

## **INFORMATION TO USERS**

**This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.**

**The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.**

**In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.**

**Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.**

**Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.**

**ProQuest Information and Learning  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
800-521-0600**

**UMI<sup>®</sup>**

**IDENTIFICATION AND CHARACTERIZATION OF  
NEUROFIBROMATOSIS TYPE 1 (*NF1*) GENE  
MUTATIONS**

by

**LI JUAN FANG**

DEPARTEMENT DE MICROBIOLOGIE ET D'INFECTIOLOGIE

FACULTE DE MEDECINE

UNIVERSITE DE SHERBROOKE

SHERBROOKE, QUEBEC, CANADA

A THESIS SUBMITTED TO THE FACULTY OF MEDICINE IN  
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR A  
PHILOSOPHIAE DOCTOR (Ph.D.) DEGREE IN MICROBIOLOGY

OCTOBER 2000



**National Library  
of Canada**

**Acquisitions and  
Bibliographic Services**

**395 Wellington Street  
Ottawa ON K1A 0N4  
Canada**

**Bibliothèque nationale  
du Canada**

**Acquisitions et  
services bibliographiques**

**395, rue Wellington  
Ottawa ON K1A 0N4  
Canada**

*Your file Votre référence*

*Our file Notre référence*

**The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.**

**The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.**

**L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.**

**L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.**

**0-612-67105-4**

**Canada**

## **I. TABLE OF CONTENTS**

<b>I. Table of contents</b> .....	<b>i</b>
<b>II. List of Figures</b> .....	<b>v</b>
<b>III. List of Tables</b> .....	<b>viii</b>
<b>IV. Glossary of Abbreviations</b> .....	<b>ix</b>
<b>V. Abstract (French)</b> .....	<b>xiv</b>
<b>VI. Abstract (English)</b> .....	<b>xvi</b>
<b>Chapter 1 Introduction</b> .....	<b>1</b>
1.1 Neurofibromatosis type 1 (NF1) .....	1
1.1.1 Clinical features .....	1
1.1.2 The <i>NF1</i> gene .....	3
1.1.3 The <i>NF1</i> gene function .....	7
1.1.4 The <i>NF1</i> gene mutations .....	9
1.2 Pre-mRNA splicing and exon definition .....	13
1.3 Linkage disequilibrium.....	19
<b>Chapter 2 Patients, Materials and Methods</b> .....	<b>23</b>
2.1 NF1 Patients and NF1 families .....	23
2.2 Collection of blood samples and separation of leucocytes from blood.....	27



2.3 Cell Culture .....	28
2.4 Extraction of DNA and RNA.....	28
2.5 PCR, cloning and sequencing of genomic 11-12a of patient 5313 .....	29
2.6 Amplification of the polymorphic region by PCR, haplotypes and statistical analysis .....	31
2.6.1 Microsatellites .....	31
2.6.2 <i>RsaI</i> RFLP .....	31
2.6.3 Haplotype analysis.....	33
2.6.4 Statistical analysis.....	35
2.7 Southern blot, <i>EcoRI</i> and <i>HincII</i> polymorphism .....	37
2.8 Protein truncation test (PTT) .....	43
2.8.1 Transcription template synthesis .....	43
2.8.2 <i>In vitro</i> transcription and translation.....	47
2.8.3 cDNA sequencing.....	47
2.9 Splicing reaction.....	48
2.9.1 <i>In vivo</i> splicing .....	48
2.9.2 <i>In vitro</i> splicing .....	52
2.9.3 Native gels.....	52
<b>Chapter 3 Results and Discussion.....</b>	<b>56</b>
3.1 A novel mutation in the <i>NFI</i> gene promotes skipping of two exons by affecting exon definition .....	58
3.1.1 Results.....	58
3.1.2 Discussion .....	72

3.2 Identification and characterization of six novel deletions in the <i>NF1</i> gene .....	77
3.2.1 Results.....	77
3.2.2 Discussion .....	93
3.3 A novel and very peculiar <i>HincII</i> polymorphism in the 5' region of the human Neurofibromatosis type 1 ( <i>NF1</i> ) gene.....	97
3.3.1 Results.....	97
3.3.2 Discusssion.....	100
3.4 Linkage disequilibrium analysis between four intragenic polymorphic microsatellites of the <i>NF1</i> gene in the Quebec population.....	102
3.4.1 .....	102
3.4.2 .....	107
<b>Chapter 4 Conclusion.....</b>	<b>112</b>
<b>References .....</b>	<b>114</b>
<b>Acknowledgments .....</b>	<b>135</b>
<b>Appendix 1 A novel and very peculiar <i>HincII</i> polymorphism in the 5' region of the human Neurofibromatosis type 1 (<i>NF1</i>) gene, <i>Annale de genetics</i>, 42, 231-233, 1999 .....</b>	<b>136</b>
<b>Appendix 2 Caculation of allele frequency for Appendix 1.....</b>	<b>137</b>
<b>Appendix 3 Pedigree and haplotypes of the 19 <i>NF1</i> families from Québec.....</b>	<b>138</b>
<b>Appendix 4 Result file of LINKAGE program .....</b>	<b>149</b>

<b>Appendix 5 Result file of HAL program .....</b>	<b>153</b>
<b>Appendix 6 Result file of DISLAMB program .....</b>	<b>174</b>
<b>Appendix 7 Result file of GASSOC program .....</b>	<b>178</b>

## **II. LIST OF FIGURES**

**Figure 1 The *NF1* gene**

**Figure 2 Schematic representation of the Ras-Raf-MAPK mitogenic signaling pathway.**

**Figure 3 Schematic representation of the two-step splicing reaction**

**Figure 4 The sequential integration of snRNPs into the spliceosome**

**Figure 5 Exon definition in pre-mRNA with small exons and large introns**

**Figure 6 Schematic chart of experimental design**

**Figure 7 The *RsaI* polymorphism at exon 5**

**Figure 8 The seven *NF1* cDNA probes**

**Figure 9 The FF13 800 bp and FF13 625 bp probes for characterization of deletion 7610**

**Figure 10 The FF1 1.0 kb and FF1 430 bp probes for characterization of mutation 5313**

**Figure 11 Schematic representation of protein truncation test**

**Figure 12 Strategy used to clone genomic 11-12a into pDUP4-1 for characterization of mutation 5313**

**Figure 13 Strategy used to clone genomic 12a into pDUP4-1 for characterization of mutation 5313**

**Figure 14 Strategy used to clone genomic 12a into pSPAd for characterization of mutation 5313**

**Figure 15 Schematic diagram of the *in vitro* splicing**

**Figure 16 Strategy for identification and characterization of *NF1* gene mutations**

**Figure 17 Protein truncation test results**

**Figure 18 Protein truncation and heteroduplex formation assays**

**Figure 19 Sequence of the mutated *NF1* allele**

**Figure 20 Pedigree and haplotype of family 5313**

**Figure 21 Southern blot results of mutation 5313**

**Figure 22 Theoretical splicing patterns of constructs with wild type *NF1* exons 11 and 12a in DUP4-1**

**Figure 23 *In vivo* splicing of *NF1* exons 11 and 12a.**

**Figure 24 Sequences of splice junctions in genomic *NF1* from exons 10c to 12b**

**Figure 25 Mutation 2056+1G→A promotes skipping of exons 11 and 12a**

**Figure 26 The 5' splice site of exon 12a plays a role in exon definition**

**Figure 27 Mechanism of loss of heterozygosity (LOH) due to deletion**

**Figure 28 Detection of the deletion breakpoints by Southern blot**

**Figure 29 Pedigree and haplotype of family 7610**

**Figure 30 Southern blot analysis of deletion 7610**

**Figure 31 Pedigree and haplotype of family 7473**

**Figure 32 Southern blot densitometry results of deletion 7473**

**Figure 33 Southern blot analysis of patient 178**

**Figure 34 Results of loss of heterozygosity analysis for non-family cases (178, 184, 236 and 237) with four microsatellite markers**

**Figure 35 Southern blot analysis of patient 184**

**Figure 36 Southern blot analysis of patient 236**

**Figure 37 Southern blot analysis of patient 237**

**Figure 38 Approximate extension and distribution of 6 *NF1* deletions *Hind*III polymorphism pattern**

**Figure 39 *Hind*III polymorphism pattern**

### **III. LIST OF TABLES**

**Table 1 Diagnostic criteria for NF1**

**Table 2 *NF1* alternative transcripts**

**Table 3 Techniques used to detect and characterize *NF1* mutations**

**Table 4 Characters of polymorphic markers used in haplotyping**

**Table 5 PCR conditions and sequences of microsatellite primers**

**Table 6 Sequences of primers used in PCR for protein truncation test**

**Table 7 Characteristics of six novel deletions**

**Table 8 Alleles sizes, frequency distributions (overall, normal and NF1), and estimated heterozygosity for each microsatellite**

**Table 9 The association between microsatellite markers**

#### IV. GLOSSARY OF ABBREVIATIONS

$\mu\text{Ci}$	microCurie
$\mu\text{g}$	microgram
$\mu\text{l}$	microlitre
$\mu\text{M}$	micromole (s)
A	adenine
ADP	adenosine diphosphate
AMP	adenosine monophosphate
ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
C	cytosine
cDNA	complementary DNA
cpm	count per minute
CRE	cAMP response element
D	Dalton
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
ddNTP	2', 3'-dideoxynucleoside triphosphate
DEAE	diethylaminoethyl
DEPC	diethyl pyrocarbonate



DGGE	Denaturing gradient gel electrophoresis
dGTP	deoxyguanosine triphosphate
DMEM	Dulbecco modified Eagle medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
dUMP	deoxyuridine monophosphate
EDTA	ethylene diamine tetra-acetic acid
EGF	epidermal growth factor
FBS	fetal bovine serum
FISH	fluorescence <i>in situ</i> hybridization
G	guanine
Gal	galactose
GAP	GTPase activating protein
GDP	guanosine diphosphate
GEFs	guanine nucleotide exchange factors
GMP	guanosine monophosphate
GRD	GAP-related domain
GTP	guanosine triphosphate
HD	heteroduplex
hnRNP	heterogeneous nuclear ribonucleoprotein

hr	hour
IPTG	isopropylthiogalactoside
IVS	intervening sequences or introns
kb	kilobase
KD, kD	kilodalton
LD	linkage disequilibrium
LOH	loss of heterozygosity
M	mole(s)
M.W.	molecular weight
MAPK	mitogen-activated protein kinase
MAPKK	mitogen-activated protein kinase kinase
MEN1	multiple endocrine neoplasia type 1
mg	milligram
min	minute
ml	millilitre
mM	millimole(s)
mRNA	messenger ribonucleic acid
mut, Mut	mutant
n.a.	not available
NE	sodium EDTA
NF1	Neurofibromatosis type 1 or von Recklinghausen syndrome
NF1-GRD	NF1 GAP related domain
NF2	Neurofibromatosis type 2

ng	nanogram
NIH	National Institutes of Health
nm	nanometer
nmol	nanomol
nts	nucleotides
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBD	phosphotyrosine binding domain
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PDGF	platelet-derived growth factor
PFGE	pulsed field gel electrophoresis
Pi	phosphate group
pmol	picomole
PMSF	phenylmethylsulfonyl fluoride
pre-mRNA	presursor mRNA
PTT	protein truncation test
Ras-GAP	Ras-GTPase Activating Protein
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rpm	rotation per minute
RT	reverse transcription
RT-PCR	reverse transcriptase-PCR

SDS	sodium dodecyl sulphate
SH2	src homology-2
SNP	single nucleotide polymorphisms
snRNA	small nuclear RNA
snRNP	small ribonuclear protein
SR	arginine- and serine-rich splicing factors
SRE	serum response element
SSCP	single strand conformation polymorphism
T	thymine
TAE	Tris-acetate-EDTA
TBE	Tris-boric acid-EDTA
TDP	thymidine diphosphate
TDT	transmission/disequilibrium test
TE	Tris-EDTA
TMP	thymidine monophosphate
tRNA	transfer RNA
TTP	thymidine triphosphate
U	uracil
u	unit
UTP	uridine triphosphate
UV	ultraviolet
wt, WT	wild type

## V. ABSTRACT (FRENCH)

La neurofibromatose 1 (NF1) affecte un individu sur 3000 et se caractérise par l'extrême variabilité des symptômes cliniques. Le gène de *NF1*, situé sur le chromosome 17q11.2, contient 350 kb et 60 exons. Le taux de mutation est de  $1 \times 10^4$ /gamète/génération. Environ 50% des mutations sont d'origine familiale et 50% sporadique. Nous avons identifié et caractérisé une mutation qui cause l'exclusion des exons 11 et 12a, six délétions et un polymorphisme *HincII* et aussi déterminé le déséquilibre de liaison dans la population québécoise.

Une mutation originale a été identifiée grâce au test de tronquation des protéines; elle serait la cause de l'exclusion d'exons. Le séquençage de la région génomique entre les introns 10c et 12a a permis d'identifier le changement d'un G en A en position 2056+1 dans le site d'épissage 5' de l'exon 12a. Ceci engendre l'excision des deux exons 11 et 12a sans changement du cadre de lecture et sans que la quantité des ARN messagers en soit affectée. Des essais d'épissage *in vivo* et *in vitro* nous ont permis de démontrer pour la première fois que le mutation naturelle supprime la définition des exons. Nos résultats montrent aussi que la séquence encodée par les exons 11 et 12a est essentielle pour l'activité de la neurofibromine.

L'haplotype de 19 familles a permis de détecter la perte d'hétérozygoté (LOH) grâce aux 4 microsatellites et aux RFLPs *RsaI* et *EcoRI*. Six délétions (2 familles: 7610 et 7473, et 4 patients: 178, 184, 236 et 237) ont été identifiées et caractérisées par la LOH et des buvardages de type Southern. La délétion dans la famille 7610 est d'origine maternelle et seul restent les exons 1-4b. La délétion dans la famille

7473 est aussi d'origine maternelle et les exons de 1 à 5 sont délétés, la délétion s'arrêtant quelque part avant l'intron 26. Avec le patient 178 la délétion commence entre les exons 23-2 et 27b pour se terminer après la région 3' de *NF1*. La délétion dans le patient 184 commence au 5' du gène et se termine entre les exons 27b-29. La délétion dans le patient 236 commence entre les exons 14-18 et se termine après l'extrémité 3' de *NF1*. Avec le patient 237, la délétion commence entre les exons 38-45 et se termine après la région 3' du gène. Ces délétions sont distribuées au hasard dans le gène *NF1*.

Un polymorphisme *HincII*, situé en position 5' du gène *NF1* a été mis en évidence. Il a été détecté grâce à une sonde faite à partir des exons 1 à 4a (nucléotides 2 à 401 du cDNA). Trois alleles ont été détectées: A1 (3.1 et 3.3 kb), A2 (2.4 et 4.0 kb), A3 (1.3 et 5.1 kb). Les hétérozygotes sont A1-A2 et A1-A3. Les homozygotes sont A1-A1 et A2-A2. Les fréquences des allèles A1, A2, A3 sont 0.545, 0.448 et 0.007 respectivement. Ce site est dans un intron. Nos résultats suggèrent aussi que la sonde révèle non pas un mais deux polymorphismes similaires.

Les 4 microsatellites: IVS26-2.3, IVS27AC28.4, IVS27AC33.1 et IVS38GT53.0 sont fortement liés. Cependant on n'observe pas de déséquilibre de liaison et pas d'effet fondateur dans la population québécoise.

Mes études ont identifié et caractérisé plusieurs mutations dans le gène *NF1*. Ces études montrent que l'analyse des mutations est complexe. Je n'ai pas identifié de mutation spécifique ou fondatrice dans la population québécoise.

## VI. ABSTRACT

Neurofibromatosis type 1 (NF1) afflicts 1 in 3,000 individuals and is characterized with variable clinical presentations. The *NF1* gene spans 350 kb on chromosome 17q11.2 with 60 exons. The gene exhibits high mutation rate of  $1 \times 10^4$ /gamete/generation and approximately 50% are sporadic new mutations. In this study, a splice site mutation and 6 gross deletions have been identified and characterized. Also, a *HincII* polymorphism was detected and the linkage disequilibrium was investigated.

Using a protein truncation assay, we have identified a exon skipping mutation. The mutation, which consists of a G to A transition at position +1 (2056+1) of the 5' splice site of exon 12a, is associated with the loss of both exons 11 and 12a in the NF1 mRNA. This loss of two exons does not change the reading frame and yields normal amounts of mRNA. Through the use of *in vivo* and *in vitro* splicing assays, we show that the mutation inactivates the 5' splice site of exon 12a. In addition, because this donor splice site is important to activate the upstream 3' splice site, the mutation prevents exon definition and leads to the skipping of both exons 11 and 12a. These results document the first example of a natural mutation that inactivates exon definition, and suggest that the 11-12a region of NF1 plays an important role in the activity of neurofibromin.

Six gross deletions (families 7610 and 7473, patients 178, 184, 236 and 236) have been identified and characterized by loss of heterozygosity analysis and Southern blot. The deletion in family 7610 arose from the maternal allele and all the *NF1* gene except exons 1 to 4b, was deleted. The deletion in family 7473 was also derived from the maternal chromosome and exons 1-5 were deleted. The breakpoint of the deletion is

located between exons 5 and 26. In the four non-family cases, for patient 178, the breakpoint of the deletion was located between exons 23-2 and 27b and the region downstream of the breakpoint was deleted. For patient 184, the breakpoint of the deletion was between exons 27b and 29, and the region upstream of the breakpoint was deleted. With patient 236, the breakpoint of the deletion was between exons 14 and 18 and the gene downstream of the breakpoint was deleted. The breakpoint of the deletion of the patient 237 was between exons 38 and 45 and the region upstream of the breakpoint was deleted. These six deletions were distributed randomly across *NF1* gene and no deletion hot spot was identified. The deletions are unique and different from those reported previously. The deletions are not associated with unusual clinical features.

The *HincII* polymorphism is located between exons 1 and 4a as defined by probe GE2-400 bp. Three alleles are detected: A1 (3.1 and 3.3 kb), A2 (2.4 and 4.0 kb), A3 (1.3 and 5.1 kb). The detected heterozygote forms are A1-A2, or A1-A3 and the homozygote forms A1-A1 or A2-A2. The allele frequencies are 0.545, 0.448 and 0.007 for A1, A2 and A3 respectively. The GE2-400 bp probe most probably detects the polymorphism in the *NF1* gene and at least in one other related locus.

Analysis haplotypes of the 19 NF1 families indicates that the four intragenic polymorphic microsatellites are strongly linked with the NF1 disease. However, no linkage disequilibrium and founder effect was observed in this Québec population collection.

In conclusion, the study shows that the analysis of *NF1* gene mutation is complex. No particular founder mutation has been observed in the Québec population.



# CHAPTER 1 INTRODUCTION

The objective of this study is to identify and characterize mutations in the neurofibromatosis type 1 (*NF1*) gene. This thesis includes three major parts: the study of *NF1* mutations, the characterization of a splicing mutation and the analysis of linkage disequilibrium using the haplotype data. Therefore, the introduction covers three parts: neurofibromatosis type 1 (NF1) (section 1.1), pre-mRNA splicing (section 1.2) and linkage disequilibrium (section 1.3).

## 1.1 Neurofibromatosis type 1 (NF1)

### 1.1.1 Clinical features

NF1 is characterized by peripheral neurofibromas, café-au-lait spots, and Lisch nodules. Patients often exhibit additional less common complications including osseous lesions, optic glioma, learning disabilities and malignancy. NF1 is known for its pleiotropic clinical manifestation and notable for its variable expressivity with clinical presentations that may vary from mild to severe even within the members of the same family (Rubsenstein and Korf, 1990; Riccardi, 1992; Carey and Viskochil, 1999). The diagnostic is according to the National Institutes of Health criteria for NF1 (Table 1) which are the most reliable means for the diagnosis (Feldkamp et al., 1998).

**Table 1. Diagnostic Criteria for NF1**

---

The diagnostic criteria of the National Institutes of Health Consensus Conference on Neurofibromatosis (July, 1987) are met by an individual who demonstrates two or more of the following:

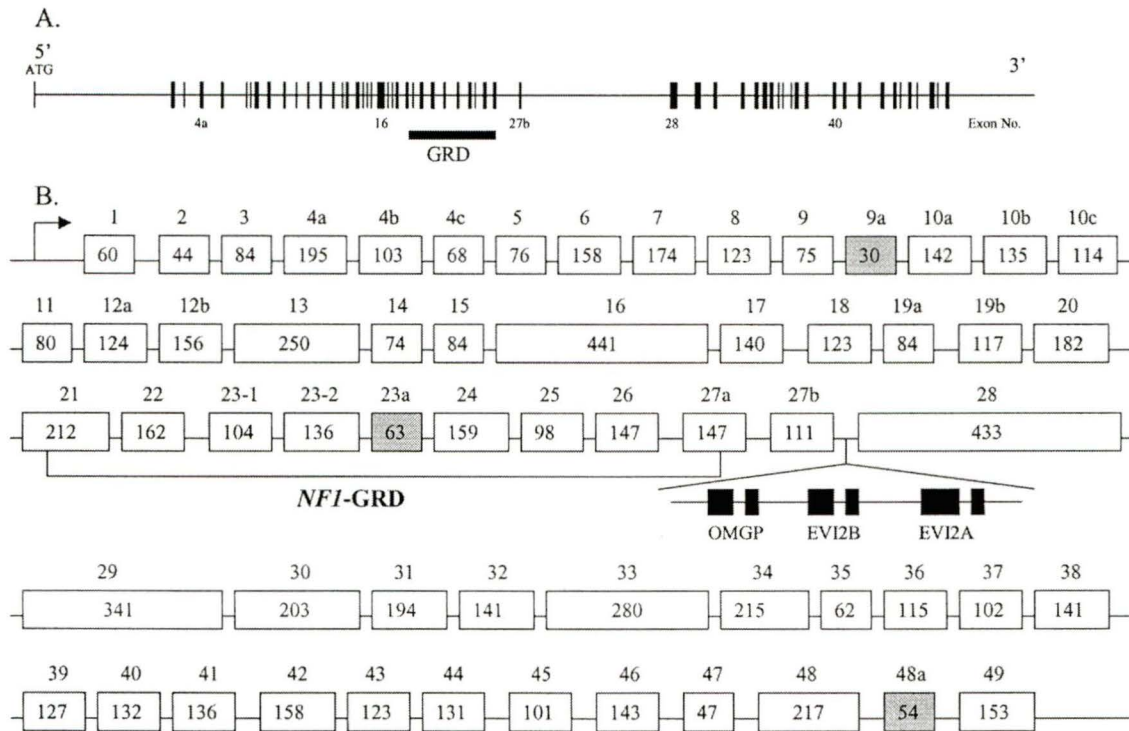
---

- 6 or more café au lait macules of over 5 mm in diameter in prepubertal individuals and over 15 mm in diameter in postpubertal individuals
  - 2 or more neurofibromas of any type or one plexiform neurofibroma
  - axillary or inguinal freckling
  - optic glioma
  - 2 or more Lisch modules (iris hamartomas of melanotic origin)
  - a distinctive osseous lesion such as sphenoid dysplasia or thinning of the long bone cortex with or without pseudoarthrosis
  - a first-degree relative (parent, sibling, or offspring) with NF1 by the above criteria
-

The inheritance pattern of NF1 is autosomal dominant. NF1 is considered completely penetrant. NF1 has an incidence of about 1:3000 (Colman and Wallace, 1994; Friedman, 1999).

### **1.1.2 The *NF1* gene**

The *NF1* gene spans approximately 350 kb of genomic DNA in chromosomal region 17q11.2. The *NF1* gene has 60 exons with intron sizes ranging from approximately 60 bp to more than 40 kb (Figure 1). It codes two transcripts of, 11 and 13 kb, presumably due to different lengths of the 3' untranslated region. The *NF1* cDNA has 8454 bp. It encodes for a 2818 amino acid protein, neurofibromin which appears to associate with microtubules and is abundant in the endoplasmic reticulum of neurons (Cawthon et al., 1990; Visochil et al., 1990; Wallace et al., 1990; Li et al., 1995; Colman and Wallace, 1994). Neurofibromin shows homology to various members of the GAP superfamily. These proteins include the mammalian GAP, the IRA1 and IRA2 gene products of *Saccharomyces cerevisiae*, the *sar1* gene product of *Schizosaccharomyces pombe* and *Drosophila* GAP1. The homology of the mammalian GAP and the *Drosophila* GAP1 is restricted to the catalytic domain of 360 amino acids. There is more extensive homology of the yeast IRA1, IRA2, and *sar1* gene proteins (Xu et al., 1990; Buchberg et al., 1990; Hall, 1992). The homology of neurofibromin with *C. elegans* has not been reported. Exons 21-27a specify a 360 amino acid domain which shows homology to the catalytic domain of the mammalian GTPase activating protein (GAP)



**Figure 1. The *NF1* gene.** A. The proximate positions of exons related to introns. The exons are indicated as boxes while introns are indicated as horizontal lines. The GAP related domain (GRD) is as indicated. B. Schematic diagram representing the *NF1* exons which are represented as rectangular boxes and the size of each exon is as indicated in bp. Introns are not shown to scale. The transcription start is depicted as a horizontal arrow upstream of exon 1. Alternative exons are shaded. The *NF1* GAP related domain (*NF1*-GRD) and the three genes: *EVI2A*, *EVI2B* and *OMGP*, embedded in intron 27b are as indicated. These three genes are transcribed in the orientation opposite to the *NF1* gene.

(p120-GAP) and the products of the yeast *IRA1* and *IRA2* genes. This region is referred to as NF1 GAP related domain (NF1-GRD) (Wallace et al., 1990; Li et al., 1995).

Three small genes, *EVI2A*, *EVI2B* and *OMGP*, whose functions are not known, are embedded within intron 27b and transcribed in the opposite orientation of *NF1* (Shen et al., 1996; Feldkamp et al., 1998; Viskochil, 1999). The search for homologous loci has shown *NF1* related loci on chromosomes 2, 12, 14, 15, 18, 20, 21, 22 (Legius et al., 1992; Cummings et al., 1993; Gasparini et al., 1993). The locus on chromosome 12 contains open reading frames homologous to *NF1* in at least two exons and is expressed in a number of tissues (Cummings et al., 1993). The complete sequence of the pseudogene on chromosome 22 consists of only the middle part of the *NF1* gene without exons 21-27a (Luijten et al., 2000).

*NF1* gene has a complex expression profile. There are multiple sites of alternative splicing within the gene affecting exons 9a, 23a and 48a. The alternative transcripts exhibit developmental and tissue-specificity. Evidence suggests that RNA splicing and RNA metabolism play a critical role in the control of *NF1* gene expression. Seven distinct *NF1* mRNA isoforms (Table 2) have been detected (Skuse and Cappione, 1997) and four neurofibromin isoforms have been identified (Feldkamp et al., 1998).

