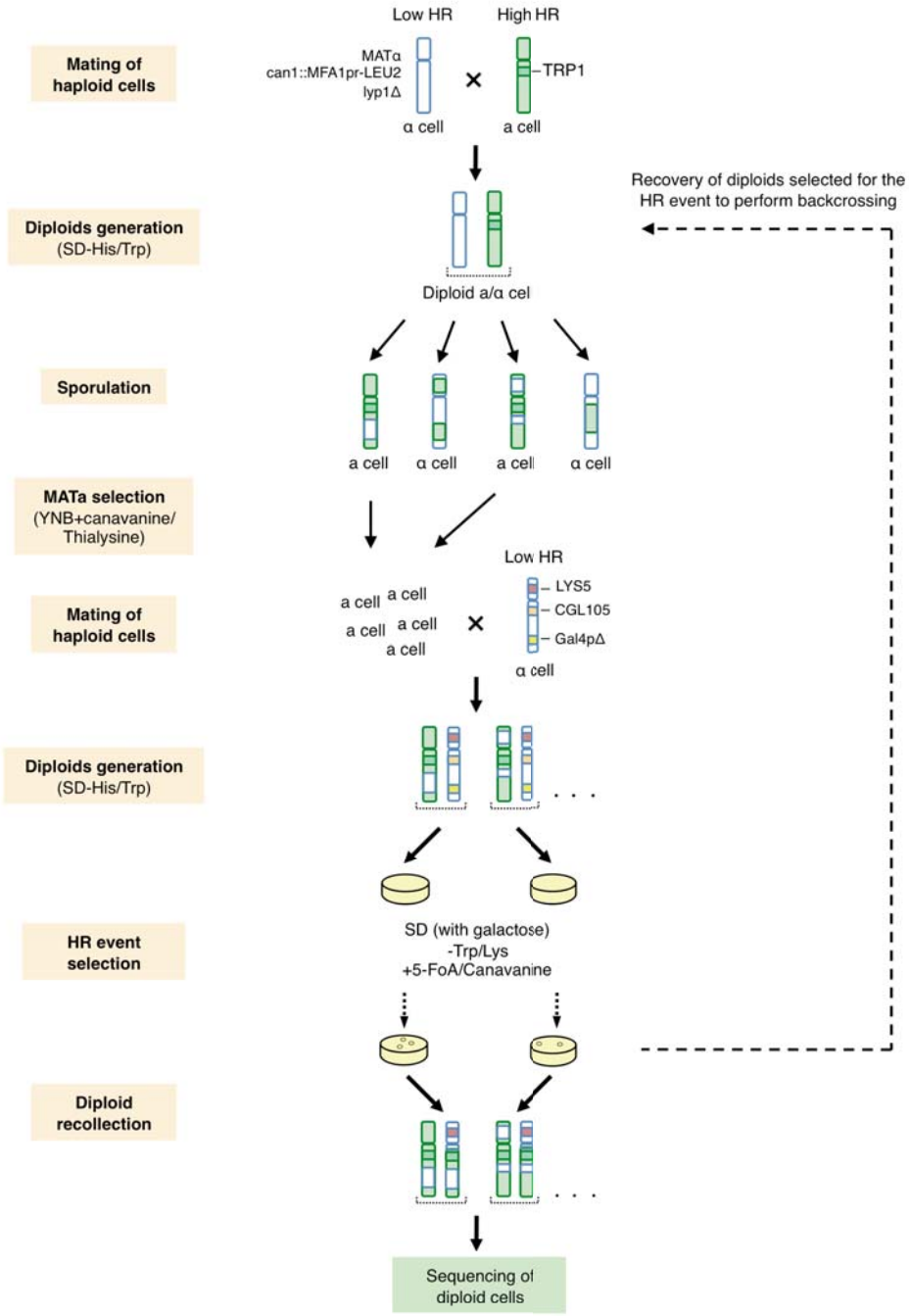


Figure 43. Phenotype-based selection and backcrossing

2.2.3.3 Surrogate assay for determining HR rate in yeast

The HoReYe assay allows for the analysis of a HR event similar to the mechanism generating LOH in dermal neurofibromas. However, its setting up is time-consuming, and multiple modifications need to be performed to the yeast strains to do the analyses described. To facilitate the analysis of multiple yeast strains in the future, a surrogate to determine the HR rate could be used. As described in the HoReYe assay, the best way to select for cells suffering a HR event is by analyzing the loss of specific selection markers. URA3 and CAN1 genes are located in the L arm of chromosome V. As described before, both genes can be negatively selected (only yeast cells losing the wild-type URA3 and CAN1 genes will grow in media containing 5-FoA and canavanine). Therefore, a haploid yeast cell bearing wild-type URA3 and CAN1 genes could be mated with another haploid yeast cell bearing both genes mutated. If a HR event occurred, segregants losing wild-type genes could be selected by growing cells with 5-FoA (selection against URA3) and canavanine (selection against CAN1) (Figure 44).

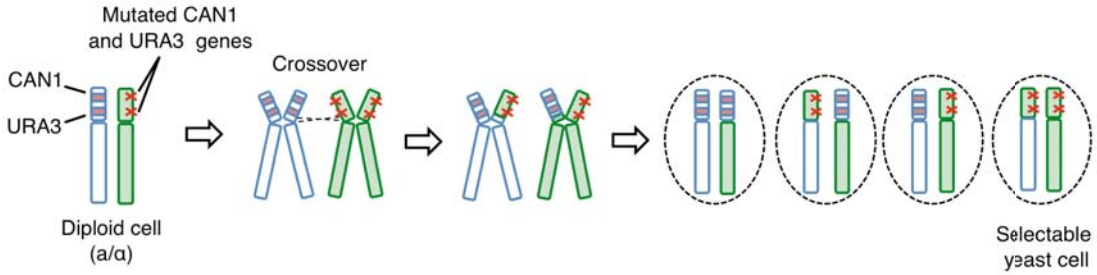


Figure 44. Surrogate for determining the HR rate

This assay should be performed for the strains already characterized using the HoReYe assay. If the HR rate obtained by both techniques were similar, we could conclude that the analysis of CAN1/URA3 loss is a good surrogate to determine the HR rate, and could be used in the analysis of future yeast strains.

Analysis of candidate genes influencing neurofibroma number

This third part summarizes the compilation of genetic variation data in candidate genes. First, Sanger sequencing was used to analyze DNA variation on candidate genes selected from the literature. Then, NGS was used to analyze a targeted exome of all hereditary cancer genes of some NF1 patients, capturing allelic variants, both common and rare.

PART 3

3. Analysis of candidate genes influencing neurofibroma number

We first planned to analyze DNA variation from three candidate genes (*BLM, NBS1* and *RAD51A)* by Sanger sequencing to obtain both common and rare variants. Then, due to capacity and cost-effectiveness, we extended the analysis to other genes by using NGS. We used NGS to analyze a targeted exome containing about 100 hereditary cancer genes for different NF1 patients.

3.1 Sanger sequencing of *BLM* gene

Among the genes accomplishing the literature-based criteria, we started sequencing the *BLM* cDNA for 12 of the 18 NF1 patients (listed in Table 4). mRNA was extracted from skin-derived primary fibroblast cell cultures obtained from each patient and treated with puromycin. Puromycin inhibits the non-sense mediated decay (NMD) mechanism that degrades mRNAs containing non-sense mutations. This guaranteed the detection of mutations affecting mRNA stability. An RT-PCR using random primers was performed, and the corresponding cDNA was amplified in 15 fragments by PCR and sequenced (see M&M section). Only variants found in both forward and reverse sequencing reactions were considered as true changes. We also sequenced 1Kb upstream the first translated ATG codon, and 1 Kb downstream of the termination codon, to gain information on the promoter, the 5' and 3'UTR regions. From this analysis we found 5 different variants in some of the NF1 patients analyzed, all of which were described as SNPs in the NCBI dbSNP database (Table 13):

1) Variant 1078 A/C (cDNA fragment 3): Corresponds to SNP rs1140119. It produces a synonymous change ACC → ACA, Thr → Thr. Non frequency data is available.

2) Variant 3199 A/G (cDNA fragment 8): Corresponds to SNP rs2227933. It produces a synonymous change ACG → ACA, Thr → Thr. Its frequency in the European population is 0.261 A / 0.739 G.

3) Variant 3628 C/A (cDNA fragment 9): Corresponds to SNP rs2227934. It produces a synonymous change GCC → GCA, Ala → Ala. Its frequency in the European

population is 0.261 A / 0.739 C.

4) Variant 4042 T/C (cDNA fragment 10): Corresponds to SNP rs1063147. It produces a synonymous change CTC → CTA, Leu → Leu. Its frequency in the European population is 0.261 T / 0.739 C.

5) Variant 4058 G/A (cDNA fragment 10): Corresponds to SNP rs7167216. It produces a missense mutation GTA → ATA, Val → Ile. Its frequency in the European population is 0.025 A / 0.975 G.

Patient	5'UTR-1	5'UTR-2	1	2	3	4	5	6
P011	-	-	-	-	1078 A>C	-	-	-
P022	-	-	-	-	1078 A>C	-	-	-
P023	-	-	-	-	1078 A>C	-	-	-
P027	-	-	-	-	1078 A>C	-	-	-
P052	-	-	-	-	1078 A>C	-	-	-
P055	-	-	-	-	1078 A>C	-	-	-
P062	-	-	-	-	1078 A>C	-	-	-
P078	-	-	-	-	1078 A>C	-	-	-
P082	-	-	-	-	1078 A>C	-	-	-
P084	-	-	-	-	1078 A>C	-	-	-
P103	-	-	-	-	1078 A>C	-	-	-
P104	-	-	-	-	1078 A>C	-	-	-

Patient	7	8	9	10	11	12	3'UTR
P011	-	3199 A/G	3628 C/A	4042 T/C	-	-	-
P022	-	-	-	4042 T/C	-	-	-
P023	-	-	3628 C/A	4042 T/C	-	-	-
P027	-	-	-	4042 T/C	-	-	-
P052	-	-	-	4042 T>C 4058 G/A	-	-	-
P055	-	-	-	4042 T>C	-	-	-
P062	-	3199 A/G	3628 C/A	4042 T/C	-	-	-
P078	-	-	-	4042 T>C	-	-	-
P082	-	-	-	4042 T>C	-	-	-
P084	-	3199 G>A	3628 C>A	-	-	-	-
P103	-	-	-	-	-	-	-
P104	-	-	-	4042 T>C	-	-	-

Table 13. Description of *BLM* cDNA fragments analyzed. (-) indicates no nucleotide change. Nucleotides showing heterozygosity are separated by a slash.

None of these nucleotide changes were near any intron-exon boundary. The only non-synonymous change affected patient P052, with a Valine → Isoleucine modification. Both Valine and Isoleucine are hydrophobic branched-chain amino acids with a similar structure, and can substitute each other with no biological consequences. Therefore, we believe this amino acid change shouldn't entail any protein function modification. All other modifications are synonymous, not affecting the protein sequence.

3.2 Next-generation sequencing of hereditary cancer sub-exome

We performed a targeted exome analysis of a set of genes involved in cancer syndromes. In total, 106 genes were analyzed, 39 involved in DNA repair pathways and 9 involved in the mechanism of HR.

3.2.1 Next-generation sequencing (NGS) methodology

Briefly, NGS first step consisted in creating a library of DNA target regions. It was decided to prepare a custom library, as commercial ones provided a low coverage in 20% of the genes (regions within genes) of interest. The DNA probes that formed the DNA library were obtained using eArray from Agilent. The second step was, using the designed library (Agilent SureSelect), to capture the targeted regions of whole DNA obtained from patients. Biotynilated RNA library baits were used to hybridize the DNA targeted sequences. Then, magnetic beads were used to capture only hybridized DNA sequences that, after performing an amplification, were used in DNA sequencing. The platform employed was IonTorrent Chip 316, which is able to generate 300Mb per run. Library preparation, target capture and sequencing was performed at the genomics facility of IMPPC.

The data obtained was analyzed by CLC Bio software. After a sequence curation (eliminating duplicates and trimming the sequences to eliminate low quality reads) an algorithm to map all the sequences was used. Then, variants were identified by the CLC Bio Probabilistic Variant Determination, which evaluated quality parameters and the probability of a nucleotide being a true variant with a defined confidence interval, in our case 95%. CLC Bio provided a list with all the variants obtained, their

localization in annotated genes, and information on the amino acid changes with potential pathogenicity.

3.2.2 NGS data analysis

NGS analysis was intended to be performed with the 18 NF1 patients characterized at clinical and tumor molecular level. At the present time, NGS was performed only for patient P027. This data, however, was received in the finalization of this thesis, and no deep analysis of the results could be performed. As a general overview of the results obtained, a total of 845 variants were identified. From these, 63 were deleterious changes, from which 48 were missense, and 15 truncated the protein.

However, we analyzed, as an illustrative example, the *NF1* germline mutation of patient P027. NGS analysis allowed us to identify a G insertion affecting aa 2088 (Figure 45). This germline mutation had been previously characterized by other means.

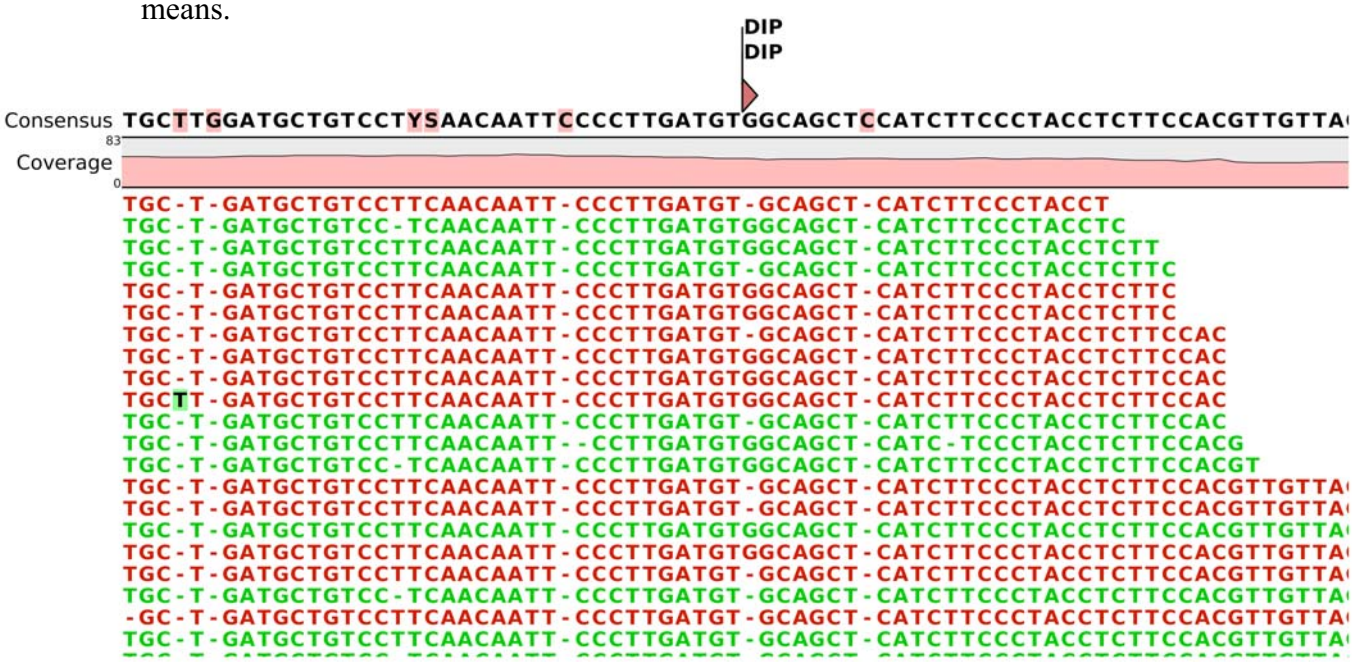


Figure 45. Identification of the *NF1* germline mutation of patient P027. DIP: deletion/insertion polymorphism.

DISCUSSION

The aim of this thesis was to identify genes involved in the variation of the number of dermal neurofibromas developed in NF1 patients. As stated in the premises of this work, we reasoned that genes involved in the somatic mutation rate of NF1 patients might be partially responsible for this number. We decided to decompose the phenotype studied to facilitate the complexity of the analysis, and focused on Schwann cells, those bearing the double *NF1* inactivation, and on genes related to the homologous recombination (HR), responsible for a high percentage of *NF1* somatic inactivation (a key event in dNF development). We divided the study in three parts. First, the phenotypic characterization of NF1 patients at both clinical and tumor molecular level. Second, the selection of candidate genes involved in the HR rate by two different approximations, one based on published literature, and the other by modeling HR mechanism in yeast. Third, the analysis of these candidate genes by direct sequencing to collect all genetic variation that might be responsible for the variation in the number of dNFs developed.

Clinical and molecular phenotype characterization of NF1 patients

Clinical phenotype: determining neurofibroma number

An adequate clinical characterization of NF1 patients was necessary to perform any present or future genotype-phenotype correlation study regarding the number of dermal neurofibromas developed. Determining the precise number of dNFs by NF1 individuals can be a difficult task, as patients can develop thousands of these benign tumors. In conjunction with the NF1 medical personnel, we implemented a standardized methodology for estimating the number of neurofibromas developed. This methodology should be important to accurately count the number of dNFs developed, and be able to compare among patients. In addition, since the number of dNFs developed is highly dependent on the age of the patient, we developed a method to calculate a value, the Risk of Dermal Tumor Burden (RDTB), to correct for age biases. RDTB is indicative of the rate of appearance of dNFs in a patient regardless patient age.

In addition to the physical and psychological burden that dNFs cause to NF1 patients,

is there any described correlation between the number of dNFs developed and the severity of the disease, like the development of malignant peripheral nerve sheath tumors (MPNSTs)? A recent study showed that there is no correlation between dNFs and other severe manifestations [159]. However, a correlation was found between subcutaneous neurofibromas and high internal neurofibroma volume [159]. In addition, a previous study showed that patients with subcutaneous neurofibromas were three times more likely to develop plexiform neurofibromas or MPNSTs [160]. Therefore, a study centered in the detection of the genetic variation causing variation in subcutaneous neurofibroma number would be relevant for the management of NF1 patients.

Molecular phenotype: characterizing the LOH-mechanisms generating NF1 second-hits in dNFs

We hypothesized that genetic variation in genes related to repair mechanisms might be involved in the variation of dNF number in NF1 patients, as somatic inactivation of the NF1 gene is a necessary and limiting step in dNF development. Therefore, we characterized and studied the importance of the different mechanisms generating LOH in these tumors, with particular interest in HR. We first determined that at least the molecular analysis of 10 dNFs per patient was required to obtain LOH prevalence with a precision of 20%. However, when possible, we increased the number of dNFs analyzed to at least 15 to improve this precision.

At a molecular level we identified the mechanisms generating LOH in dNFs of NF1 patients. The presence/absence of LOHwas analyzed, together with the assessment of the copy-number of the *NF1* gene, and a general picture of the mutational mechanisms generating the LOHs was obtained. We analyzed 606 dNFs from 117 NF1 patients. From these, 144 (23.7%) showed LOH, in agreement with previous studies [138,161].

In 53 dNFs (37% of tumors exhibiting LOHs), LOH was generated by a copy-loss, with almost all deletions ranging from 80Kb to 8Mb. Importantly, while performing the whole dNF analysis we identified one neurofibroma (P082-1N) bearing a type-2 NF1 deletion, caused by a non-allelic homologous recombination (NAHR) between the SUZ12 gene and its pseudogene, with the breakpoint located in the SUZ12 intron

4. This mechanism has been identified to be responsible for *NF1* first-hit mutations in patients with somatic mosaicism, but it had never been found to cause *NF1* second-hit mutations in dNFs. However, not all type 2 microdeletions can be PCR amplified by a primer set for common breakpoints, and thus we cannot rule out the involvement of NF1-REPs or the SUZ12 sequences in other tumor deletions, like the one identified in tumor P004-7N showing MLPA results consistent with this type of deletion.