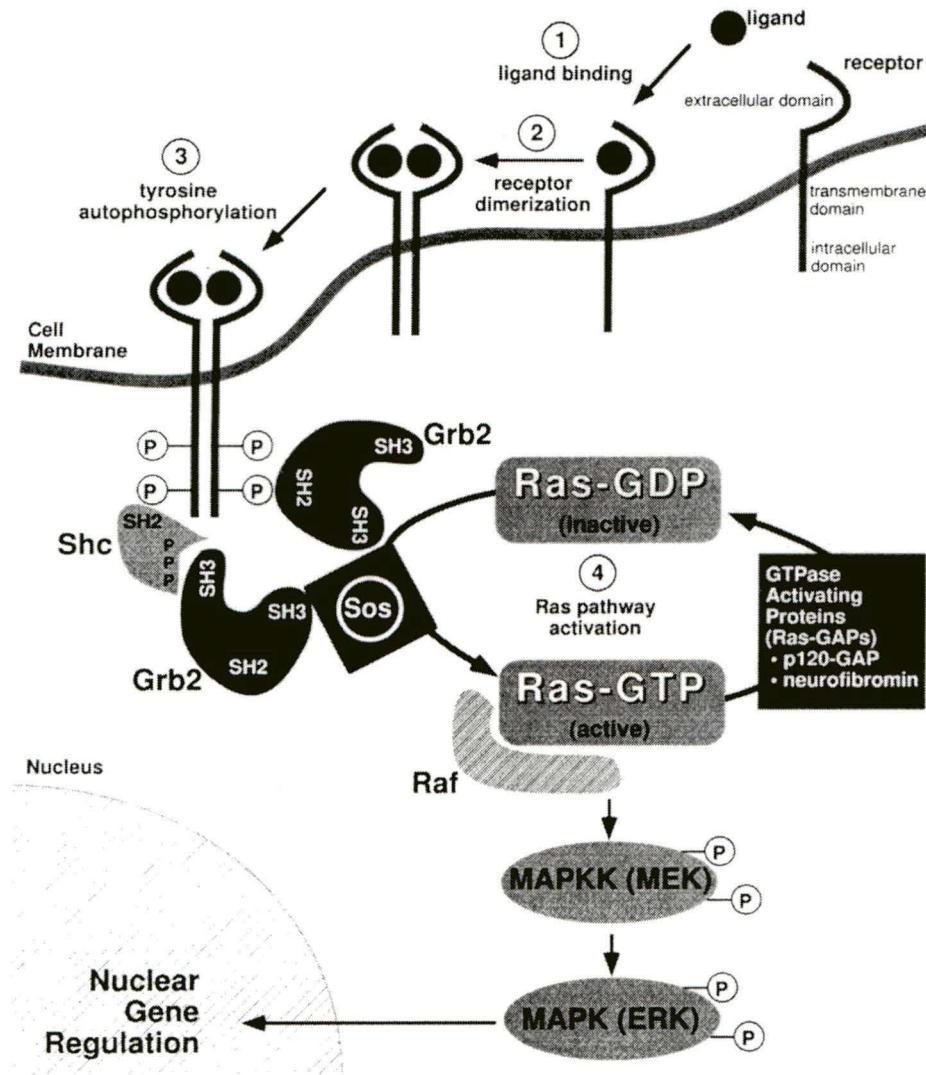
**Table 2. NF1 alternative transcripts**

Transcript name	Alternative exon included	Consequence in neurofibromin
9br	9br	Addition of 10 amino acids
Type II	23a	Addition of 21 amino acids
Type III (rodent)	23a and 23b	Introduction of frame shift
Type IV (rodent)	23b	Introduction of frame shift
Type 3	48a	Addition of 21 amino acids
Type 4	23a and 48a	Addition of 21 amino acids in GRD and 18 amino acids at carboxy terminus
N-isoform	Excludes exon 11 to most of exon 49	Excludes amino acids 548-2815

### **1.1.3 The *NF1* functions**

It is believed that neurofibromin acts in the cellular signal transduction pathway by down-regulating Ras via its GAP-related domain (GRD) (Feldkamp et al., 1998). As schematized in Figure 2, in the mitogenic signaling pathway, growth factors such as platelet-derived growth factor (PDGF) and epidermal growth factor (EGF) bind to and activate their cognate receptors by autophosphorylation of specific tyrosine residues of the intracellular domains. Then adaptor proteins such as Grb2 and Shc recognize the specific phosphotyrosine residues by their SH2 (src homology-2) domains or PTB (phosphotyrosine binding) domains, and in turn recruit GEFs (guanine nucleotide exchange factors) such as Sos to the proximity of its substrate, Ras-GDP. Ras functions as molecular switch which is tightly regulated by cycling between an active GTP-bound conformation (Ras-GTP) and an inactive GDP-bound state (Ras-GDP). Sos exchanges Ras bound GDP for GTP, resulting in activation of Ras and initiation of downstream signaling pathway, in particular the mitogenic cascade through Raf kinase, MAPKK (mitogen-activated protein kinase kinase), and MAPK (mitogen-activated protein kinase) (Feldkamp et al., 1998; Weiss et al., 1999).

Ras has a slow intrinsic GTPase activity, converting active Ras-GTP to inactive Ras-GDP, that is enhanced by GTPase activating proteins (GAPs) which thereby act as negative regulators of Ras. The *NF1*-GAP related domain of neurofibromin accelerates the conversion of Ras-GTP to Ras-GDP, and thus down regulates Ras activity (Shen et al., 1996; Weiss et al., 1999).



**Figure 2. Schematic representation of the Ras-Raf-MAPK mitogenic signaling pathway.** This cascade is activated in normal cells when ligand (a growth factor such as platelet-derived growth factor) binds its cognate surface receptor (Step 1). Ligand-receptor interaction results in receptor dimerization (Step 2), and trans-autophosphorylation of tyrosine residues on the intracellular domain of the receptor (Step 3). This allows signaling molecules (Shc, Grb2) to interact with the phosphotyrosine residues, bringing the nucleotide exchange factor Sos in proximity to the cell surface, where it exchanges GDP with GTP, activating Ras. Raf interacts with activated Ras-GTP, and phosphorylates MAPKK. MAPKK then subsequently phosphorylates MAPK. MAPK translocates to the nucleus where it participates with other molecules in activating transcription by the transcription factors such as Fos and Jun, resulting in the increased transcription of genes involved in cell division and other functions (Step 4). MAPPK: mitogen-activated protein kinase kinase; MAPK: mitogen-activated protein kinase; MEK: mitogen extracellular signal related kinase; ERK: extracellular signal related kinase.



The *NF1* gene knockout studies support that *NF1* gene functions as a tumor suppressor gene. The heterozygous *Nf1* +/- mice which possess a single functional *Nf1* allele, mice succumbed to tumors appeared earlier and with higher frequency compared to wild type animals (Jacks et al., 1994; Cichowski et al., 1999; Vogel et al., 1999). The chimeric mice composed in part of *Nf1* -/- cells develop neurofibromas (Cichowski et al., 1999), and mice that carry linked germline mutations in *Nf1* and *p53* develop malignant peripheral nerve sheath tumors (Vogel et al., 1999). The homozygous *Nf1* -/- knockout mice were embryonically lethal between days 12.5 and 14 of gestation, demonstrating the importance of neurofibromin in cellular growth and differentiation (Jacks et al., 1994; Brannan et al., 1994). A neurofibrosarcoma from a NF1 patient has a homozygous inactivation of *NF1* (Legius et al., 1993). Both alleles of the *NF1* gene are inactivated in leukemic cells in 5 out of 18 children with NF1, explaining that the risk of malignant myeloid disorder in children with NF1 is 200-500 times of the normal risk (Side et al., 1997). A NF1 patient with a germline deletion of at least exons 39-45 of the paternal allele and the maternal copy deleted in the tumor suggests that *NF1* inactivation is involved in the development or progression of some tumors (Martinsson et al., 1997).

#### **1.1.4 The *NF1* gene mutations**

About 30-50% of NF1 patients represent *de-novo* germline mutations, implying a high rate of spontaneous mutation in the *NF1* gene. This mutation rate has been estimated at 1 mutation per 10,000 alleles per generation, a rate 10-100 times higher than the usual

mutation rate for a single locus. This may be due to (1) the large gene size, (2) palindromes, symmetrical elements, and runs of repeated sequences with preferential occurrence of insertions and deletions, (3) the homonucleotide repeats or CpG dinucleotides (because 5-methylcytosine in CpG dinucleotides often undergo spontaneous deamination), (4) gene conversion between homologous pseudogenes on chromosomes 2, 12, 14, 15, 18, 20, 21 and 22 and *NFI* (Colman and Wallace, 1994), (5) germ-line mosaicisms which would elevate the apparent but not actual mutation rate (Friedman, 1999).

The techniques for the detection of *NFI* mutations are listed in Table 3. Mutation analysis in the *NFI* gene is difficult due to several factors: (1) the large gene size, (2) the presence of the normal allele, (3) the high mutation rate and subsequent mutations in most families, (4) wide variations in the size and type of mutations (from complete deletion of the locus to missense point changes), and (5) homologous loci on other chromosomes (Colman and Wallace, 1994).

To date, over 100 *NFI* germline mutations have been reported to the International NF1 Genetic Analysis Consortium (Shen et al., 1996). A total of 190 germline mutations have been reported to the Human Gene Mutation Database at Cardiff. In these 190 mutations, 52 are missense or nonsense mutations, 28 splicing mutations, 50 small deletions, 27 small insertions, 1 small insertion-deletion, 26 gross deletions, 2 gross insertions and duplications, and 4 complex rearrangements (including inversions). The occurrence of an identical nonsense mutation in several unrelated individuals has been

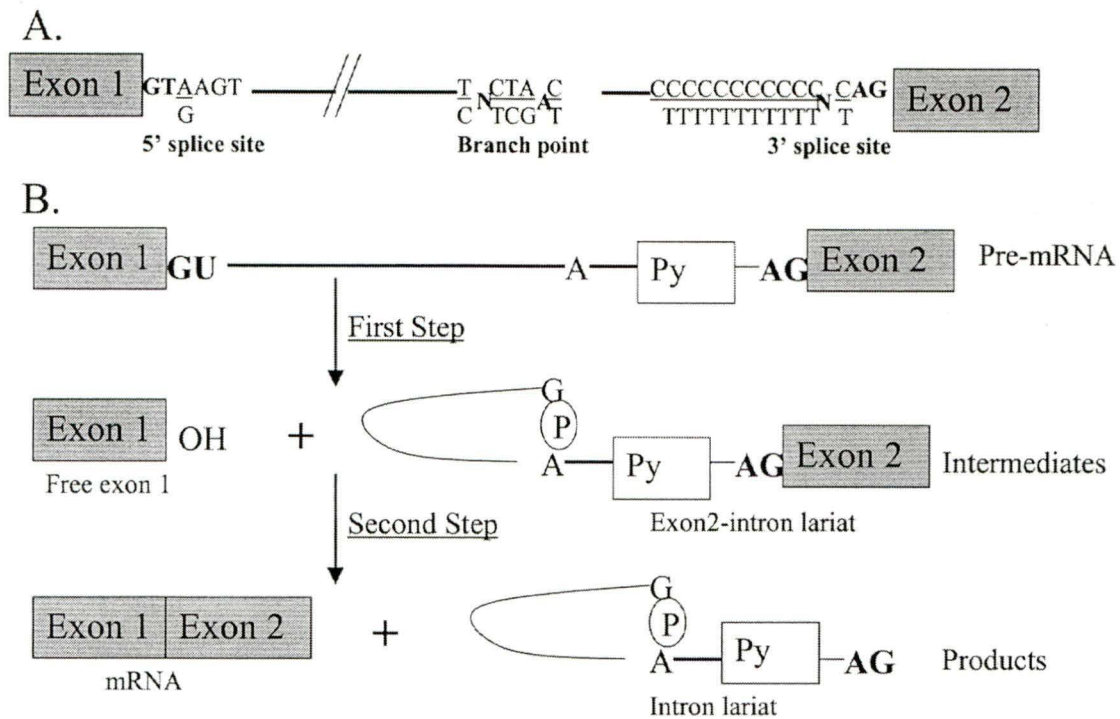
**Table 3. Techniques used to detect and characterize *NF1* mutations**

Techniques	Types of mutation
Microsatellite and RFLP (Restriction fragment length polymorphism) marker analysis	Deletion
Southern blot hybridization	Deletion
Cloning and sequencing	Point mutation
FISH (Fluorescence <i>in situ</i> hybridization) Somatic cell hybrids PFGE (Pulsed field gel electrophoresis)	Gross deletion and chromosome rearrangement
Heteroduplex (HD) and SSCP (Single strand conformation polymorphism) analysis	Point mutation, deletion
Protein truncation test (PTT)	Truncated neurofibromin
DGGE (Denaturing gradient gel electrophoresis)	Point mutation, deletion
Mismatch cleavage: chemical or enzymatic	Point mutation, deletion

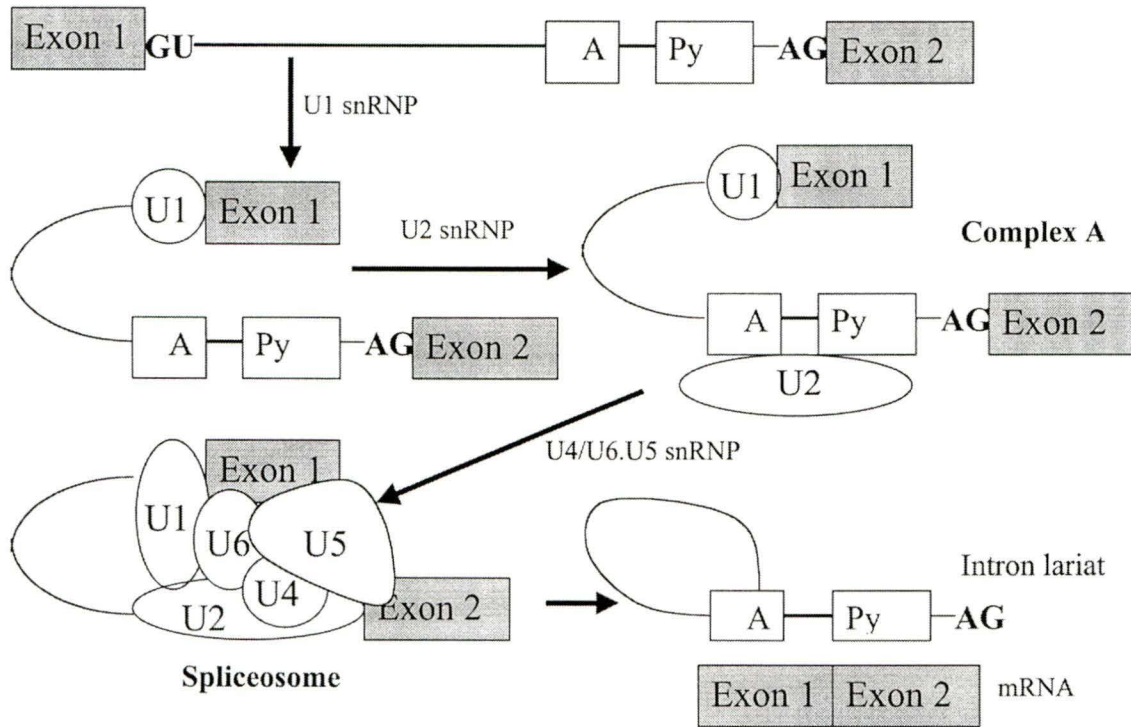
reported. All these patients have the same C→T transition in exon 31, changing an arginine codon to a stop codon. However, mutations at this locus occur relatively infrequently (Colman and Wallace, 1994). About 60% of the *NF1* germline mutations are predicted to result in truncated neurofibromin presumably due to nonsense or frameshift mutations. Large deletions of the gene, which account for about 4-7% of cases, are predominantly of maternal origin due to recombinations during oogenesis (Carey and Viskochil, 1999). About 90% of all *NF1* mutations are of paternal origin including the majority of point mutations and small deletions. This is probably due to male germ cells undergoing more divisions during spermatogenesis. It has been suggested that gross deletions are commonly associated with distinct manifestations such as learning disability/mental retardation, dysmorphic features, neurofibroma, and developmental delay (Upadhyaya et al, 1998; Upadhyaya et al, 1996; Cnossen et al, 1997; Tonsgard et al, 1997; Wu et al, 1997; Leppig et al, 1997; Valero et al, 1997; Riva et al, 1996; Kayes et al, 1994; Kayes et al, 1992; ). However, large *NF1* deletions may not always be associated with unusual clinical features (Rasmussen et al, 1998). So far, no mutation hotspot has been identified in the *NF1* gene. Also no significant genotype-phenotype correlation based on the location and type of mutation has been documented.

## 1.2 Pre-mRNA splicing and exon definition

Splicing of nuclear pre-mRNAs occurs in two-steps (Figure 3) (Lührmann et al., 1990). It involves snRNP (small nuclear RNA-protein) complexes and non-snRNP splicing factors (proteins only). In mammalian cells, the major snRNPs are U1, U2, U4, U5 and U6. One of the roles of some spliceosomal snRNPs is the recognition and binding of the short consensus sequences in the intron (Lührmann et al., 1990). The snRNPs assemble in an apparently ordered fashion onto the pre-mRNA to form a large complex, the spliceosome, where splicing occurs. Spliceosome formation (Figure 4) is initiated by two pre-mRNA/snRNP interactions: U1 snRNP binds to 5' splice site, while U2 binds to the region of the intron encompassing the branch point, both before and after the branch has been formed. U1 snRNP binds to 5' splice site in an ATP-independent manner. Base-pairing between the 5' end of U1 snRNA and the complementary 5' splice site is one essential element of this recognition process. SnRNP proteins are also required for efficient U1 snRNP/5' splice site complex formation. The binding of U2 RNP to the branch point is ATP-dependent and requires at least the U2AF protein and possibly several others. U2AF consists of two polypeptides of a parent molecular weight 35 and 65 kd. U2AF<sup>65</sup> preferentially binds to the polypyrimidine tract at the 3' splice site. U2 snRNP is then recruited to the U2AF/pre-mRNA complex, probable through direct interaction with pre-mRNA-bound U2AF. This is followed by the stable interaction of U2 snRNP with the branch point. Following interaction of snRNPs U1 and U2 with the intron, snRNPs U4/U6 and U5 assemble into the spliceosome, most likely in the form of a 25S U4/U5/U6 tri-snRNP complex. U4 and U6 primarily interact via base pairing,



**Figure 3. Schematic representation of the two-step splicing reaction.** (A) Consensus sequences in the mammalian intron, including 5' splice site with an invariant GU, 3' splice site of an invariant AG, and the branch point of an invariant A. N represents any nucleotides and Py the pyrimidine tract. (B) The two steps of the splicing reaction. The first step involves cleavage at the 5' splice site and the formation of a 2'-5' phosphodiester bond between the 5'-terminal G residue of the intron and an A residue normally located between 18 and 40 nucleotides upstream of the 3' splice site, namely the branch point. Thus, the intermediates of the splicing reaction, the cleaved exon 1 and the intron-exon 2 lariat are formed. The second step involves cleavage at the 3' splice site, ligation of the two exons, and release of the intron which is still in the form of a lariat.

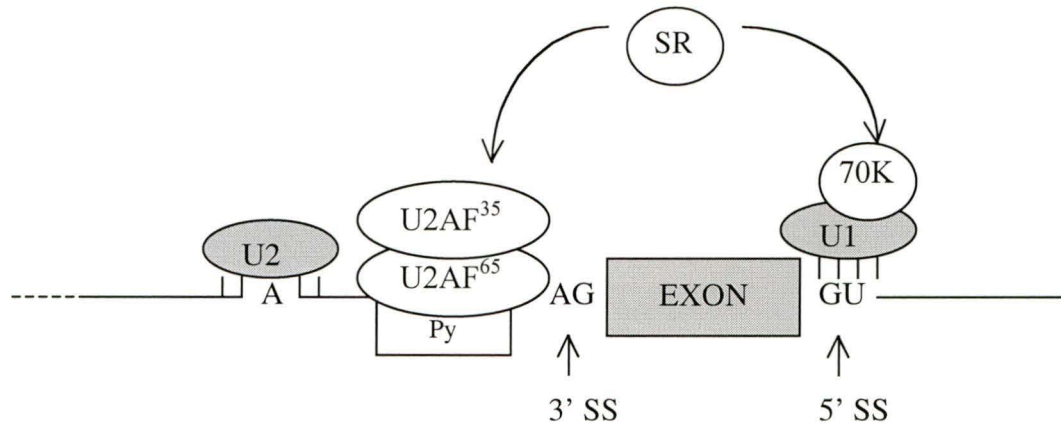


**Figure 4. The sequential integration of snRNPs into the spliceosome.** The U1 snRNP binds to the 5' splice site of the pre-mRNA and U2 binds to the 3' splice site region encompassing the branch point, forming the complex A, then followed by tri-snRNPs U4/U6.U5 assembled into the spliceosome. A represents branch point, and Py pyrimidine tract.

while U5 snRNP probably interacts with U4/U6 through protein-protein contacts. Only when all snRNPs are integrated into the mature spliceosome can splicing occur (Lührmann et al., 1990).

Splicing of small exons from large introns in vertebrates has been postulated to occur via exon definition. The exon definition model (Figure 5) proposes that in pre-mRNAs with large introns, the splicing machinery searches for a pair of closely spaced splice sites in an exonic polarity. When such a pair is encountered, the exon is defined by the binding of U1 and U2 snRNPs and associated factors to 5' and 3' splice sites respectively. The binding of U1 to 5' splice site can help define the upstream 3' splice site. U2AF<sup>65</sup> binds to the polypyrimidine tract at the 3' splice site, then recruits U2 snRNP probable through direct interaction. U2 snRNP forms stable interaction by base-pairing with the branch point. This process requires SR proteins which mediate an interaction between U2AF<sup>35</sup> and U1 snRNP 70 k protein. The protein-protein contacts involved in the 70K-SR-U2AF<sup>35</sup> bridging interaction are mediated by serine-arginine domain (SR). This process defines an exon. The best-characterized mammalian SR proteins are U2AF (U2 auxiliary factor), ASF/SF2 (alternative splicing factor/splicing factor 2), and SC35 (35-KD spliceosomal component). All three proteins have two functional subdomains: RNA-recognition motif (RRM) and serine-arginine rich domain (SR). RRM mediates the SR protein binding to RNA (Moore et al., 1993; Berget, 1995; Chabot, 1996).





**Figure 5. Exon definition in pre-mRNA with small exons and large introns.** U1 snRNP base-pairing with the 5' splice site stimulates U2AF<sup>65</sup> bound to the polypyrimidine tract, and recruiting U2 snRNP to the branch site. U2 snRNP base-pairs with the branch site. This process is mediated by SR proteins through interaction with U2AF<sup>35</sup> and U1 snRNP 70K protein. This process defines an exon. SS, splice site; Py, polypyrimidine tract; SR, serine-arginine-rich proteins.

The efficient recognition of splicing sites is crucial for maintaining the precision and order of intron removal in constitutive splicing units. Evidences support exon definition in vertebrate from several observations. (1) Introns preceding artificially long exons are not spliced. (2) U1 snRNP is required for initial recognition of 3' splice sites. (3) The presence of a 5' splice site within an exon affects splicing of the upstream intron. (4) Mutations that decreases the match of splice sites to consensus sites often promote exon skipping or the use of aberrant splice junctions (Treisman et al., 1983; Moore et al., 1993; Berget, 1995; Chabot, 1996). In contrast, evidence exists to suggest that pre-mRNAs with small introns use the intron, rather than the exon, as the initial mode of pairing between splice sites (Berget, 1995).

### **1.3 Linkage disequilibrium**

Genes present on the same chromosome are said to show linkage. Linked genes tend to be inherited together. We assume that the particular alleles for different loci on a chromosome are independent of one another. The alleles at linked loci will be associated randomly is known as linkage equilibrium. When alleles are in linkage equilibrium, the allele frequency distribution follows Hardy-Weiberg equation:  $(p+q)^2 = p^2+2pq+q^2=1$ . However, some particular sets of alleles are more likely to be associated nonrandomly. The nonrandom association of alleles at linked loci is referred to as linkage disequilibrium (LD). Linkage disequilibrium is a population-based phenomenon and reflects the effects of recombination over many past generations in the genomic region of interest (Feingold, 1991; Zubay, 1987; Korf, 1996). Three types of LD are considered: LD between polymorphic markers, LD in the whole population, LD within a family that is usually refered as linkage. Both linkage and linkage-disequilibrium measures a co-segregation between a genetic marker and the disease. Linkage focuses on a locus and linkage disequilibrium on an allele. Linkage analysis has the potential to localize a gene to a region of about 1-2 cM, while linkage disequilibrium operates over a smaller region and may be observed in a region of 0.5 Mb in RFLP systems or 1 Mb for highly polymorphic microsatellite markers. Linkage analysis examines the evidence of co-segregation of a marker and the disease within the families, while linkage disequilibrium analysis examines association between a marker allele and the disease in the whole population. Linkage results from recombination events in the last 2-3 generations.

Linkage disequilibrium results from much earlier, ancestral recombination events (Li, 1998; Higgins et al., 1998; Martin, 2000).

Linkage disequilibrium may be created when a disease is first introduced into a population either by a new mutation or an immigrant individual. At this time, linkage disequilibrium is complete because there is exactly one haplotype containing the disease gene, so for every marker locus in the haplotype there is a unique allele associated with the disease. Linkage disequilibrium may also arise in the absence of a founder individual due to either population admixture or genetic drift. The latter may be important in small, old, isolated population. As generation pass and the number of disease chromosomes descended from the founder individual grows, recombinations occur between the disease and marker loci, decreasing disequilibrium. Since the recombination factor is approximately proportional to physical distance, linkage disequilibrium decays faster for markers further from the disease locus. It is this fact that allows the relative position of the disease gene to be estimated and the founder to be traced (Martin et al., 2000; MacLean et al., 2000).

Computer programs have been developed for linkage and linkage disequilibrium analysis. They are based on the basic statistical philosophy: hypothesis test or significance test. First a null hypothesis is proposed. Then some appropriate test criterions are calculated from the sample values. The criterion distribution is also calculated under the null hypothesis. By comparison the criterion distribution, the significance level will become clear. The most used criterion value is p-value. It is a

common (but arbitrary) convention to consider p-value greater than 0.05 as 'not significant'. We usually consider  $0.01 < p < 0.05$  as significant;  $0.001 < p < 0.01$  as highly significant;  $p < 0.001$  as very highly significant (Elston and Johnson, 1990).

Recently, whole-genome linkage disequilibrium studies to map common disease genes have been investigated. It uses a dense map of single nucleotide polymorphisms (SNP) to detect association between markers and disease. Construction of SNP maps is currently underway and approximately 500,000 SNPs will be required for whole-genome studies (Kruglyak, 1999).

Linkage disequilibrium has been investigated in many diseases, such as X-linked juvenile retinoschisis (Huopaniemi et al., 1997), multiple endocrine neoplasia type 1 (Bassett et al., 1997), and Niemann-Pick disease (Greer et al., 1999). Linkage disequilibrium has been applied for mapping many genes, such as, Salla disease gene (Schleutker et al., 1995), Ataxia with vitamin E deficiency gene (Doerflinger et al., 1995), Werner Syndrome gene (Goddard et al., 1996), dominant optic atrophy locus (Votruba et al., 1998), and Batten disease gene (Mitchison et al., 1995). Some anonymous genomic regions have also been studied by linkage disequilibrium distributions which provide information for genetic mapping and for investigation, at the population level, of the processes such as recombination, selection, and mutation (Peterson et al., 1995; Kendler et al., 1999).

Linkage disequilibrium among RFLPs in the *NF1* region has been previously reported in Spanish and Italian population (Jorde et al., 1993; Messiaen et al. 1993; Valero et al., 1996, Purandare et al., 1996; Natacci et al., 1999). A high degree of disequilibrium is maintained among seven *NF1* intragenic polymorphisms ( $r > .82$ ,  $p < 10^{-7}$ ), even though they are separated by as much as 340 kb (Jorde et al., 1993). Three intragenic polymorphic microsatellite markers, AAAT Alu repeat, IVS27ACI27.2 and IVS38GT53.0, exhibit linkage disequilibrium between each other (Valero et al., 1996). Five polymorphisms in the *NF1* gene also showed lack of independence (Messiaen et al., 1993).

Here in this thesis, I report the identification and characterization of a splicing mutation which causes skipping of both exons 11 and 12a, and six gross deletions in the *NF1* gene. In addition, a *HincII* polymorphism has been detected in the 5' region of the *NF1* gene and linkage disequilibrium in the Québec population has been analyzed using the haplotype data.

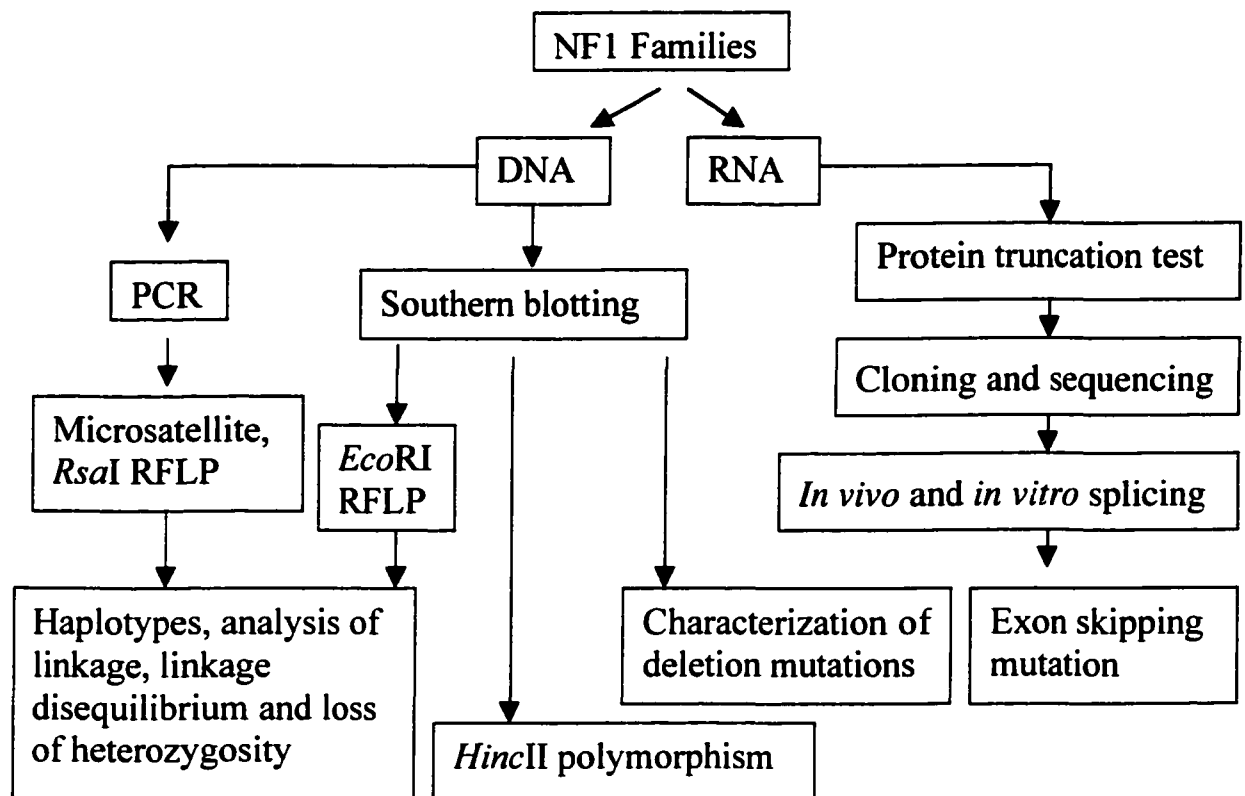
## CHAPTER 2 MATERIALS AND METHODS

The experimental design is depicted in Figure 6. Blood samples were collected from the NF1 patients and the members of their families. DNA or RNA was extracted. The haplotypes were constructed with six polymorphic markers (Table 4) and analyzed statistically for linkage disequilibrium. The loss of heterozygosity (LOH) analysis was performed to identify deletions. The deletions were further characterized by Southern blot analysis. At the same time, the protein truncation test (PTT) was performed to detect mutations affecting neurofibromin. One neurofibromin mutation (5313) was further characterized by sequencing, *in vivo* and *in vitro* splicing analysis.

### 2.1 NF1 Patients and NF1 families

All patients diagnosed with NF1 fulfilled the diagnostic criteria as described in the NIH Consensus Development Conference Statement (Stumpf et al., 1988) and gave informed consent for a molecular study of their DNA.

A total of 19 family and 20 non-family cases from Québec, Canada was used. The 19 NF1 families (total 87 persons: 45 NF1 patients and 42 normal individuals) include 15 family cases and 4 sporadic cases. In these 19 families, 16 families are two-generations and 3 are three-generations. All families are unrelated and come from different regions of Québec, representing an apparently random sample of the population. The 20 non-family cases were made of individual NF1 patients with no family records available.



**Figure 6. Schematic chart of the experimental design.** Blood samples were obtained from the NF1 patients and the members of the NF1 families. DNA or RNA was extracted. Polymorphic markers were amplified by PCR for linkage and loss of heterozygosity analysis. DNA was analysed by Southern blotting. RNA was used in the protein truncation test. The exon skipping mutation (5313) was further characterized by cloning, sequencing, *in vivo* and *in vitro* splicing analysis.



**Table 4. Characters of polymorphic markers used in haplotyping**

<b>Polymorphic type</b>	<b>Markers</b>	<b>Genomic localisations</b>	<b>Fragment lengths</b>
RFLP	EcoRI	Exons 1-4a	7.3 kb, 4.2 kb in Southern blots
RFLP	RsaI	Exon 5	500 bp, 450 bp, 300 bp, 50 bp
Microsatellite	IVS26-2.3	Intron 26	≈206 pb
Microsatellite	IVS27AC28.4	Intron 27b	≈213-225 pb
Microsatellite	IVS27AC33.1	Intron 27b	≈104-124 bp
Microsatellite	IVS38GT53.0	Intron 38	≈171-187 bp

For the *HincII* polymorphism study, 67 unrelated persons from Québec including 14 NF1 patients and 53 unaffected individuals (who are not from the above mentioned NF1 families) were analyzed for allelic frequencies by *HincII* digestion of the genomic DNA.

NF1 patients and families were identified in the following hospitals and clinics: Dr. Bernard Lemieux from Département de Pédiatrie du Centre universitaire de santé de l'Estrie (CUSE), Drs. Ortenberg and Rouleau from Montreal Children's Hospital, and Dr. Ruest from Centre hospitalier de Rimouski.

Three NF1 families: 7610, 7473 and 5313 were studied extensively. Their clinically features are as follows. With family 7610, a 5 year old boy (II-1) and his mother (I-1) were diagnosed with NF1. When the mother was examined at 36 years old, more than 6 café-au-lait spots with sizes larger than 15 mm were present on the trunk and the limbs. Bulky café-au-lait spots around the neck were excised for cosmetic reason. Multiple under-cutaneous nodules and neurofibromas, but not plexiform neurofibroma were observed. Many ephelises on axillary bilateral region, Lisch nodules and a light scoliosis of the column in the dorsal part were observed. When she was 33 years old, arterial hypertension started to appear. In etiologic surgery, detachment of the bottom part of the retina was observed. The son (II-1) is the only child of the woman (I-1) who has not had an abortion. The pregnancy and the labour were both normal. A convulsion happened at the second day of the birth. One week after birth, the boy was noted to have a few café-au-lait spots. At 5 years old, the growth and development of the boy was

normal except for some pronunciation troubles. Cephalalgia was absent and there was no visual trouble. The cranial nerves including the back of the eye, the visual field and the pupil reflex were normal. No Lisch nodule was observed. More than 6 café-au-lait spots with sizes larger than 5 mm were observed on the trunk and the limbs. Café-au-lait spots were also present on the ephelises of auxiliary inguinal bilateral region. Three under-cutaneous neurofibroma on posterior trunk and fold of buttock were noticed. No plexiform neurofibroma, bone lesion, limbs and face abnormality, proptosis and scoliosis were observed. The grandfather and one aunt of the boy were also reported possibly with NF1. The father (I-2) was normal. In family 7473, a 36-year-old woman (II-1) and her two daughters (III-1 and III-2) were diagnosed with NF1 at the age of 32 for the woman and 3 and 4 for the two daughters. The woman had multiple café-au-lait spots from birth. The first neurofibroma appeared during her puberty. Her grand parents, parents, sister and brother were healthy. In family 5313, the mother (I-2) and her two sons (II-1 and II-2) were NF1 patients. Numerous café-au-lait spots and small cutaneous neurofibromas were diagnosed from these three NF1 patients. The father (I-1) was normal.

## **2.2 Collection of blood samples and separation of leucocytes from blood**

About 30-40 ml of blood was collected per person: 10 ml in a sterile acid citrate dextrose tube for cell transformation, 20-30 ml in either EDTA or ACD tube for DNA isolation. For children, 4 or 5 ml of blood samples were collected. Blood samples were kept at room temperature or at 4<sup>0</sup>C. Leucocytes were isolated from whole blood by Ficoll-hypaque gradient centrifugation. About 5 ml of blood was laid on the top of 5 ml

of HISTOPAQUE-1077 (Sigma) solution and centrifuged. After removing the top plasma layer, leucocyte cells were transferred to a fresh tube. Red cells were lysed in hypotonic 0.2 % NaCl solution and then treated in hypertonic 1.6 % NaCl to restore the ionic strength. The leucocytes were collected by centrifugation, resuspended in PBS buffer and kept at  $-20^{\circ}\text{C}$ .

### **2.3 Cell Culture**

Lymphoblastoid cell lines from members of NF1 family were established by Louise LeBuis in Dr. Rouleau's laboratory (Montreal General Hospital) by Epstein-Barr virus transformation of mononuclear cells from the peripheral blood in the presence of cyclosporine A ( $0.1\ \mu\text{g}/\text{ml}$  in culture). Cells were cultured in ISCOVE medium (Gibco BRL) with 10 % fetal bovine serum (FBS) and 5 %  $\text{CO}_2$  at  $37^{\circ}\text{C}$ . Adherent cells (HeLa) were grown in Dulbecco modified Eagle medium (DMEM; Gibco-BRL) supplemented with 10% newborn calf serum in 5%  $\text{CO}_2$  at  $37^{\circ}\text{C}$ . Cells were seeded at  $0.7 \times 10^6$  cell/plate in 100-cm-diameter plates. Cells were split 1:5 every 3 days in 100-mm-diameter plates.

### **2.4 Extraction of DNA and RNA**

Genomic DNA was extracted either from about  $2 \times 10^6$  leucocyte cells or from lymphoblastoid cell lines by standard phenol-chloroform method (Sambrook et al., 1989). Cells were first treated with  $200\ \mu\text{g}/\text{ml}$  of proteinase K at  $55^{\circ}\text{C}$  overnight in NE buffer

(105 mM NaCl, 26.38 mM EDTA, pH 8). Then an equal volume of phenol was added to the solution. After shaking and centrifugation, the top phase was transferred to a fresh tube and extracted once with equal volume of phenol-chloroform-isoamyl alcohol (25:24:1), and twice with chloroform. DNA was precipitated with 2 volume of 95 % ethanol, washed and suspended in TE buffer or bidistilled water. The concentration of DNA was measured at O.D. 260 nm as 50 µg DNA/ml/OD and examined by agarose gel electrophoresis with  $\lambda$ DNA/*Hind*III as a standard.

Total cellular RNA was extracted with TRIzol reagent (Gibco BRL). About 10<sup>6</sup> cells were collected by centrifugation and resuspended in 1 ml of TRIzol. After incubation for 5 mins at room temperature, 0.2 ml of chloroform was added and the mixture vortexed. After 5 mins at room temperature, the mixture was centrifuged. The RNA was precipitated with isopropanol and resuspended in DEPC-treated H<sub>2</sub>O. The concentration of the RNA was measured at O.D. 260 as 40 µg RNA/ml.

## **2.5 PCR, cloning and sequencing of the genomic 11-12a region of patient 5313**

With patient 5313, three *NF1* genomic fragments: exon 11, exon 12a and exon 11-12a, were amplified by PCR with intronic primers (Li et al., 1995). Exon 11 was amplified with primers: forward 5' GTA CTC CAG TGT TAT GTT TAC C 3' and reverse 5' TAA AGT TGA AAT TTA AAA ATT AAA GTA C 3'. Exon 12a was amplified with primers: forward 5' AAA CCT TAC AAG AAA AAC TAA GCT 3' and reverse 5' ATT ACC ATT CCA AAT ATT CTT CCA 3'. The 11-12a region (from

intron 10c to intron 12a) was amplified by primers: forward 5' GTA CTC CAG TGT TAT GTT TAC C 3' and reverse 5' ATT ACC ATT CCA AAT ATT CTT CCA 3'. The amplifications of exons 11 and 12 were 5 mins at 96<sup>0</sup>C, 35 cycles of 1 min at 93<sup>0</sup>C, 1 min at 53<sup>0</sup>C, and 1 min at 72<sup>0</sup>C, and finally 10 min at 72<sup>0</sup>C. The amplification protocol of 11-12a was the same as for exon 11 and exon 12a except that the annealing temperature was 57<sup>0</sup>C.

The PCR products of the genomic 11-12a fragment (from intron 10c to intron 12a) of both wild type (wt) and mutant (mut) were blunted with DNA polymerase I Klenow fragment and ligated into pUC18 at *Sma*I site to generate pUC11-12a. The ligation was in a total volume of 15  $\mu$ l containing no less than 100 ng of insert, 50 - 100 ng of plasmid, 1 mM ATP, 1 mM dNTP, and 1 u T4 DNA ligase. The ligation mixture was incubated at room temperature for at least 6 hr. *E. coli* DH5 $\alpha$  was transformed by heat shock at 42<sup>0</sup>C for 2 min. Recombinants were selected on plates with 100  $\mu$ g/ml of ampicillin, X-Gal and IPTG. White colonies were grown in LB media with 100  $\mu$ g/ml of ampicillin and with shaking overnight at 37<sup>0</sup>C. Plasmid DNA was prepared by mini-preps (Sambrook et al., 1989). The insertion was checked by restriction digestion and sequencing. Sequencing was performed using a T7 Sequencing kit (Amersham) following manufacturer's instructions. Briefly, templates were denatured with 10 N NaOH at room temperature for 10 min. Sequencing primers were annealed to templates at 65<sup>0</sup>C for 10 min. The elongation was with T7 polymerase at room temperature for 5-7 min with incorporation of [ $\alpha$ -<sup>32</sup>P]dCTP and terminated with ddNTP. The radio-labeled products were separated by electrophoresis in a 8% denaturing acrylamide-urea gel with

1 x TBE using a constant power of 45 W. Gels were autoradiographed at room temperature for 1-16 hr.

## **2.6 Amplification of the polymorphic regions by PCR, haplotyping and statistical analysis**

### **2.6.1 Microsatellites**

The microsatellite allelic regions at intron 26 (IVS26-2.3), intron 27 (IVS27AC28.4 and IVS27AC33.1) and intron 38 (IVS38GT53.0) were carried out as described (Andersen et al., 1993; Lazaro et al, 1994; Lazaro et al, 1993) with minor modifications. The PCR was in a total volume of 20  $\mu$ l containing 0.5  $\mu$ M dNTP, 0.5  $\mu$ M of each primer, 1 u Taq polymerase, 5  $\mu$ Ci of [ $\alpha$ -<sup>32</sup>P]dCTP (Amersham 3000 Ci/mmol) and 100 ng of genomic DNA. The amplification was 5 mins at 94<sup>0</sup>C, then 27 cycles of 20 sec at 94<sup>0</sup>C, 20 sec at the appropriate annealing temperature (see Table 5), and 30 sec at 72<sup>0</sup>C plus 1 second per cycle followed by a final elongation for 7 min at 72<sup>0</sup>C. The amplified fragments were analyzed on a 40 cm 8 % denaturing acrylamide-urea gel. The sizes of the allelic fragments were determined by comparison with a pUC18 sequence ladder prepared with a T7 Sequencing kit (Amersham).

### **2.6.2 *Rsa*I RFLP**

**Table 5. PCR Conditions and sequences of the microsatellite primers**

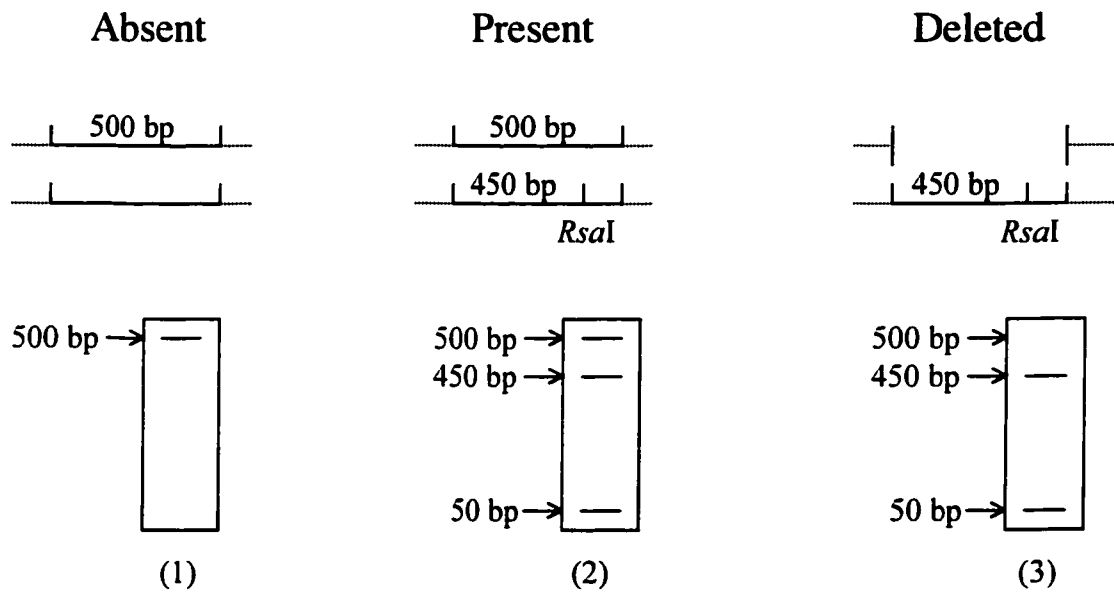
Locus	Primers	Type of repeats	Annealing temperature	Allelic sizes (bp)
IVS26-2.3	F 5'-AGGCCAGGAGTTCAAGACCA-3' R 5'-ATGAGCCACTGTGCCCAATC-3'	(TAGA), (TAGG), (TAAGA)	58 <sup>o</sup> C	204-226
IVS27AC28.4	F 5'-TGAAGTATGCAGTTTTCCAG-3' R 5'-GGCTAAGTGTAACGCAAAG-3'	(CA)	55 <sup>o</sup> C	215-231
IVS27AC33.1	F 5'-TAGATTATATGGGACAGAAAATG-3' R 5'-CTTGAGGTGATGACAGGATG-3'	(CA)	55 <sup>o</sup> C	104-116
IVS38GT53.0	F 5'-CAGAGCAAGACCCTGTCT-3' R 5'-CTCCTAACATTTATTAACCTTA-3'	(CA)/(GT)	52 <sup>o</sup> C	171-189



The *RsaI* RFLP site of exon 5 was amplified with primers 5E-1 (5'-CAT GTG GTT CTT TAT TTA TAG GC-3') and 5E-2 (5'-CCA AGT CAA ATA GCT TTT CTG C-3') to yield a 800 bp PCR product which covers exon 5, intron 5 and part of exon 6 (Hoffmeyer and Assum, 1994). PCR was performed in a total volume of 50  $\mu$ l containing 0.4 mM dNTP, 0.2  $\mu$ M each primers, 2 units of Taq polymerase and 100 ng of genomic DNA. The amplification was 94<sup>0</sup>C for 5 min, then 35 cycles of 1 min at 92<sup>0</sup>C, 1 min at 52<sup>0</sup>C and 1 min at 72<sup>0</sup>C, and a final extension at 72<sup>0</sup>C for 10 min. All PCR products were digested with *RsaI* at 37<sup>0</sup>C. The digested fragments were separated on a 1 % agarose gel and the sizes of the fragments were determined by comparison with a 1 kb marker. The *RsaI* digests yield a 500 bp and a 300 bp fragments if the restriction site is absent. If present, *RsaI* cuts the 500 bp fragment into a 450 bp and a 50 bp fragments. Since the 50 bp fragment runs out of the gel, the presence of the 450 bp fragment is indicative of the presence of the *RsaI* site. However, in some cases if there is a deletion (Figure 7), then the 450 bp and 50 bp fragments may be present, while the 500 bp fragment is missing as with families 7610 and 7473.

### 2.6.3 Haplotype analysis

Haplotypes were constructed manually, assuming a minimum number of recombinations in each family. Haplotypes were also redrawn and confirmed using the GENEHUNTER program. For all families the markers were informative at different loci, and for each marker the disease-bearing chromosome could be identified.



**Figure 7. The *RsaI* polymorphism in exon 5.** Exon 5 region was amplified by PCR and digested with *RsaI* (represented by small vertical lines). If the *RsaI* site is not present, 500 bp band is detected (1). When the site is present, *RsaI* cut the 500 bp fragment to 450 bp and 50 bp fragments (2). However, if one of the sites is deleted, 450 bp and 50 bp fragments are present, while 500 bp fragments are missing.

#### 2.6.4 Statistical analysis

The allele frequency was calculated. We sampled alleles only in "founders" (parents and married-ins). Children do not provide independent samples, since they derive their alleles from the parents. A total of 79 'founders' were used in the analysis. The frequency distribution of alleles on disease-bearing and non-disease-bearing chromosomes was analysed using all 19 pedigrees. Affected individual bearing the full disease haplotype from 19 families was selected for analysis. This gave rise to a total of 27 affected haplotypes (chromosomes) to be included in the analysis. The 52 unaffected individuals were taken as controls for the assessment of normal population allele frequency, giving rise to a maximum of 104 chromosomes.

Linkage analysis between the disease and markers was carried out using the computer program LINKAGE [Terwilliger and Ott, 1994] and GENEHUNTER [Kruglyak, et al, 1996]. Such a linkage analysis is necessary to ensure that the regions being focused on are indeed related to the disease. Logarithm (base 10) of the likelihood ratio, between that of linkage under the given disease model (e.g., dominant) and that of no linkage, is called the LOD score. A LOD score of 3 is usually used as a strong evidence for linkage. LINKAGE program is used for a two-point analysis (i.e., each marker is analyzed separately); and GENEHUNTER is used for multi-point analysis (i.e., all markers are considered at the same time).

Unlike linkage analysis which examines the evidence of co-segregation of a marker and the disease within the families, linkage disequilibrium analysis examines association between a marker allele and the disease in the whole population. Such association is present when there is only one or a few disease haplotype at the time the mutation was first created (founder effect), and if the marker is close enough to the disease locus. We use the computer program HAL (currently renamed TRIMHAP) [MacLean et al., 2000] to examine whether there is evidence for such founder haplotype. The null hypothesis that founder haplotype does not exist is tested, and the test result is presented by a p-value.

Another population-based association analysis is carried out by the DISEQ program [Terwilliger, 1995]. In this approach, the assumption that some alleles are abundant is tested by a likelihood ratio test as well as the standard  $\chi^2$  test. Again, p-value for the null hypothesis of no association between the disease and a marker is obtained. Association can be established only when the p-value is small.

In-between the linkage analysis within a family and linkage disequilibrium analysis within a population, there is a third type of analysis called family-based association analysis. Family-based association examines both linkage and linkage disequilibrium in parents-affected-child triads. Either lack of the co-segregation between the marker and the disease within the pedigree or lack of association between the marker alleles and the disease in the population will render the test insignificant. We used the GASSOC program [Schaid, 1996]. Generalizing the popular transmission/disequilibrium

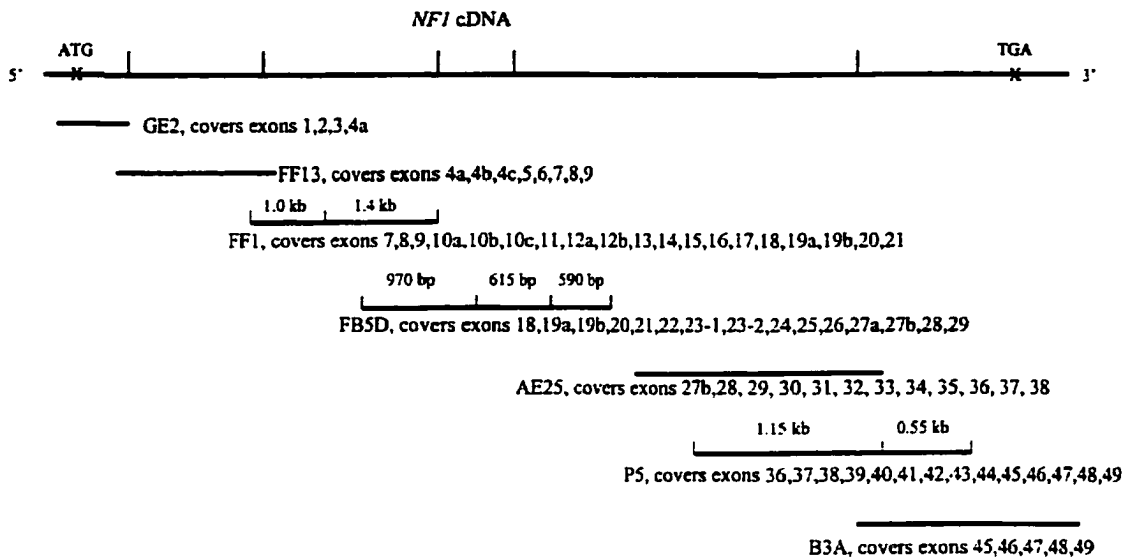
test (TDT) [Spielman et al., 1993], this program uses a likelihood ratio test with the incorporation of a disease model. As a result, tests under dominant and recessive disease models can be analyzed separately (on the other hand, TDT is a non-parametric test without the specification of the disease model).

Linkage disequilibrium between markers is determined by a Monte Carlo version of the Fisher's exact test. In a Fisher's test, the joint count table of alleles at two markers is constructed from the haplotype data. All other possible joint count tables with the same column and row margins can be in principle enumerated, and from the distribution of these tables, p-value of any test statistic can be calculated. Rather than enumerating all tables, we randomly sample 100,000 tables, and an approximate p-value is determined. The corresponding p-value for a  $\chi^2$  test is also included, which is correct in the large sample limit.

All computer programs mentioned above are listed in the webpage: <http://linkage.rockefeller.edu/soft/>, and their distribution addresses (URL's) can be traced accordingly.

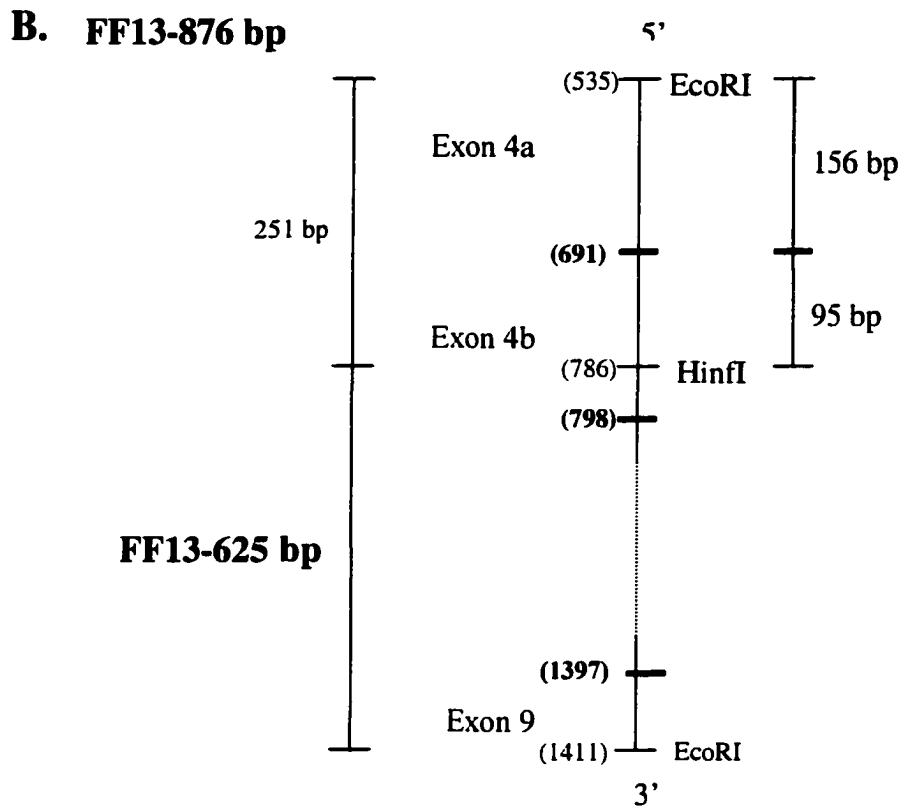
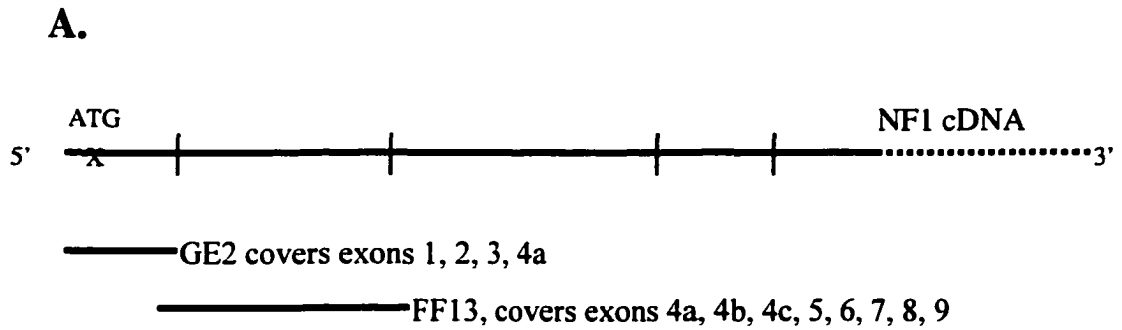
## **2.7 Southern blot, *EcoRI* and *HincII* polymorphism**

The seven cDNA probes: GE2, FF13, FF1, FB5D, P5, and B3A, made of portions of the *NFI* cDNA and encompassing the entire coding region (Figure 8), were used in



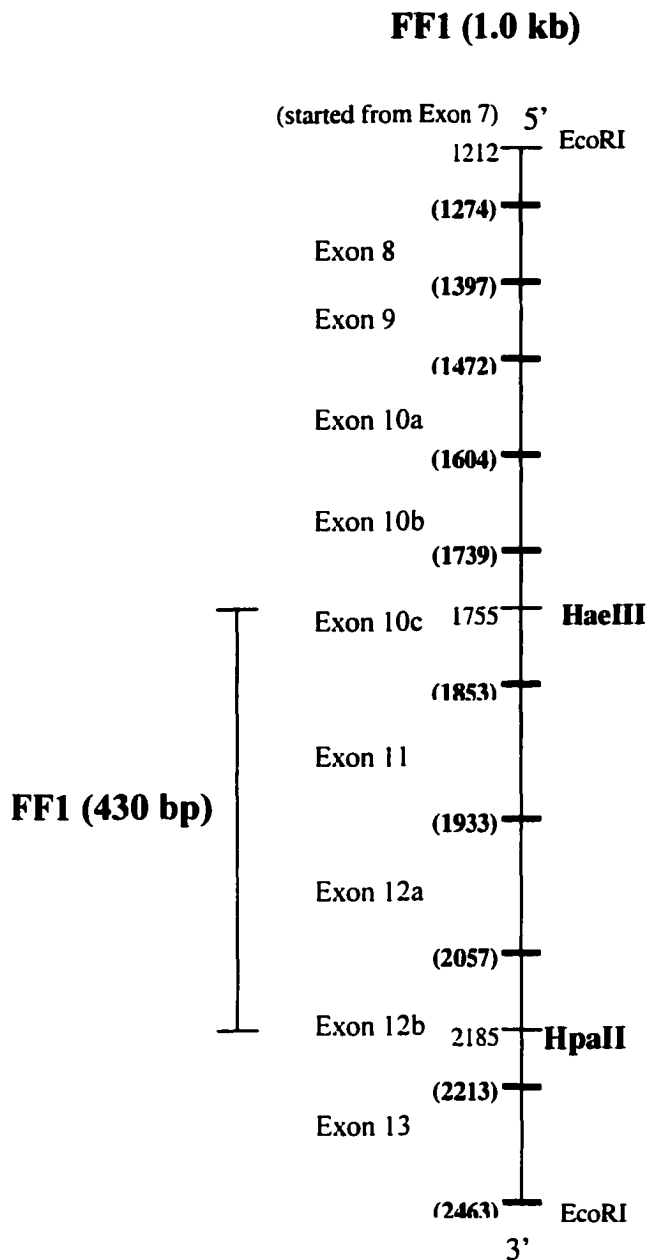
**Figure 8. The seven *NF1* cDNA probes.** The probes encompass the entire coding region with overlapping. The proximate positions of the probes related to the *NF1* cDNA are as depicted. Vertical lines on the *NF1* cDNA represent *EcoRI* sites. This diagram is not quite to scale.