As a whole, the mechanism of homologous recombination was responsible for LOHs in 91 dNFs (63% of tumors exhibiting LOH). All HR events were generated by a single crossover between the centromere and the *NF1* gene, and reducing the constitutional *NF1* mutation to homozygosity. LOH extended until the end of the 17th chromosome, as shown in all dNFs caused by mitotic recombination and analyzed by SNP-arrays, in which 17q isodisomy extended to the 17q telomere. For five dNFs (three from the same patient) there were no informative 17p microsatellite markers raising the possibility of a loss of the complete chromosome 17 with an endoreduplication of the remaining chromosome. However, we believe this is not a very plausible possibility because this type of event has only been reported in the literature once for a dNF [138], and the loss of a whole chromosome would imply a degree of genomic instability not observed in dermal and plexiform neurofibroma genomic studies.

Our analysis shows that HR is the most important mechanism generating LOH in dNFs, which accounts for ~15% of somatic inactivations of the *NF1* gene. Deletions account for ~9% of these somatic mutations. The remaining ~76% might be generated by point mutations (framsehift, nonsense...), or even small deletions, as our approach identified DNA alterations affecting only microsatellite locations, so small deletions occurring between these markers might be underestimated in our study.

Genome integrity in dNFs

There have been different attempts to characterize the whole genome of dNFs in order to search for additional genetic alterations that could help to find other genes implicated in dNF development. However, although there is some disparity in different studies, reports using different methodologies, such us comparative genomic hybridization (CGH) [162], cytogenetic karyotyping of dermal neurofibroma-derived

Schwann cells [163], cytogenetic karyotyping of short term neurofibroma cultures [164], array-CGH [165,166], Eric Legius [167], and SNP-array analyses (Nancy Ratner and the NF1 Microarray Consortium, unpublished results), indicated a paucity of recurrent gross alterations in dNFs. SNP-array analysis of 19 dNFs allowed us to have a better idea of the global genome integrity of dNFs beyond the NF1 region. We identified two dNFs with large deletions in two distinct chromosomal regions, one in 2q24.2–q31.1 (observed in tumor P090-3N) and the other in 3q11.2–q22.1 (in tumor P072-1N). These findings are consistent with the most recent findings obtained using array-CGH, where no gross alterations or just a few and sporadic copy number alterations were identified in dermal neurofibromas [165-167]. The genomic alterations found in the present work appeared to be random events, as deduced from their low frequency and the lack of recurrence among different dNFs. It is unlikely that these somatic events are necessary for neurofibroma growth. Thus, the majority of discrete dermal neurofibromas do not frequently exhibit additional gross alterations in other genomic regions except for 17q. However, it remains to be investigated whether mutations such as small deletions, small insertions, and point mutations in other genes or regulatory regions exist that would contribute to dNF growth. Nevertheless, the lack of any recurrent gross alteration in the genome of dNFs favors the idea that, genetically, NF1 inactivation in neurofibroma-initiating cells (precursor cells that have the capacity to differentiate into Schwann cells) might be sufficient for dermal neurofibromas to develop (in an NF1 (+/-) background). Recent results in a dermal neurofibroma mice model favors this view [77].

Individual frequencies of dNF LOH and dNF HR-mediated LOH

In our study the individual frequencies of dNF LOH were possible to be estimated for a reduced number of NF1 patients (n=18), so any suggestion or conclusion has to be taken as preliminary. Individual LOH frequencies in dNFs show a high variability, with patients showing from 0 to 55.5% of LOH. In addition, many patients show a major LOH-causing mechanism in their dNFs, being HR the most frequent.

Why patients show this degree of LOH variability? LOH results from errors non-correctly repaired by DNA repair mechanisms, in the case of HR, errors like DSBs. DSBs can be repaired by three different mechanisms, Non-Homologous End Joining

(NHEJ), Single Strand Annealing (SSA) or HR. These mechanisms might be in some sort of equilibrium to repair the DSBs. So, a worst performance of one of these mechanisms might promote the activation of the others. Then, the dNF LOH variability shown among NF1 patients might be indicative of an overall better or worse performance of these repair mechanisms, also in relation to how the other repair mechanisms are functioning in each individual. That is, there could be different biological sources of variation on the resulting importance and performance of the HR mechanism in each individual in relation to the formation of LOH in dNFs.

The correlation study performed in just 16 NF1 patients with at least 10 dNFs analyzed (excluding patient P027) suggested that patients with the highest rates of HR frequency in dNFs showed the highest number of dNFs developed (with a p value close to significance). Again, our analysis was performed with few NF1 patients, but despite being necessary the study of a higher number of NF1 individuals, we think this could be indicative of a relationship between both parameters. Does having a higher frequency of HR means a better or a worse performance of this repair mechanism? In this case two scenarios can be proposed. Either NHEJ and SSA perform worst, in relative terms, being the DSBs mostly repaired by HR. Another possibility is that the mechanism of HR is in some way altered, repairing DSBs that should be repaired by the other mechanisms. There is no reason to think that different repair mechanisms might operate in a different way in Schwann cells than in other cell types. Therefore, it could be proposed that the HR frequency generating LOH in dNFs of NF1 patients could be related to the constitutional HR rate of the individual. If, as suggested by our results, patients with the highest HR rates were developing higher number of dNFs, then the number of dNFs developed at a certain age could be indicative of the HR rate of the NF1 individual. One way to confirm this hypothesis could be to measure *in vivo* constitutional HR rates of NF1 patients, and check whether or not patients showing higher frequencies of HR-mediated LOH in dNFs are also those with the highest constitutional HR rates. Different techniques have been used to evaluate the somatic mutation rate of individuals, analyzing the function of individual gene products, like the first studies of radiation exposure analyzing Glycophorin A (GPA), an erythrocyte membrane protein [168]; the analysis of Hypoxanthine phosphorybosyltransferase (HPRT), a T cell membrane protein [169]; or the study of PIG-A protein, linked to hematopoietic stem cells [170]. During the course

of this thesis we tried to evaluate the *in vivo* somatic mutation rate of some NF1 patients using the GPA assay [171] to evaluate the display of this protein on the surface of erythrocytes of patients heterozygous for the *GPA* gene. We chose this assay because it provided information not only on the somatic mutation rate of the individual, but also on the mutational mechanisms generating it, as GPA assay could evaluate both the presence/absence of each GPA allelic form and the relative quantity of the GPA protein. However, due to methodological problems, no results could be obtained. We will try to gain this information in the near future by repeating GPA assay, or by performing some of the other techniques available. If the *in vivo* constitutional HR rate of NF1 patients perfectly correlates with the frequencies of HR-mediated LOH in dNFs of the same individuals, and a significant correlation is also identified for high HR-mediated LOH and high number of dNFs developed, then the number of dNFs developed could be indicative of their individual somatic mutation rate.

Making the molecular characterization of dNFs from NF1 patients feasible

The determination of mutational mechanisms responsible for LOH in dNFs required LOH analysis and copy-number assessment of the *NF1* gene and regions surrounding it. Today, a combination of at least two techniques, or the use of techniques like SNP-array, are required. To perform our study it was necessary to analyze a high number of NF1 patients and dNFs to obtain an accurate estimation of LOH frequencies, and the use of two different techniques was expensive and time-consuming. However, we realized that microsatellite analysis could provide all the necessary information regarding LOH and locus copy-number assessment in dNFs. We developed a new methodological approach, called Microsatellite Multiplex PCR Analysis (MMPA) that improved and facilitated neurofibroma analysis. MMPA was based on the comparison of microsatellites from control-tumor sample pairs amplified in one PCR run. With this technique it was possible to obtain: data regarding the tumor sample allelic imbalance (AI) status and extension, the percentage of normal cells present in the tumor sample, the copy-number status of specific alleles of heterozygous loci showing AI and the mechanisms generating these AIs.

All information that can be obtained by applying an MMPA is based on the use of a constant of proportionality (Kμ) between a multiplex PCR reaction of a tumor DNA

and its matching control pair. Thus, one of the key factors when using MMPA calculations is to design and set up a robust multiplex reaction controlling all aspects that affect Kμ to minimize variation in its value and in that way minimize the errors in the estimation of expected peak heights. DNA quality is also an important aspect for the correct performance of MMPA, as it could affect DNA amplification and increase Kμ dispersion. Bad quality DNA can make MMPA calculations impossible. In addition, it is important to precisely quantify and always analyze the same amount of DNA for all samples. K values from microsatellites with no AI considerably different from 1 could indicate genetic alterations causing copy-number gains or losses without generating allelic imbalances, as happens in cases where the two copies of a given locus were lost (e.g. nullosomy) or gained (e.g. tetrasomy).

Since Kμ is calculated from control microsatellites, it is desirable to have at least three informative control microsatellites in the MMPA. Previous knowledge about the degree of genetic instability and the recurrence of altered and non-altered genomic regions would facilitate the inclusion of pre-established control markers in the MMPA design. This should not be a problem when analyzing specific genomic regions in tumors with a low or intermediate degree of recurrent copy-number alterations. However, in tumors with a highly altered genome, it will not be possible to determine upfront which markers could be used as control microsatellites, and only after doing MMPA calculations, those not exhibiting AI (and Ks close to 1) could be used as control markers. In addition, during the MMPA analysis, special care should be paid to heterozygous dinucleotide microsatellites when both alleles only differ in one repeat (2bp). In these cases, the slippage of the DNA polymerase during PCR amplification can hamper the correct analysis of peak height values of smaller alleles, thus affecting the calculations of copy-number and percentage of AI-cells in markers showing AI. In the MMPA setup presented here, we decided not to use heterozygous microsatellite markers with a difference in length of only 2bp for the MMPA calculations. However in the case that these microsatellites were required, a correction on the peak height values of the smaller alleles could be applied.

The interpretation of the information produced by an MMPA is based on a simple model in which tumors are composed of two cell populations: normal 2n cells and cells containing allelic imbalances. A limitation of the MMPA in relation to the ascertainment of the percentage of normal cells present within a tumor is that this

information can only be obtained from microsatellite markers showing AI due to copy-number loss or copy-neutral events. Then, this information can be used to analyze microsatellite markers showing AI due to copy-number gains present in the same tumor, but again, only under the assumption that all genetic alterations causing AI are present in the same population of cells. In tumors with a higher cellular and genetic heterogeneity not fitting with this simple model, AI could be detected but difficult to interpret in the context of copy-number and percentage of AI-cells.

The re-analysis of 29 neurofibromas showed a good agreement between the information generated by MMPA (copy-number status; percentage of AI/non-AI cells in the tumor) and the data generated for the same tumors by MLPA, PRA and SNP-array analysis. MMPA showed a high degree of sensitivity and accuracy. For the study of neurofibromas and the MMPA setup described here, most microsatellites (82%) showing AI were correctly determining the copy-number status and the AI-mechanism. 15% of microsatellites generated out of range values and thus, were finally not used in the analysis. The number of microsatellites indicating an incorrect AI-mechanism was very low (2%). Considering these data, using a criterion that allows the discrepancy of 1 AI-microsatellite, for neurofibromas with at least 3 informative AI-microsatellites (27/29) the MMPA analysis correctly determined the AI-mechanism in all of them. In addition, the percentage of non-AI cells determined by MMPA didn't differ more than 8% for all samples but one when compared with the results obtained by SNP-array. MMPA can be implemented as a common laboratory technique to analyze, in an easier and cheaper way, dermal neurofibromas obtained from NF1 patients. In addition, it can be adapted to any microsatellite multiplex PCR already developed to analyze any other disease conferring changes in the AI status of microsatellites.

Selection of candidate genes influencing neurofibroma number

The second part of the work consisted in identifying candidate genes involved in the HR rate as possible modifiers for the number of dermal neurofibromas. Techniques like GWAS have been widely used to identify genes responsible for complex traits. However, these studies have not been as successful as expected because, for most analyses, only variants accounting for a small fraction of the variation on the

phenotype studied have been identified. In addition, GWAS-selected genes have to be evaluated at a functional level, which can be a complicated task. We were aware of our limitations in the possibilities of characterizing the clinical and molecular phenotype of a large number of NF1 patients. That is why we though in two alternative strategies to obtain genes involved in the variation of dNFs: a literature-based candidate gene selection and the development of a strategy using yeast as a model organism.

Modeling HR in yeast

As the main hypothesis of the present work was that patients with a high HR rate would generate a high number of dNFs, we used the yeast model as a second strategy to identify genes influencing the *in vivo* HR rate. We thought the yeast modeling approximation was advantageous over the direct analysis of candidate genes, the literature-based approximation, for different reasons: first, the genetic architecture behind the HR rate was not known. Genes involved in the mechanism of HR itself might not be the only ones affecting its rate, or the most important in generating variation in HR rate among humans. Therefore, the candidate genes obtained through the yeast model could provide unbiased data not dependent on previous information. Second, yeast allowed us to functionally evaluate the effect of the variants in the HR rate by, for instance, swapping alleles considered to contribute to high or low HR rate among different yeast strains.

Why did we choose specifically the yeast model to analyze the trait of HR rate? HR has been widely studied in yeasts, and a high conservation between both yeast and human genes implicated in the HR mechanism has been found. Examples are the high number of yeast orthologs present in human HR machinery. However, little was known about the genes important for modifying the rate of HR, those involved in an increased frequency of this mechanism over other repair mechanisms. In our analysis, we were interested in those genes and genetic variation implicated in producing high rates of HR in NF1 patients.

We created the HoReYe assay to analyze the *in vivo* HR rate in yeast, in a process identical to the one generating HR-mediated LOH in dNFs, that is, one cross-over

between the centromere and a gene of interest generating LOH until the chromosome telomere. We performed the HoReYe assay, in duplicate, with the BMA64 yeast strain and obtained, on average, a HR rate of $1.21 \cdot 10^{-4}$ events/cell/generation, while other studies obtained an average HR rate of 1.2 to $1.4 \cdot 10^{-5}$ events/cell/division [172,173]. Our results should be taken as preliminary for different reasons. First because, as explained in the results section, early mutations in yeast cultures can bias the results obtained, and to overcome this issue a fluctuation analysis should be performed. Therefore, the HoReYe assay should be repeated several times to obtain reliable HR rate results. And second, because mutation rate can vary across the yeast genome up to 6-fold [174], as the distance between the selection marker and the centromere, together with the methodology used, affects the rate of HR obtained. It should be taken into account, however, that the aim of the HoReYe assay was to identify yeast strains with high and low HR rates, rather than measuring exact yeast HR rates. In addition, HoReYe assay could also be used to study the variation in genes modifying the rate of other repair mechanisms responsible for related to point mutations, deletions, or loss of a chromosome and endoreduplication.

HoReYe assay should be performed with different yeast strains containing different genetic backgrounds in order to find strains with high and low HR rates, showing phenotypes as different as possible. The more yeast strains characterized, the more genetic variation could be analyzed for the trait studied. However, the setting up of each strain was time-consuming, as it required numerous genomic modifications and rounds of yeast selection. In consequence, we proposed to find an alternative methodology, similar to other assays already used to determine yeast HR rates [172,173] that perfectly correlated with the HoReYe assay for a group of several yeast strains. This assay would simply analyze the loss of endogenous selection markers, and act as a surrogate of the HoReYe assay, allowing the easy characterization of a great number of other yeast strains. We proposed the analysis of the loss of two endogenous genes, URA3 and CAN1, both closely located in the L arm of chromosome V. In a diploid yeast cell URA3/ura3 CAN1/can1, the negative selection of both genes would indicate the loss of the wild-type selection markers.

Once HR rates of different yeast strains were correctly characterized, we would use a modified X-QTL assay to determine genetic variation responsible for high or low HR rates. We think that, today, X-QTL is one of the most powerful approximations to

identify genetic variation involved in complex traits, as demonstrated by the high percentage of the heritability of certain traits explained using this technique [120,175]. The success of X-QTL methodology would lie on the correct characterization of yeast strains by HoReYe assay, and on the analysis of strains showing extreme phenotypes (high and low HR rates in our case).

Analysis of candidate genes influencing neurofibroma number

In addition to the HR yeast modeling approximation, we analyzed human genes based on a literature-based selection. The criteria used in the literature-based selection followed the rationale exposed in the premises of this work. Genes involved in the mechanism of HR, related to cancer syndromes, and with corresponding ortholog in yeast, were selected. In addition, we took advantage of a setting up developed by the genetic testing unit of our laboratory to analyze, by Next-generation sequencing (NGS), a targeted exome of 106 genes related to cancer syndromes, 39 involved in DNA repair pathways, and 9 known to be related to the mechanism of HR.

Sanger sequencing of *BLM* gene

We analyzed cDNA of *BLM* gene for 12 NF1 patients (Table 4) by Sanger sequencing. 5 variants were found, four of which synonymous, and one (variant 4058 G/A) non-synonymous, generating a missense mutation. This variant, however, produced a Valine-Isoleucine change, not predicted to be affecting the general structure of the protein (both are hydrophobic amino acids, which are similar in their molecular structure). For this sequencing approach we had to start from mRNA material in order to sequence the complete coding region of the gene. For that, primary fibroblast cultures had to be performed. Altogether made this analysis time-consuming and expensive. These limitations, together with the advent of NGS technologies, made us to get over Sanger sequencing technology, and apply a target sub-exome strategy. This new approximation allowed the analysis of more than one hundred gens for each NF1 patient directly from DNA, reducing costs and producing much more information.

Next-generation sequencing of hereditary cancer sub-exome

The NGS setting up was developed at the end of this thesis, and only DNA from patient P027 was sequenced: 845 variants were identified. 63 were deleterious, from which 48 were missense, and 15 truncated the protein. No further analysis could be performed for these variants. However, as a proof of concept, we correctly identified P027 *NF1* germline mutation, previously determined by other means. NGS should be performed to the remaining 17 NF1 patients characterized.

Concluding remarks

We were aware the overall work planned to be developed in the present thesis was an ambitious task. It included human molecular genetics, tumor analysis, yeast genetics, analysis of complex traits, etc. The final goal of the thesis was to identify genes and genetic variants important for modifying the number of dNFs developed by NF1 patients. To rich it, different experimental procedures are still required. However, this work has either set up or defined all the different components of the global experimental design to achieve this goal, putting the conceptual and experimental basis to complete the whole project in the near future: it has established the basis to accurately estimate the number of dNFs developed by NF1patients and correct for age biases; it has developed a new methodology to improve dNF molecular analysis and facilitate the acquisition of LOH and HR-driven LOH frequencies in dNFs from a significant group of NF1 patients; it has established the basis to identify candidate genes influencing the *in vivo* HR rate using yeast as a model organism; and it has established the basis to characterize genetic variation of candidate genes in the group of NF1 patients well characterized for the clinical and molecular phenotypes.

For accomplishing the work, it will be necessary to increase the number of NF1 patients and dNFs analyzed. It will be necessary to finish the identification of candidate genes in yeast modifying the *in vivo* HR rate and test their functional involvement. Finally, the sequence of all candidate genes will need to be performed for the largest cohort of NF1 patients possible with the clinical and molecular phenotype well characterized.

CONCLUSIONS

1. Concerning the clinical phenotype of NF1 patients, we have developed a standardized methodology for estimating the number of dermal neurofibromas (dNFs) developed, and to calculate the Risk of Dermal Tumor Burden (RDTB) in order to avoid the bias generated by patient age.

2. We have obtained a general picture of the importance of loss of heterozygosity (LOH) as an evidence of second hit in neurofibromas, but also dissected the frequencies of mutational mechanisms generating LOH in our cohort of NF1 patients.