Southern blotting. The cDNA fragments cloned at the *EcoRI* site in pBluescript were obtained from Dr. Wallace's and Dr. Vidaud's laboratories. The plasmids were prepared by mini-preps and purified by equilibrium centrifugation in CsCl-ethidium bromide gradients (Sambrook et al., 1989). The cDNA probes were obtained by restriction digestions. *EcoRI* and *SacII* double digestion of GE2 plasmid gave a 150 bp piece (untranslated region, very G/C rich) and a 400 bp fragment (which includes the ATG and spans exons 1-4a, overlaps FF13 by 82 bp). Digestion of the FF13 plasmid with *EcoRI* gave the 876 bp FF13 probe. FF13 covers exons 4a-9 and overlaps FF1 by 220 bp. Further digestion of FF13 (876 bp) with *HinfI* gave the FF13 (625 bp) probe which was used to fine localize the deletion 7610 breakpoint. The FF13 (625 bp) fragment covered 12 bp of exon 4b and exons 4c-9 (Figure 9). Digestion of FF1 plasmid with *EcoRI* and *HincII* yielded a 1.4 kb (3' half) and a 1.0 kb (5' half) fragments, which were used separately. The entire FF1 fragment is 2625 bp and covers exons 7-21. FF1 overlaps the 970 bp FB5D fragment by 530 bp. Double digestion of the FF1 1.0 kb fragment with *HaeIII* and *HpaII* yielded the FF1 (430 bp) probe (Figure 10) which was used in the characterization of mutation 5313. FF1 (430 bp) covers exon 10c (98 bp), exons 11-12a, and exon 12b (128 bp). The entire FB5D fragment is 2177 bp and covers exons 18-29 which contain the *NF1*-GRD domain. Digestion of FB5D plasmid with *EcoRI* yielded three fragments: 970, 615 and 590 bp. The 970 bp fragment is the 5'-most, 615 bp is in the middle, and 590 bp is at the 3' end. The 970 bp fragment was used independently. The 615 bp and 590 bp fragments could not be separated and were used together. *EcoRI* digestion of plasmid P5 released two inserts: 1.15 and 0.55 kb (5' and 3' respectively). The



**Figure 9. The FF13 876 bp and FF13 625 bp probes for characterisation of deletion 7610.** (A) The relative position of probe GE2 and probe FF13 to the *NF1* cDNA. Probes GE2 and FF13-876 bp overlap by 82 bp in exon 4a. (B) Digestion of FF13 (876 bp) with *HinfI* excludes exon 4a and 95 bp of exon 4b to give FF13 625 bp probe.





**Figure 10. The FF1 1.0 kb and FF1 430 bp probes for characterisation of mutation 5313.** Double digestion of FF1 1.0 kb with *HaeIII* and *HpaII* gave the FF1 430 bp probe. The FF1 430 bp probe covers exon 10c to exon 12b and was used to characterise mutation 5313.

entire P5 fragment is 1700 bp and covers exons 36-49. It overlaps the B3A fragment by 500 bp. Digestion of B3A plasmid released a 800 bp fragment which covers exons 45-49, the stop codon and some 3' untranslated sequence.

The digested fragments were fractionated on 1% low melting agarose gel in 1x TAE buffer. Bands were excised under UV and purified with Wizard<sup>TM</sup> PCR DNA Purification kit (Promega) following the manufacturer's instruction. Briefly, the agarose slice was melted at 70<sup>0</sup>C, and mixed with the resin. Then, the resin/DNA mix was loaded onto a Wizard<sup>TM</sup> minicolumn, spun for 12 sec, and washed with 2 ml of 80% isopropanol. The bound DNA was eluted with 50  $\mu$ l of water.

The probes were labeled using a Multiprime DNA labeling Kit (Amersham) following the manufacturer's instructions. Briefly, about 25 ng of DNA was denatured for 5 min and then labeled at 37<sup>0</sup>C for 1 hr in a total volume of 50  $\mu$ l containing dATP, dGTP, dTTP, and random hexanucleotide primers in the presence of [ $\alpha$ -<sup>32</sup>P]dCTP (Amersham 3000 Ci/mmol, 25  $\mu$ Ci/ $\mu$ l), with 2 u of DNA polymerase I Klenow fragment. The labeled probes were purified with a Sephadex G-50 spun-column in the presence of salmon sperm DNA as carrier (Sambrook et al., 1989).

Blotting, hybridization and washing were as follows. About 10  $\mu$ g of digested genomic DNA were separated in a 0.7 % agarose gel, depurinated with 0.25 M HCl, and denatured with 0.5 M NaOH. The DNA was alkali blotted to Hybond<sup>TM</sup>-N<sup>+</sup> nylon membrane (Amersham). The transfer was at room temperature with 0.4 M NaOH for 12-

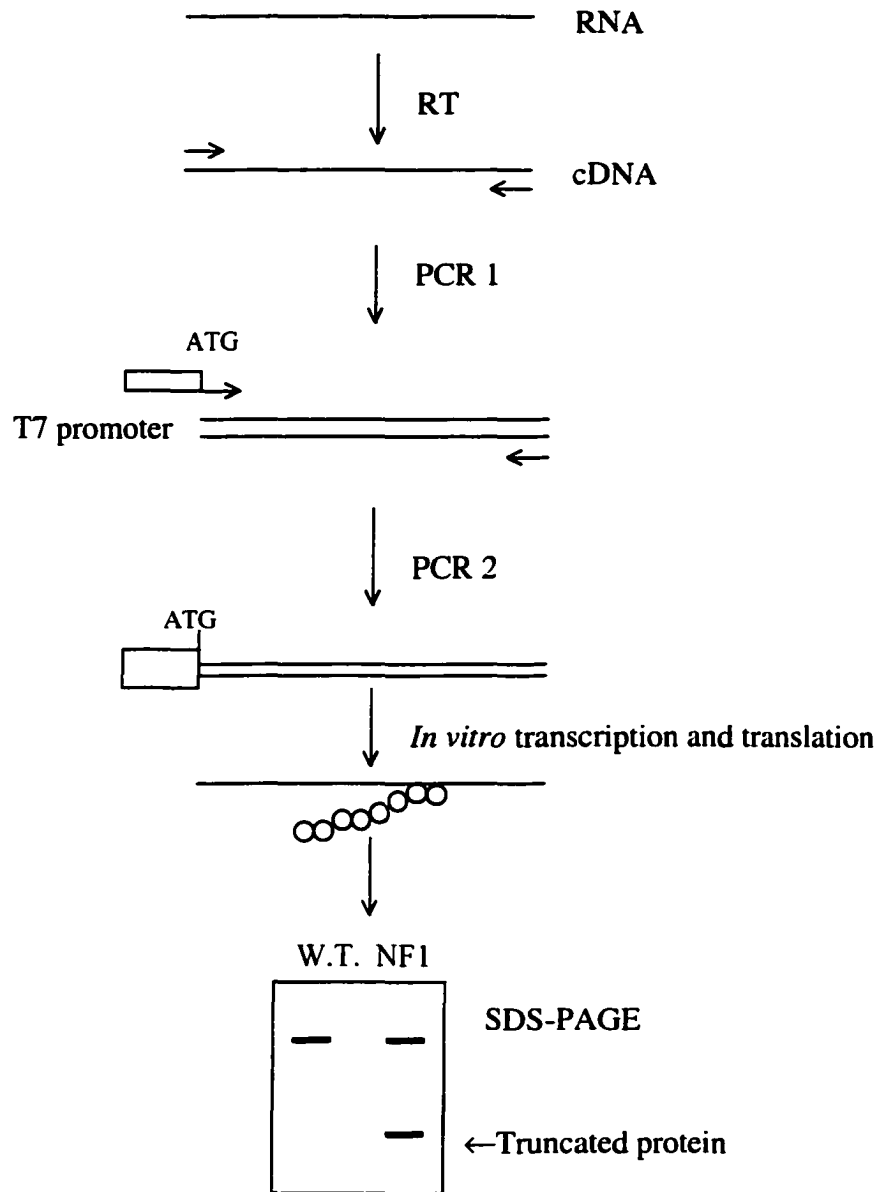
24 hr. The membrane was pre-hybridized in a solution containing 5x SSC, 5x Denhardt's solution, 0.5% SDS and 20 µg/ml of denatured salmon sperm DNA at 65°C in a hybridization bag. Hybridization was carried out with labeled probe overnight at 65°C. After hybridization, the membrane was washed with 2x SSC for 20 min at room temperature and then with 0.1 x SSC containing 0.1% SDS at 65°C for 30-60 min, and autoradiographed. For rehybridization, the membrane was stripped with 0.5% SDS (w/v).

Both *EcoRI* and *HincII* polymorphisms were detected by Southern blotting with probe GE2-400 bp. The *EcoRI* polymorphism detects two-allele fragments of 7.3 kb and 4.2 kb (Reyniers et al., 1993). For the *HincII* polymorphism analysis, genomic DNA was extracted from 67 unrelated individuals. About 10-15 µg of genomic DNA was digested with *HincII* at 37°C for at least 2 hr.

## **2.8 Protein truncation test (PTT)**

The PTT is based upon a coupled *in vitro* transcription and translation assay (Figure 11). It was performed as described (Heim et al., 1995) with minor modifications (Bahau et al., 1998). Mainly, the entire coding region of the NF1 gene was amplified in nine instead of five overlapping segments by RT-PCR. This has the advantage to produce fragments of size between 928 and 1484 bp and thus make the localization of the mutation easier.

### **2.8.1 Transcription template synthesis**



**Figure 11. Schematic representation of protein truncation test.** RNA is extracted and reverse transcribed into cDNA. Specific fragments are amplified by PCR and used as templates for transcription. After *in vitro* transcription and translation, protein products are labelled by  $^{35}\text{S}$ -methionine and separated on a SDS-PAGE. If there is a mutation affecting the protein size, a truncated product will be detected.

First, the cDNA was synthesized. About 5 µg each RNA was incubated 10 min at 65°C followed by 5 min at 25°C in a 50 µl reaction mix containing 0.25 nmol of oligo d(T)<sub>16</sub> (Perkin-Elmer) and 2.5 mM of random hexamers (Pharmacia). Then, the above reaction mix was adjusted to a final volume of 100 µl by adding 1x Superscript buffer (GibcoBRL), 0.5 mM dNTP, 10 mM dithiothreitol (DDT), 20 u RNasin (Promega), and 100 u Superscript<sup>®</sup> reverse transcriptase (Gibco-BRL). The reverse transcription was incubated at 42°C for 40 min and stopped by heating at 90°C for 5 min.

Second, a two-step PCR procedure was used to produce 9 overlapping fragments (A to I) covering the entire neurofibromin cDNA (Table 6). Step one used nine pairs of primers (A1-A2, B1-B2, C1-C2, etc.) to amplify the entire cDNA. Step two used another nine pairs of primers (A3-NFPTT, B3-NFPTT, C3-NFPTT, etc) to further amplify each segments. NFPTT primer is common to all fragments (A-I) and contains a T7 promoter and a translation initiation site within a Kozak sequence for efficient transcription and translation. The primers (A3, B3, C3, etc.) are internal (nested) to the step-one PCR product to promote specific amplifications. Step 1 PCR involved a 50 µl reaction mix containing 10 µl of the RT product (about 250 ng cDNA), 1 x PCR buffer II (Perkin-Elmer), 1.5 mM MgCl<sub>2</sub>, 0.1 µM each primer, 1 u Taq polymerase and 0.2 µl TaqStart<sup>®</sup> Antibody (Clontech). The amplification was 5 min at 94°C, 20 cycles of 30 sec at 94°C, 30 sec at 55°C, and 1 min at 72°C, and finally 7 min at 72°C. Step 2 PCR was performed in a 100 µl reaction mix containing 1 µl of the above PCR product, 1x PCR buffer I (Perkin-Elmer), 0.2 mM dNTP, 0.1 µM each primer, 2 u Taq polymerase (Perkin-Elmer),

**Table 6. Sequences of primers used in PCR for protein truncation test.**

The primers used in PCR1 (e.g. A1-A2, B1-B2, etc.) include a 3' sequence specific for the NF1 cDNA and a 5' sequence correspondent to a Kozak sequence and part of the T7 RNA polymerase promoter (underlined).

Fragments	Exons	PCR primers	Sizes of PCR products (bp)
A	1-8	A1 5' <u>CACTATAGGAACAGACCACCATGGCCGCGCACAGGCCGGTGGAA</u> 3'	A1-A2 : 1175
		A2 5'ACGAAAGCAAGAAACAAGGCAGTCA3'	
		A3 5'CATTAGATCCACATCTGCAGGCTGA3' NFPTT 5'GGATCCTAATACGACTCACTATAGGAACAGACCACC3'	A3-NFPTT: 1164
B	6-12a	B1 5' <u>CACTATAGGAACAGACCACCATGAATAAGAAGTTATTTCTGGAC</u> 3'	B1-B2 : 940
		B2 5'AACCACTTGAGAATTTCTGTGCT3'	
		B3 5'GCATTTGATGACTAGTTAATTC3'	B3-NFPTT: 928
C	10b-17	C1 5' <u>CACTATAGGAACAGACCACCATGGTGAAACTAATTCAT</u> 3'	C1-C2 : 1492
		C2 5'GCTTGCCCTAGATGTTTCAGA3'	
		C3 5'GCCTTCAGTATGATTATCTAGCA3'	C3-NFPTT: 1485
D	16-23	D1 5' <u>CACTATAGGAACAGACCACCATGGGTCCAGTCAGTGAA</u> 3'	D1-D2 : 1367
		D2 5'TTGCCAATCAGAGGATGTGA3'	
		D3 5'TCGTAATAAAGGATCCAGGAGTT3'	D3-NFPTT: 1356
E	21-27b	E1 5' <u>CACTATAGGAACAGACCACCATGGGTGATCAAGGAGA</u> ACTCC3'	E1-E2 : 1336
		E2 5'CGATTGCTAGGCCCGGTATG3'	
		E3 5'CATTGATTTGACCAGTTTTGAA3'	E3-NFPTT: 1252
F	27a-32	F1 5' <u>CACTATAGGAACAGACCACCATGGCAACACTTCTTGCATAC</u> 3'	F1-F2 : 1528
		F2 5'CCAGAAGCCAAAGCTACAGC3'	
		F3 5'CCTGTTGCACTGGTTTTGATGA3'	F3-NFPTT: 1484
G	31-38	G1 5' <u>CACTATAGGAACAGACCACCATGACTCCATGGCTGTCAA</u> 3'	G1-G2 : 1186
		G2 5'AGTGCGGTACCTGCTGAATAC3'	
		G3 5'GCCACAGCTACCCAAAAGAAG3'	G3-NFPTT: 1154
H	35-44	H1 5' <u>CACTATAGGAACAGACCACCATGAGAGATATTC</u> CAACGTG3'	H1-H2 : 1199
		H2 5'CCTGGATCTTCGGATCAGTAA3'	
		H3 5'CGTAAATGTGGGTGCTGTTGT3'	H3-NFPTT: 1147
I	42-49	I1 5' <u>CACTATAGGAACAGACCACCATGGGGCAATCTTCTCAGG</u> 3'	I1-I2 : 1090
		I2 5'CAAGGACAGGGAAGGGGTCACT3'	
		I3 5'GTGAAGAGCCCATGTTAAGTTG3'	I3-NFPTT: 1083

and 0.4  $\mu$ l TaqStart<sup>®</sup> Antibody (Clontech). The amplification was the same as in step 1 except for the number of cycles (25x).

### 2.8.2 *In vitro* transcription and translation

*In vitro* transcription was performed using T7 Cap-Scribe<sup>®</sup> kit (Boehringer Mannheim). The entire PCR step 2 product (100 $\mu$ l) was purified with a column Microcon<sup>®</sup> 30 (Amicon), recovered with 14  $\mu$ l H<sub>2</sub>O/DEPC (0.1% v/v), with addition of 4  $\mu$ l 5x CapScribe<sup>®</sup> buffer and 40 u T7 RNA polymerase for a final volume of 20  $\mu$ l. The reaction mix was incubated 1 hr at 37<sup>0</sup>C. Reaction was stopped by placing the reaction tube on ice. *In vitro* translation was performed with 2  $\mu$ l of each of the above transcription product using the Reticulocyte Type I<sup>®</sup> kit (Boehringer Mannheim), following the manufacturer's instructions, and labeled with <sup>35</sup>S-methionine (1,000 Ci/mmol, Amersham). The translation products were separated by SDS-PAGE. The gel was dried and autoradiographed.

### 2.8.3 cDNA sequencing

The cDNA sequencing was carried out using the ABI PRISM<sup>®</sup> Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer), with AmpliTaq<sup>®</sup> DNA polymerase FS, following the manufacturer's instruction. The sequencing template was 100 ng RT-PCR product purified with Microcon<sup>®</sup> 30 (Amicon) columns. Oligonucleotides NFEx16A and NFEx16B were used as extension primers. Reaction

products were precipitated in a 70% ethanol/0.5 mM MgCl<sub>2</sub> solution, dried, and resuspended in 9 µl of loading buffer (5 volumes of deionized formamide, 1 volume of 25 mM EDTA pH 8, and 50 mg/ml dextran blue), denatured 2 min at 90°C, and placed on ice. A 1.5 µl of each sample was loaded onto a 6% acrylamide-urea gel for electrophoresis, using a ABI PRISM<sup>®</sup> 377 DNA sequencer (Perkin-Elmer), followed by analysis with DNA Sequencing Analysis Software.

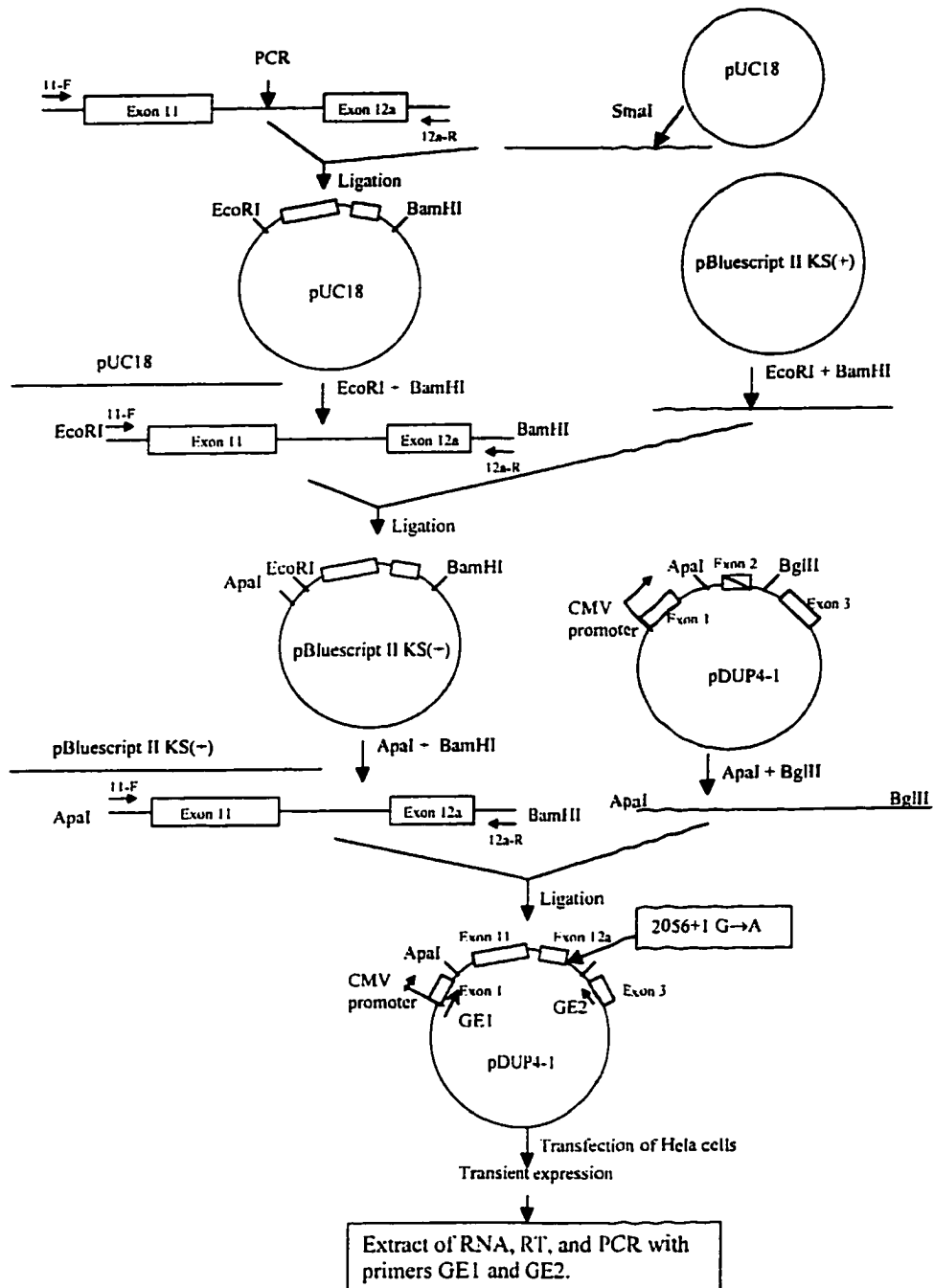
## 2.9 Splicing reaction

### 2.9.1 *In vivo* splicing

Plasmid DUP4-1, a chimeric β-globin minigene (Modafferi and Black, 1997; Wei et al., 1997), was used for *in vivo* splicing. The globin sequences were fused to the CMV immediately-early promoter. Between exon 1 and exon 3 (from globin), there is a 33-nucleotide hybrid exon, exon 2. The exon 2 is always spliced out and thus DUP4-1 was used as control. The NF1 fragments were cloned at *Apa*I site (47 nucleotide upstream from exon 2) and *Bgl*II site (46 nucleotide downstream from the exon 2) to replace exon 2 (Fig. 12, also referred to Fig. 22 and Fig. 23).

Three types of NF1 genomic fragments were used: DUP11-12a, DUP12aL and DUP12aS of both the wild type (wt) and mutant (mut). (1) The DUP11-12a fragment (Figure 12) was from digestion of pUC11-12a with *Eco*RI and *Bam*HI. This *Eco*RI-*Bam*HI fragment was transiently cloned into pBluescript II KS(+) to obtain an *Apa*I site



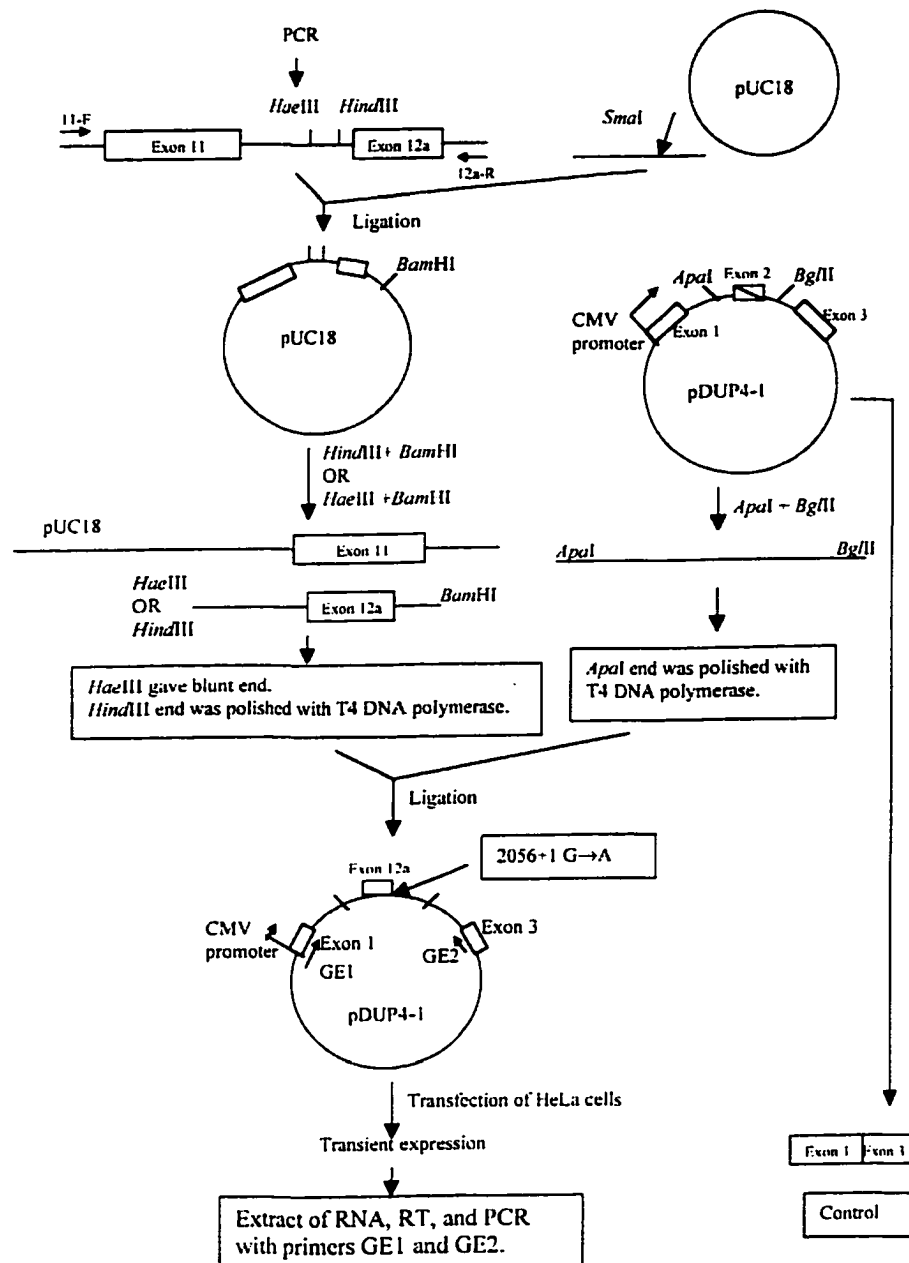


**Figure 12. Strategy used to clone genomic 11-12a into pDUP4-1 for characterization of mutation 5313.** The 11-12a was obtained by digestion of pUC11-12a with *EcoRI* and *BamHI*. The *EcoRI* and *BamHI* fragment was cloned into pBluescript II KS(+). The *ApaI*-*BamHI* fragment was then cloned into pDUP4-1. DUP11-12a was used to transfect HeLa cells. After transient expression, RNA was extracted and reverse transcribed. Splicing products were amplified by PCR.

at the 5' end. The *ApaI*-*Bam*HI fragment from pBluescript II KS(+) was recloned into DUP4-1 at *ApaI* and *Bg*III site. (2) The DUP12aL fragment (550 bp) was obtained by digestion of pUC11-12a with *Hae*III and *Bam*HI (Figure 13). The *Hae*III site is at intron 11 (1934-371, accession number M82814). The *Bam*HI site is a pUC18 site. This *Hae*III-*Bam*HI fragment was cloned in DUP4-1 at the *ApaI* and *Bg*III site with *ApaI* site blunted with T4 DNA polymerase. (3) The DUP12aS fragment (287 bp) was prepared similarly to DUP12L by digestion of pUC11-12a with *Hind*III (at intron 11, 1934-105, accession number M82814) and *Bam*HI (Figure 13). The *Hind*III site (in NF1 fragment) and the *ApaI* site (in DUP4-1) were both blunted with T4 DNA polymerase.

Plasmid DNA was purified by banding in CsCl gradients twice. Lipofectamine (GibcoBRL) was used to transfect HeLa cells following the manufacturer's instructions. Briefly, 10-20 µg of each plasmid DNA was mixed with 20-50 µl of Lipofectamine and incubated in room temperature for 45 min to allow DNA-liposome complexes to form. The DNA-liposome complexes were laid onto PBS rinsed cells. Transfection was carried out for 24 hr. After removing the DNA-liposome complexes, cells were allowed to transiently express for another 24 hr.

Cytoplasmic RNA from transfected cells was prepared by TRIzol (GibcoBRL) and treated with DNase I. Reverse transcription was carried out with random hexamer primers using 10 units of AMV reverse transcriptase in 20 µl of total volume containing up to 5 µg of total RNA, 10 mM DTT, 1 mM dNTP and 4 u of RNasin (Promega). The mixture was incubated for 10 min at 20°C, then 1 hr at 42°C. PCR reactions were in



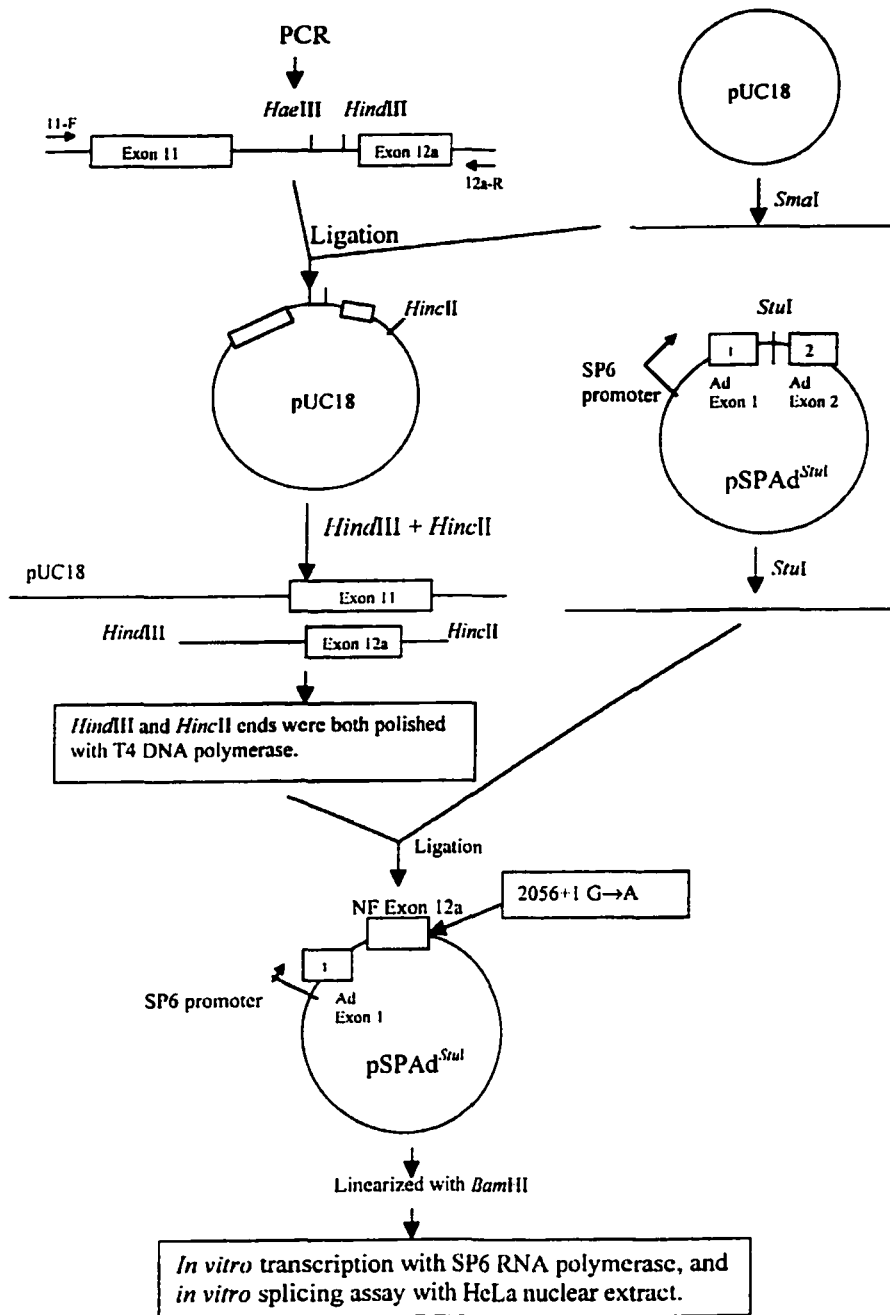
**Figure 13. Strategy used to clone genomic 12a into pDUP4-1 for characterization of mutation 5313.** The 12aL was obtained by digestion of pUC11-12a with *Hae*III and *Bam*HI which was then cloned into pDUP4-1. The 12aS was obtained by digestion of pUC11-12a with *Hind*III and *Bam*HI and cloned into pDUP4-1. The constructs were used to transfect HeLa cells. After transient expression, RNA was extracted and reverse transcribed. Splicing products were amplified by PCR with DUP primers GE1 and GE2. The parent pDU4-1 was used as control.

25 µl containing 2.5 µl of cDNA, 1x Taq buffer (Amersham), 2 mM dNTP, 5 pmol each primers GE1: 5' ACA CAA CTG TGT TCA CTA GC 3' and GE2: 5' AGT GGA CAG ATC CCC AAA GG 3', 2.5 u Taq polymerase. The amplification was 5 min at 96<sup>0</sup>C, 35 cycles of 1 min at 93<sup>0</sup>C, 1 min at 55<sup>0</sup>C, and 1 min at 72<sup>0</sup>C, and finally 10 min elongation at 72<sup>0</sup>C. PCR products were electrophoresed in 1% agarose gel and bands were excised. DNA were electroeluted, cloned into pUC18 at the *SmaI* site, and sequenced.

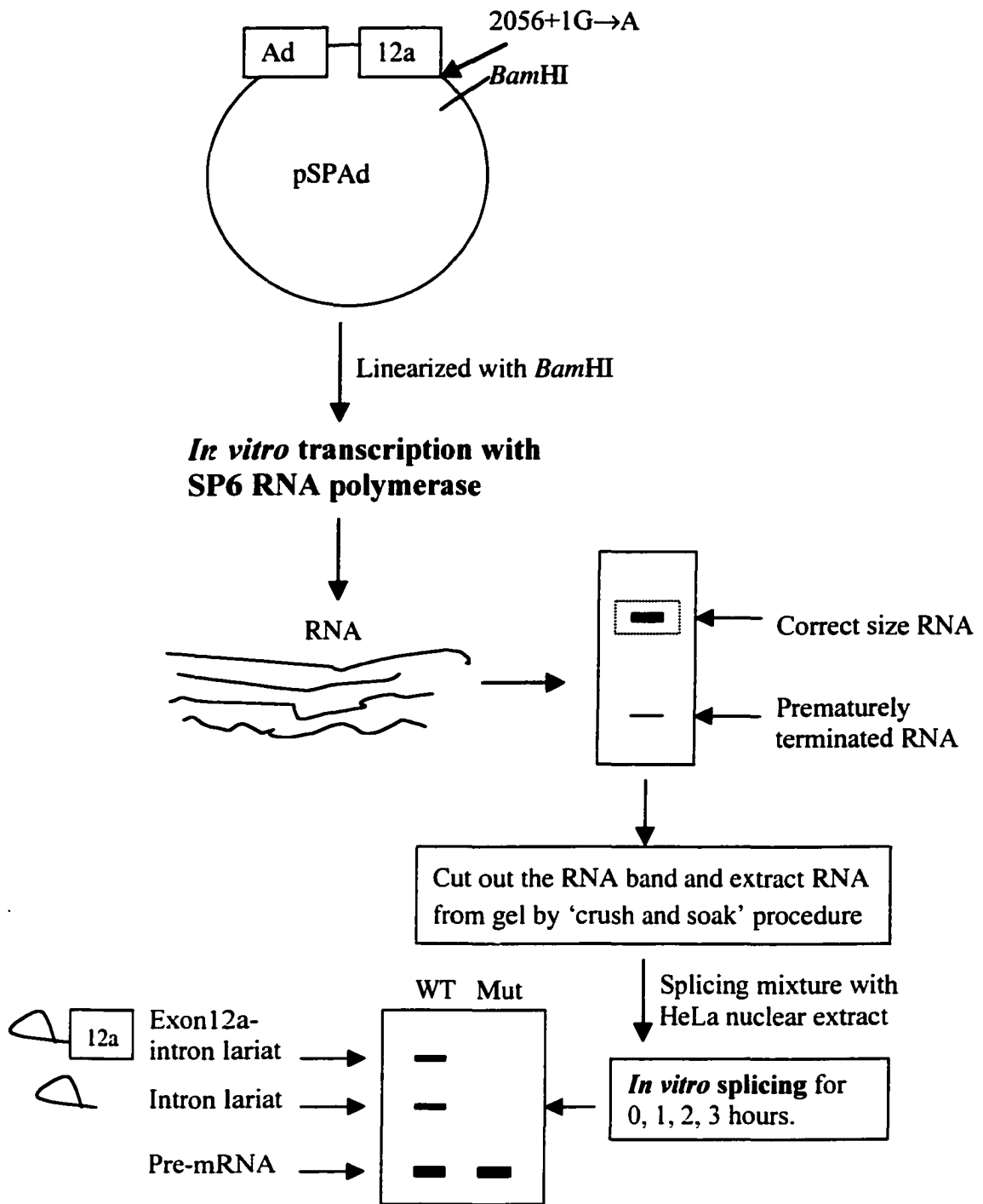
### 2.9.2 *In vitro* splicing

The plasmid pSPAd<sup>StuI</sup> was described previously (Lavigneur et al., 1993) with the *HindIII* site at position 225 of the intron modified to a *StuI* site. A 280 bp fragment was prepared by digestion of pUC11-12a with *HindIII* and *HincII*, and cloned into pSPAd<sup>StuI</sup> at the *StuI* site (Figure 14). The ends of *HindIII* (at intron 11, position 1934-105, accession number M82814) and *HincII* (a pUC18 site) were both blunted. Splicing substrates were produced from plasmids linearized with *BamHI* and transcribed with SP6 RNA polymerase (Pharmacia) in the presence of cap analog and [ $\alpha$ -<sup>32</sup>P]UTP (Amersham) (Figure 15). Transcript RNA purification (Chabot, 1994), HeLa nuclear extracts preparation (Dignam et al., 1983) and splicing reaction (Chabot et al., 1997) were carried out as described.

### 2.9.3 Native gels- RNA mobility-shift assays



**Figure 14. Strategy used to clone genomic 12a into pSPAd for characterization of mutation 5313.** The 12a fragment was obtained by digestion of pUC11-12a with *Hae*III and *Hinc*II. The *Hae*III and *Hinc*II sites were blunted and then cloned into pSPAd at the *Stu*I site. The construct was linearized with *Bam*HI and used as template for *in vitro* transcription with SP6 RNA polymerase.



**Figure 15. Schematic diagram of production of pre-mRNA template and the *in vitro* splicing reaction.** The exon 12a was cloned into pSPAd and linearized with *Bam*HI to give the template DNA. The 'run-off' transcription used SP6 RNA polymerase. The nascent RNA of the correct size was cut from the gel, extracted, and used for *in vitro* splicing in a HeLa nuclear extract.

RNA mobility-shift assays were performed by incubating RNAs for 15 min on ice in splicing conditions prior to addition of 1 mg/ml heparin and incubation for 2 min on ice. The reaction products were run on a 5% native acrylamide gel, 5% glycerol, 50 mM Tris, 50 mM glycine in Tris-glycine running buffer (50 mM Tris, 50 mM glycine).

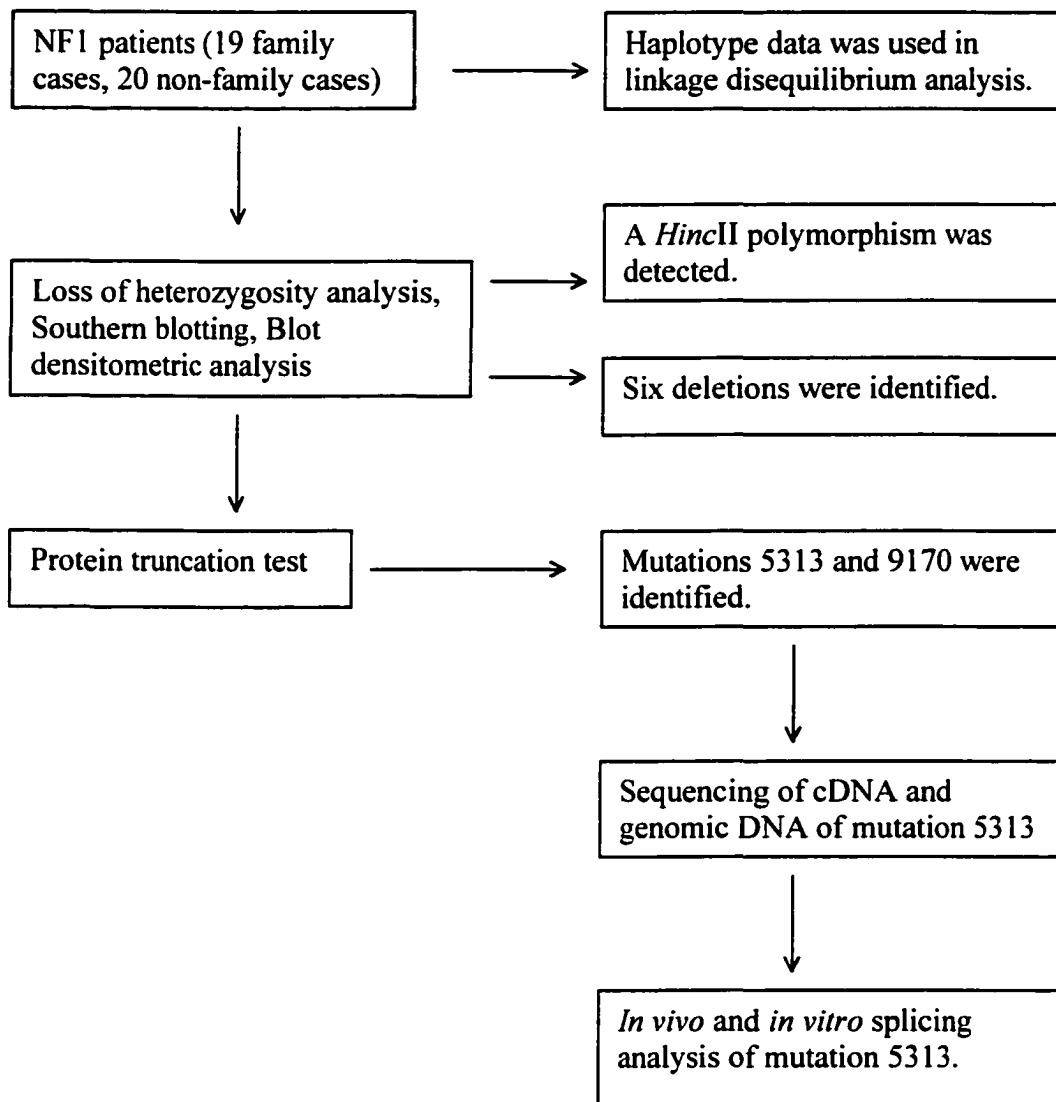
## CHAPTER 3 RESULTS AND DISCUSSION

This chapter is divided into four sections, namely:

3.1 A novel mutation in the <i>NF1</i> gene promotes skipping of two exons by affecting exon definition.....	58
3.2 Identification and characterization of six novel deletions in the <i>NF1</i> gene .....	77
3.3 A novel and very peculiar <i>HincII</i> polymorphism in the 5' region of the human Neurofibromatosis type 1 ( <i>NF1</i> ) gene .....	97
3.4 Linkage disequilibrium analysis between four intragenic polymorphic microsatellites of the <i>NF1</i> gene in the Québec population.....	102

The strategy for the *NF1* mutation study is as outlined in Figure 16. Firstly, the segregation of disease-bearing chromosome was analyzed by haplotyping the *NF1* families. Then, the analysis of loss of heterozygosity (LOH) and Southern blotting were used to study deletions. Meanwhile, protein truncation test (PTT) was used to detect mutations with truncated neurofibromin. A splicing mutation was identified in family 5313 (Section 3.1). Six gross deletions were detected in two families (7610 and 7473) and four patients (178, 184, 236 and 237) (Section 3.2). During the screening deletion by Southern blotting, a *HincII* polymorphism was detected in the 5' region of the *NF1* gene (Section 3.3). Linkage disequilibrium at the *NF1* locus between four intragenic microsatellites was investigated in the Québec population by using the haplotype data (Section 3.4). The work was collaborated with Drs. D. Vidaud and Jean-Paul Thirion (protein truncation test), and Mr. Martin J. Simard (*in vitro* splicing).





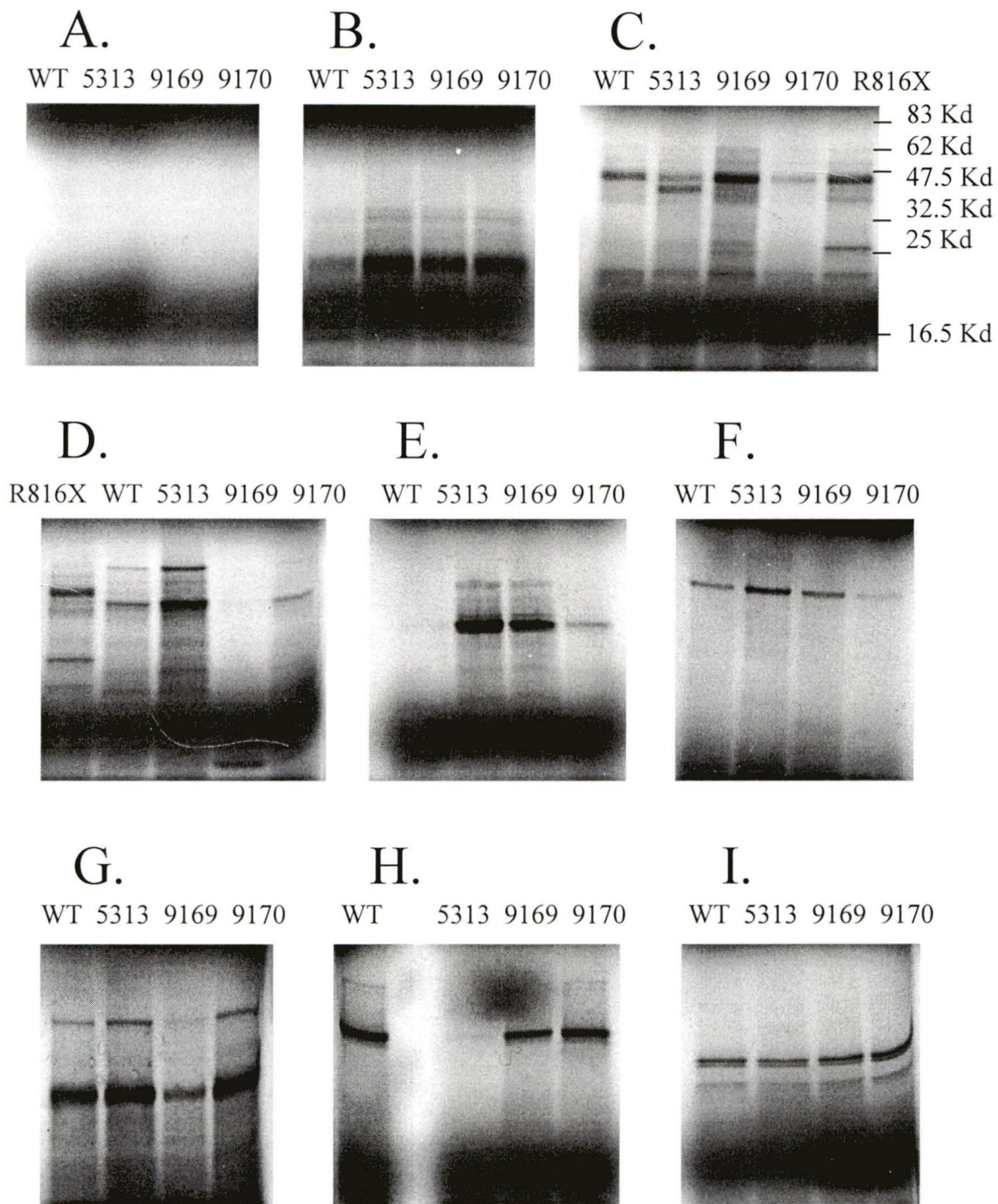
**Figure 16. Strategy for identification and characterization of *NF1* gene mutations.**

### **3.1 A novel mutation in the *NFI* gene promotes skipping of two exons by affecting exon definition**

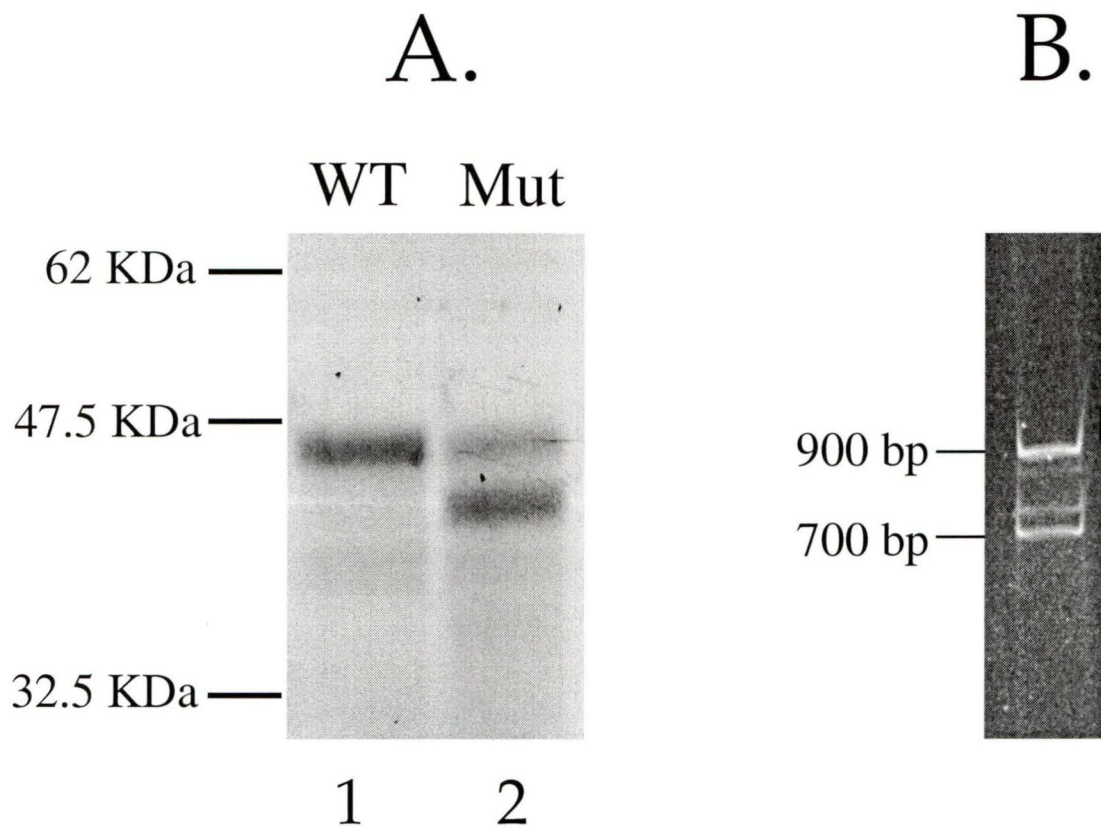
#### **3.1.1 Results**

To identify the relative position of the mutation in the *NFI* gene of patients Rou 5313, Rou 9169 and Rou 9170 (thereafter called 5313, 9169 and 9170 for simplification), a protein truncation test was performed using RNAs from these patients and from a normal individual as wild type control. These RNA samples were reverse transcribed and amplified to generate nine nested sets of overlapping cDNA fragments (Fig. 17) (Bahau et al., 1998) which were individually subjected to *in vitro* transcription and translation. On SDS gels there was no significant difference between the translation products obtained with patient 5313 and the control except for the polypeptides generated with fragment C (Fig. 17, Fig. 18A). Whereas the control sample yielded a 46 Kd product, fragment C from patient 5313 produced both the 46 Kd product and a shorter peptide of approximately 42.5 Kd. Because fragment C spans 1459 bp of coding sequences from exons 10b to 17, it could theoretically yield a 486 amino acid peptide. The truncated polypeptide was estimated to be around 35-50 amino acids shorter than the wild type polypeptide. No truncated protein was detected with patients 9169 and 9170 (Fig. 17).

Similar PTT experiments were carried out with patients 9169 and 9170 by using different primers dividing the coding sequence into 5 overlapping segments (Laboratory Dr. Kam-Morgan, Laboratory Corporation of America, Research Triangle Park, USA)



**Figure 17. Protein truncation test results.** The translation products from fragments A-I of the wild type (WT) and patients R816X (control), 5313, 9169 and 9170 were labeled with  $^{35}\text{S}$ -methionine and fractionated by SDS-PAGE. Molecular weights were indicated along fragment C by comparison with colored protein markers. A protein truncation mutant R816X was used as positive control. No difference was observed except for fragment C of patient 5313 and control R816X. Fragment A did not work for some unknown reasons.



**Figure 18. Protein truncation and heteroduplex formation assays.** (A) Protein truncation assay. The translation products from fragment C of the wild type (WT) and mutant (Mut) NF1 cDNA were labeled with  $^{35}\text{S}$ -methionine and fractionated by SDS-PAGE. The size of the wild type and the mutant proteins are 46 Kd and 42.5 Kd, respectively. Molecular weights were determined by comparison with colored protein markers (B) Gel electrophoresis of the RT-PCR products derived from fragment C of patient 5313. The 900 bp and 700 bp fragments correspond to the wild-type and mutated alleles, respectively, while the 750 bp fragment most probably corresponds to an heteroduplex.

(Heim et al., 1995). No truncated polypeptide was detected with patient 9169. With patient 9170, an approximate 29 Kd truncated protein was detected in segment 2, which encodes into a full length polypeptide product of 78.2 Kd (data not shown). Segment 2 covers nucleotides 1468 to 3583 of the cDNA sequence, spanning exons 10b to 21 (Heim et al., 1995). This suggests that the mutation in patient 9170 is located between nucleotide 1468 and 2252, thus between exons 10b and 14. However, no further characterization was carried out with mutation 9170.

To confirm that the *NFI* mutation of patient 5313 was located in cDNA fragment C, we monitored heteroduplex formation during PCR amplification of this fragment. Figure 18B indicates that in contrast to the wild type sample, patient 5313 yielded two bands of almost equal intensities of approximately 900 bp and 700 bp each, as well as a third band slightly superior to the 700 bp. This third band most likely represents an heteroduplex formed between one wild type and one mutant DNA strand. This result suggests that the mutant mRNA lacks approximately 200 nucleotides, a value that would correspond to a loss of about 66 amino acids, in agreement with the result of the protein truncation assay.