3. Homologous recombination is the most frequent mechanism generating LOH in dNFs accounting for 63% of the identified LOH

4. Copy number loss of the *NF1* locus accounts for 37% of dNF-LOH, with deletions ranging from 80Kb to 8Mb.

5. Although rare, dNF can bear second hits in the *NF1* locus caused by non-allelic homologous recombination generating type-2 microdeletions.

6. Genomic gross alterations in other genomic regions apart from the NF1 locus are found in a very low frequency in dNFs with no single recurrence identified out from 20 dNF analyzed. Thus, other gross genomic alterations do not seem to be involved in dermal neurofibroma development.

7. Individual LOH frequencies in dNFs show a high variability, with patients exhibiting a dNF LOH frequencies ranging from 0 to 55.5%.

8. Many patients show a major or preferential LOH-causing mechanism in their dNFs, HR being the most frequent.

9. Our results suggest a possible correlation between having a high percentage of neurofibromas with HR-mediated LOH as a second hit and developing a high number of dNFs in NF1 patients.

10. A new methodology has been developed, the Microsatelite Multiplex PCR Analysis (MMPA), which makes more feasible the genetic analysis of tumors and thus the molecular characterization of dNFs in NF1 patients.

11. The MMPA allows obtaining accurate information not only regarding the AI status of tumors, but also the percentage of tumor-infiltrating normal cells, the locus copy-number status and the mechanism involved in AI.

12. A new assay (HoReYe Assay) has been developed to determine the *in vivo* HR rate, among other repair mechanisms, in different yeast strains. This assay mimics the exact HR-mediated LOH process identified in dNFs. In addition, a new strategy has been experimentally designed to identify gene candidates influencing the HR rates in yeast strains.

BIBLIOGRAPHY

1. Stratton, M.R., Campbell, P.J. & Futreal, P.A. The cancer genome. in *Nature* Vol. 458 719-24 (2009).
2. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. in *Cell* Vol. 144 646-74 (2011).
3. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
4. Fearon, E.R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. in *Cell* Vol. 61 759-67 (1990).
5. Vogelstein, B. & Kinzler, K.W. Cancer genes and the pathways they control. in *Nat Med* Vol. 10 789-799 (2004).
6. Pietras, K. & Ostman, A. Hallmarks of cancer: interactions with the tumor stroma. *Exp Cell Res* **316**, 1324-31 (2010).
7. Fearon, E.R. Molecular genetics of colorectal cancer. in *Annual review of pathology* Vol. 6 479-507 (2011).
8. Pardal, R., Clarke, M.F. & Morrison, S.J. Applying the principles of stem-cell biology to cancer. in *Nat Rev Cancer* Vol. 3 895-902 (2003).
9. Visvader, J.E. Cells of origin in cancer. in *Nature* Vol. 469 314-22 (2011).
10. Fialkow, P.J., Denman, A.M., Jacobson, R.J. & Lowenthal, M.N. Chronic myelocytic leukemia. Origin of some lymphocytes from leukemic stem cells. in *J Clin Invest* Vol. 62 815-23 (1978).
11. Vermeulen, L. *et al.* Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. in *Nat Cell Biol* Vol. 12 468-76 (2010).
12. Wang, X. *et al.* A luminal epithelial stem cell that is a cell of origin for prostate cancer. in *Nature* Vol. 461 495-500 (2009).
13. Chen, J. *et al.* A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* **488**, 522-6 (2012).
14. Driessens, G., Beck, B., Caauwe, A., Simons, B.D. & Blanpain, C. Defining the mode of tumour growth by clonal analysis. in *Nature* (2012).
15. Mackay, T.F. The genetic architecture of quantitative traits. in *Annu Rev Genet* Vol. 35 303-39 (2001).
16. Wayte, N., Da Silva, L., Chenevix-Trench, G. & Lakhani, S.R. What's in a cancer syndrome? Genes, phenotype and pathology. in *Pathology* Vol. 40 247-59 (2008).
17. Knudson, A.G. Mutation and cancer: statistical study of retinoblastoma. in *Proc Natl Acad Sci USA* Vol. 68 820-3 (1971).
18. Cavenee, W.K. *et al.* Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. in *Nature* Vol. 305 779-84 (1983).
19. Chen, Z. *et al.* Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725-30 (2005).
20. Cook, W.D. & McCaw, B.J. Accommodating haploinsufficient tumor suppressor genes in Knudson's model. in *Oncogene* Vol. 19 3434-8 (2000).
21. Berger, A.H., Knudson, A.G. & Pandolfi, P.P. A continuum model for tumour suppression. in *Nature* Vol. 476 163-9 (2011).
22. Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D. & Lowe, S.W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* **88**, 593-602 (1997).
23. Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. in *Science* Vol. 266 66-71 (1994).

24. Thompson, M.E., Jensen, R.A., Obermiller, P.S., Page, D.L. & Holt, J.T. Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression. in *Nat Genet* Vol. 9 444-50 (1995).
25. McClatchey, A.I. Neurofibromatosis. in *Annual review of pathology* Vol. 2 191-216 (2007).
26. Carroll, S.L. & Ratner, N. How does the Schwann cell lineage form tumors in NF1? in *Glia* Vol. 56 1590-605 (2008).
27. Tidyman, W.E. & Rauen, K.A. The RASopathies: developmental syndromes of Ras/MAPK pathway dysregulation. in *Current Opinion in Genetics & Development* Vol. 19 230-6 (2009).
28. Riccardi, V. *Neurofibromatosis: Phenotype, Natural History, and Pathogenesis (Johns Hopkins Series in Contemporary Medicine and Public Health)*, (The Johns Hopkins University Press; 3rd edition, 1992).
29. Jett, K. & Friedman, J.M. Clinical and genetic aspects of neurofibromatosis 1. in *Genet Med* Vol. 12 1-11 (2010).
30. Huson, S.M., Compston, D.A., Clark, P. & Harper, P.S. A genetic study of von Recklinghausen neurofibromatosis in south east Wales. I. Prevalence, fitness, mutation rate, and effect of parental transmission on severity. in *Journal of Medical Genetics* Vol. 26 704-11 (1989).
31. DeBella, K., Szudek, J. & Friedman, J.M. Use of the national institutes of health criteria for diagnosis of neurofibromatosis 1 in children. in *Pediatrics* Vol. 105 608-14 (2000).
32. Cichowski, K. & Jacks, T. NF1 tumor suppressor gene function: narrowing the GAP. in *Cell* Vol. 104 593-604 (2001).
33. Wallace, M.R. *et al.* Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science* **249**, 181-6 (1990).
34. Cawthon, R.M. *et al.* A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. *Cell* **62**, 193-201 (1990).
35. Viskochil, D. *et al.* The gene encoding the oligodendrocyte-myelin glycoprotein is embedded within the neurofibromatosis type 1 gene. in *Mol Cell Biol* Vol. 11 906-12 (1991).
36. Cawthon, R.M. *et al.* cDNA sequence and genomic structure of EV12B, a gene lying within an intron of the neurofibromatosis type 1 gene. *Genomics* **9**, 446-60 (1991).
37. Hajra, A. *et al.* DNA sequences in the promoter region of the NF1 gene are highly conserved between human and mouse. *Genomics* **21**, 649-52 (1994).
38. Dorschner, M.O., Sybert, V.P., Weaver, M., Pletcher, B.A. & Stephens, K. NF1 microdeletion breakpoints are clustered at flanking repetitive sequences. in *Hum Mol Genet* Vol. 9 35-46 (2000).
39. López-Correa, C. *et al.* Recombination hotspot in NF1 microdeletion patients. in *Hum Mol Genet* Vol. 10 1387-92 (2001).
40. Jenne, D.E. *et al.* Molecular characterization and gene content of breakpoint boundaries in patients with neurofibromatosis type 1 with 17q11.2 microdeletions. in *Am J Hum Genet* Vol. 69 516-27 (2001).

41. Luijten, M. *et al.* Duplication and transposition of the NF1 pseudogene regions on chromosomes 2, 14, and 22. in *Hum Genet* Vol. 109 109-16 (2001).
42. Yu, H. *et al.* Expression of NF1 pseudogenes. *Hum Mutat* **26**, 487-8 (2005).
43. Daston, M.M. *et al.* The protein product of the neurofibromatosis type 1 gene is expressed at highest abundance in neurons, Schwann cells, and oligodendrocytes. *Neuron* **8**, 415-28 (1992).
44. Heim, R.A., Silverman, L.M., Farber, R.A., Kam-Morgan, L.N. & Luce, M.C. Screening for truncated NF1 proteins. in *Nat Genet* Vol. 8 218-9 (1994).
45. Marchuk, D.A. *et al.* cDNA cloning of the type 1 neurofibromatosis gene: complete sequence of the NF1 gene product. *Genomics* **11**, 931-40 (1991).
46. Nishi, T. *et al.* Differential expression of two types of the neurofibromatosis type 1 (NF1) gene transcripts related to neuronal differentiation. *Oncogene* **6**, 1555-9 (1991).
47. Teinturier, C. *et al.* The neurofibromatosis 1 gene transcripts expressed in peripheral nerve and neurofibromas bear the additional exon located in the GAP domain. *Biochem Biophys Res Commun* **188**, 851-7 (1992).
48. Gutman, D.H., Andersen, L.B., Cole, J.L., Swaroop, M. & Collins, F.S. An alternatively-spliced mRNA in the carboxy terminus of the neurofibromatosis type 1 (NF1) gene is expressed in muscle. *Hum Mol Genet* **2**, 989-92 (1993).
49. Skuse, G.R., Cappione, A.J., Sowden, M., Metheny, L.J. & Smith, H.C. The neurofibromatosis type I messenger RNA undergoes base-modification RNA editing. in *Nucleic Acids Res* Vol. 24 478-85 (1996).
50. Maertens, O. *et al.* Comprehensive NF1 screening on cultured Schwann cells from neurofibromas. in *Hum Mutat* Vol. 27 1030-40 (2006).
51. DeClue, J.E., Cohen, B.D. & Lowy, D.R. Identification and characterization of the neurofibromatosis type 1 protein product. in *Proc Natl Acad Sci USA* Vol. 88 9914-8 (1991).
52. Gutmann, D.H., Wood, D.L. & Collins, F.S. Identification of the neurofibromatosis type 1 gene product. in *Proc Natl Acad Sci USA* Vol. 88 9658-62 (1991).
53. Izawa, I., Tamaki, N. & Saya, H. Phosphorylation of neurofibromatosis type 1 gene product (neurofibromin) by cAMP-dependent protein kinase. *FEBS Lett* **382**, 53-9 (1996).
54. Hackstein, J.H. The lethal prune/Killer-of-prune interaction of Drosophila causes a syndrome resembling human neurofibromatosis (NF1). *Eur J Cell Biol* **58**, 429-44 (1992).
55. Scheffzek, K. *et al.* Structural analysis of the GAP-related domain from neurofibromin and its implications. in *EMBO J* Vol. 17 4313-27 (1998).
56. D'angelo, I., Welti, S., Bonneau, F. & Scheffzek, K. A novel bipartite phospholipid-binding module in the neurofibromatosis type 1 protein. in *EMBO Rep* Vol. 7 174-9 (2006).
57. Casey, P.J., Solski, P.A., Der, C.J. & Buss, J.E. p21ras is modified by a farnesyl isoprenoid. in *Proc Natl Acad Sci USA* Vol. 86 8323-7 (1989).
58. Gibbs, J.B., Sigal, I.S., Poe, M. & Scolnick, E.M. Intrinsic GTPase activity distinguishes normal and oncogenic ras p21 molecules. in *Proc Natl Acad Sci USA* Vol. 81 5704-8 (1984).

59. Vogel, U.S. *et al.* Cloning of bovine GAP and its interaction with oncogenic ras p21. *Nature* **335**, 90-3 (1988).
60. Roux, P.P. & Blenis, J. ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. in *Microbiol Mol Biol Rev* Vol. 68 320-44 (2004).
61. Rasmussen, S.A., Yang, Q. & Friedman, J.M. Mortality in neurofibromatosis 1: an analysis using U.S. death certificates. in *Am J Hum Genet* Vol. 68 1110-8 (2001).
62. Kehrer-Sawatzki, H. & Cooper, D.N. Mosaicism in sporadic neurofibromatosis type 1: variations on a theme common to other hereditary cancer syndromes? in *Journal of Medical Genetics* Vol. 45 622-631 (2008).
63. De Raedt, T., Maertens, O., Serra, E. & Legius, E. Somatic NF1Mutations in Tumors and
Other Tissues. 11 (2008).
64. Colman, S.D., Williams, C.A. & Wallace, M.R. Benign neurofibromas in type 1 neurofibromatosis (NF1) show somatic deletions of the NF1 gene. in *Nat Genet* Vol. 11 90-2 (1995).
65. Sawada, S. *et al.* Identification of NF1 mutations in both alleles of a dermal neurofibroma. in *Nat Genet* Vol. 14 110-2 (1996).
66. Serra, E. *et al.* Confirmation of a double-hit model for the NF1 gene in benign neurofibromas. in *Am J Hum Genet* Vol. 61 512-9 (1997).
67. Cichowski, K. *et al.* Mouse models of tumor development in neurofibromatosis type 1. in *Science* Vol. 286 2172-6 (1999).
68. Skuse, G.R., Kosciolek, B.A. & Rowley, P.T. The neurofibroma in von Recklinghausen neurofibromatosis has a unicellular origin. in *Am J Hum Genet* Vol. 49 600-7 (1991).
69. Stefansson, K. *et al.* Myelin-associated glycoprotein in human retina. *Nature* **307**, 548-50 (1984).
70. Menon, A.G. *et al.* Chromosome 17p deletions and p53 gene mutations associated with the formation of malignant neurofibrosarcomas in von Recklinghausen neurofibromatosis. in *Proc Natl Acad Sci USA* Vol. 87 5435-9 (1990).
71. Kluwe, L., Friedrich, R. & Mautner, V.F. Loss of NF1 allele in Schwann cells but not in fibroblasts derived from an NF1-associated neurofibroma. in *Genes Chromosomes Cancer* Vol. 24 283-5 (1999).
72. Serra, E. *et al.* Schwann cells harbor the somatic NF1 mutation in neurofibromas: evidence of two different Schwann cell subpopulations. in *Hum Mol Genet* Vol. 9 3055-64 (2000).
73. Zhu, Y., Ghosh, P., Charnay, P., Burns, D.K. & Parada, L.F. Neurofibromas in NF1: Schwann cell origin and role of tumor environment. in *Science* Vol. 296 920-2 (2002).
74. Yang, F.-C. *et al.* Nf1-dependent tumors require a microenvironment containing Nf1+/-- and c-kit-dependent bone marrow. in *Cell* Vol. 135 437-48 (2008).
75. Joseph, N.M. *et al.* The loss of Nf1 transiently promotes self-renewal but not tumorigenesis by neural crest stem cells. in *Cancer Cell* Vol. 13 129-40 (2008).

76. Wu, J. *et al.* Plexiform and dermal neurofibromas and pigmentation are caused by Nf1 loss in desert hedgehog-expressing cells. in *Cancer Cell* Vol. 13 105-16 (2008).
77. Le, L.Q., Shipman, T., Burns, D.K. & Parada, L.F. Cell of origin and microenvironment contribution for NF1-associated dermal neurofibromas. *Cell Stem Cell* **4**, 453-63 (2009).
78. Miller, S.J. *et al.* Integrative genomic analyses of neurofibromatosis tumours identify SOX9 as a biomarker and survival gene. in *EMBO Mol Med* Vol. 1 236-48 (2009).
79. Friedman, J.M. Epidemiology of neurofibromatosis type 1. in *Am J Med Genet* Vol. 89 1-6 (1999).
80. Messiaen L, W.K. NF1 Mutational Spectrum. *Monografic Human Genetics. Basel, Karger, 2008* **16**, 63-77 (2008).
81. Wimmer, K., Callens, T., Wernstedt, A. & Messiaen, L. The NF1 gene contains hotspots for L1 endonuclease-dependent de novo insertion. in *PLoS Genet* Vol. 7 e1002371 (2011).
82. Forbes, S.H., Dorschner, M.O., Le, R. & Stephens, K. Genomic context of paralogous recombination hotspots mediating recurrent NF1 region microdeletion. in *Genes Chromosomes Cancer* Vol. 41 12-25 (2004).
83. Serra, E. *et al.* Somatic NF1 mutational spectrum in benign neurofibromas: mRNA splice defects are common among point mutations. in *Hum Genet* Vol. 108 416-29 (2001).
84. Thomas, L., Kluwe, L., Chuzhanova, N., Mautner, V. & Upadhyaya, M. Analysis of NF1 somatic mutations in cutaneous neurofibromas from patients with high tumor burden. in *Neurogenetics* (2010).
85. Thomas, L. *et al.* Exploring the somatic NF1 mutational spectrum associated with NF1 cutaneous neurofibromas. in *Eur J Hum Genet* Vol. 20 411-9 (2012).
86. Easton, D.F., Ponder, M.A., Huson, S.M. & Ponder, B.A. An analysis of variation in expression of neurofibromatosis (NF) type 1 (NF1): evidence for modifying genes. in *Am J Hum Genet* Vol. 53 305-13 (1993).
87. Szudek, J., Joe, H. & Friedman, J.M. Analysis of intrafamilial phenotypic variation in neurofibromatosis 1 (NF1). in *Genet Epidemiol* Vol. 23 150-64 (2002).
88. Sabbagh, A. *et al.* Unravelling the genetic basis of variable clinical expression in neurofibromatosis 1. in *Hum Mol Genet* Vol. 18 2768-78 (2009).
89. Upadhyaya, M. *et al.* An absence of cutaneous neurofibromas associated with a 3-bp inframe deletion in exon 17 of the NF1 gene (c.2970-2972 delAAT): evidence of a clinically significant NF1 genotype-phenotype correlation. in *Am J Hum Genet* Vol. 80 140-51 (2007).
90. Pasmant, E. *et al.* NF1 microdeletions in neurofibromatosis type 1: from genotype to phenotype. in *Hum Mutat* Vol. 31 E1506-18 (2010).
91. Niccum, B.A. *et al.* On an unbiased and consistent estimator for mutation rates. in *Journal of Theoretical Biology* Vol. 300 360-7 (2012).
92. Galhardo, R.S., Hastings, P.J. & Rosenberg, S.M. Mutation as a stress response and the regulation of evolvability. *Crit Rev Biochem Mol Biol* **42**, 399-435 (2007).

93. Hodgkinson, A. & Eyre-Walker, A. Variation in the mutation rate across mammalian genomes. in *Nat Rev Genet* Vol. 12 756-66 (2011).
94. Jackson, A.L. & Loeb, L.A. On the origin of multiple mutations in human cancers. *Semin Cancer Biol* **8**, 421-9 (1998).
95. Negrini, S., Gorgoulis, V.G. & Halazonetis, T.D. Genomic instability--an evolving hallmark of cancer. in *Nat Rev Mol Cell Biol* Vol. 11 220-8 (2010).
96. Moynahan, M.E. & Jasin, M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. in *Nat Rev Mol Cell Biol* Vol. 11 196-207 (2010).
97. Lord, C.J. & Ashworth, A. The DNA damage response and cancer therapy. in *Nature* Vol. 481 287-94 (2012).
98. David, S.S., O'shea, V.L. & Kundu, S. Base-excision repair of oxidative DNA damage. in *Nature* Vol. 447 941-50 (2007).
99. Cleaver, J.E., Lam, E.T. & Revet, I. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. in *Nat Rev Genet* Vol. 10 756-68 (2009).
100. Jiricny, J. The multifaceted mismatch-repair system. in *Nat Rev Mol Cell Biol* Vol. 7 335-46 (2006).
101. Lieber, M.R. NHEJ and its backup pathways in chromosomal translocations. in *Nat Struct Mol Biol* Vol. 17 393-5 (2010).
102. Frankenberg-Schwager, M. *et al.* Single-strand annealing, conservative homologous recombination, nonhomologous DNA end joining, and the cell cycle-dependent repair of DNA double-strand breaks induced by sparsely or densely ionizing radiation. in *Radiat Res* Vol. 171 265-73 (2009).
103. Jackson, S.P. & Bartek, J. The DNA-damage response in human biology and disease. in *Nature* Vol. 461 1071-8 (2009).
104. Jackson, S.P. Sensing and repairing DNA double-strand breaks. in *Carcinogenesis* Vol. 23 687-96 (2002).
105. Heyer, W.-D., Ehmsen, K.T. & Liu, J. Regulation of homologous recombination in eukaryotes. in *Annu Rev Genet* Vol. 44 113-39 (2010).
106. Dudley, D.D., Chaudhuri, J., Bassing, C.H. & Alt, F.W. Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. in *Adv Immunol* Vol. 86 43-112 (2005).
107. Cole, F., Keeney, S. & Jasin, M. Evolutionary conservation of meiotic DSB proteins: more than just Spo11. in *Genes & Development* Vol. 24 1201-7 (2010).
108. Smith, J.A. *et al.* Accurate homologous recombination is a prominent double-strand break repair pathway in mammalian chromosomes and is modulated by mismatch repair protein Msh2. *Mol Cell Biol* **27**, 7816-27 (2007).
109. Bishop, A.J.R. & Schiestl, R.H. Role of homologous recombination in carcinogenesis. in *Exp Mol Pathol* Vol. 74 94-105 (2003).
110. Gusella, J.F. *et al.* A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234-8 (1983).
111. Risch, N.J. & Zhang, H. Mapping quantitative trait loci with extreme discordant sib pairs: sampling considerations. *Am J Hum Genet* **58**, 836-43 (1996).
112. The International HapMap Project. *Nature* **426**, 789-96 (2003).

113. Gibson, G. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**, 135-45 (2011).
114. Eichler, E.E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. in *Nat Rev Genet* Vol. 11 446-50 (2010).
115. Metzker, M. Sequencing technologies - the next generation. in *Nat Rev Genet* (2009).
116. Schneeberger, K. & Weigel, D. Fast-forward genetics enabled by new sequencing technologies. in *Trends in Plant Science* Vol. 16 282-8 (2011).
117. Bamshad, M.J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. in *Nat Rev Genet* Vol. 12 745-55 (2011).
118. Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. in *Nat Genet* Vol. 44 623-30 (2012).
119. Aitman, T.J. *et al.* The future of model organisms in human disease research. in *Nat Rev Genet* Vol. 12 575-82 (2011).
120. Ehrenreich, I.M. *et al.* Dissection of genetically complex traits with extremely large pools of yeast segregants. in *Nature* Vol. 464 1039-42 (2010).
121. Gaertner, B.E. & Phillips, P.C. Caenorhabditis elegans as a platform for molecular quantitative genetics and the systems biology of natural variation. in *Genet. Res.* Vol. 92 331-48 (2010).
122. Mackay, T.F.C., Stone, E.A. & Ayroles, J.F. The genetics of quantitative traits: challenges and prospects. in *Nat Rev Genet* Vol. 10 565-77 (2009).
123. Peters, L.L. *et al.* The mouse as a model for human biology: a resource guide for complex trait analysis. in *Nat Rev Genet* Vol. 8 58-69 (2007).
124. Heinicke, S. *et al.* The Princeton Protein Orthology Database (P-POD): a comparative genomics analysis tool for biologists. in *PLoS One* Vol. 2 e766 (2007).
125. Tong, A.H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. in *Science* Vol. 294 2364-8 (2001).
126. Aylon, Y. & Kupiec, M. DSB repair: the yeast paradigm. in *DNA Repair (Amst)* Vol. 3 797-815 (2004).
127. Fogel, S. & Hurst, D.D. Meiotic gene conversion in yeast tetrads and the theory of recombination. in *Genetics* Vol. 57 455-81 (1967).
128. Strand, M., Prolla, T.A., Liskay, R.M. & Petes, T.D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274-6 (1993).
129. Kataoka, T. *et al.* Functional homology of mammalian and yeast RAS genes. *Cell* **40**, 19-26 (1985).
130. Steinmetz, L.M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. in *Nature* Vol. 416 326-30 (2002).
131. Deutschbauer, A.M. & Davis, R.W. Quantitative trait loci mapped to single-nucleotide resolution in yeast. in *Nat Genet* Vol. 37 1333-40 (2005).
132. Romano, G.H. *et al.* Different sets of QTLs influence fitness variation in yeast. in *Molecular Systems Biology* Vol. 6 346 (2010).
133. Holt, D. *et al.* Interindividual variation in mitotic recombination. in *Am J Hum Genet* Vol. 65 1423-7 (1999).
134. Dermitzakis, E.T. Cellular genomics for complex traits. in *Nat Rev Genet* Vol. 13 215-20 (2012).

135. Lander, E.S. The new genomics: global views of biology. *Science* **274**, 536-9 (1996).
136. Göransson, H. *et al.* Quantification of normal cell fraction and copy number neutral LOH in clinical lung cancer samples using SNP array data. in *PLoS One* Vol. 4 e6057 (2009).
137. Solomon, E. *et al.* Chromosome 5 allele loss in human colorectal carcinomas. *Nature* **328**, 616-9 (1987).
138. De Raedt, T. *et al.* Somatic loss of wild type NF1 allele in neurofibromas: Comparison of NF1 microdeletion and non-microdeletion patients. in *Genes Chromosomes Cancer* Vol. 45 893-904 (2006).
139. Assié, G. *et al.* SNP arrays in heterogeneous tissue: highly accurate collection of both germline and somatic genetic information from unpaired single tumor samples. in *Am J Hum Genet* Vol. 82 903-15 (2008).
140. Li, A. *et al.* GPHMM: an integrated hidden Markov model for identification of copy number alteration and loss of heterozygosity in complex tumor samples using whole genome SNP arrays. in *Nucleic Acids Res* (2011).
141. Serra, E. *et al.* Mitotic recombination effects homozygosity for NF1 germline mutations in neurofibromas. in *Nat Genet* Vol. 28 294-6 (2001).
142. Freund. Mathematical statistics. *Englewood Cliffs, NJ: Prentice Hall, Inc.* (1992).
143. Cerbinskaite, A., Mukhopadhyay, A., Plummer, E.R., Curtin, N.J. & Edmondson, R.J. Defective homologous recombination in human cancers. in *Cancer Treatment Reviews* Vol. 38 89-100 (2012).
144. De la Torre, C., Pincheira, J. & Lopez-Saez, J.F. Human syndromes with genomic instability and multiprotein machines that repair DNA double-strand breaks. *Histol Histopathol* **18**, 225-43 (2003).
145. Kass, E.M. & Jasin, M. Collaboration and competition between DNA double-strand break repair pathways. in *FEBS Lett* Vol. 584 3703-8 (2010).
146. Payne, M. & Hickson, I.D. Genomic instability and cancer: lessons from analysis of Bloom's syndrome. in *Biochem Soc Trans* Vol. 37 553-9 (2009).
147. Luo, G. *et al.* Cancer predisposition caused by elevated mitotic recombination in Bloom mice. in *Nat Genet* Vol. 26 424-9 (2000).
148. Gangloff, S., McDonald, J.P., Bendixen, C., Arthur, L. & Rothstein, R. The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. in *Mol Cell Biol* Vol. 14 8391-8 (1994).
149. Digweed, M. & Sperling, K. Nijmegen breakage syndrome: clinical manifestation of defective response to DNA double-strand breaks. *DNA Repair (Amst)* **3**, 1207-17 (2004).
150. Krogh, B.O. & Symington, L.S. Recombination proteins in yeast. in *Annu Rev Genet* Vol. 38 233-71 (2004).
151. Hurst, D.D., Fogel, S. & Mortimer, R.K. Conversion-associated recombination in yeast (hybrids-meiosis-tetrads-marker loci-models). *Proc Natl Acad Sci U S A* **69**, 101-5 (1972).
152. Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. & Stahl, F.W. The double-strand-break repair model for recombination. *Cell* **33**, 25-35 (1983).

153. Spell, R.M. & Jinks-Robertson, S. Determination of mitotic recombination rates by fluctuation analysis in Saccharomyces cerevisiae. in *Methods Mol Biol* Vol. 262 3-12 (2004).
154. Luria, S.E. & Delbruck, M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **28**, 491-511 (1943).
155. Foster, P.L. Methods for determining spontaneous mutation rates. in *Meth Enzymol* Vol. 409 195-213 (2006).
156. Storici, F. & Resnick, M.A. The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol* **409**, 329-45 (2006).
157. Michelmore, R.W., Paran, I. & Kesseli, R.V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. in *Proc Natl Acad Sci USA* Vol. 88 9828-32 (1991).
158. Earley, E.J. & Jones, C.D. Next-generation mapping of complex traits with phenotype-based selection and introgression. in *Genetics* Vol. 189 1203-9 (2011).
159. Plotkin, S.R. *et al.* Quantitative assessment of whole-body tumor burden in adult patients with neurofibromatosis. in *PLoS One* Vol. 7 e35711 (2012).
160. Tucker, T., Wolkenstein, P., Revuz, J., Zeller, J. & Friedman, J.M. Association between benign and malignant peripheral nerve sheath tumors in NF1. *Neurology* **65**, 205-11 (2005).
161. Serra, E. *et al.* Tumor LOH analysis provides reliable linkage information for prenatal genetic testing of sporadic NF1 patients. in *Genes Chromosomes Cancer* Vol. 46 820-7 (2007).
162. Mechtersheimer, G. *et al.* Analysis of chromosomal imbalances in sporadic and NF1-associated peripheral nerve sheath tumors by comparative genomic hybridization. in *Genes Chromosomes Cancer* Vol. 25 362-9 (1999).
163. Wallace, M.R. *et al.* Culture of cytogenetically abnormal schwann cells from benign and malignant NF1 tumors. in *Genes Chromosomes Cancer* Vol. 27 117-23 (2000).
164. Mertens, F. *et al.* Cytogenetic characterization of peripheral nerve sheath tumours: a report of the CHAMP study group. in *J. Pathol.* Vol. 190 31-8 (2000).
165. Mantripragada, K.K. *et al.* High-resolution DNA copy number profiling of malignant peripheral nerve sheath tumors using targeted microarray-based comparative genomic hybridization. in *Clin Cancer Res* Vol. 14 1015-24 (2008).
166. Mantripragada, K.K. *et al.* Genome-wide high-resolution analysis of DNA copy number alterations in NF1-associated malignant peripheral nerve sheath tumors using 32K BAC array. in *Genes Chromosomes Cancer* Vol. 48 897-907 (2009).
167. Beert, E. *et al.* Atypical neurofibromas in neurofibromatosis type 1 are premalignant tumors. *Genes Chromosomes Cancer* **50**, 1021-32 (2011).
168. Langlois, R.G. *et al.* Evidence for increased somatic cell mutations at the glycophorin A locus in atomic bomb survivors. *Science* **236**, 445-8 (1987).

169. Albertini, R.J., O'Neill, J.P., Nicklas, J.A., Heintz, N.H. & Kelleher, P.C. Alterations of the hprt gene in human in vivo-derived 6-thioguanine-resistant T lymphocytes. *Nature* **316**, 369-71 (1985).
170. Araten, D.J. *et al.* A quantitative measurement of the human somatic mutation rate. in *Cancer Res* Vol. 65 8111-7 (2005).
171. Grant, S.G. The GPA in vivo somatic mutation assay. in *Methods Mol Biol* Vol. 291 179-95 (2005).
172. Hartwell, L.H. & Smith, D. Altered fidelity of mitotic chromosome transmission in cell cycle mutants of S. cerevisiae. in *Genetics* Vol. 110 381-95 (1985).
173. Klein, H.L. Spontaneous chromosome loss in Saccharomyces cerevisiae is suppressed by DNA damage checkpoint functions. in *Genetics* Vol. 159 1501-9 (2001).
174. Lang, G.I. & Murray, A.W. Mutation rates across budding yeast chromosome VI are correlated with replication timing. in *Genome Biology and Evolution* Vol. 3 799-811 (2011).
175. Ehrenreich, I.M. *et al.* Genetic Architecture of Highly Complex Chemical Resistance Traits across Four Yeast Strains. in *PLoS Genet* Vol. 8 e1002570 (2012).
176. Ferner, R.E. Neurofibromatosis 1. *Eur J Hum Genet* **15**, 131-8 (2007).

RESUM EN CATALÀ

INTRODUCCIÓ

Càncer

La cèl·lula és una entitat complexa. Processos com la transcripció de l'ADN, la divisó cel·lular, etc... estan regulats per sistemes i vies que preserven la integritat i homeostasi cel·lular. El càncer defineix un grup de malalties complexes causades per desregulacions en els processos que mantenen la homeostasi cel·lular, portant a una divisió cel·lular descontrolada, invasió i metàstasi. La disrupció d'aquesta homeostasi cel·lular és el resultat d'alteracions en la regulació de múltiples gens, els quals poden incrementar, disminuir o perdre la seva funció. S'han descrits certes capacitats que una cèl·lula hauria d'adquirir per a ser tumorigènica [2]: mantenir senyals proliferatius; evasió de la supressió del creixement; evitar la destrucció inmunològica; desregulació de les capacitats energètiques de la cèl·lula; alteració de la maquinària de reparació; alteració de la resposta inflamatoria; resistència a mort cel·lular; activació d'invasió, metàstasi i angiogènesi; i una replicació inmortal.

Les cèl·lules han desenvolupat certs mecanismes per protegir-les d'esdevenir tumorigèniques, molts regulats per, al menys, tres tipus de gens: oncogens, relacionats amb proliferació cel·lular, gens supressors de tumors (TSGs, en anglès), els quals controlen la proliferació cel·lular i promouen la seva mort, i gens d'estabilitat, relacionats sobre tot a mecanismes de reparació [4,5]. Una cèl·lula esdevè tumorigènica quan sofreix alteracions, tant genètiques com epigenètiques, en aquests gens (incrementant, disminuint o perdent la seva funció). La formació tumoral també depèn del microambient, éssent el tumor el resultat d'alteracions en les cèl·lules, i la seva interacció amb l'estroma que les rodeja, la qual pot estar composada per diferents tipus cel·lulars.

Els últims estudis han mostrat que cèl·lules mare, o cèl·lules progenitores que desenvolupen capacitats de cèl·lula mare, deuen ser les cèl·lules patint les mutacions originals en la formació tumoral.

Càncer com a tret complexe

El càncer és considerat un tret complexe ja que és el resultat de l'afectació de múltiples gens, i de la seva interacció amb l'ambient. La combinació de l'herència de l'ADN (seqüència genètica, estructura genòmica, modificacions a l'ADN...) i les agresions ambientals conferiran el risc global d'un individu a desenvolupar càncer. Considerant l'aparició de càncer a famílies, es poden dividir en: càncer esporàdic, quan afecta a un sól individu en una família; càncer familiar, quan hi ha certa agregació d'un tipus de càncer en una família; càncer hereditari, quan l'afectació d'individus en una família presenta una clara herència Mendeliana.

Sídrome de Càncer Hereditari

Un síndrome de càncer hereditari (CS en anglès) és una malaltia genètica amb herència Mendeliana causada per alteracions en gens que confereixen una alta predisposició a l'individu a desenvolupar càncer [16]. Els gens responsables de CSs poden classificar-se en oncogens, TSGs i gens d'estabilitat (explicats prèviament). La majoria de CSs són causats per TSGs. Els individus que presenten aquests CS requereixen la presència d'un al·lel mutat per a desenvolupar la malaltia. Malgrat tot, els tumors es desenvolupen, quasi sempre, per una segona mutació en l'al·lel no afecte. Aquest procés va ser explicat per Knudson [17] i la seva validesa va ser demostrada per Cavenee [18]. Els pacients amb CS degut a mutacions en un TSG presenten una mutació germinal en un al·lel del gen, mentre que els tumors es desenvolupen a partir d'aquelles cèl·lules que pateixen una segona mutació en l'altre al·lel. Aquesta teoria ha evolucionat fins a funció dosi-depenent de TSG, ja que depenent de en quina via es trobi el TSG, pot no ser necessària la segona mutació, o second hit, per a desenvolupar el tumor.

Neurofibromatosis

Les neurofibromatosis, Neurofibromatosi tipus 1 (NF1), Neurofibromatosi tipus 2 (NF2), i Schwannomatosis, són un grup de malalties hereditàries caracteritzades per l'aparició de tumors en el sistema nerviós perifèric iniciats en cèl·lules de Schwann revisat en [25,26]. Malgrat tot, les 3 malalties difereixen en la seva patogènesi genètica i

molecular, ja que els gens i vies afectats per aquestes malalties són clarament diferents.

Neurofibromatosi tipus 1

La neurofibromatosi tipus 1 (NF1) és un CS caracteritzat per la generació de tumors derivats de la cresta neural [28,29]. Té una prevalència en adults de 1/3000 – 1/4000, amb una penetrància virtualment completa als 8 anys [31]. És una malaltia autosòmica dominant, amb el 50% dels casos *de novo*. Un dels aspectes més importants és la seva variabilitat clínica i la seva pleiotropia.

Bases moleculars de la Neurofibromatosi tipus 1

El gen *NF1* es troba localitzat en la regió pericentromèrica del cromosoma 17 (17q11.2). El gen *NF1* està rodejat de seqüències repetitives anomenades NF1-REP A, NF1-REP B i NF1-REP C, responsables de l'aparició de microdelecions pel mecanisme de recombinació homòloga no-al·lèlica com a mutacions germinals. El gen *NF1* s'expressa en tots els teixits, però la seva expressió és especialment significativa al cervell, la medul·la espinal i el sistema nerviós perifèric [33,43]. El gen *NF1* codifica per una proteïna de 11-13Kbs, generant principalment dues isoformes, tipus 1 i tipus 2. La isoforma tipus 2 conté l'exó 23a, el qual està localitzat al domini relacionat amb GAP (GRD en anglès). L'inserció d'aquest exó disminueix l'activitat de la proteïna comparat amb la isoforma tipus 1, que no el conté. El gen *NF1* codifica per la neurofibromina, que té 2818 aminoàcids. El domini més ben caracteritzat és el GRD, el qual interacciona amb la proteïna RAS afavorint la capacitat GTPàsica d'aquesta. La neurofibromina actua com un GTPasa activadora de proteïnes (GAP, en anglès). La proteïna RAS cicla entre la forma activa, unida a GTP, i la forma inactiva, unida a GDP. Aquesta proteïna està involucrada en vies relacionades amb proliferació cel·lular, diferenciació, supervivència i mort cel·lular. La neurofibromina inactiva la proteïna RAS.

Patogènesi molecular de la Neurofibromatosi tipus 1

La NF1 té una gran variació clínica. Els trets més freqüents són l'aparició de pigmentació a la pell, nòduls de Lisch, i múltiples neurofibromes dèrmics benignes [29,176]. Els pacients NF1 també poden desenvolupar altres manifestacions clíniques com retard mental, vasculopaties, tumors en el seistema nerviós central o tumors malignes de les veines dels nervis perifèrics (MPNST, en anglès). Els neurofibromes són el tret més característic de la NF1. Són tumors benignes que apareixen en les veines dels nervis perifèrics, els quals poden ser dividits en neurofibromes discrets, o dèrmics, i neurofibromes plexiforms. Els neurofibromes dèrmics (dNF, en anglès), poden aparèixer en gran nombre, arribant a milers de tumors desenvolupats en un pacient. Presenten una alta heterogeneïtat cel·lular, éssent composats per cèl·lules de Schwann (SC en anglès), cèl·lules perineurals, fibroblasts, cèl·lules mast i cèl·lules endotelials.