To identify the mutation, the PCR mixture containing cDNA fragment C from patient 5313 was sequenced (Fig. 19A). The arrow indicates the first position where the profile changes from a unique sequence to the superposition of two different sequences. Left of the arrow, the sequence AGCAATGGAG corresponds to the 3' end of wild type exon 10c. This sequence continues with a mixture of two sequences: GCTCTGCT, which



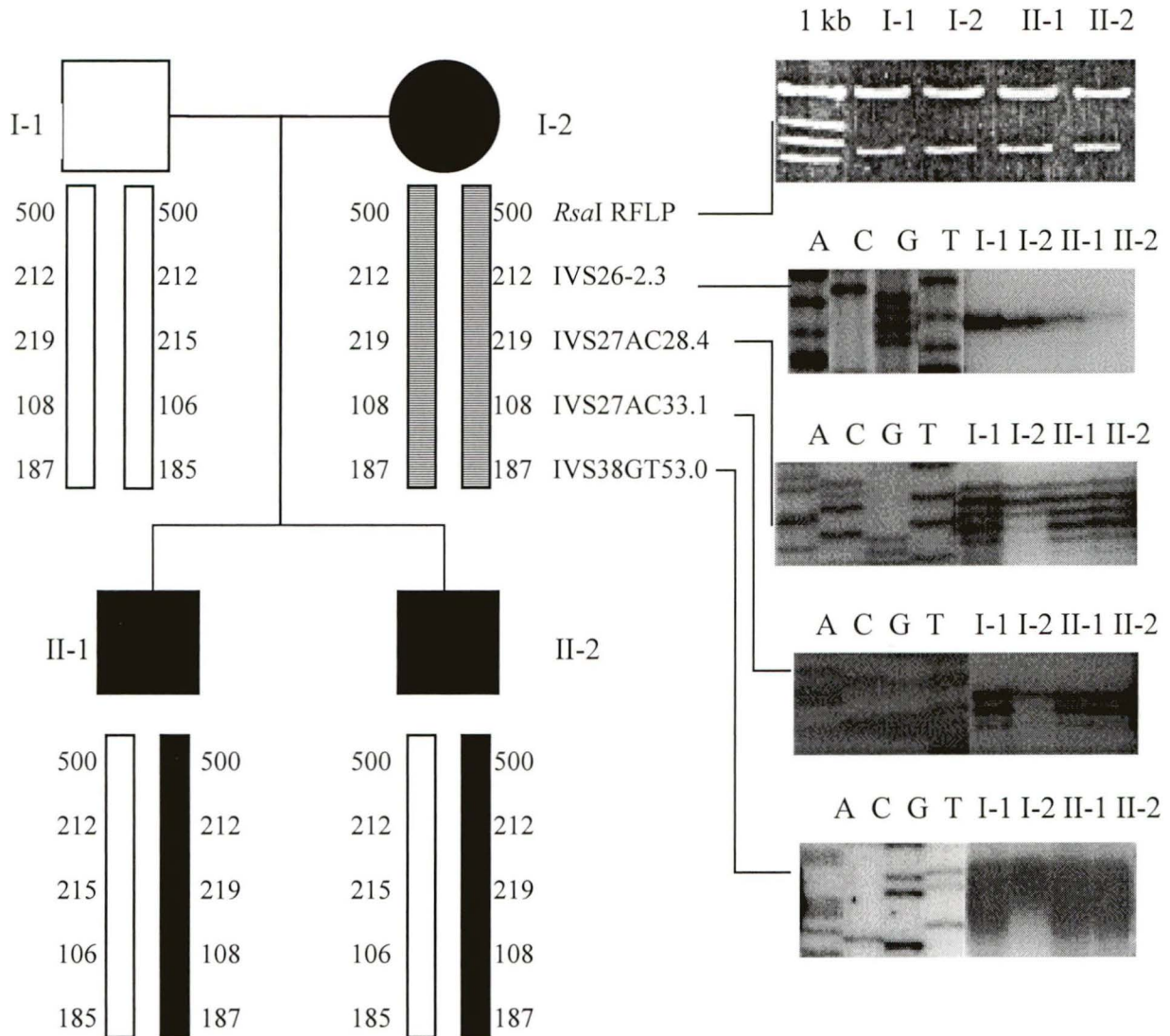
**Figure 19. Sequence of the mutated *NF1* allele.** (A) Sequence of the fragment C portion of the mutated cDNA from patient 5313. Left of the arrow is the normal sequence for the 3' end of exon 10c; this followed by the superposition of the sequences corresponding to exons 11 and 12b. (B) Genomic sequence of the non-coding strand corresponding to the intron 12a-exon 12a boundary for the mutant and wild type alleles. These sequences differ by a C to T transition at the 5' splice site of intron 12 (position 2056+1).



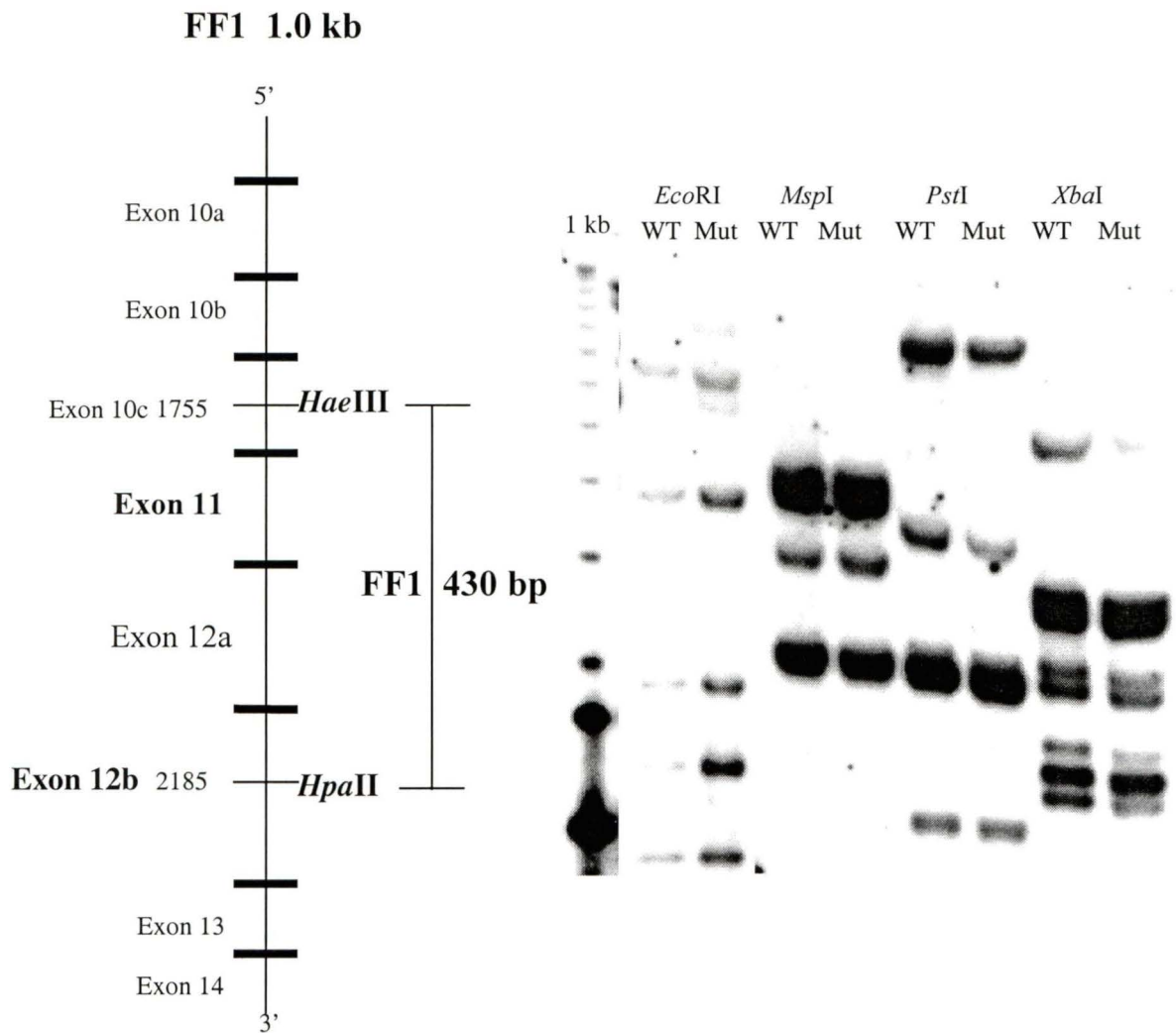
matches the 5' end of exon 11, and CAGGCAGA, which corresponds to the 5' end of exon 12b. This result indicates that the mutated mRNA produced in patient 5313 lacks exons 11 and 12a. Notably, the absence of exons 11 and 12a maintains the proper reading frame in the coding sequence of the mutant mRNA. These observations are therefore entirely consistent with the results of the protein truncation test and heteroduplex formation assay since the loss of exons 11 and 12a corresponds to 204 nucleotides in the NF1 mRNA and 68 amino acids in neurofibromin.

To distinguish between a genomic deletion and a mutation that promotes aberrant pre-mRNA splicing, firstly the patient 5313 family was haplotyped and analyzed for the loss of heterozygosity. Figure 20 shows that the mother (I-2) who could be hemizygous at the *NF1* loci is indeed homozygous by analysis of the segregation of the markers in the whole family at these loci. Then, genomic DNA of patient 5313 was digested with *EcoRI*, *MspI*, *PstI* and *XbaI*, and submitted to Southern blot analysis with probes covering the exons 11 to 12 region. A genomic deletion from exon 11 to exon 12a would remove at least 1.5 kb of DNA. However, no difference was observed between wild type and mutant genomic DNA digests (Fig. 21) suggesting that the mutation that led to exon skipping cannot be a large deletion. To identify the mutation in the gene, the genomic DNA of patient 5313 from intron 10c to intron 12a was amplified by PCR. Ten independent clones derived from this region, which should contain either the normal or the mutated allele were sequenced. For three clones the sequences corresponded to the wild type allele and for seven clones the sequences contained a C to T transition at position 2056+1 (or a G to A on the coding strand) (Fig. 19B). This +1 position





**Figure 20 Pedigree and haplotype of family 5313.** Haplotypes were determined by analysis of 5 polymorphic loci. Microsatellite fragments and *RsaI* polymorphism are indicated as bp at loci along each chromosome. The NF1 patient and the mutation-carrying chromosome are in black. For the mother, it was not possible to identify the mutation-carrying chromosome and the chromosomes are in grey. Results are shown on the right.

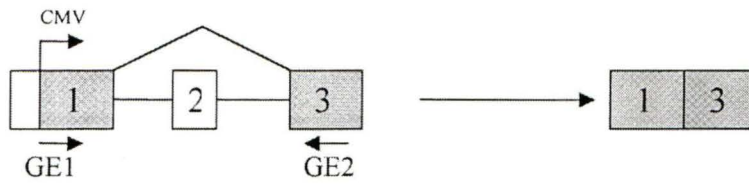


**Figure 21. Southern blot analysis of mutation 5313.** Genomic DNA of the patient 5313 (Mut) and a control wild-type(WT) were digested with *EcoRI*, *MspI*, *PstI* and *XbaI* and sequentially hybridised with probes FF1 (1.0 kb) and FF1 (430 bp). No abnormal band in the mutant sample was detected in both hybridizations, indicating no deletion in this region. The schematic drawing of the probes FF1 (1.0 kb) and FF1 (430 bp) is shown in the left side. The Southern results with probe FF1 (1.0 kb) is shown on the right side.

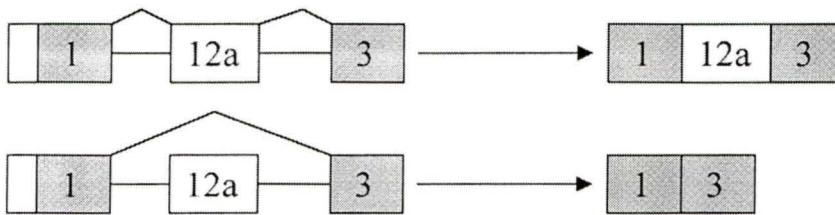
corresponds to the first intron nucleotide of the 5' splice site of exon 12a (wild type sequence is AAG/***GTAAGC***, where / is the exon/intron junction and position +1 is indicated in bold and italic).

Mutations that affect position +1 of an intron often promote the use of aberrant (cryptic) splice sites or lead to skipping of the exon (Krawczak et al., 1992). To investigate the type of splicing anomaly created by the presence of the mutation in the *NF1* intron, we conducted *in vivo* splicing assays using a 2.2 kb genomic portion extending from intron 10c to intron 12a. The wild type and mutated 11-12a segments were inserted in between the two human  $\beta$ -globin exons of a reporter construct (DUP4-1) (Fig. 22). Following transfection of HeLa cells and transient expression, total RNA was isolated. The splicing products were reversed transcribed and amplified by PCR using globin-specific primers. To confirm the identity of these products, DNA bands were excised from the gel, cloned and sequenced. Because the small size of exon 2 leads to exon 2 skipping (Dominski and Kole, 1991), transfection with the parent DUP4-1 always yielded a 192 bp product corresponding to splicing between exon 1 and exon 3 (Fig. 23B, lane 7). The expression vector carrying the wild type 11-12a segment yielded products containing exon 12a but also a product carrying both exons 11 and 12a (Fig. 23B, lane 6). Because this latter product was less abundant and because no product containing the *NF1* exon 11 alone was detected, inclusion of exon 11 is less efficient than inclusion of exon 12a. Comparison of the 3' and 5' splice site sequences TTTTAAAAAATTCAG/ and /GTATAT surrounding exon 11 with respective consensus sequences (Py)<sub>n</sub>NCAG/ and /GT(A/G)AGT (Chabot, 1996) suggests that this is due to weak splice sites (Fig. 24).

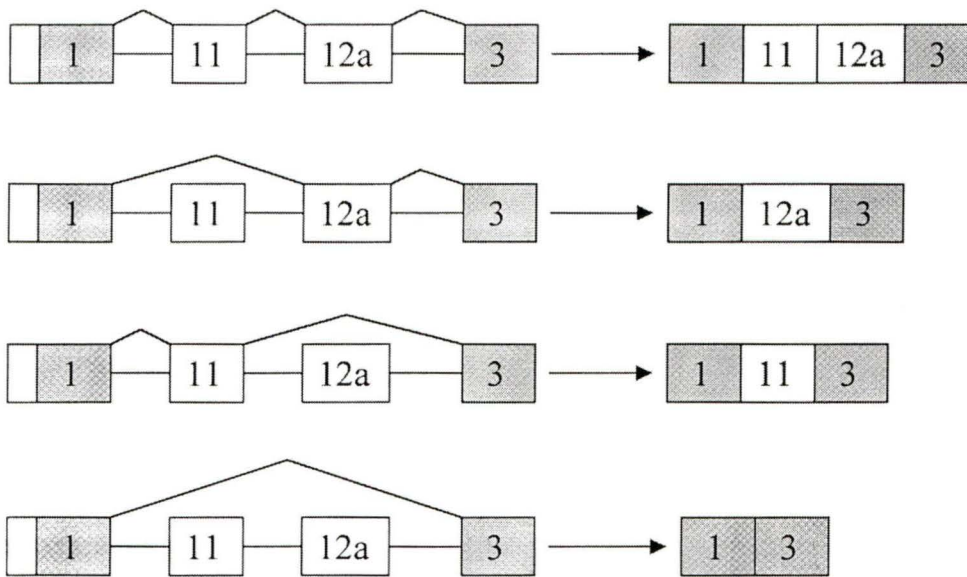
A. DUP4-1



B. 12a in DUP4-1

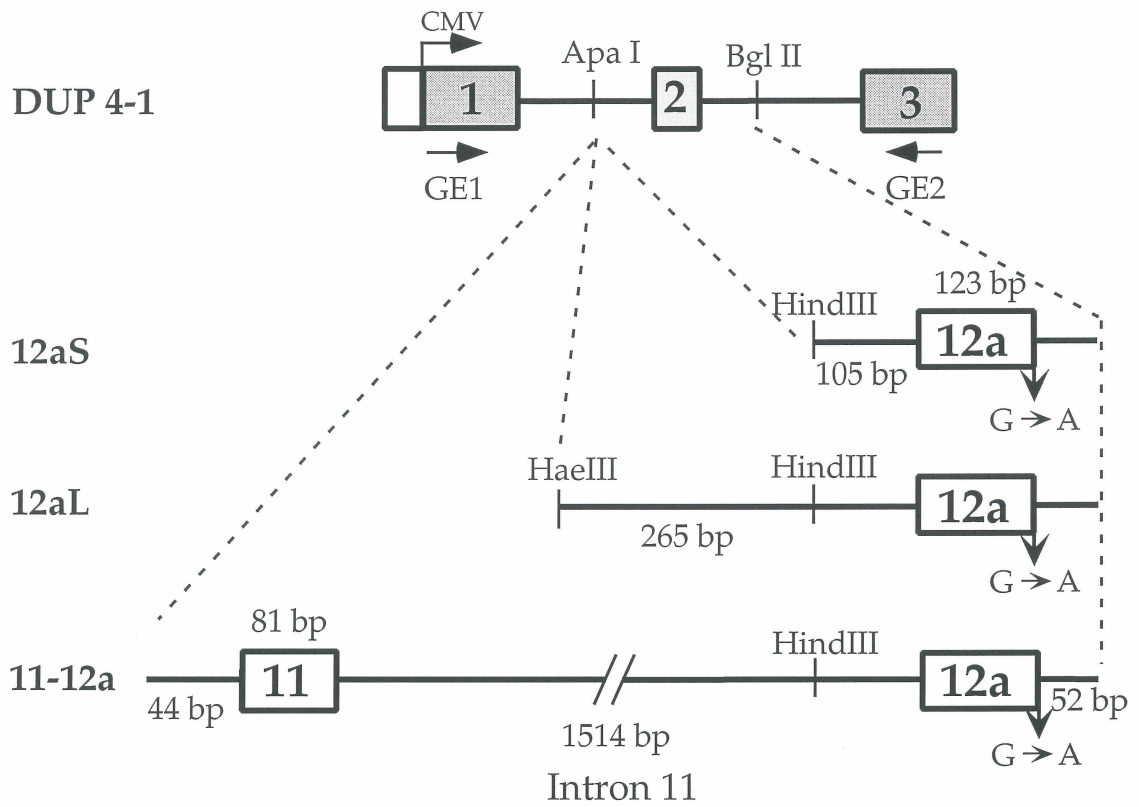


C. 11-12a in DUP4-1



**Figure 22. Theoretical splicing patterns of constructs with wild type *NFI* exons 11 and 12a in DUP4-1.** (A) The parent DUP4-1 always gives exclusion product without exon 2. (B) The DUP12L and DUP12S have two possible splicing patterns: exon 12a inclusion, exon 12a exclusion. (C) The DUP11-12a has four possible splicing patterns: exon 11 and exon 12a both inclusion, exon 12a inclusion, exon 11 inclusion and exon 11 and exon 12a both exclusion.

# A. DUP 4-1

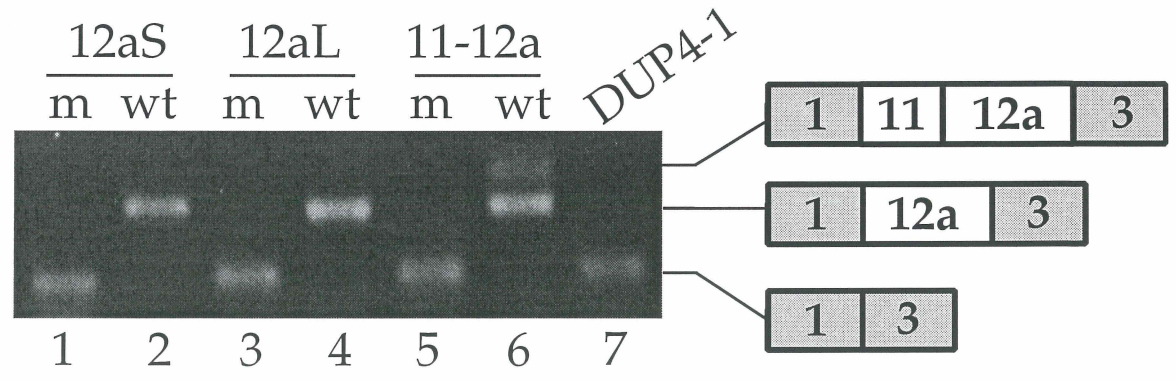


Transfection of HeLa cells

Transient expression

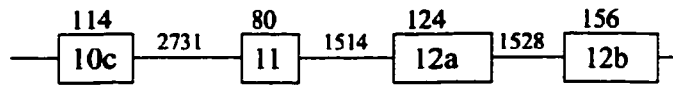
Extract of RNA, RT, and PCR with DUP primers GE1 and GE2

# B.

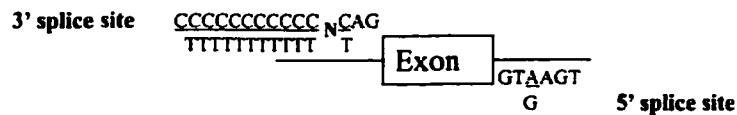


**Figure 23. *In vivo* splicing of NF1 exons 11 and 12a.** (A) Map of the NF1 fragments and the globin minigene. Three NF1 fragments were cloned into the human globin reporter gene DUP4-1 at the *Apa*I and *Bgl*III sites. The position of the G→A mutation is indicated by the arrow. Exons are shown as boxes, the introns as lines. The transcription start site in DUP4-1 is as indicated by the arrow. The positions of PCR primers GE1 and GE2 are as indicated. (B) Migration of the RT-PCR products on agarose gels containing ethidium bromide after amplification with GE1 and GE2 primers. The assay was performed on total cellular RNA after transient transfection of HeLa cells with the wild-type (wt) or the mutated (m) versions of each construct. The amplification products were electroeluted and sequenced (not shown). The identity of the splicing products as indicated on the right. Notice the absence of exon 11 inclusion product.

**A. Size**



**B. Consensus sequences**



**C. Sequences of *NF1* splice sites**



**Figure 24. Sequences of splice junctions in genomic *NF1* from exons 10c-12b.**

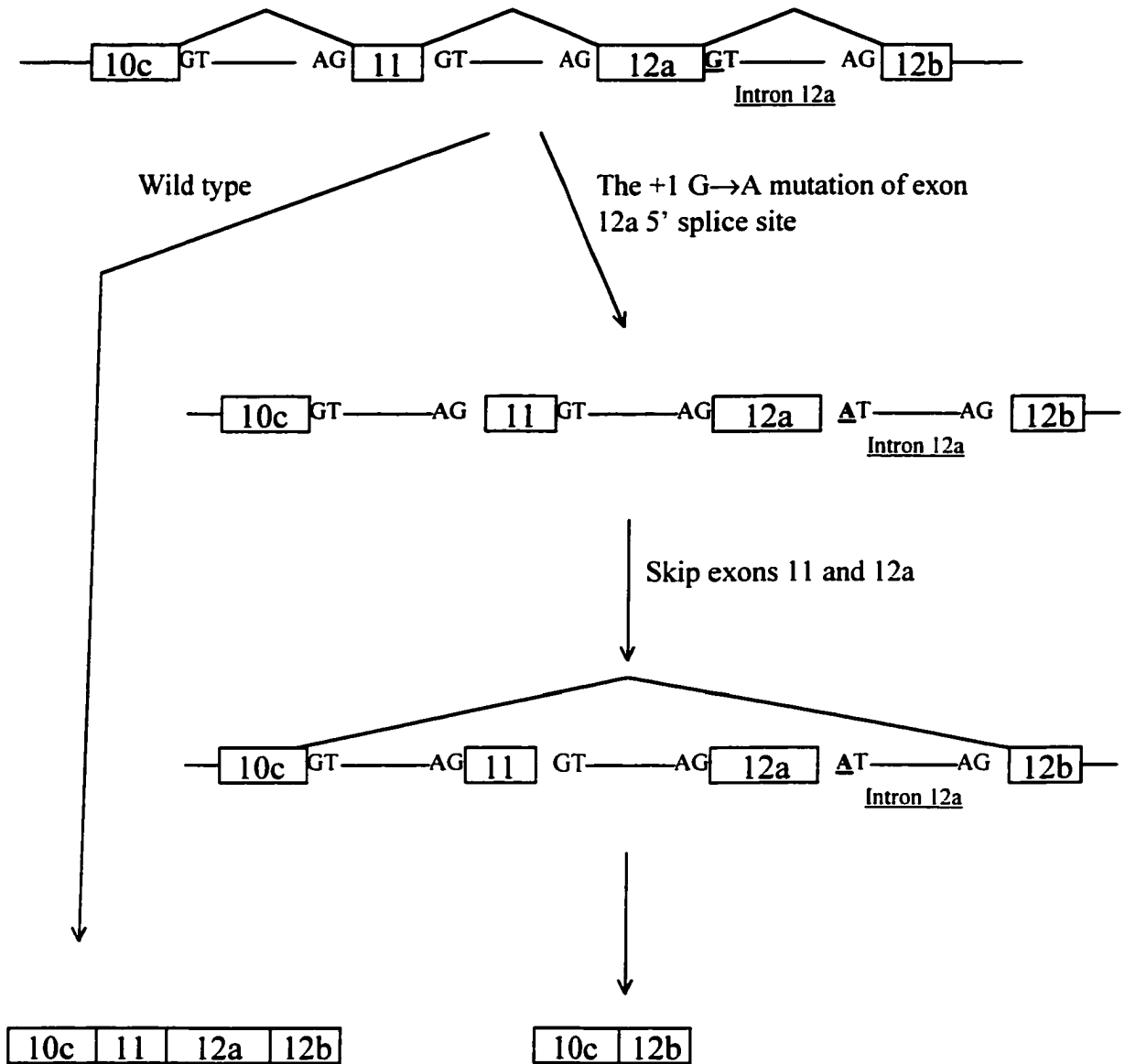
(A) Exon and intron sizes from intron 10b to intron 12 (not to scale); (B) The consensus splice site sequences; and (C) the splice site sequences of exons 11 and 12a. Because position +5 of the 5' splice site of exon 11 is not a G, this donor site is most probably weak. Likewise, the pyrimidine stretches at the 3' splice sites of both exons 11 and 12a are interrupted by purines, suggesting weak sites.

Notably, expression of the mutated version led to complete skipping of both *NF1* exons, thereby reproducing in HeLa cells the effects observed with patient 5313. The effect of the mutation on the inclusion of exon 12a was confirmed by testing DUP versions containing exon 12a and varying portions of flanking intron sequences (Fig. 23A, DUP12aL and DUP12aS). In this case, while wild type DUP12aL and DUP12aS both yielded a product containing exon 12a (315 bp fragment in Fig. 23B, lanes 2 and 4), the mutated versions yielded the exclusion products only (192 bp fragment in lanes 1 and 3). This result confirms that the G→A mutation at position +1 of the 5' splice site of intron 12a prevents the use of exon 12a. In addition, the fact that globin exon 1 is never spliced to the mutant NF1 exon 12a suggests that the 3' splice site of exon 12a is intrinsically weak.

According to the model of exon definition (Berget, 1995), a mutation in the downstream 5' splice site is predicted to reduce the use of the upstream 3' splice site. Thus, the G→A mutation at position +1 has two consequences: first, it inhibits exon 12a/12b splicing, an effect common to many mutations that affect 5' splice sites (Krawczak et al., 1992); second, it prevents exon definition across exon 12a, thereby compromising the use of the 3' splice site of exon 12a. Because inclusion of exon 11 is intrinsically inefficient and may normally rely on the 3' splice site of exon 12a, the mutation promotes skipping of both exon 11 and exon 12a (Fig. 25).

To verify whether the 5' splice site of exon 12a is indeed important for the use of upstream 3' splice site, we relied on an *in vitro* splicing assay using a chimeric





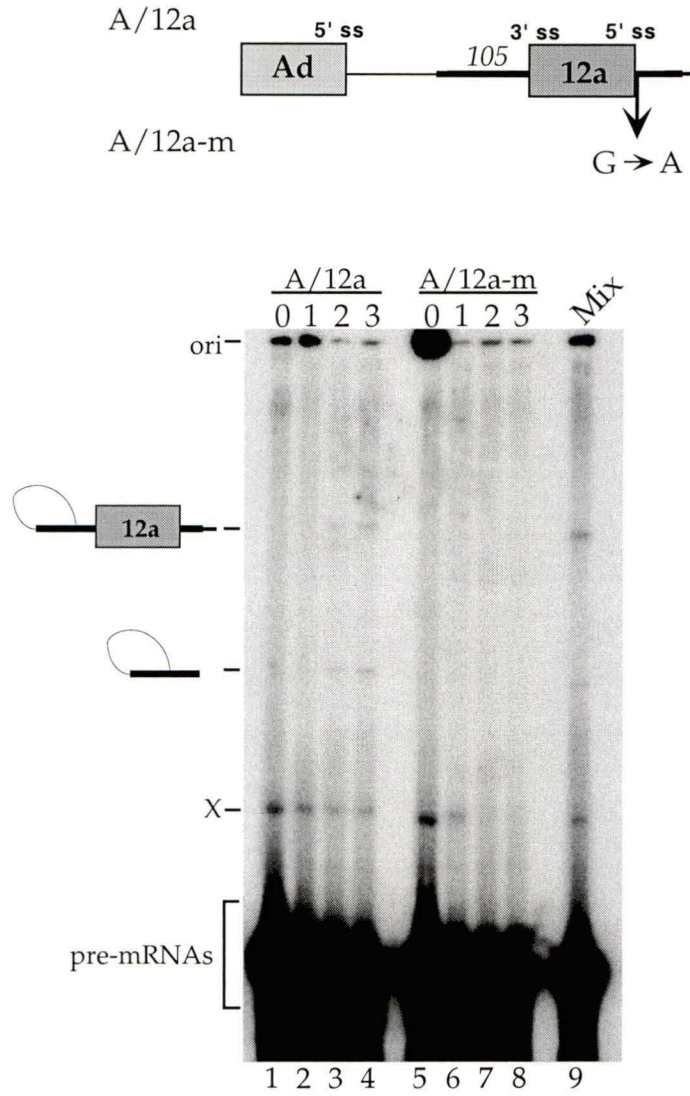
**Figure 25. Mutation 2056+1G→A promotes skipping of exons 11 and 12a.** Exons 10c, 11, 12a and 12b are depicted as boxes and introns thin lines. The conserved dinucleotide at splice site, 5' GT and 3' AG, are shown.

adenovirus/NF1 pre-mRNA (Fig. 26A). Following a time-course incubation in a HeLa nuclear extract, lariat products diagnostic of splicing were detected with the pre-mRNA containing the wild type 5' splice site of exon 12a (Fig. 26A, lanes 1-4). In contrast, splicing products derived from the pre-mRNA carrying the mutated 5' splice site of exon 12a could not be detected (lanes 5-8). This result demonstrates that an intact 5' splice site of exon 12a is essential for the use of the upstream 3' splice site, consistent with the exon definition model. As confirmed by native gel analysis, lack of splicing was associated with inefficient assembly of splicing complexes on the 3' splice site of exon 12a (Fig. 26B). Overall, our results demonstrate that the mutation at the 5' splice site of exon 12a inactivates this site and prevents the use of the upstream 3' splice site. Combined with the intrinsic weakness of exon 11 splice sites, the G→A mutation yields mRNAs lacking both exons 11 and 12a.