Per a la generació de dNFs és necessària una doble inactivació del gen *NF1*. El tumor té un origen unicel·lular, éssent les cèl·lules de Schwann les que sofreixen la doble inactivació. Malgrat tot, l'estroma té un paper important en la tumorigèncesi, ja que és necessària la presència de cèl·lules amb la mutació germinal (+/-) per a la generació del neurofibroma. En particular s'ha comprovat que les cèl·lules mast (+/-) tenen un paper molt important en la tumorigènesi. A més, diversos estudis han intentat identificar en quin estadi cel·lular les SC desenvolupen les capcitats tumorigèniques. Estudis semblen indicar que cèl·lules precursores derivades de la pell (SPKs en anglès), que són semblants a cèl·lules mare de la cresta neural, serien les cèl·lules originàries del neurofibroma (**Le** et al. 2009).

El gen *NF1* presenta una de les taxes de mutació més elevades [79]. Entre elles han estat descrites 3 tipus de delecions, les quals abarquen el 5% de delecions germinals. Aquestes es generen per recombinació homòloga no al·lèlica, i hi intervenen els NF1-REP 1 i NF1-REP C per les delecions de Tipus 1, el gen *SUZ12* i el pseudogen *SUZ12p* per les delecions de Tipus 2, i els NF1-REP B i NF1-REP C per les delecions Tipus 3. A més, existeixen les delecions atípiques, els punts de trencament de les quals són variables.

Per a determinar quina part de la variació fenotípica té un component genètic, i quin paper podria tenir l'ambient en la tumorigènesi, es van realitzar diferents estudis [86-88]. Tots tres estudis van estudiar trets clínics en famílies amb NF1. Així, es va demostrar

que la variació clínica es pot explicar en gran mesura per factors genètics. A més, la mutació germinal no afectaria a l'expressió fenotípica dels pacients. Tan sols s'han trobat dues relacions genotip-fenotip en NF1, una deleció que no afecta el marc de lectura en exó 17, relacionat amb l'absència de neurofibromes [89], i pacients amb microdelecions tipus 1 i 2, els quals presenten un fenotip sever de la malaltia [90].

Taxa de mutació somàtica en humans i resposta al dany en l'ADN

La taxa de mutació somàtica d'un organisme es defineix com el ràtio en el que les mutacions s'acumulen en el genoma. Aquesta taxa és el resultat d'un equilibri entre la generació de mutacions i la seva reparació. Per això els organismes presenten la resposta al dany de l'ADN (DDR en anglès).

Una de les alteracions més importants que pot afectar l'ADN són els trencaments de doble hebra (DSB, en anglès). Els DSBs poden ser reparats per 3 mecanismes diferents: Unió final no homòloga (NHEJ en anglès), Unió de hebra simple (SSE en anglès), i recombinació homòloga (HR en anglès). La HR és un mecanisme de reparació lliure d'errors, actua principalment a les fases S i G2 del cicle cel·lular, quan hi ha present l'ADN replicat, i utilitza seqüències homòlogues com a motlle per a reparar [96]. També intervè en la resolució dels entrecreuaments durant la meiosi.

Aquest mecanisme consta de diferents passos: 1) Eliminació dels extrems 5' del DSB, generant cues 3' mono-hebra; 2) Formació de filaments en la mono-hebra 3', eliminant possibles estructures secundàries; 3) Invasió de la cadena d'ADN, en el qual la mono-hebra s'uneix a la seva homòloga formant el que s'anomena gir-D (D-loop en anglès); 4) Resolució de D-loop, el qual, segons com es resolgui, generarà diferents hebres amb diferents càrregures gèniques.

Al ser la HR tan important en la reparació, el seu mal funcionament pot generar inestabilitat genòmica. Alteracions en gens involucrats en HR podrien inhibir el seu funcionament, afavorint els mecanismes de NHEJ o SSA, o augmentar-lo, provocant la generació de grans fragments d'homologia. La importància de la HR està demostrada pel fet que molts gens involucrats en aquest mecanisme són TSGs, responsables de CS com Bloom o el Càncer de mama i ovari familiar.

El llevat com a model per a diseccionar trets complexes

La majoria de trets biològics són complexes. En humans s'han utilitzat, principalment, dues aproximacions per a estudiar aquests trets, els anàlisis de lligament i els estudis d'associació genòmics (GWAS en anglès). Tots dos mètodes han aportat informació molt valuosa en la identificació de gens responsables de malalties. Els anàlisi de lligament han servit sobretot en la identificació de gens responsables de malalties Mendelianes, mentre que els estudis d'associació han identificat gens involucrats en trets complexes, explicant, però, una petita part de l'heretabilitat. En els últims anys s'ha utilitzat la seqüènciació de próxima generació (NGS en anglès), la qual té l'avantatge d'aportar informació de totes les variants presents en el genoma estudiat. Malgrat tot, estudiar trets complexes en humans segueix éssent una tasca molt complicada. És per això que s'utilitzen models animals, per a superar els problemes tècnics i ètics que es poden presentar en estudis amb humans.

El llevat és un molt bon sistema per a estudiar trets complexes ja que, entre d'altres raons, és un organisme unicel·lular amb complexitat eurcariota, presenta múltiples tècniques per a manipular fàcilment el seu ADN, i té un elevat grau de conservació en processos biològics, fins i tot amb humans. El llevat presenta un cicle cel·lular amb dues formes sexuals, a i α, i pot estar en forma haploid (a o α), o diploid (a/α). Amb el llevat s'utilitzen marcadors de selecció, els quals permeten realitzar fàcilment assajos genètics.

El llevat s'ha utilitzat per a estudiar trets complexes. Els estudis s'han basat principalment en l'anàlisi de la progènie obtinguda de creuar soques de llevat amb fenotips extrems. Els primers estudis analitzaven la progènie amb arrays, detectant regions que mostraven una baixa probabilitat de segregació aleatòria [130]. Més tard es va utilitzar la repetició de creuaments seguits de selecció pel tret estudiat. Amb aquest cicle, creuament i selecció repetits, es mantenien les variants al·lèliques responsables pel tret, mentre s'enriquia l'ADN amb el genoma no responsable del tret [131]. La última metodologia emprada ha estat la tècnica de loci de trets quantitatius extrem (X-QTL en anglès). Aquesta tècnica es basa en 3 aspectes: 1) La generació d'una progènie amb fenotip extrem elevada, de fins a 10^7 individus; 2) La selecció d'aquesta progènie, fent crèixer una part en medi de selecció, i l'altra sense selecció (aquesta última serà la població control); 3) Anàlisi de les freqüències al·lèliques, en les quals

milers d'individus són analitzats per identificar al·lels sobrerepresentats en la població crescuda en medi de selecció [120].

PREMISES

El tret "nombre de neurofibromes desenvolupats" és un tret complexe, i com a tal es veu afectat tant per factors genètics com ambientals. En aquest cas, però, els factors ambientals tindran un rol minoritari, ja que el sistema nerviós perifèric no està molt exposat a agresions externes. Nosaltres hipotetitzem que bona part de la variació en el nombre de neurofibromes serà deguda a factors genètics, mostrant així una elevada heretabilitat. Fins ara, tan sols dues correlacions entre la mutació germinal del gen *NF1* i el fenotip dels pacients han estat determinades, afectant a una baix nombre de pacients. A més, tres estudis han demostrat que, per alguns trets com el nombre de neurofibromes, són gens modificadors (gens diferents del gen *NF1*) els responsables de bona part del fenotip observat [86-88].

Els neurofibromes dèrmics estan composats per diferents tipus cel·lulars (cèl·lules perineural, fibroblasts, cèl·lules de Schwann, macròfags). Aquests tumors es desenvolupen degut a una doble inactivació del gen *NF1* a les cèl. de Schwann, sempre per mutació. El nostre raonament és que gens involucrats en la taxa de mutació somàtica dels individus podrien ser gens candidats com a modificadors del nombre de neurofibromes desenvolupats. D'entre els mecanismes de reparació de l'ADN, el de recombinació homòloga és responsable d'un alt percentatge de les inactivacions somàtiques del gen *NF1*. A més, aquest mecanisme mostra variabilitat interindividual [133]. Dins la complexitat que comporta el desenvolupament d'un neurofibroma, nosaltres ens centrarem en estudiar les cèl. de Schwann i el mecanisme de recombinació homòloga.

La forma més emprada per a estudiar l'arquitectura genètica de trets complexes en humans han estat els estudis d'associació genòmica utilitzant arrays de SNPs. Aquests estudis, però, han permès identificar majoritàriament loci responsables d'una petita part de la variació en el fenotip estudiat. A més, amb aquesta tècnica no és possible identificar relacions funcionals de les variants trobades. Sabíem que obtenir les freqüències de LOH i el mecanisme de RH en un elevat nombre de pacients NF1 era

complicat. Així, degut a que el mecanisme de recombinació homòloga està molt conservat entre humans i llevats, vam decidir utilitzar aquest organisme model per a identificar gens involucrats en la taxa de recombinació homòloga. Utilitzar *Saccharomices cerevisiae* com a model és molt ventatjós, ja que presenta un ventall de tècniques de manipulació gènica molt ampli, a més de permetre realitzar estudis funcionals per confirmar correlacions genotip-fenotip. Pel que fa a l'anàlisi pròpiament dels gens candidats, ens vam decidir per la seqüenciació, ja que preteníem analitzar tant variants comunes com rares, a diferència de les aproximacions realitzades en els últims temps sobretot a arrel dels GWAS. Vam utilitzar primer la seqüenciació per Sanger, i més tard les noves tècniques d'ultra-seqüenciació establertes en els últims anys.

Nosaltres creiem que gens involucrats en la taxa de mutació somàtica de l'individu seran responsables del nombre de neurofibromes dèrmics en pacients NF1. Si fos així, el nombre de neurofibromes podria ser indicatiu de la taxa de mutació somàtica *in vivo*. Hi ha altres síndromes de càncer que presenten l'aparició de múltiples tumors benignes. Si la nostra hipòtesi fos certa, els gens influenciant el nombre de neurofibromes podrien també influenciar l'aparició de tumors benignes en d'altres síndromes cancerígens.

MATERIAL I MÈTODES

Neurofibromes dèrmics

Els neurofibromes dèrmics (dNFs) van ser extrets amb una petita intervenció quirúrgica. El teixit envoltant el dNF es va extreure, i el tumor va ser tallat a trocets. Les mostres es van conservar a -80ºC.

PCR Multiplex de Microsatèl·lits (MMP)

Els primers utilitzats es troben descrits a la Taula annex A1. Els primers es van marcar amb els fluoròfors FAM (blau), NED (groc), VIC (verd), i PET (vermell) (Invitrogen). La reacció es va realitzar utilitzant el kit Multiplex PCR Kit (Qiagen), en 25 µl de volum final amb 80 ng d'ADN: es va fer un cicle inicial de desnaturalització

a 95ºC durant 15 min, seguit de 25 cicles de desnaturalització a 94ºC, hibridació a 56ºC, i extensió a 72ºC durant 30sec, 3 min i 1.5 min, respectivament, amb un cicle final de 60ºC durant 30 min. 2 µl de producte final es van barrejar amb 7.8 µl de formamida desionitzada (Applied Biosystems, Bedford, MA), i 0.2 µl de LIZ500 Size Standard (Applied Biosystems).

Anàlisi de PCR Multiplex de Microsatèl·lits (MMPA)

Les condicions per la MMPA eren les mateixes que per la MMP, amb la única diferència de realitzar 22 cicles d'amplificació.

Anàlisi de LOH/AI

Pèrdues d'heterozigositat (LOH en anglès) i desequilibris al·lèlics (AI en anglès) s'utilitzen per a referir-se a la pèrdua d'un al·lel (LOH), o a desequilibris generats per la pèrdua o el guany d'un al·lel (AI). Per a determinar la presència de LOH/AI vam utilitzar l'expressió Q^{AI} [136], que representa el quocient entre el ràtio dels al·lels del tumor d'un microsatèl·lit heterozigot i el ràtio dels al·lels de la mostra control, pel mateix marcador [137]. Vam considerar que un microsatèl·lit presentava LOH/AI si hi havia una diferència entre els ràtios superior a 0.2.

Amplificació de sondes depenents d'una multiplex de lligació (MLPA)

Vam utilitzar la tècnica MLPA per analitzar el nombre de còpies de la regió del gen *NF1* present en les mostres d'ADN. Aquesta tècnica es basa en l'amplificació de sondes, que prèviament han estat lligades, localitzades en el gen *NF1*, en regions al voltant del gen i en altres cromosomes (sondes control). La comparació en les intensitats de les sondes control amb les sondes localitzades a la regió del gen *NF1* ens indicarà quines regions han patit una deleció. Consideràvem deleció valors menors de 0.8.

Anàlisi de ràtios paràlegs (PRA)

Aquesta tècnica es va utilitzar per a determinar el nombre de còpies de la regió de l'exó 22 del gen *NF1*, comparat amb la mateixa regió del pseudogen de *NF1* localitzat

al cromosoma 15 [138]. La reacció de PRA es va realitzar en 10 µl de volum final amb 80ng d'ADN: un cicle inicial de desnaturalització a 94ºC durant 3 min, seguit de 26 cycles de desnaturalització a 94ºC, hibridació a 55ºC, i extensió a 72ºC durant 30sec cada pas. Al final es va realitzar un cicle a 72ºC durant 7min.

SNP-array

Vam utilitzar la tecnologia Illumina Infinium technology (Illumina, San Diego, CA), amb la qual es podien analitzar fins a ~592,000 SNPs, ~16,600 localitzats en el cromosoma 17. Vam utilitzar dos mètodes per a calcular el percentatge de cèl·lules no portadores d'AI, el desenvolupat per Assie et al. [139], i l'algoritme Global Parameter Hidden Markov Model (GPHMM) [140].

Protocol de transformació d'alta eficiència de llevat

Per a transformar les soques de llevat vam utilitzar el mètode de Gietz. Aquest es basa en l'ús d'Acetat de Liti per a que, després de realitzar un xoc tèrmic a 42ºC, es pugui introduir a les cèl·lules ADN forani.

Delitto Perfetto

Aquesta tècnica es va utilitzar per a delecionar gens. Es basa en la introducció, primer, d'un constructe amb dos gens, un de selecció positiva i l'altre de selecció negativa. Aquest constructe s'introduirà a la regió a delecionar mitjançant la maquinària de recombinació de la cèl·lula de llevat, i serà seleccionat de forma positiva. Després, s'introdueix un altre constructe, el qual serà tan sols la seqüència que romandrà finalment en la regió delecionada. En aquest pas es seleccionarà de forma negativa pel primer constructe. Aquest mètode no deixa cap tipus de rastre a l'ADN.

Construcció del cassette

Vam construir un cassette (CGL105) per una inserció seqüencial de gens al plàsmid pRS415. El mètode utilitzat va ser el de transformació d'alta eficiència, en el qual el plàsmid linearitzat i el gen a insertar eren transformats conjuntament en la soca de

llevat. La maquinària de recombinació del llevat s'encarregava d'unir tots dos fragments, generant el plàsmid amb el gen inserit.

RESULTATS

Caracterització clínica i molecular del fenotip dels pacients NF1

Caracterització del nombre de dNFs en pacients NF1

En aquest estudi es van obtenir dNFs de fins a 117 pacients NF1, dels quals es van recollir dades clíniques com l'edat, el sexe, el nombre de neurofibromes dèrmics (dNF en anglès) desenvolupats, i la presència d'altres manifestacions com neurofibromes plexiformes o MPNSTs.

Vam estudiar el comportament del nombre de dNFs en la nostre cohort de pacients. Degut al baix nombre de pacients amb les dades recopilades, vam decidir utilitzar altres estudis en els que es mostraven les dades clíniques, com un estudi de pacients del Regne Unit i el sud de Gales [86], i un de la població francesa [88], ja que vam considerar que la distribució de nº de dNFs no variaria molt entre poblacions (Figura 8). Totes dues poblacions presenten una distribució similar (Figura 8), i es mostra la clara relació existent entre l'edat i el nº de dNFs desenvolupats (Figura 9). A més, com volíem compara dades clíniques dels pacients, era necessari normalitzar el nº de dNFs desenvolupats per l'edat, ja que pacients amb edat avançada podrien mostrar més nº de dNFs que pacients més joves, tot i tenir la mateixa taxa de desenvolupament de dNFs. És per això que vam desenvolupar un valor que mostrava el risc de desenvolupar dNFs (RDTB en anglès). La forma de calcular-ho està resumit a la Figura 11.

Caracterització del fenotip molecular dels dNFs en pacients NF1

La caracterització tumoral es basà en l'estimació de les freqüències de pèrdues d'heterozigositat (LOH en anglès) implicades en la generació de neurofibromes, i la identificació del mecanisme mutacional responsable.

Es va desenvolupar una PCR multiplex de microsatèl·lits (MMP) per tal d'amplificar 15 microsatèl·lits localitzats al llarg del cromosoma 17, incloent el gen *NF1*, en una sóla reacció d'amplificació. Es van analitzar 606 neurofibromes de 117 pacients. D'aquests, 144 presentaven LOH, un 23.7% dels neurofibromes analitzats (Taula 1). També es va determinar el nombre de còpies del gen *NF1* utilitzant les tècniques de MLPA i PRA (Veure Materials i Mètodes). Amb aquesta informació, LOH i nombre de còpies, va ser possible inferir el mecanisme mutacional causant de les LOHs. A més 19 dNFs van ser analitzats per SNP-array, el qual va proporcionar informació sobre LOH i extensió, el nombre de còpies de la regió analitzada, el mecanisme mutacional generador de la LOH, el percentatge de cèl·lules portadores de la LOH, i la presència d'alteracions a altres zones del genoma.

LOH amb canvis en el nombre de còpies

De tots els neurofibromes analitzats, 53 (~37% LOHs) mostraven pèrdua en el nombre de còpies del gen *NF1* (Taula 1 i exemple a Figura 14). En tots els casos, la pèrdua estava localitzada al gen *NF1* i les regions del voltant, amb un tamany d'entre ~80Kb i ~8Mb (Figura 15). En 13 neurofibromes es va trobar una LOH localitzada, però amb la presència de les dues còpies del gen *NF1* (Figura 16). Semblaria que el mecanisme causant de les LOHs en aquests tumors seria una recombinació homòloga amb dos entrecreuaments. Vam comprovar, però, que analitzant els ràtios obtinguts en l'assaig de microsatèl·lits, i a través de l'estudi de SNP-array per alguna d'elles, les mostres contenien un alt percentatge de cèl·lules normals, no portadores de l'alteració. El cultiu i anàlisi de cèl·lules de Schwann *NF1* (-/-) va confirmar que, en aquestes mostres, la deleció del gen *NF1* estava enmascarat per l'elevada presència de cèl·lules normals *NF1* (+/-).

En dues mostres analitzades (P004-7N, P082-1N) els extrems de la LOH estaven localitzats en les zones repetitives *NF1*-REPs i SUZ12, properes al gen *NF1*. Es va identificar una deleció de tipus 2 en el neurofibroma P082-1N (Figura 17), causada per un mecanisme de recombinació no al·lèlica entre el gen SUZ12 i el seu pseudogen. Aquest tipus d'alteració es troba en pacients mosaïcs com a mutació germinal. És la primera vegada que s'identifica una deleció de tipus 2 com a mutació somàtica en un pacient NF1.

LOHs amb canvis neutrals

Es van identificar 91 neurofibromes (~63% LOHs) amb una LOH que s'extenia desde el gen *NF1*, fins el final del cromosoma 17, i que mostrava la presència de dues còpies en tota la regió analitzada (Figura 18). Aquestes LOHs eren degudes al mecanisme de recombinació homòloga, amb un entrecreuament entre el centròmer i el gen *NF1*, que s'extenia, en tots els casos, fins el final del cromosoma 17 (Figura 19).

Altres alteracions

L'estudi de neurofibromes amb la tècnica d'SNP-array ens va permetre analitzar, a més de la regió *NF1* possibles alteracions en d'altres zones del genoma. Dels 19 neurofibromes analizats, en dos es van trobar altres alteracions (a part de la mutació al gen *NF1*). Es va trobar una deleció en el cromosoma 2 (regió 2q24.2-q31.1) del pacient P090-3N, de ~15.3Mb, i una altra deleció en el cromosoma 3 (3q11.2-q22.1) del pacient P072-1N, de ~39.8Mb (Figura 20). No creiem, però, que cap d'aquestes alteracions estiguin relacionades amb la generació dels neurofibromes.

Freqüències de LOH en pacients NF1

Per tal de caracteritzar els pacients per les freqüències de LOH individuals era necessari determinar quants dNFs s'haurien d'analitzar per pacient per a determinar la LOH amb certa confiança. Assumint una prevaleça a la població de 20%, vam determinar que era necessari analitzar, al menys, 10 dNFs per pacient, per a obtenir una precisió d'un 20% (Figura 21). Vam poder obtenir com a mínim 10 dNFs en 17 pacients NF1 (Taula 2).

Si analitzem les freqüències obtingudes (Taula 3), veiem que aquestes varien de menys d'un 10% fins a més d'un 50%. Fins i tot hi ha un pacient que no mostra LOH als seus tumors. Veiem pacients amb LOHs causades tant per HR com per deleció, i pacients amb un mecanisme mutacional predominant (Taula 3). A més, tenim dos pacients amb la mateixa mutació germinal, i freqüències de LOH diferents, la qual cosa podria indicar que la mutació germinal no afecta a la freqüència de LOH dels pacients (Taula 3).

Si estudiem la distribució del nº de dNFs desenvolupats en la nostre cohort de pacients (Figura 22), veiem que es comporta de forma similar a les distribucions obtingudes per les poblacions del Regne Unit amb el Sud de Gales, i la Francesa. A més, vam decidir estudiar possibles correlacions entre les freqüències de LOH/HR dels pacients, i el nº de dNFs desenvolupats. Sabíem, però, que la nostre cohort contè pocs pacients, i que qualsevol resultat s'ha de considerar com a preliminar fins que la població de pacients NF1 analitzats pugui augmentar. Primer, vam agrupar els pacients amb alt i baix nombre de RDTB, agafant com a punt de tall el valor de 0.5 de forma arbitrària (Figura 23). Així, vam comprovar que, en la nostra cohort de pacients, no hi havia diferències apreciables entre la població d'alt i baix RDTB per les freqüències de LOH. En canvi, sí que s'apreciava certa diferència entre aquestes dues poblacions amb la freqüència de HR. No hi havia cap tipus de diferència si fem referència a les freqüències de deleció (Figura 23). Aleshores vam decidir analitzar si els pacients mostrant altes freqüències de LOH o HR eren els pacients mostrant un RDTB més elevat. Un estudi de Chi-quadrat ens va mostrar que pacients amb un elevat RDTB estaven molt a prop d'estar significativament relacionats amb una elevada freqüència de HR ($p = 0.058$) (Taula 5). Una altra vegada, els nostres estudis han estat realitzats amb pocs pacients, i tota conclusió que se'n pugui extreure és preliminar, però pensem que aquest resultat pot ser indicatiu d'una relació entre una elevada freqüència de HR generant LOH en dNFs, i un elevat nº de dNFs desenvolupats.

Anàlisi de la PCR multiplex de microsatèl·lits (MMPA)

L'anàlisi per microsatèl·lits proporcionava informació sobre l'estat dels desequilibris al·lèlics (AI en anglès), i l'extensió d'aquests. Utilitzant altres tècniques per a determinar el nombre de còpies del gen *NF1*, es va poder inferir el mecanisme mutacional generador de les alteracions. Per a facilitar i abaratir l'estudi de neurofibromes vam desenvolupar una tècnica, la MMPA, que ens permetia obtenir, amb una sóla reacció de PCR, informació sobre la presència/absència d'AI, la seva extensió, el percentatge de cèl·lules normals (no portadores de l'alteració), el nombre de còpies de la zona analitzada, i el mecanisme mutacional responsable.

Per a realitzar els càlculs amb MMPA, les amplificacions s'haurien d'analitzar en la fase exponencial de la PCR (22 cicles d'amplificació per la nostra reacció) (Figura

25). La proporcionalitat obtinguda en l'amplificació dels amplicons en la fase exponencial permetia la comparació de dues reaccions independents de PCR. Així, una constant de proporcionalitat (K) podia ser calculada per cada microsatèl·lit (Figura 26). Aquesta K hauria de ser semblant en tots els al·lels utilitzats. Amb les Ks obtingudes dels microsatèl·lits sense AI es pot calcular una K mitjana (Kμ). Aquesta Kμ s'utilitzarà en els càlculs proposats. La variació de Kμ (CV) hauria de ser el més baixa possible (Figures 27 i 28). Els càlculs de MMPA es basen en la comparació de les intensitats dels pics observats amb una situació hipotètica en que, per aquell microsatèl·lit, no hi hagués AI, o sigui, que el 100% de cèl·lules no tinguessin l'alteració, el que anomenem pic esperat. Obtenim el pic esperat per "$Peak_{exp} = peak_C \cdot K\mu$" (Figura 30).

Les intensitats dels pics d'un microsatèl·lit amb AI depenen del nombre de còpies d'aquell al·lel en les cèl·lules amb AI, i del percentatge d'aquestes cèl·lules en el tumor (Figura 31). Depenent del mecanisme mutacional, els pics observats seran majors, iguals, o inferiors als pics esperats. Comparant els pics podem saber si l'al·lel ha guanyat, mantingut o perdut càrrega gènica. Per a fer aquesta comparació, necessitarem un intèrval de pics esperats utilitzant un valor obtingut empíricament (ε).

$Peak_{exp} [\pm\varepsilon] = peak_C \cdot (K\mu \pm \varepsilon)$

Càlcul de percentatge de cèl·lules portadores de l'AI

El percentatge de cèl·lules portadores de l'AI es calcula a partir de l'al·lel que mostra una disminució del pic observat respecte l'esperat (Figura 32). La diferència entre els pics observat i esperat és proporcional al percentatge que busquem, ja que l'amplificació del pic observat és deguda a les cèl·lules que no porten l'alteració.

$\%\ \text{non-AI cells} = peak_{obs} / peak_{exp} \cdot 100$

Càlcul del nombre de còpies en cèl·lules amb AI

Considerant el mateix microsatèl·lit, el pic que no disminueix repecte l'esperat serà utilitzat per a determinar el nombre de còpies i el mecanisme mutacional causant de l'AI (Figura 32). Utilitzant el percentatge de cèl·lules amb AI, i tenint en compte les càrregues gèniques que resulten de cada mecanisme mutacional (Figura 31), es poden

calcular pics esperats per cada mecanisme interrogat, en aquest cas per pèrdua o per recombinació homòloga.

Pèrdua

$Peak_{exp}$ cl (±ε) = [$peak_{exp}$ (±ε) · % non-AI cells] + [$peak_{exp}$ (±ε) · % AI cells]

Recombinació homòloga

$Peak_{exp}$ cn (±ε) = [$peak_{exp}$ (±ε) · % non-AI cells] + [$peak_{exp}$ (±ε) · 2 · % AI cells]

El pic observat estarà entre un dels dos intèrvals de pics esperats, indicant el mecanisme mutacional responsable de l'AI. En el cas que el pic observat no es trobés entre cap dels dos intèrvals, es consideraria un pic fora de rang (OOR en anglès), i no s'utilitzaria per als càlculs de MMPA.

Anàlisi de neurofibromes aplicant la tècnica de MMPA

Vam validar la tècnica de MMPA re-analitzant neurofibromes dèrmics prèviament caracteritzats per les tècniques d'MMP pel que fa a la LOH, i MLPA, PRA o SNP-array pel que fa a l'estimació del nombre de còpies del gen *NF1*. En 26 del 29 neurofibromes la tècnica de MMPA va determinar correctament el número de còpies i el mecanisme mutacional, considerant tots el microsatèl·lits informatius (Taula 6). A més, el mecanisme mutacional es va estimar correctament en tots els neurofibromes que tenien al menys 3 microsatèl·lits informatius, considerant l'estimació correcte d'almenys un 60% dels microsatèl·lits. En total, 144 microsatèl·lits van ser analitzats, dels quals 116 (80.5%) van determinar correctament el mecanisme, 25 (17.4%) van ser OOR, i tan sols 3 (2.1%) van errar en el mecanisme estimat. Pel que fa al percentatge de cèl·lules sense AI en el tumor, dades obtaingudes per SNP-array van ser comparades amb les obtaingudes per MMPA en els mateixos tumors (Taula 7). Així, tots els tumors analitzats mostraven valors semblants entre les dues tècniques, diferint tan sols per a una mostra amb més d'un 8% de diferència.

Selecció de gens candidats influenciant el nº de dNFs

L'objectiu d'aquest treball era trobar gens responsables del nombre de neurofibromes pels pacients NF1, i nosaltres vam centrar els nostres estudis en el mecanisme de

recombinació homòloga. Es van dissenyar dues estratègies per a identificar gens candidats relacionats amb la HR involucrats en el nombre de neurofibromes: gens seleccionats a partir de la literatura, i guiant-nos de certs criteris, i gens obtinguts a partir d'utilitzar el llevat per a modelar el mecanisme de HR.

Gens basats en la literatura

Vam decidir estudiar el gens relacionats amb el mecanisme de recombianció homòloga, que fossin responsables de l'aparició de CSs, i que tinguessin ortòlegs al llevat. D'aquesta manera, vam decidir començar l'anàlisi pels gens *BLM*, *NBS1* i *RAD51A*.

En primer lloc es va plantejar obtenir la variació genètica d'aquests gens amb seqüenciació per Sanger. En el transcurs d'aquesta tesi es va desenvolupar la tecnologia de NGS, i es va decidir utilitzar aquesta tecnologia per a sequenciar gens relacionats amb mecanismes de reparació i de recombinació.

Gens seleccionats per modelar el mecanisme de HR en llevats

El mecanisme de recombinació homòloga ha estat àmpliament estudiat en llevats [151,152], però es té poca informació sobre el gens que poden influenciar la taxa de recombinació homòloga *in vivo*. Vam desenvolupar un assaig, HoReYe (de l'anglès Homologous Recombination Yeast assay), per a determinar la taxa de HR en diferents soques de llevat. El mecanisme modelat és similar al mecanisme mutacional responsable de LOH en neurofibromes. Després s'utilitzaria un altra assaig per a determinar les variants responsables de les diferents taxes obtingudes. Després de evaluar funcionalment les variants obtingudes en llevat, s'analitzarien en els ortòlegs dels pacients NF1.

Assaig Homologous Recombination Yeast (HoReYe)

Aquest assaig permetia determinar la taxa de recombinació homòloga en llevat. Per a fer-ho, es va modelar el mateix mecanisme que causa LOH per HR en els pacients NF1 (Figura 36). Per a realitzar l'assaig, es van introduir marcadors de selecció al braç llarg del cromosma XV, i a prop el centròmer i al telòmer del braç curt del mateix cromosoma XV.

Gens marcadors de selecció

Es va crear un constructe amb 4 marcadors de selecció, anomenat CGL105, que s'introduiria a prop del centròmer del cromosoma XV, en una posició anàloga a on està localitzat el gen *NF1* en el cromsoma 17 humà. Aquest cassette estava format pels gens URA3 (utilitzat per selecció negativa), CAN1 (utilitzat per selecció negativa), ADE2 (utilitzat per selecció per color), i HIS3 (utilitzat per selecció positiva). A més, es van introduir el marcador NAT1 al braç curt, i el marcador KANMX al telòmer del braç llarg del cromsoma XV (Figura 37).

Preparació de soques

L'assaig d'HoReYe estava basat en la incorporació de marcadors de selecció al genoma del llevat. Com alguns d'aquests eren gens auxotròfics, estaven, en alguns casos, presents en altres localitzacions del genoma (Taula 9). Va ser necessari delecionar-los abans d'incorporar els marcadors de selecció al cromosoma XV. El mètode emprat va ser el Delitto Perfetto [156]. Per altra banda, els gens es van introduir al genoma del llevat amb el mateix mètode amb què es va construir el cassette de gens.

Realització de l'assaig HoReYe

La figura 39 mostra un resum de l'assaig. Es tracta de crear soques diploids que continguin els marcadors de selecció, crèixer-les en un medi sense selecció per un temps de duplicació (en aquest període trencaments a la doble hebra i la seva reparació tindrien lloc, alguns en el cromosoma XV on s'han inserit els marcadors de selecció), i plaquejar-les en diferents medis de selecció. Veient on creixen i on no creixen les colònies és possible determinar quins marcadors de selecció s'han perdut, i estimar el mecanisme mutacional responsable (Taula 10 i Figura 40). D'aquesta manera la taxa de recombinació homòloga pot ser estimada. Es va realitzar l'assaig per la soca BMA64, i es va obtenir una taxa de recombinació de $1.21 \cdot 10^{-4}$.

Dissecció genètica de la taxa de recombinació homòloga en llevat

Encara que es va obtenir la taxa de recombinació per tan sols una soca, es va planejar com s'hauria de realitzar l'assaig per a obtenir variants relacionades amb la taxa de recombinació. Primer es va pensar aplicar l'assaig X-QTL, el qual es basa en l'anàlisi de milions d'individus descendents de dues soques amb fenotips oposats extrems, seleccionats pel tret d'interès. La força d'aquest assaig rau en el poder estadístic que dóna analitzar milions d'individus alhora de fer assajos d'associació. Aquesta tècnica no podia ser aplicada directament per a analitzar el nostre tret per dues raons: primer, X-QTL analitza soques haploids, i el nostre fenotip, la recombinació homòloga, requereix de soques diploids. Segon, el nostre tret és molt poc freqüent, pel que un elevat nombre d'individus, fins a 10^9 cèl·lules de llevat, s'haurien d'utilitzar. Per tant, vam modificar la tècnica X-QTL per un anàlisi basat en la selecció del fenotip per retroencreuament. L'assaig es basaria en obtenir cèl·lules haploids resultants de creuar dues soques amb fenotips oposats, en aquest cas, una soca amb una elevada taxa de recombinació homòloga i una altra amb baixa taxa de recombinació. Aquestes soques es creuarien amb la soca amb baixa recombinació, i es realitzaria l'assaig HoReYe amb aquests diploids. Un cop seleccionats els diploids amb més alta recombinació, s'esporularia i es seleccionaria per les cèl·lules haploids MATa, i es tornaria a fer el procés. D'aquesta manera, els retroencreuaments enriqueixen el genoma dels descendents amb DNA parental de la soca amb baixa recombinació, mantenint els al·lels responsables per tret d'alta recombinació. Així, l'assaig de freqüències al·lèliques final requeriria de menys individus per a trobar les variants importants en el tret estudiat.

Anàlisi de gens candidats influenciant el nº de dNFs

Vam decidir analitzar els gens candidats utilitzant la tecnologia de seqüenciació. D'aquesta manera podríem detectar tant variants comunes com rares, el qual ens interessava per a no perdre cap variant que pogués afectar el nostre tret.

En referència als gens seleccionats a partir de la literatura, vam analitzar el gen *BLM* en 12 pacients NF1. Es van trobar 5 variants, 4 de sinònimes i una que generava una mutació de canvi d'amino àcid (missense en anglès). Aquest canvi, però, era d'una

Valina per una Isoleucina, les quals tenen una estructura mol·lecular molt similar, no afectant així al fnucionament general de la proteïna.

A més, en el transcurs d'aquesta tesi es va desenvolupar la tecnologia d'ultra-seqüenciació. Vam aprofitar la posada a punt d'aquesta tècnica en el nostre laboratori per a analitzar pacients NF1. La ultra-seqüenciació ens permetia analitzar fins a 106 gens relacionats amb CSs i RASopaties, dels quals 39 estaven relacionats amb mechanismes de reparació, i 9 amb el mecanisme de recombinació homòloga. Durant la finalització d'aquesta tesi, es van obtenir les dades per a un pacient, el P027. Com a il·lustració de la identificació de variants, vam detectar, d'entre 845 variants, de les quals 63 deletèrees, la mutació germinal ja descrita pel pacient P027, la inserció d'una G provocant el truncament de la proteïna.

DISCUSIÓ

L'objectiu de la present tesi era identificar gens involucrats en la variació en el nombre de dNFs desenvolupats en pacients NF1. L'estudi es va dividir en 3 parts: Primer, la caracterització fenotípica dels pacients a nivell clínic i molecular tumoral. Segon, la selecció de gens candidats relacionats amb la taxa de HR utilitzant dues aproximacions, una basada en la literatura, i l'altra en el modelatje de la HR en llevats. Tercer, l'anàlisi dels gens candidats per seqüenciació, obtenint totes les variants que poden afectar a la variació en el nº de dNFs desenvolupats.

Caracterització fenotípica clínica i molecular dels pacients NF1

Una correcte caracterització clínica era necessària per a estudiar qualsevol present o futura correlació genotip-fenotip del pacients NF1 relacionat amb el nombre de dNFs desenvolupats. És per això que es va implementar un mètode per a standaritzar el contatge de dNFs desenvolupats, i es va desenvolupar un mètode per a corregir el nombre de dNFs per l'edat, el valor de RDTB. Ara mateix no s'ha determinar cap relació entre el nº de dNFs i altres manifestacions severes, com MPNSTs. Malgrat tot, sí que es va trobar una correlació entre ens neurofibromes subcutanis i un alt volum intern de neurofibromes [159]. Per tant, seria interessant afegir la caracterització clínica

de neurofibromes subcutanis, el que podria ajudar a identificar informació genètica rellevant per els pacients.

Vam caracteritzar els diferents mecanismes generadors de LOH en els pacients. Dels 606 dNFs analitzats, un 23.7% mostraven LOH, en concordança amb altres estudis [138,161]. 53 dNFs (~37% de tumors amb LOH) la LOH era generada per deleció, amb tamanys de ~80Kb fins a ~8Mb. Durant aquest estudi es va identificar un neurofibroma (P082-1N) amb una deleció tipus 2. Aquest tipus de deleció s'havia identificat com a mutació germinal en pacients mosaïc, éssent la primera vegada que es troba com a segon hit en el gen *NF1*. 91 dNFs (~63% de tumors amb LOH) la LOH era generada per HR, amb un sól entrecreuament entre el centròmer i el gen *NF1*, reduïnt la mutació NF1 constitucional a homozigosi. La LOH s'extenia, en tots el casos, fins al final del cromosoma 17. Aquest anàlisi mostra que la HR és el mecanisme mutacional més important generador de LOHs amb un ~15%, mentre que les delecions generent un ~9% de les mutacions somàtiques. La resta de mutacions deuen ser degudes a mutacions puntuals, o petites delecions, ja que la nostra metodologia no permet identificar delecions més petites que l'espai present entre els microsatèl·lits analitzats.

L'anàlisi per SNP-arrays va mostrar que les alteracions en d'altres zones del genoma no són comunes en els neurofibromes dèrmics. De 19 dNFs analitzats, tan sols es van trobar alteracions, a part de les pròpies del gen *NF1*, en dos: una deleció en la regió 2q24.2-31.1 en el tumor P090-3N, i una altra deleció en la regió 3q11.2-q22.1 en el tumor P072-1N. No pensem que aquestes alteracions tinguin cap relació amb la generació del neurofibroma.