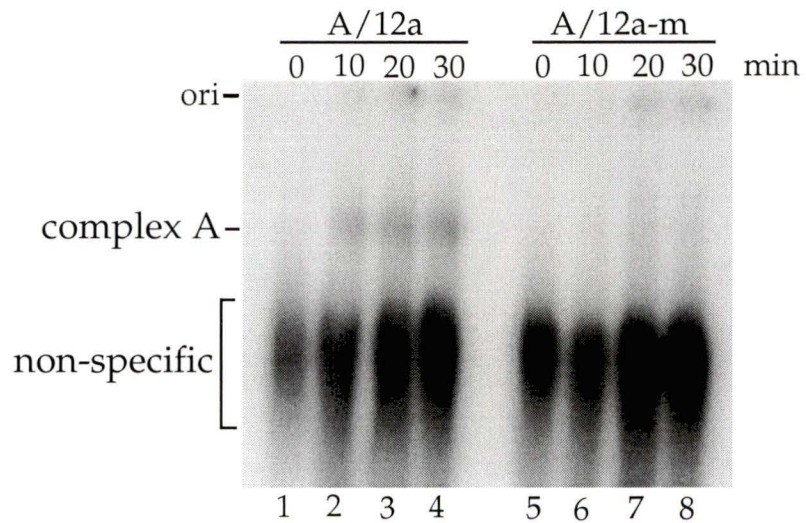
### 3.1.2 Discussion

It is estimated that 15% of all point mutations that cause genetic diseases affects pre-mRNA splicing (Krawczak et al., 1992). In NF1, this proportion has been reported to be even much higher, of the order of 50% (Ars et al., 2000). Among them one finds a 5' splice site mutation affecting the invariable GT dinucleotide in intron 16 in two unrelated patients (Maynard et al., 1992). Also, a donor splice site mutation in intron 18 was reported to result in skipping of *NF1* exon 18 (Purandare et al., 1995). Three other splice site mutations have been found in the region encompassing exons 20 to 27a which encodes the GAP-related domain (GRD). These mutations affected the conserved GT or

A.



B.



**Figure 26. The 5' splice site of exon 12a plays a role in exon definition.** (A) *In vitro* splicing assay. A <sup>32</sup>P-labeled adenovirus/NF1 chimeric pre-mRNA was incubated in a HeLa nuclear extract for the time indicated above the lanes (in hours). Both the wild-type and the mutated version carrying the G→A lesion in the 5' splice site of exon 12a were tested. Splicing products were fractionated on denaturing acrylamide gel. Only the portion of the gel that contains lariat intermediate and product molecules is shown. A control reaction containing a mixture of Ad/12a and Ad/12a-m RNAs was set up to rule out the presence of a non-specific inhibitor in the Ad/12a-m RNA preparation. The identity of band X is unknown but is unrelated to splicing because it is seen at time 0. (B) Splicing complex formation in a native gel. Splicing reactions were set up with labeled pre-mRNAs and aliquots were harvested at different times (indicated in minutes above lanes). Fractionation of splicing complexes was accomplished in a non-denaturing 5% acrylamide gel. The identity of complex A (which is formed by the binding of U2 to the 3' splice site region of the pre-mRNA, refer to Fig. 4) was confirmed by performing an assay using a HeLa extract where the U2-snRNA was specifically inactivated by oligonucleotide-targeted RNase H degradation. In this case, complex A was not formed following incubation with Ad/12a pre-mRNA (not shown). (This gel was done by Mr. Martin Simard.)

AG dinucleotide at the 3' splice site of exon 22 and exon 23.2 and the 5' splice site of exon 23.2 (Upadhyaya et al., 1997). An A→G mutation was reported at intron position +4 and was responsible for the skipping of exon 33 (Hutter et al., 1994).

Mutations that inactivate a splice junction either promote exon skipping or cryptic splice site utilization. In some cases, the mutation directly creates a novel splicing signal. We have identified a new mutation which affects the invariable GT dinucleotide of the 5' splice site of exon 12a. In contrast to known 5' splice site mutations which always promote skipping of the exon immediately preceding the lesion, this *NFI* mutation promotes skipping of both exons 11 and 12a, without changing the reading frame or reducing the amount of mutant mRNA. We find that this mutation alters the contribution that the donor splice site makes toward improving the use of the upstream 3' splice site. Thus, while the mutation as expected prevents exon 12a/exon 12b splicing, it also affects exon 11/exon 12a splicing. Because sites of exon 11 are intrinsically weak, skipping of both exon 11 and exon 12a was the ultimate consequence of the G→A point mutation. To our knowledge, this is the first report of a point mutation that causes skipping of multiple exons and interferes with the process of exon definition. A point mutation that creates a novel 5' splice site which leads to the activation of an upstream cryptic 3' splice site was described in one thalassemic patient (Treisman et al., 1983). Although this phenotype is consistent with the theory of exon definition, this interpretation has yet to be confirmed.

The neurofibromin protein is an exceptionally large protein which must probably have multiple functions. It has been reported (Fahsold et al., 2000) that missense or

single-amino-acid-deletion mutations are clustered in two distinct regions, the GAP-related domain and an upstream gene segment comprising exons 11-17. The latter region encodes a cysteine/serine-rich domain with three cysteine pairs suggestive of ATP binding, as well as three potential cAMP-dependent protein kinase (PKA) recognition sites. The functional importance of this region encoded by exons 12a, 15 and 16 is suggested by its high degree of conservation between human and *Drosophila*, and by its high occurrence of mutations. Our result with mutation 5313 which promotes skipping of exons 11 and 12a also supports the biological relevance of this region for the function of neurofibromin.

NF1 is notable for its wide range of clinical presentations that may vary from mild to severe even within members of the same family. These observations have raised the possibility that the expression of disease-causing mutations may be modulated by other genetic (Easton et al., 1993) or epigenetic factors (Riccardi, 1993). A statistical analysis of the clinical features in a group of NF1 patients concluded that the observed variability was most likely attributable to somatic mutations, modifying gene or genes that modulate NF1 gene expression (Eisenbarth et al., 2000; Easton et al., 1993). Gene dosage, methylation, mRNA editing, and alternative RNA splicing have been proposed to play potential roles in the pathogenesis of NF1, and therefore contribute to the variability of the clinical manifestations. The neurofibromin protein being an exceptionally large protein with most probably multiple functions, mutations that affect one or several of these functions may contribute to the pleiotropic effects. Current evidence suggests that RNA splicing and RNA metabolism play a critical role in the normal control of NF1

expression (Skuse and Cappione, 1997). When the splicing profile of the NF1 pre-mRNA is altered as with patient 5313, truncated forms of neurofibromin may exert dominant negative effects that may vary depending on the affected region. Because splicing decisions are controlled by a variety of cell-specific, sex-specific and possibly age-specific factors, it would not be surprising that subtle differences in NF1 RNA splicing between affected individuals may lead to large differences in the clinical presentation of the NF1 disease.

### 3.2 Identification and characterization of six novel deletions in the *NF1* gene

#### 3.2.1 Results

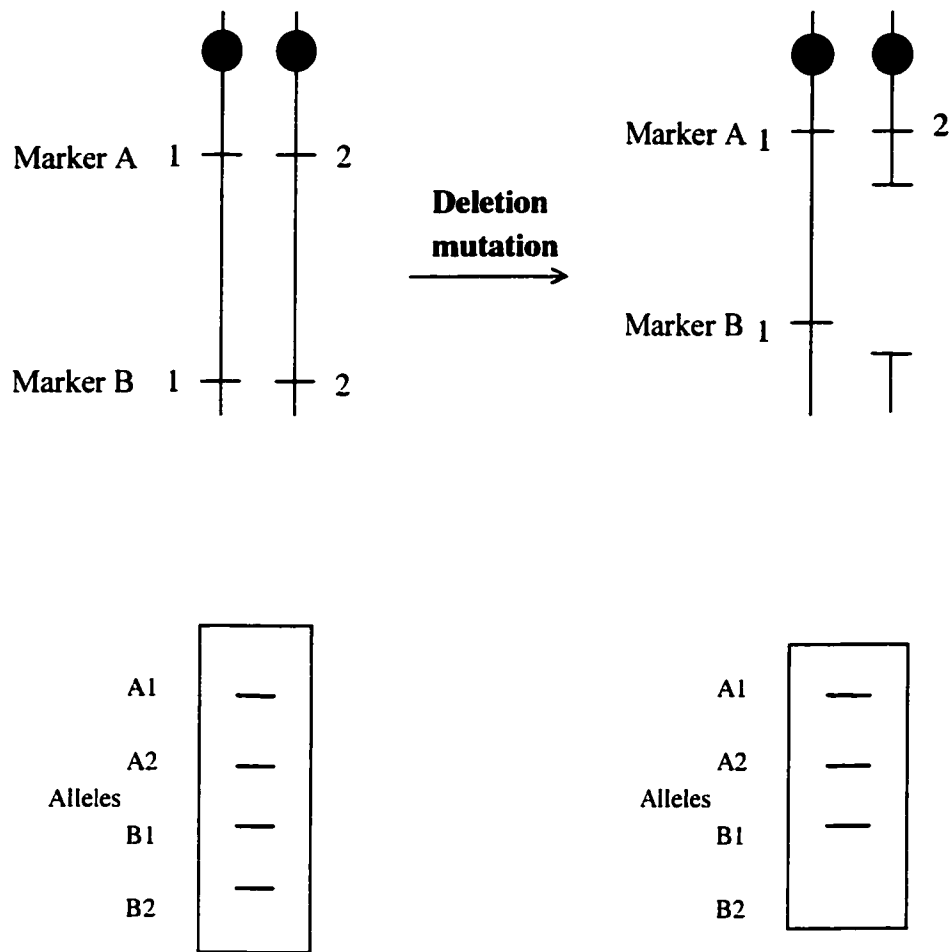
Besides the identification of mutations by the protein truncation test, deletions were searched by loss of heterozygosity analysis and Southern blotting. Six gross deletions (Table 7) have therefore been identified and further characterized in order to correlate between phenotype and genotype.

**Table 7. Characteristics of the six novel deletions**

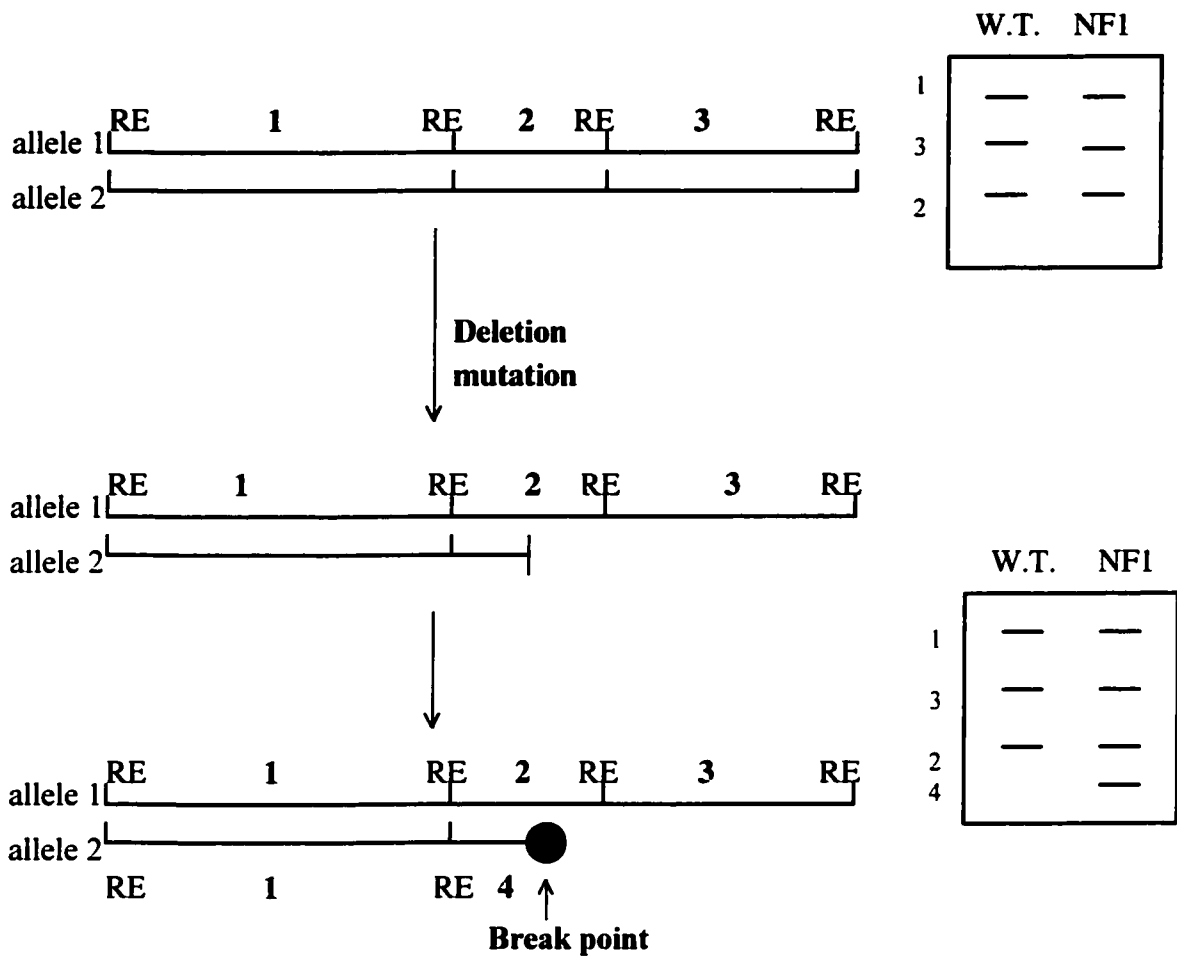
Deletions	Breakpoint location	Deleted regions
Rou 7610	Between exons 4a and 4b	Breakpoint to 3' end of the gene
Rou 7473	Between exons 5 and 26	5' end of the gene to the breakpoint
178	Between exons 23-2 and 27b	Breakpoint to 3' end of the gene
184	Between exons 27b and 29	5' end of the gene to the breakpoint
236	Between exons 14 and 18	Breakpoint to 3' end of the gene
237	Between exons 38 and 45	5' end of the gene to the breakpoint

In loss of heterozygosity analysis (Fig. 27), the failure to inherit one parental allele usually indicates a deletion which can be verified by Southern blots. With a deletion, extra bands are detected in the blots (Figure 28). Segregation analysis of haplotypes constructed with two RFLPs (*EcoRI* and *RsaI*) and four microsatellites (IVS26-2.3, IVS27AC28.4, IVS27AC33.1 and IVS38GT53.0) revealed possible





**Figure 27. Mechanism of Loss of heterozygosity (LOH) due to deletion.** Loss of heterozygosity (LOH) is the loss of alleles on one chromosome as detected by haplotyping using markers for which an individual is constitutively heterozygous. Markers A and B are both heterozygous (A1A2 and B1B2). Loss of allele A1, A2, B1 or B2 indicates a deletion. In the figure, allele B2 is lost, indicating a deletion in the region.

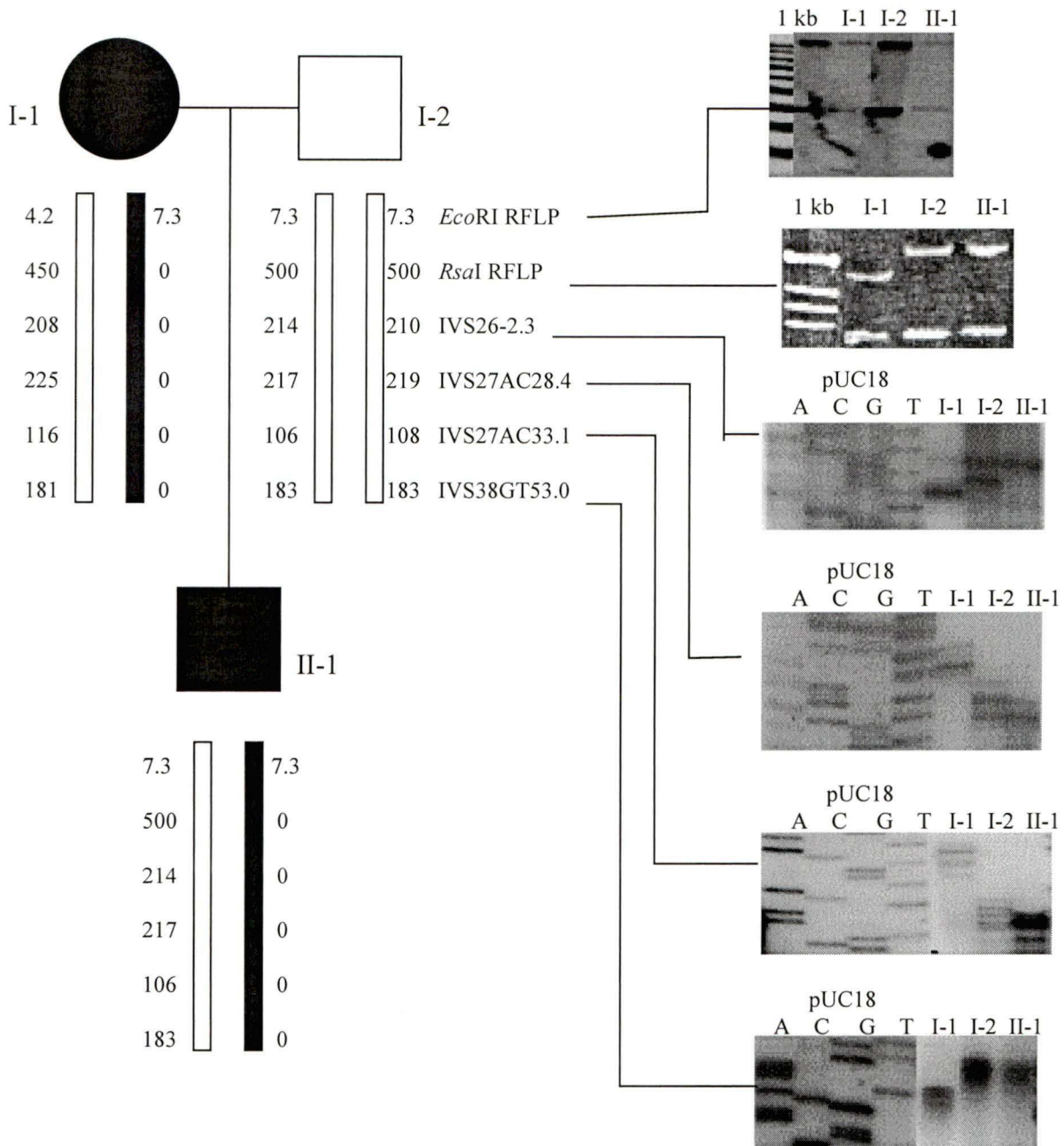


**Figure 28. Detection of the deletion breakpoint by Southern blot.** Genomic DNA digested with restriction enzyme (RE) gives 3 fragments (1, 2, 3) when there is no deletion. When there is a deletion in one allele, a new fragment (4) is created, suggesting the location of one of the breakpoints for the deletion. Then, sequences either downstream (represented in the figure) or upstream of the breakpoint are deleted in this allele.

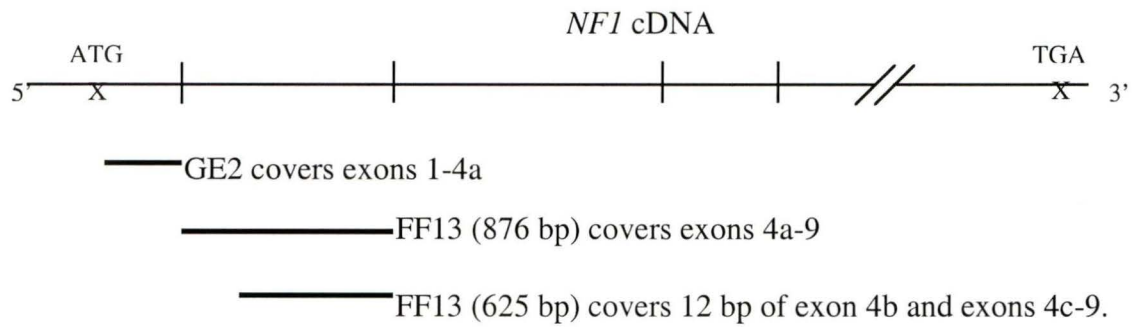
hemizyosity in affected individuals of 2 unrelated families (families 7610 and 7473) and four patients (178, 184, 236 and 237). Then, Southern hybridization followed by blot densitometric analysis which investigates the copy number of different region of the *NFI* gene was used to confirm the presumed deletions and to define the extent of the deletions.

By analysis of the haplotypes and segregation of the family 7610 (Figure 29), both the mother (I-1) and the son (II-1) were found to be hemizygous for markers at exon 5 (*RsaI* RFLP), introns 26, 27 and 38, however the *EcoRI* locus was heterozygous. When the genomic DNA was digested with *HindIII*, *BamHI*, *SacI* or *XbaI*, extra bands of about 18 kb, 10 kb, 9.5 kb, 8 kb were detected respectively with GE2-400 bp probe (Figure 30A). The same extra bands of *HindIII* (18 kb) and *BamHI* (10 kb) were also detected with FF13 (876 bp) probes (Figure 30B). Probe GE2 (400 bp) covers exons 1 to 4a including the starting codon (ATG) and overlaps with FF13 (876 bp) by 82 bp in exon 4a. Probe FF13 (876 bp) covers exons 4a-9. When 95 bp of exon 4b and exons 4c-9 were removed from FF13 (876 bp) by digestion with *HinfI* (refer to Figure 9, FF13 (625 bp)), no abnormal band was detected with the probe FF13 (625 bp). This indicates that the breakpoint of the deletion is between exons 4a and 4b. Desitometry analysis indicates that the rest of the *NFI* gene is in all probability deleted. In this family, the deletion arose from the maternal allele and all *NFI* gene except exons 1 to 4b was deleted.

With family 7473 (Figure 31), markers at introns 26, 27 and 38 of the mother (II-1) and the two daughters (III-1 and III-2) are heterozygous. Examination of the segregation of the *RsaI* locus (exon 5) indicates that the mother (II-1) is either

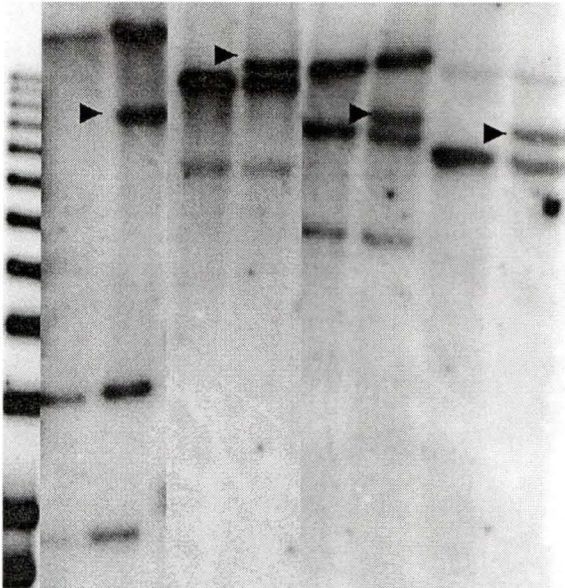


**Figure 29. Pedigree and haplotype of family 7610.** Haplotypes were determined by analysis of 6 polymorphic loci. Microsatellite fragments and *RsaI* polymorphism are indicated as bp at loci along each chromosome. *EcoRI* polymorphism is indicated as fragments in kb. The *EcoRI* polymorphism detects two-allele fragments of 7.3 kb and 4.2 kb. The deleted loci are indicated as 0. The NF1 patient and the mutation-carrying chromosome are in black. Results are shown on the right.



### A. GE2-400 bp

*Bam*HI    *Hind*III    *Sac*I    *Xba*I  
1 kb WT 7610    WT 7610    WT 7610    WT 7610



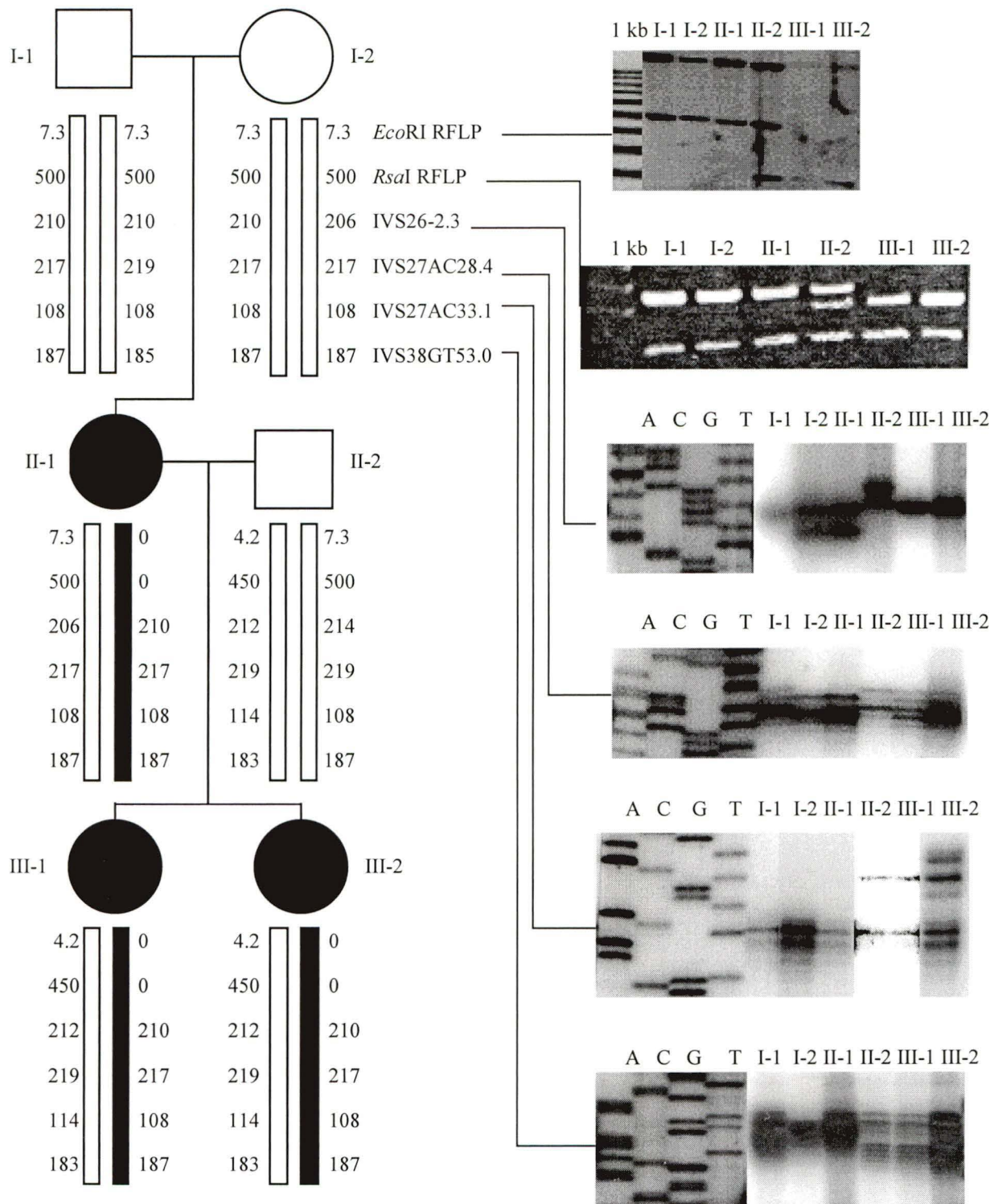
### B. FF13-876 bp

*Bam*HI    *Hinc*II    *Hind*III  
1 kb WT 7610    WT 7610    WT 7610



**Figure 30. Southern blot analysis of deletion 7610.** Genomic DNA of deletion 7610 and wild type (WT) was digested with restriction enzymes as indicated and hybridized with probes GE2 (400 bp) (A) and FF13 (876 bp) (B). The approximate positions of probes GE2 (400 bp), FF13 (876 bp) and FF13 (625 bp) are as depicted above. Vertical lines represent *Eco*RI sites. Extra bands (indicated with arrow haeds) are detected with probe GE2 (400 bp) and FF13 (876 bp), but not with probe FF13 (625 bp) (not shown), indicating that the breakpoint of deletion 7610 is in between exons 4a and 4b.