En el nostre anàlisi es van poder estudiar fins a 18 pacients per les freqüències de LOH. Aquestes mostraven certa variabilitat, amb pacients mostrant mensy de 10%, i d'altres mostrant més de 50%. A més, la majoria de pacients mostraven un mecanisme predominant en la generació de LOHs. Aquesta variablitat pot ser expliacada per la forma de reparar l'ADN. La HR s'activa per a reparar DSBs. Aquests, a més de per HR, poden ser reparats per NHEJ i SSA, els quals generen pèrdua en el material genètic. És possible que la variabilitat en les freqüències de LOH en els pacients NF1 sigui deguda a diferències en la capacitat de reparar els DSBs per els diferents mecanismes.

Els estudis de correlació van suggerir que els pacients mostrant una freqüència més elevada de HR generadora de LOH, podrien generar un nombre més elevat de dNFs. Degut al baix nombre de pacients analitzats, aquests resultats s'han de prendre com a preliminars. Tenir una alta freqüència de HR, però, implica tenir un millor o pitjor funcionament del mecanisme? Una alta freqüència de HR podria ser deguda a un mal funcionament dels altres mecanismes responsables de reparar DSBs, o un mal funcinament del propi mecanisme de HR, el qual podria reparar en excés alteracions que podrien ser reparades més eficaçment per altres mecanismes. Si considerem que la freqüència de HR en cèl·lules de Schwann serà similar que en altres cèl·lules, podríem pensar que la freqüència de HR generadora de LOHs podria estar relacionada amb la taxa de HR del pacient. D'aquesta manera, si els pacients amb més altes freqüències de HR són els que generen un nº elevat de dNFs, podríem concloure que la taxa de HR està relacionada amb el nº de dNFs desenvolupats.

Es va crear una tècnica, la MMPA, per a analitzar el fenotip molecular del dNFs d'un forma més ràpida i barata. Tots els càlculs en l'anàlisi de MMPA estan basats en la Kμ, la constant de proporcionalitat en la reacció de PCR. Un dels factors més importants és preparar la PCR amb una variació de Kμ mínima. S'haurien de tenir al menys 3 microsatèl·lits control per a realitzar l'anàlisi, els quals podrien ser pre-establerts, o utilitzar microsatèl·lits sense AI. A més, s'ha d'anar amb compte amb aquells microsatèl·lits amb tan sols 2pb de diferència entre els pics, ja que la proximitat entre ells pot alterar les intensitats en l'anàlisi, afectant als resultats obtinguts.

El re-anàlisi de 29 neurofibromes amb la tècnica de MMPA va donar uns resultats concordants amb la informació obtinguda a partir de tècniques com MLPA, PRA, o SNP-array per a les mateixes mostres. Pel que fa al percentage de cèl·lules sense l'alteració, tots els tumors, menys un, tenien valors obtainguts amb la tècnica de MMPA que no diferien més d'un 8% amb els obtinguts a partir de l'anàlisi de SNP-array. Pel que fa a la determinació del mecanisme mutacional, de tots els microsatèl·lits analitzats, un 82% van mostrar correctament el mecanisme mutacional, un 15% eren fora de rang, i tan sols un 2% van errar en l'estimació (Taula 6).

Selecció de gens candidats influenciant el nombre de neurofibromes

Vam utilitzar el llevat com a model per a estudiar la taxa de HR. Vam crear l'assaig HoReYe, el qual ens permetia determinar la taxa de HR en diferents soques de llevat. La tècnica es va poder utilitzar en una única soca, per la qual es va obtenir una taxa de HR de $1.21 \cdot 10^{-4}$ esdeveniments/cèl·lula/generació. Aquest resultat és preliminar, ja que s'hauria d'aplicar l'anàlisi de fluctuació per evitar biaixos degut a mutacions primerenques. Aquesta tècnica s'hauria d'aplicar a diferents soques de llevat, per tal d'obtenir soques mostrant altes i un baixes taxes de HR. L'assaig d'HoReYe, però, és laboriós, i requereix de múltiples modificacions genètiques. És per això que vam idear un assaig alternatiu que ens podria proporcionar el mateix tipus d'informació, una tècnica sustitutiva. Aquesta consistiria en analitzar la pèrdua per HR de dos marcadors de selecció endògens, els gens URA3 i CAN1. Comparant els resultats entre ambdues tècniques en differents soques de llevat ens demostraria si realment aquesta aproximació alternativa seria fiable per a determinar la taxa de HR en el llevat.

Una vegada les soques estiguessin caracterizades per la taxa de HR, es procediria a realitzar l'assaig de X-QTL per a identificar les variants gèniques responsables de l'elevada taxa de HR.

Anàlisi de gens candidats influenciant el nombre de neurofibromes

En afegit a l'aproximació del modelatge en llevat, vam analitzar gens obtinguts d'una selecció a partir de la literatura. A més, es va utilitzar la seqüenciació de pròxima generació per a estudiar 106 gens, dels quals 39 estaven involucrats en vies de reparació de l'ADN, 9 relacionats amb el mecanisme de HR.

Vam seqüenciar el gen de *BLM* de 12 pacients (Taula 4). Es van trobar 5 variants, de les quals 4 eren sinònimes, i tan sóls una no ho era. Aquesta última, però, era un canvi Valina-Isoleucina, el qual no afecta a l'estructura de la proteïna.

A més es van obtenir resultats en l'anàlisi per NGS del pacient P027. Aquests resultats, però, van arribar en la finalització de la present tesi, i és per això que no es va poder fer cap tipus d'anàlisi. Com a prova del bon funcionament de la tècnica, però, es va identificar la mutació germinal a *NF1*, la inserció d'una G creant un corriment de lectura. L'anàlisi per NGS s'hauria de realitzar en la resta de 17 pacients NF1 caracteritzats.

CONCLUSIONS

1. En referència al fenotip clínic dels pacients NF1, hem desenvolupat una metodologia estandaritzada per estimar el nombre de neurofibromes dèrmics (dNFs) desenvolupats, i hem determinat el valor RDTB (Risk of Dermal Tumor Burden, en anglès), per tal de evitar els biaixos generats per l'edat dels pacients.

2. Hem obtingut una imatge general de la importància de la pèrdua d'heterozigositat com a evidència de second hits a neurofibromes, i també hem disseccionat les freqüències dels mecanismes mutacionals generadors de LOH en la nostra cohort de pacients.

3. La recombinació homòloga és el mecanisme generador de LOH més freqüent en dNFs, éssent responsable d'un 63% d'aquestes LOHs.

4. Un 37% de dNF-LOHs són degudes a la pèrdua del nombre de còpies del locus *NF1*, amb delecions que van de 80Kb a 8Mb.

5. Encara que en pocs casos, dNF pot patir un second hit en el gen *NF1* causat per una recombinació homòloga no al·lèlica, generant una microdeleció tipus 2.

6. Grans alteracions genomiques en altres regions, apart de en la regió del gen *NF1*, es troben en molt poca freqüència en els dNFs, sense cap recurrència identificada en els 20 dNFs analitzats. A més, les grans alteracions trobades no sembla que tinguin relació amb el desenvolupament del neurofibroma dèrmic.

7. Les freqüències de LOH individuals en els dNFs mostren una alta variabilitat, amb pacients mostrant freqüències de LOH que van de 0% fins a 55.5%.

8. Molts pacients mostren una mecanisme causant de LOH preferencial en els seus dNFs, éssent la recombinació homòloga el més freqüent.

9. Els nostres resultats suggereixen una possible correlació entre tenir un alt percentatge de neurofibromes amb una LOH generada per recombinació homòloga, i desenvolupar un elevat nombre de neurofibromes en pacients NF1.

10. Una nova metodologia ha estat desenvolupada, la Microsatelite Multiplex PCR Analysis (MMPA), la qual facilita l'anàlisi genètic de tumors i la caracterització molecular d'aquests en pacients NF1.

11. La tècnica de MMPA ens permet obtenir informació acurada, no solament sobre l'estatus de desequilibris al·lèlics (AI en anglès) dels tumors, sino també del percentatge de cèl·lules sense AI, el nombre de còpies del gen *NF1* i el mecanisme mutacional responsable de l'AI.

12. Un nou assaig (HoReYe Assay) ha estat desenvolupat per a determinar la taxa de recombinació homòloga *in vivo*, d'entre altres mecanismes, en diferents soques de llevat. Aquest assaig imita el mecanisme de recombinació homòloga generador de LOH identificat en els pacients NF1. A més, una nova estratègia ha estat disenyada experimentalment per tal d'identificar gens candidats afectant les taxes de recombinació homòloga en les soques de llevat.

ANNEX

BLM gene analysis

Primer name	Primer sequence
BLM PRIMER D1 5'UTR	GTACCTCGCACTTACAGACAG
BLM PRIMER R1 5'UTR	GTCTCTGTTTCACCCGTACC
BLM PRIMER D2 5'UTR	CATGGAGGCATCTGAGTGTG
BLM PRIMER R2 5'UTR	AGCTTGCCTATTCCGATTGG
BLM PRIMER D 3'UTR	GCAAGAGCTTCTGAGCATAAC
BLM PRIMER R 3'UTR	AACTCCTGACCTCGTGATCC
BLM PRIMER D1	GGATTATGGCTGCTGTTCCTC
BLM PRIMER R1	CTTTGGGGTGGTGTAACAAATG
BLM PRIMER D2	ACTCCGAAGGAAGTTGTATGC
BLM PRIMER R2	CTGGAGAAGGTGGAACAAAATC
BLM PRIMER D3	AGATGCTCAGGAAAGTGACTC
BLM PRIMER R3	GTTCGTTCCCACAATCCAAAAG
BLM PRIMER D4	CAGCAGCAGCTTATTCATGTG
BLM PRIMER R4	GTGGAAGATTTGCTGGCTGC
BLM PRIMER D5	TGATGATGATGACTGGGAAGAC
BLM PRIMER R5	TGAGTCAGTCTTATCACCTGTC
BLM PRIMER D6	GGTCACTGTTGTCATTTCTCC
BLM PRIMER R6	GCCACCTTTTTAGGCTTTTTCG
BLM PRIMER D7	AAGATTCTCAGACCTCAGGTG
BLM PRIMER R7	CCAAAAGCTGTATTCTCCTGC
BLM PRIMER D8	CACTGCCTGCTTTTCTATACC
BLM PRIMER R8	CAGTTTGGGCTTTATTTCCGAG
BLM PRIMER D9	ACGACACAATGCCGAAAGAC
BLM PRIMER R9	CTTCCGCACCATATTTTTCCAG
BLM PRIMER D10	TACCGTCACTCTCAAGAAGC
BLM PRIMER R10	TTTGCTATTGGCTCCTGATGTC
BLM PRIMER D11	CTTCCAAAACGAAATCCTCCAG
BLM PRIMER R11	CTCTTTTCCCCGTCATTTGTC

PRA analysis

Primer name	Primer sequence
Pseudo NF1 F	TCTACCAACTGCTCTGGAACAT
Pseudo NF1 R-2	GAATGATACAAACCTTGAAACAG

Tables annex A1. Primers used in different processes in this thesis

Yeast genes transfomation

Primer name	Primer sequence
LLextFcen	GCGCGCGTAATACGACTCACTATAGGGCGAATTGGGTACCGGCGCGCCTATATGGTGCCTT TCGCAATG
LLextRcen	CAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGGCGCGCCAGTCATTCTTCTT GTTGCTGC
LLintFcen	TCGCTTAGCTACGCGTATCGGCGGGCATATTGAATTG
LLintRcen	GCCCGCCGATACGCGTAGCTAAGCGATAGCTCTTTC
URA Fcen	ACCTTTTTTTTAAAATTTTTGAAAGAGCTATCGCTTAGCTACGCGTGCACCATACCACAGCTT TTC
URA Rcen	GAAATGTTCATCTACTGCGCAATTCAATATGCCCGCCGATGCGGCCGCCGCATCTGTGCGG TATTTC
CAN1cen F-3	AGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGGCGGCCGCTTGTGTGTATGGG CACAAACC
CAN1cen R-3	GAAATGTTCATCTACTGCGCAATTCAATATGCCCGCCGATGTTTAAACAGTGTATGTTGCGCT TACTAC
ADE Fcen-2	ATGCTAAGCTATTTCCAAGGGTTTGTGCCCATACACACAAGTTTAAACCCGCATAGATCTTAT GTATG
ADE Rcen-2	GAAATGTTCATCTACTGCGCAATTCAATATGCCCGCCGATGAGCTCTAGATCTGAATTAATTC TTG
HIS Rcen-2	GAAATGTTCATCTACTGCGCAATTCAATATGCCCGCCGATCTCGAGAGATCCGTCGAGTTCA AGAG
HIS Fcen-3	AAGAAAAGTTATGTATTATTCAAGAATTAATTCAGATCTAGAGCTCCGTTTTAAGAGCTTGGTG AG
TRP F3cen	ACCTTTTTTTTAAAATTTTTGAAAGAGCTATCGCTTAGCTACGCGTGCACCATAAACGACATTA CT
TRP R3cen	GAAATGTTCATCTACTGCGCAATTCAATATGCCCGCCGATGCGGCCGCAGGCAAGTGCACA AACAATAC
Scree TRP1 F	CTTTGCTCACCGAATGGCAC
Scree TRP1 R	ATGGCAATCACCGAAGACG
Nat1braçp F	CTTACGTTAGCGAGTACTCGAATGAGGGTGAAATGGCATCGTCATTGATTTAAGGCGCGCCA GATCTGTT
Nat1braçp R	TCAGTAAGGCCAACCATAAACACCTCAATGTTGGGGAGTCTGTCCTCGGCCATCGATGAATT CGAGCTCG
Natscreen-right F	ACCTCTGGCTGGAGGTCAC
Natscreen-rightt R	CTCCACGTTCACGATTTAAGG
Natscreen-left F	TGACTCAGAAACCGTACTGC
Natscreen-left R	GTACGAGACGACCACGAAGC
Kan braçllarg F	AACGCTCTATACTAATGCGCATGAAGATTTTACTTGCGAATCAAACTGATCGGATCCCCGGG TTAATTAA
Kan braçllarg R	CATACGGCGAGATGTAAGGTAGGTATAGCAGACTTAGATATTGCGTGCGAGAATTCGAGCTC GTTTAAAC
ScreeKan braçllarg-F	ACTCACGCAAGGGATTATGG
ScreeKan braçllarg-R	TTCTGCGTCGTTCCTTCAC

Tables annex A1. Primers used in different processes in this thesis

CGL105 insertion screening

Primer name	Primer sequence
Scree pCGL105 1 F	AGACGGTCACAGCTTGTCTG
Scree pCGL105 1 R	TTCCAACAGTACCACCGAAG
Scree pCGL105 2 F	CATTTATAAAGTTTATGTAC
Scree pCGL105 2 R	CCACCATTGCCTATTTGGTC
Scree pCGL105 3 F	GTTCTTGATGGTTTCCTCCAC
Scree pCGL105 3 R	CAAGAAATCACAGCCGAAGC
Scree pCGL105 4 F	CTTGACATTGGAACGAACATC
Scree pCGL105 4 R	GTAATTGGTTGTTTGGCCGAG
Scree pCGL105 5 F	ACGATACCTGAGTATTCCCAC
Scree pCGL105 5 R	TGCTTTACGGCACCTCGAC
Scree pCGL105 6 F	GCCCACTACGTGAACCATC
Scree pCGL105 6 R	GTGCACGCGTAGCTAAGCGA
Scree pCGL105 7 F	CCATAATCATTCCAATTCAG
Scree pCGL105 7 R	GTAATGTCTGCCCATTCTGC
Scree pCGL105 8 F	GGTAATACAGTCAAATTGCAG
Scree pCGL105 8 R	TCTGTGCGGTATTTCACACC
Scree pCGL105 9 F	CTAAACTCACAAATTAGAGC
Scree pCGL105 9 R	CAATGCATATTCAGATGCAG
Scree pCGL105 10 F	GGAATCGATGTCGACATCTC
Scree pCGL105 10 R	GCCACATATCTTCAACGCTG
Scree pCGL105 11 F	TGCGGCAGAAATAATGGTTG
Scree pCGL105 11 R	TGCCCTGGAACTTAGTGTAG
Scree pCGL105 12 F	CGCCAGTGGAACTTTGTACG
Scree pCGL105 12 R	AGACGCCGACATAGAGGAG
Scree pCGL105 13 F	GGTTGTGACCGGCTCATTG
Scree pCGL105 13 R	AACGGCAAACAGCAAAGG
Scree pCGL105 14 F	GCACTTGCCAGTAAGTGATTG
Scree pCGL105 14 R	TCGGTTACGAAGCTTATCTAG
Scree pCGL105 15 F	AAGGACACCTGTAAGCGTTG
Scree pCGL105 15 R	ACTCTGACTTGCCGGTAATG
Scree pCGL105 16 F	TGTGGAGTTCTATGAGCAGAG
Scree pCGL105 16 R	ATGTTATGCGCCTGCTAGAG
Scree pCGL105 17 F	CGGGAAAAGATTTGATTGC
Scree pCGL105 17 R	TGGGACGTATGATTGTTGAGG
Scree pCGL105 18 F	CAGCATCTAGTATTACCGTC
Scree pCGL105 18 R	GTGTACTAGAGGAGGCCAAG
Scree pCGL105 19 F	CGGCATTAGTCAGGGAAGTC
Scree pCGL105 19 R	CAAACCAAGTTCGACAACTGC
Scree pCGL105 20 F	ATGAGGCACTTTCCAGAGC
Scree pCGL105 20 R	TGACCATGATTACGCCAAGC
Scree pCGL105 21 F	CAGCAACAAGAAGAATGACTGG
Scree pCGL105 21 R	GCTTCCCGAAGGGAGAAAGG
Scree pCGL105 22 F	AGCTCCCTCGTGCGCTCTCC
Scree pCGL105 22 R	GGATCTAGGTGAAGATCCT
Scree pCGL105 23 F	GTCTGACGCTCAGTGGAACG
Scree pCGL105 23 R	ACGGATGGCATGACAGTAAG
Scree pCGL105 24 F	AAGTTGGCCGCAGTGTTATC
Scree pCGL105 24 R	CCTAGAGTCTTTTACATCTTC
Scree pCGL105 25 F	GGTCCTTTTCATCACGTGC
Scree pCGL105 25 R	CCTGACGGGCTTGTCTGCTC

Tables annex A1. Primers used in different processes in this thesis

Delitto Perfetto

Primer name	Primer sequence
P1 CAN1 DP	GCATAGCAATGACAAATTCAAAAGAAGACGCCGACATAGAGGAGAAGCATGAGCTCGTTTTCGACACTGG
P2 CAN1 DP	CGTCATATCTATGCTACAACATTCCAAAATTTGTCCCAAAAAGTCTTTGGTCCTTACCATTAAGTTGATC
P3 CAN1 DP	GCATAGCAATGACAAATTCAAAAGAAGACGCCGACATAGAGGAGAAGCAT
P4 CAN1 DP	CGTCATATCTATGCTACAACATTCCAAAATTTGTCATGCTTCTCCTCTAT
P1 ADE2 DP	GGACAAAACAATCAAGTATGGATTCTAGAACAGTTGGTATATTAGGAGGGGAGCTCGTTTTCGACACTGG
P2 ADE2 DP	CTTATATATTACTTGTTTTCTAGATAAGCTTCGTAACCGACAGTTTCTAATCCTTACCATTAAGTTGATC
P3 ADE2 DP	GGACAAAACAATCAAGTATGGATTCTAGAACAGTTGGTATATTAGGAGGG
P4 ADE2 DP	CTTATATATTACTTGTTTTCTAGATAAGCTTCGTACCCTCCTAATATACC
Scree CAN1DP1 L-F	AGAGTGGTTGCGAACAGAG
Scree CAN1DP1 L-R	AACCTAAGAACCTCCCTTCG
Scree CAN1DP1 R-F	GCTGGTGAAGCTGCAAAC
Scree CAN1DP1 R-R	GAGAATGCGAAATGGCGTGG
LYS5-DP P1	ACTAAAATGAGAATGAGGAGAAAGAGAATCACTCAGCAAAAAAACCGTGGGAGCTCGTTTTCGACACTGG
LYS5-DP P2	AGAGTCTATCGATTACATAAATGTGAGCAAGCGAAAAAAAAAAATTGGCATCCTTACCATTAAGTTGATC
LYS5-DP P3	ACTAAAATGAGAATGAGGAGAAAGAGAATCACTCAGCAAAAAAACCGTGG
LYS5-DP P4	CATAAATGTGAGCAAGCGAAAAAAAAAAATTGGCACCACGGTTTTTTTGC
Scree LYS5 F	TCATCGCCTACGATACCGTC
Scree LYS5 R	GAACAAGTTCGCTGATACGTG

Tables annex A1. Primers used in different processes in this thesis

Nº of dNFs	Percentage of NF1 patients			
0	85	42	7	0
<10	15	46	13	0
11-99	0	12	55	12
100-500	0	0	23	59
+500	0	0	2	29
Rank of ages	<10 (n=20)	10-19 (n=43)	20-39 (n=56)	>40 (n=41)

Table annex A2. Data of dNFs developed in relation to the age of the NF1 patients for UK and Southern Sales populations.

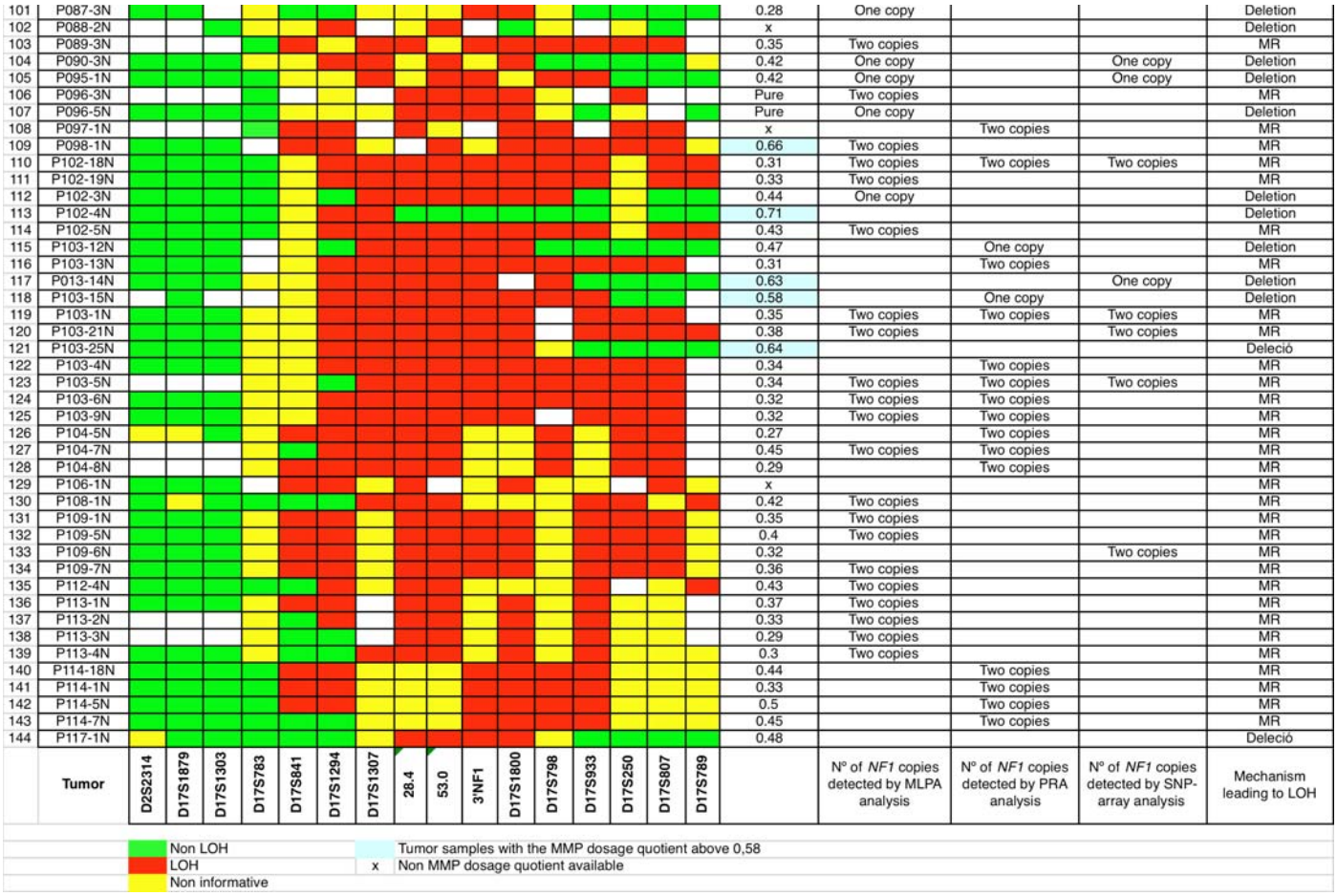
Nº of dNFs	Percentage of NF1 patients			
0	93	64	18	7
<10	7	28	28	15
11-99	0	8	38	37
>100	0	0	17	41
Rank of ages	<10 (n=135)	10-19 (n=156)	20-39 (n=245)	>40 (n=214)

Table annex A3. Data of dNFs developed in relation to the age of the NF1 patients for French population.

Nº of dNFs	Percentage of NF1 patients								
0	82	66	5	25	0	0	0	0	0
<10	18	34	55	55	10	0	0	0	0
11-99	0	0	35	20	40	15	55	0	0
100-500	0	0	5	0	45	85	30	60	80
+500	0	0	0	0	5	0	15	40	20
Rank of ages	5-10	10-15	15-20	20-30	30-40	40-50	50-60	60-70	70-80

Table annex A4. Data of dNFs developed in relation to the age of the NF1 patients for Southern Wales population.

		17p			17q				NF1 gene			17q									
	Tumor	D2S2314	D17S1879	D17S1303	D17S783	D17S841	D17S1294	D17S1307	28.4	53.0	3'NF1	D17S1800	D17S798	D17S933	D17S250	D17S807	D17S789	MMP Q^{LOH} median value	Nº of *NF1* copies detected by MLPA analysis	Nº of *NF1* copies detected by PRA analysis	Nº of *NF1* copies detected by SNP-array analysis	Mechanism leading to LOH
1	P001-1N																	0.29	Two copies			MR
2	P002-1N																	0.2	Two copies	Two copies		MR
3	P002-2N																	0.22		Two copies		MR
4	P004-5N																	0.46	One copy	One copy		Deletion
5	P004-7N																	0.27	One copy			Deletion
6	P007-1N																	0.35	Two copies			MR
7	P008-1N																	x				Deletion
8	P009-1N																	0.74				Deletion
9	P010-2N																	x				Deletion
10	P011-11N																	Pure		Two copies		MR
11	P011-16N																	0.52	Two copies		Two copies	MR
12	P011-26N																	0.64		Two copies		MR
13	P011-28N																	0.7		Two copies		Deletion
14	P011-37N																	0.66				Deletion
15	P011-39N																	0.47				MR
16	P011-4N																	0.37	Two copies	Two copies		MR
17	P011-7N																	0.43				Deletion
18	P011-9N																	0.32	Two copies	Two copies		MR
19	P013-5N																	0.52		Two copies		MR
20	P016-2N																	x				Deletion
21	P018-1N																	x				Deletion
22	P018-3N																	x		One copy		Deletion (SSCP)
23	P020-2N																	x		Two copies		MR
24	P020-9N																	0.57		Two copies		MR
25	P022-11N																	0.26	Two copies	Two copies		MR
26	P022-15N																	0.28				MR
27	P022-17N																	0.6		One copy		Deletion
28	P022-19N																	0.33	Two copies			MR
29	P022-1N																	0.23	Two copies	Two copies		MR
30	P022-21N																	0.39	Two copies			MR
31	P022-2N																	0.41	One copy	One copy		Deletion
32	P022-6N																	0.26		Two copies		MR
33	P022-8N																	Pure	Two copies			MR
34	P023-14N																	0.29		Two copies	Two copies	MR
35	P023-16N																	x	Two copies	Two copies		MR
36	P023-21N																	0.33	Two copies	Two copies		MR
37	P023-25N																	0.27	Two copies	Two copies		MR
38	P023-38N																	x				Deletion
39	P023-42N																	0.25		Two copies		MR
40	P023-4N																	0.22	Two copies	Two copies		MR
41	P023-51N																	0.3	Two copies	Two copies		MR
42	P023-52N																	0.31	Two copies	Two copies		MR
43	P023-5N																	0.26	Two copies	Two copies		MR
44	P023-62N																	0.42	One copy			Deletion
45	P023-6N																	0.36	One copy	One copy	One copy	Deletion
46	P023-97N																	0.4		One copy	One copy	Deletion
47	P023-99N																	0.43	One copy			Deletion
48	P024-1N																	x				MR
49	P027-4N																	0.48	Two copies		Two copies	MR
50	P029-1N																	x		One copy		Deletion (SSCP)
51	P029-2N																	x		One copy		Deletion (SSCP)
52	P030-2N																	0.26	Two copies			MR
53	P033-1N																	0.52				Deletion
54	P034-1N																	x	Two copies			MR
55	P034-4N																	0.56		One copy		Deletion
56	P039-1N																	0.48	One copy		One copy	Deletion
57	P040-1N																	Pure	One copy			Deletion
58	P041-2N																	Pure	Two copies			MR
59	P044-3N																	0.57	Two copies			MR
60	P044-4N																	0.56	One copy			Deletion
61	P045-1N																	0.41	One copy			Deletion
62	P047-1N																	0.37	Two copies			MR
63	P047-5N																	0.32		Two copies		MR
64	P047-6N																	0.46		Two copies		MR
65	P047-8N																	0.34		Two copies		MR
66	P052-4N																	0.52	One copy			Deletion
67	P054-1N																	0.45		Two copies		MR
68	P055-2N																	0.45		Two copies		MR
69	P056-1N																	0.71	Two copies			MR
70	P058-10N																	0.39	Two copies			MR
71	P058-7N																	0.35	One copy			Deletion
72	P062-11N																	0.35	One copy			Deletion
73	P062-1N																	0.35	Two copies	Two copies		MR
74	P062-3N																	0.42		Two copies		MR
75	P062-4N																	Pure	One copy			Deletion
76	P062-6N																	x		Two copies		MR
77	P062-8N																	0.45	Two copies			MR
78	P065-1N																	x	One copy			Deletion
79	P067-2N																	Pure	One copy			Deletion
80	P068-2N																	0.28	Two copies			MR
81	P068-3N																	0.38	Two copies			MR
82	P072-1N																	0.35	One copy		One copy	Deletion
83	P078-1N																	0.32		Two copies		MR
84	P078-23N																	0.3		Two copies		MR
85	P078-31N																	0.36	Two copies	Two copies		MR
86	P078-34N																	Pure		Two copies		MR
87	P078-39N																	0.3		Two copies		MR
88	P078-45N																	0.3	Two copies			MR
89	P079-1N																	Pure	One copy		One copy	Deletion
90	P081-1N																	0.55	One copy			Deletion
91	P081-2N																	0.45	One copy			Deletion
92	P082-11N																	0.66		One copy		Deletion
93	P082-12N																	0.71		One copy	One copy	Deletion
94	P082-1N																	0.53		One copy		Deletion
95	P082-6N																	0.6		One copy	One copy	Deletion
96	P082-8N																	x		One copy		Deletion
97	P084-2N																	0.33	Two copies			MR
98	P084-5N																	0.35		Two copies		MR
99	P085-2N																	0.74	Two copies			MR
100	P086-11N																	0.34		Two copies		MR
101	P087-3N																	0.28	One copy			Deletion
102	P088-2N																	x				Deletion



	Tumor	D2S2314	D17S1879	D17S1303	D17S783	D17S841	D17S1294	D17S1307	28.4	53.0	3'NF1	D17S1800	D17S798	D17S933	D17S250	D17S807	D17S789		N° of *NF1* copies detected by MLPA analysis	N° of *NF1* copies detected by PRA analysis	N° of *NF1* copies detected by SNP-array analysis	Mechanism leading to LOH
101	P087-3N																	0.28	One copy			Deletion
102	P088-2N																	x				Deletion
103	P089-3N																	0.35	Two copies			MR
104	P090-3N																	0.42	One copy		One copy	Deletion
105	P095-1N																	0.42	One copy		One copy	Deletion
106	P096-3N																	Pure	Two copies			MR
107	P096-5N																	Pure	One copy			Deletion
108	P097-1N																	x		Two copies		MR
109	P098-1N																	0.66	Two copies			MR
110	P102-18N																	0.31	Two copies	Two copies	Two copies	MR
111	P102-19N																	0.33	Two copies			MR
112	P102-3N																	0.44	One copy			Deletion
113	P102-4N																	0.71				Deletion
114	P102-5N																	0.43	Two copies			MR
115	P103-12N																	0.47		One copy		Deletion
116	P103-13N																	0.31		Two copies		MR
117	P013-14N																	0.63			One copy	Deletion
118	P103-15N																	0.58		One copy		Deletion
119	P103-1N																	0.35	Two copies	Two copies	Two copies	MR
120	P103-21N																	0.38	Two copies		Two copies	MR
121	P103-25N																	0.64				Deleció
122	P103-4N																	0.34		Two copies		MR
123	P103-5N																	0.34	Two copies	Two copies	Two copies	MR
124	P103-6N																	0.32	Two copies	Two copies		MR
125	P103-9N																	0.32	Two copies	Two copies		MR
126	P104-5N																	0.27		Two copies		MR
127	P104-7N																	0.45	Two copies	Two copies		MR
128	P104-8N																	0.29		Two copies		MR
129	P106-1N																	x				MR
130	P108-1N																	0.42	Two copies			MR
131	P109-1N																	0.35	Two copies			MR
132	P109-5N																	0.4	Two copies			MR
133	P109-6N																	0.32			Two copies	MR
134	P109-7N																	0.36	Two copies			MR
135	P112-4N																	0.43	Two copies			MR
136	P113-1N																	0.37	Two copies			MR
137	P113-2N																	0.33	Two copies			MR
138	P113-3N																	0.29	Two copies			MR
139	P113-4N																	0.3	Two copies			MR
140	P114-18N																	0.44		Two copies		MR
141	P114-1N																	0.33		Two copies		MR
142	P114-5N																	0.5		Two copies		MR
143	P114-7N																	0.45		Two copies		MR
144	P117-1N																	0.48				Deleció

Non LOH
LOH
Non informative

Tumor samples with the MMP dosage quotient above 0,58
x Non MMP dosage quotient available

Table annex A5. Summary of neurofibromas showing LOH

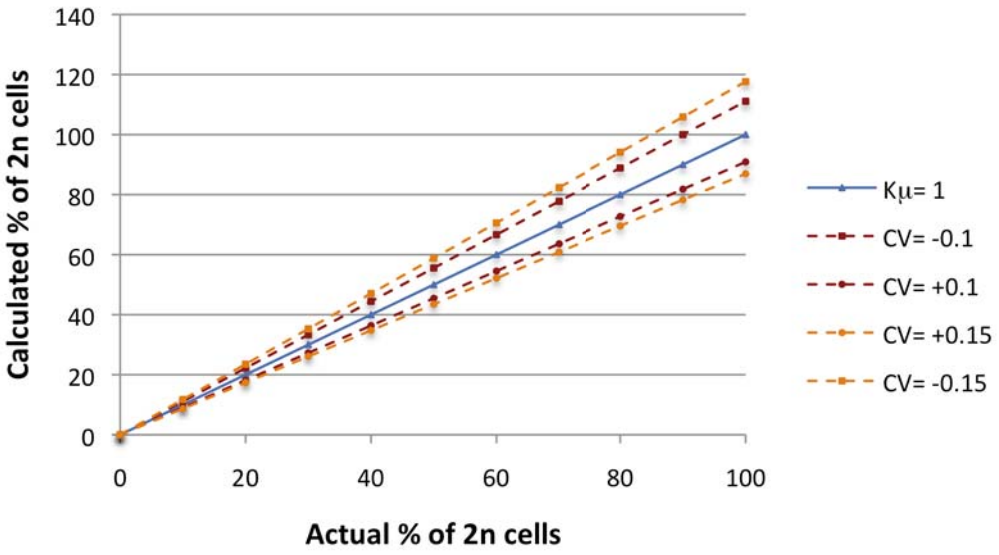
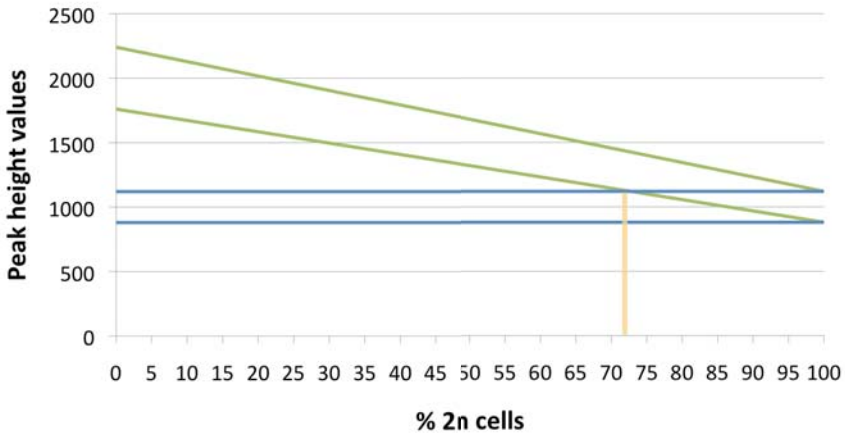


Figure annex A1. illustration of how variation on Kμ (by applying different CVs) affects the calculation of the % of 2n cells present in a tumor sample in a hypothetical reaction, for AI-microsatellites with one observed allele peak height lower than expected (copy-loss and copy-neutral events). In the presented hypothetical scenario the parameters used are: a Kμ=1; a microsatellite with a control allele peak height of 1000 (fluorescence intensity); a tumor peak height representing a % of 1000 (fluorescence intensity) proportional to the % of non-AI cells present in the tumor. Solid line: % of 2n cells for CV=0. Dashed lines: % of 2n cells for the different CVs applied. This figure shows that it is important to set up MMPA reactions with a low degree of variation on Kμ, since the greater the CV on Kμ the less accurate is the calculation of the percentage of 2n cells in the tumor.



$$\text{Theoretical limit \% of 2n cells within tumors for locus copy number determination} = \frac{2 - \frac{(K\mu + \varepsilon)}{(K\mu - \varepsilon)}}{0{,}01}$$

Figure annex A2. To determine the locus copy number of AI-microsatellites, two different intervals of expected allele peak height values are calculated, one for a copy-number loss scenario and another considering a copy-neutral event. The observed allele peak height will fit into one of the two expected intervals. However, depending on the percentage of normal cells present in the tumor and the φ used, the two different intervals can overlap, making it impossible to discern which is the mechanism generating AI. Figure A2 shows at which percentage of 2n cells these two intervals overlap for an hypothetical MMPA reaction. The parameters used were: Kμ=1, Φ=0.12, control peak height =1000 (fluorescence intensity), tumor peak height =1000 for copy-loss (one copy of the locus) and 2000 for copy-neutral (two copies of the locus). Blue and green lines define the intervals (K$\mu \pm \varepsilon$; ε= K$\mu \cdot \Phi$) of expected allele peak height values for copy-loss and copy-neutral mechanisms respectively, under different percentages of 2n cells. This figure shows that the lower the Φ of the MMPA reaction set up, the higher the sensitivity for differentiating between copy-loss and copy-neutral events. This is important when analyzing tumors with high percentages of infiltrating normal cells.