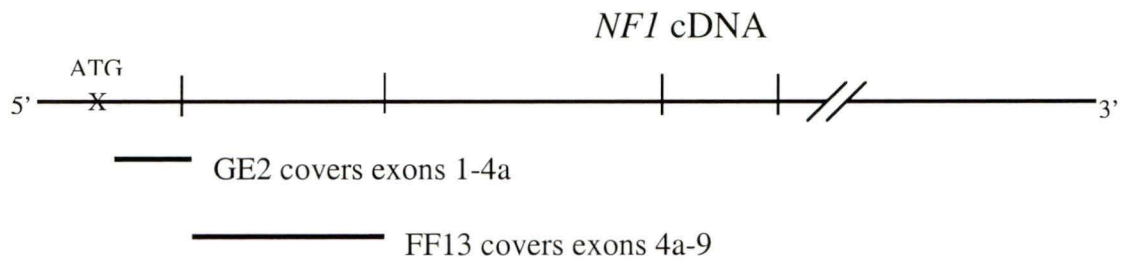




**Figure 31. Pedigree and haplotype of family 7473.** Haplotypes were determined by analysis of 6 polymorphic loci. Microsatellite fragments and *Rsa*I polymorphism are indicated as bp at each loci along each chromosome. *Eco*RI polymorphism was indicated as fragments in kb. The *Eco*RI polymorphism detects two-allele fragments of 7.3 kb and 4.2 kb . Deletion loci are indicated as 0. The NF1 patient and the mutation-carrying chromosome are in black. Results are shown on right.

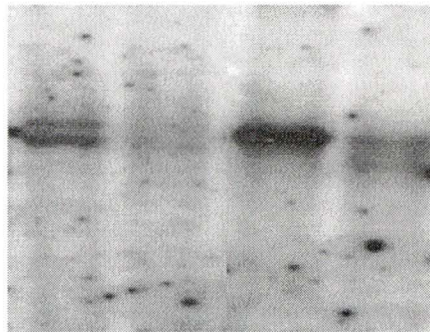
homozygous or hemizygous because of the presence of 500 bp and the absence of 450 bp fragment. Examination of the *RsaI* locus of the two daughters (III-1 and III-2) indicates that the mother is hemizygous, therefore suggesting a deletion at this locus. The deletion was confirmed by the analysis of *EcoRI* RFLP. At the *EcoRI* loci, the mother (II-1) and the two daughters (III-1 and III-2) were apparently homozygous, with the mother 7.3 kb and the two daughters 4.2 kb. Segregation analysis indicated that the 4.2 kb fragments of the daughters came from their father (II-1). This suggests that the mother and the two daughters are hemizygous. This indicates that the deletion covers at least exons 1 to 5, most probably begins upstream from the initiation codon (ATG) and ends before exon 26. When further characterization of the deletion by southern blotting was carried out, no extra band was detected when the genomic DNA of family 7473 was digested with 15 different restriction enzymes including *EcoRI*, *EcoRV*, *HindIII*, *PstI*, *BamHI*, *HincII*, *SacI*, *Sall*, *SmaI*, *XbaI*, *BglI*, *BglII*, *BclI*, *MspI*, *PvuII*. Southern blot densitometry results indicated that the region covered by GE2 (400 bp, exons 1-4a) was deleted, however two copies were detected by probe FF13 (876 bp, exons 4a-9) (Figure 32). Therefore, in family 7473, the deletion was derived from maternal chromosome and exons 1-5 were deleted. The breakpoint of the deletion is located between exons 5 and 26.

In the four non-family cases, with patient 178, when genomic DNA was digested with *EcoRI*, an extra band of 2.0 kb was detected using probe FB5D-615 bp (covering exons 23-2 to 27b) (Figure 33), indicating that the breakpoint of the deletion with this patient is in between exons 23-2 and 27b. Microsatellite markers at introns 26, 27, and 38 were hemizygous with amplified fragments of 208 bp, 214 bp, 117 bp and 180 bp



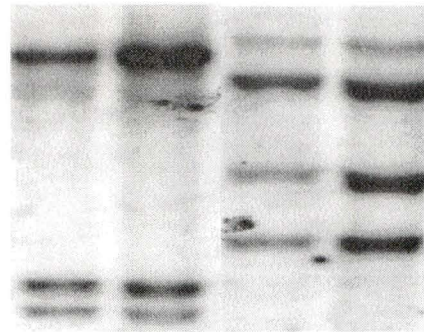
**A. GE2-400 bp**

*Bgl*III      *Eco*RV  
W.T.   7473    W.T.   7473



**B. FF13-876 bp**

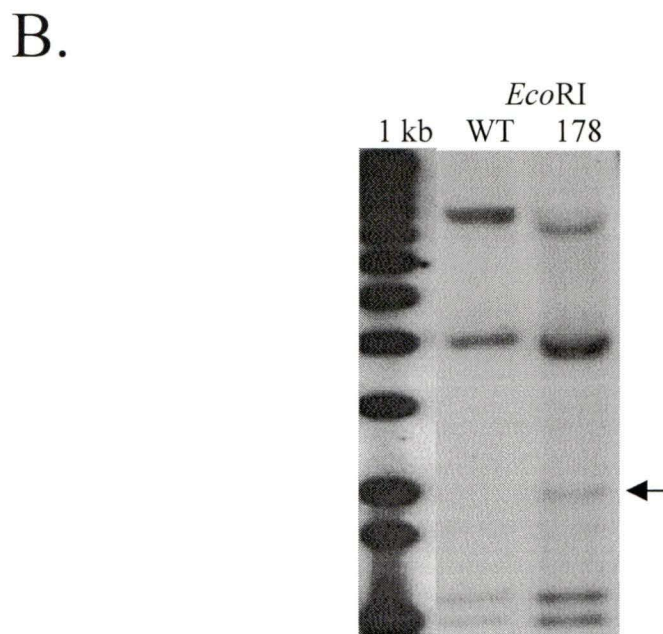
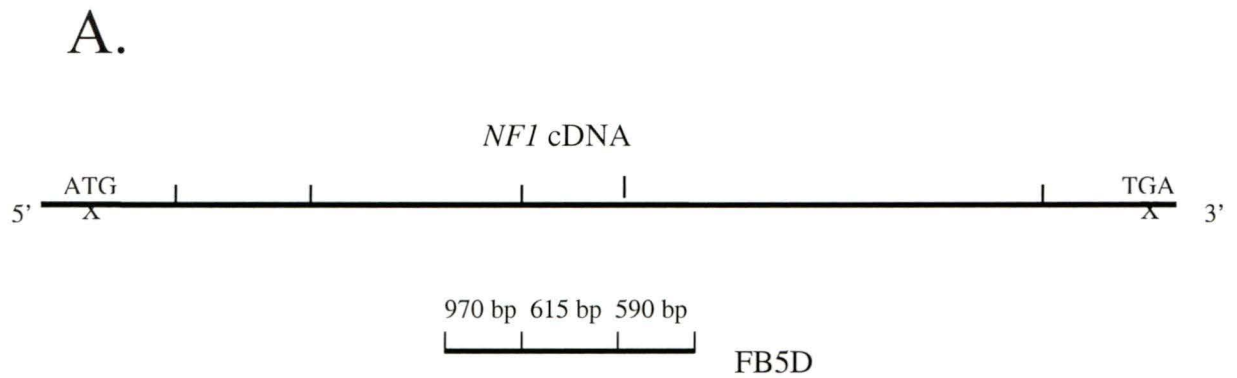
*Bgl*III      *Eco*RV  
W.T.   7473    W.T.   7473



7473/WT	7473/WT	7473/WT	7473/WT
0.32	0.22	2.0	1.6
		2.0	1.8
		1.7	2.2
		1.2	2.0

**Figure 32. Southern blot densitometry results of deletion 7473.** The same membrane was hybridized sequentially with (A) probe GE2 (400 bp) and (B) FF13 (876 bp). Approximate positions of probes related to *NF1* cDNA are depicted at the top. Southern blots are in the middle. Bottom table lists out the densitometrical ratio between deletion 7473 and wild type (WT) of each band. The wild type controls are arbitrarily assigned a value of 1.0.



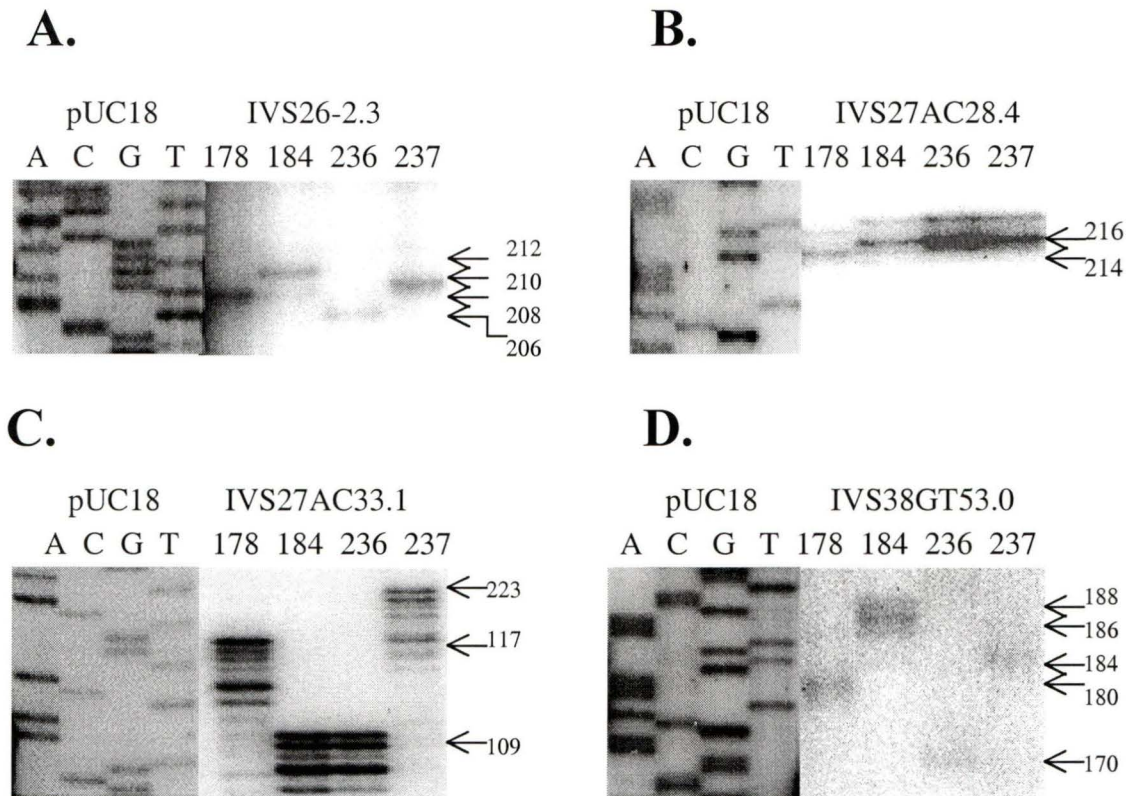


**Figure 33. Southern blot analysis of the patient 178.** (A) The approximate positions of probe FB5D are as depicted. Vertical lines represents *EcoRI* sites. (B) Genomic DNA of 178 was digested with *EcoRI* and hybridized with FB5D (615 bp). An extra band with size of 2.0 kb was detected as indicated by arrow, indicating the breakpoint of the deletion is in between exons 23-2 and 27b. The molecular weight marker 1 kb was run along side. The wild types (WT) sample was treated the same as the deletion sample.

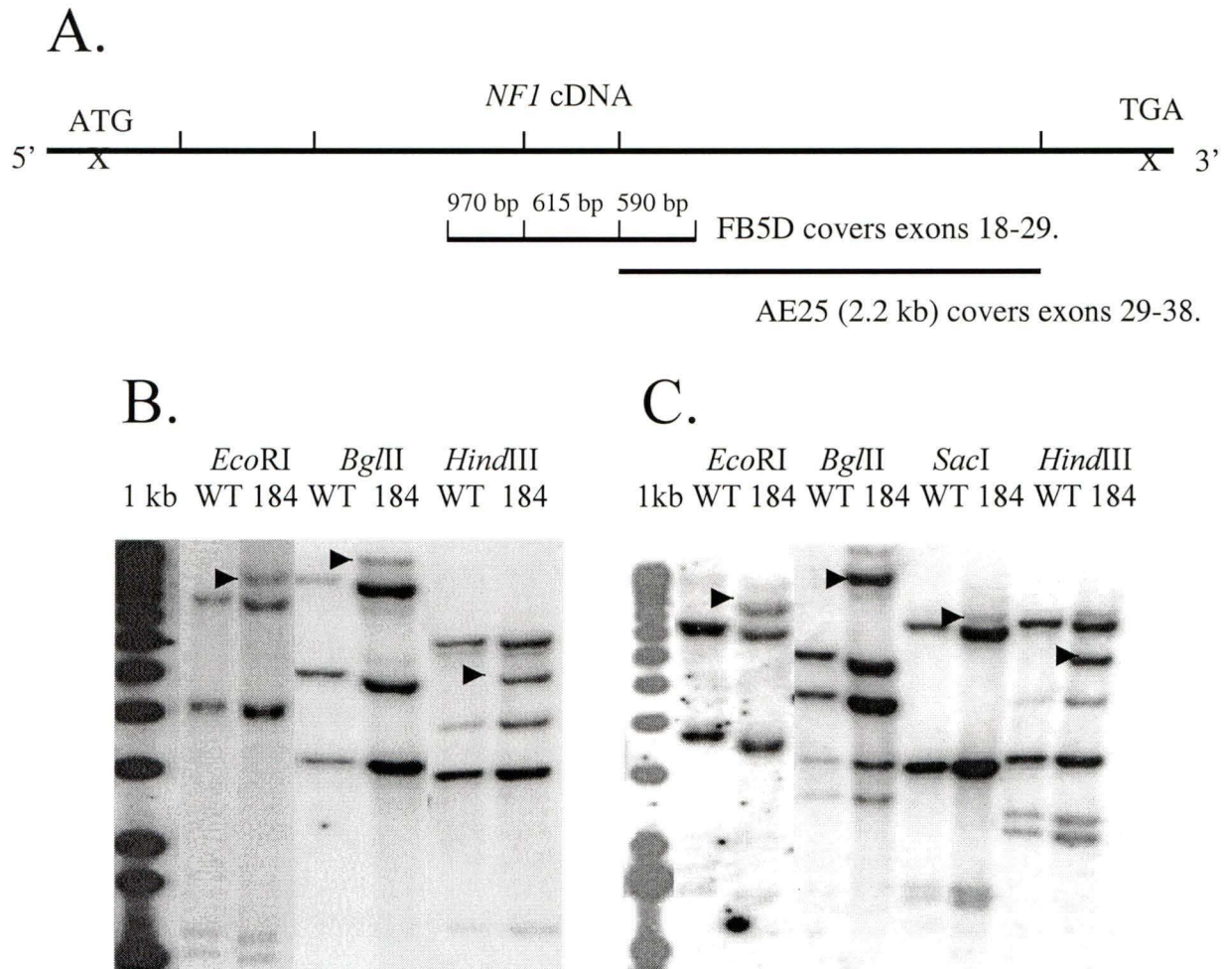
respectively (Figure 34), suggesting that these loci are deleted. Southern blot densitometry analysis confirmed that sequences downstream of the breakpoint were deleted.

With patient 184, extra bands of 9 kb, 16 kb and 6 kb were detected with probe FB5D (590 bp) when genomic DNA was digested with *EcoRI*, *BglII* and *HindIII* (Figure 35B), respectively. The same extra bands (9 kb, 16 kb and 6 kb) were detected with probe AE25 (Figure 35C). The overlapping region for these two probes is exons 27b to 29. Therefore, the breakpoint for deletion 184 was in between exons 27b and 29. Hemizygosity was detected with microsatellite markers at introns 26, 27 (Figure 34) with fragments of 212 bp, 216 bp and 109 bp at loci IVS26-2.3, IVS27AC28.4 and IVS27AC33.1 respectively. The microsatellite marker at intron 38 (IVS38GT53.0) was heterozygous with amplified fragments of 186 bp and 188 bp (Figure 34). These suggest that the region upstream of exons 27b-29 is deleted. Southern densitometry confirmed that the region upstream of exons 27b-29 was deleted.

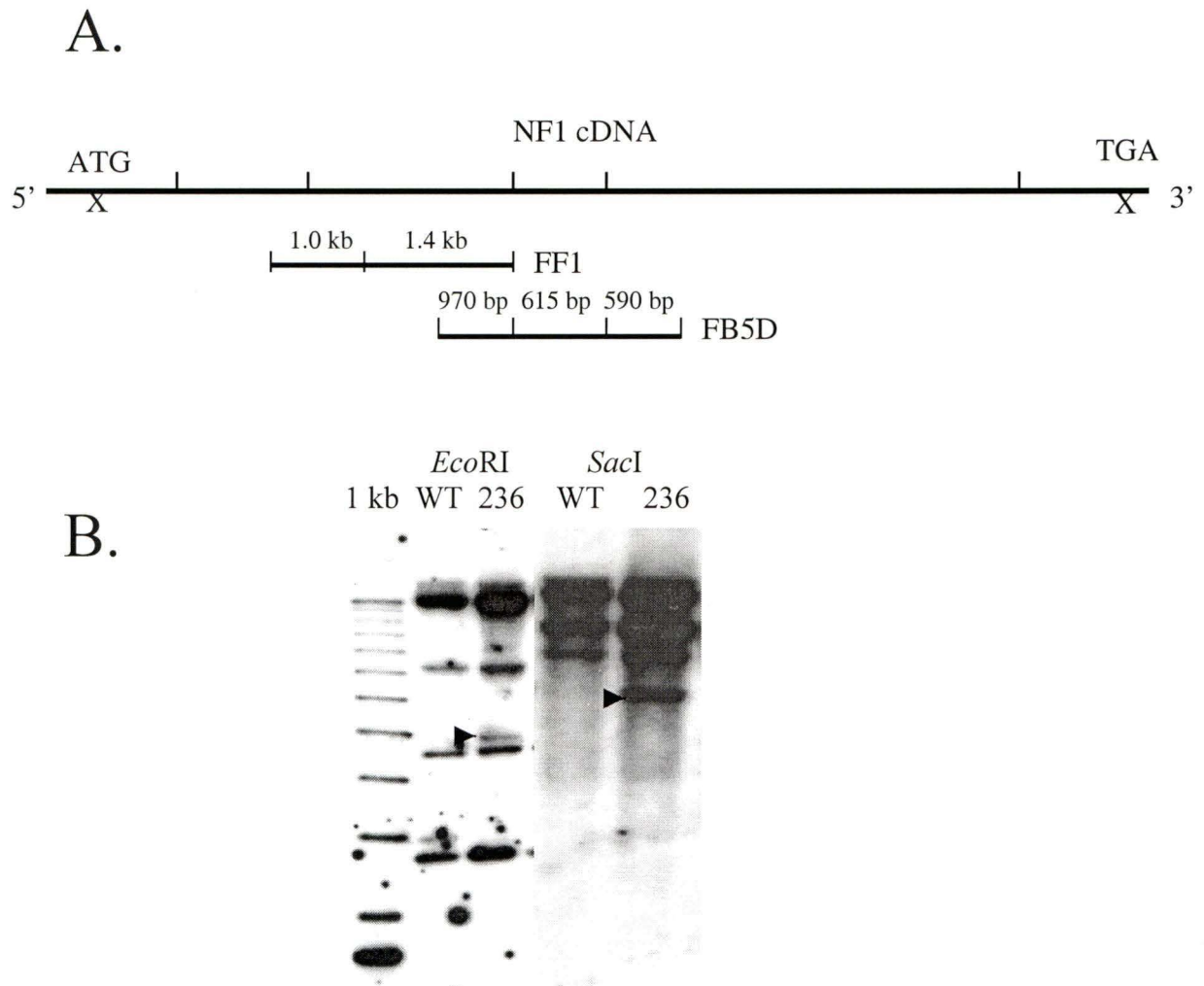
For patient 236, extra bands (5.0 kb with *EcoRI* and 6.2 kb with *SacI*) were detected with FF1 (1.4 kb) (Figure 36B), but not with FB5D (970 bp) and other probes. Probe FF1 (1.4 kb) covers exons 14-21 and probe FB5D (970 bp) covers exons 18-21. Therefore, the breakpoint of the deletion is located between exons 14 and 18. Microsatellite markers at intron 26, 27 and 38 were hemizygous with amplified fragments of 206 bp, 216 bp, 109 bp and 170 bp at loci IVS26-2.3, IVS27AC28.4, IVS27AC33.1



**Figure 34. Results of loss of heterozygosity analysis for non-family cases (178, 184, 236 and 237) at microsatellite markers IVS26-2.3 (A), IVS27AC28.4 (B), IVS27AC33.1 (C) and IVS38GT53.0 (D).** Hemizyosity was detected at all loci except IVS38GT53.0 for patient 184, where heterozygous 186 and 188 bp fragments were identified. The sequencing ladder of pUC18 was used as molecular weight marker shown on the left of each microsatellite marker. Amplified fragments were indicated as bp by arrows on the right of each microsatellite marker.



**Figure 35. Southern blot analysis of patient 184.** The approximate positions of probes FB5D and AE25 are as depicted at the top (A). Vertical lines represent *Eco*RI sites. Genomic DNA of 184 was digested with restriction enzymes as indicated and sequentially hybridized with probe FB5D (590 bp) (B) and probe AE 25 (C) using the same membrane. Extra bands were detected with both probes FB5D (590 bp) and AE 25, indicating that the breakpoint of the deletion is in between exons 27b and 29. The molecular weight marker 1 kb was run along side. The wild types (WT) sample was treated the same as the deletion sample. Extra bands are indicated by arrow aheads.



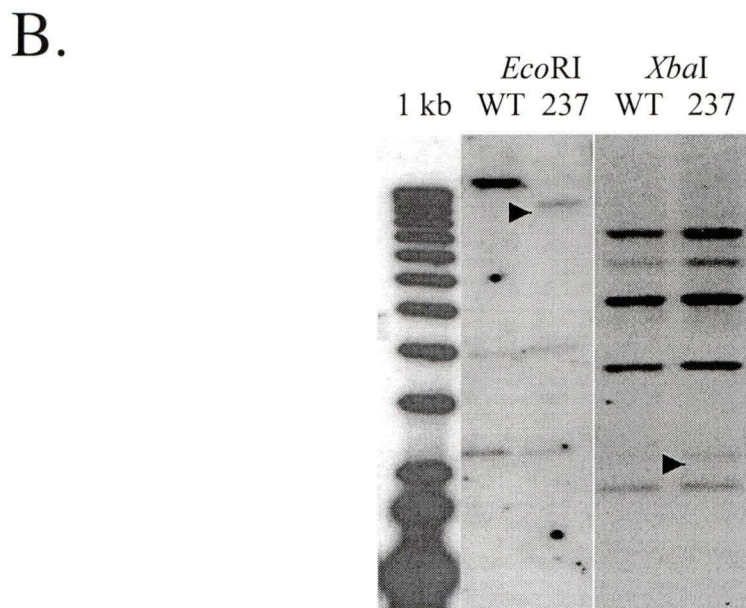
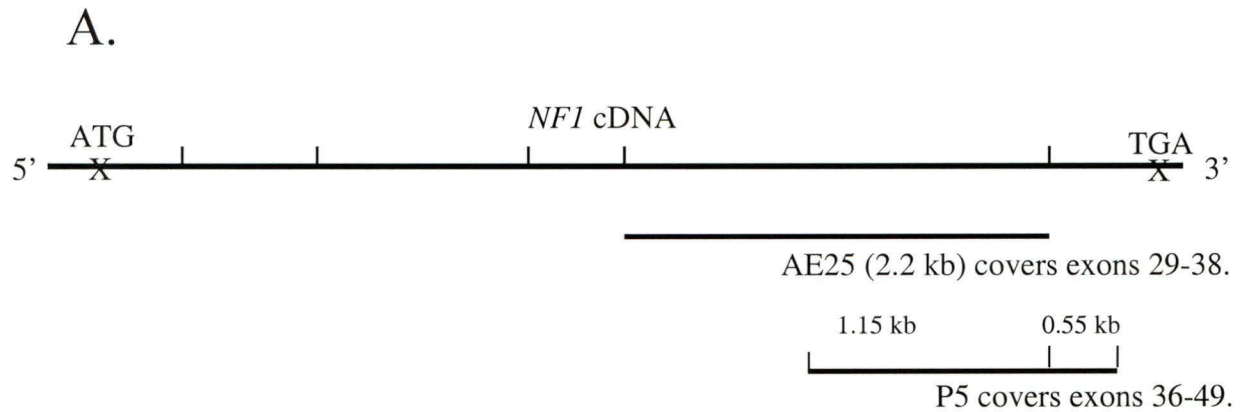
**Figure 36. Southern blot analysis of the patient 236.** (A) The approximate positions of probes are as depicted. Vertical lines represent *EcoRI* sites. (B) Genomic DNA of 236 was hybridized with probe FF1 (1.4 kb). Restriction enzymes used are as indicated. Extra bands were detected with probe FF1 (1.4 kb), but not with probe FB5D (970 bp), indicating that the breakpoint of the deletion 236 is in between exons 14 and 18. The molecular weight marker 1 kb was run along side. The wild types (WT) sample was treated the same as the deletions sample. Extra bands are indicated by arrow heads.



and IVS38GT53.0 respectively (Figure 34). This suggests that the region downstream of the breakpoint is deleted.

For patient 237, when genomic DNA was digested with *EcoRI* and probed with P5 (1.15 kb), a smaller fragment (10 kb) was detected concomitant with the absence of the wild type 15 kb fragment (Figure 37B). The missing of wild type 15 kb fragment may suggest there is a homozygote deletion in this region. When genomic DNA of patient 237 was digested with *XbaI*, an extra band with size of 2.0 kb was detected (Figure 37B). No abnormality was detected with upstream probe AE25 and other probes (data not shown). Probe AE25 covers exons 27b-38 and probe P5 (1.15 kb) covers exons 36-45. This indicated the breakpoint for this deletion was located between exons 38 and 45. Microsatellite markers at introns 26, 27, and 38 were hemizygous (Figure 34), suggesting these regions are deleted. Southern blot densitometry confirmed that upstream of exons 38-45 was deleted.

In summary (Table 7), the deletion in family 7610 arose from the maternal allele and all the *NFI* gene except exons 1 to 4b was deleted. The deletion 7473 was also derived from maternal chromosome and exons 1-5 were deleted. The breakpoint of the deletion is located in between exons 5 and 26. In the four non-family cases, the breakpoint of deletion 178 was located in between exons 23-2 and 27b and the sequences downstream of the breakpoint were deleted. For deletion 184, the breakpoint was in between exons 27b and 29, and the region upstream of the breakpoint was deleted. With deletion 236, the breakpoint was in between exons 14 and 18 and the region downstream



**Figure 37. Southern blot analysis of patient 237.** (A) The approximate positions of probes AE25 and P5 are as depicted. Vertical lines represent *EcoRI* sites. (B) Genomic DNA of 237 was hybridized with probe P5 (1.15 kb). Notice that the absence of wild type 15 kb fragment in the patient 237 with *EcoRI* digestion. An extra band was detected with *XbaI* digestion using probe P5 (1.15 kb). No abnormality was detected with probe AE25 (not shown), indicating that the breakpoint of the deletion is between exons 38 and 45. Restriction enzymes used are as indicated. The molecular weight marker 1 kb was run along side. The wild types (WT) sample was treated the same as the deletion sample. Abnormal bands are indicated by arrow heads.

of the breakpoint was deleted. The breakpoint of deletion 237 was in between exons 38 and 45 and the sequences upstream of the breakpoint were deleted. These six gross deletions were distributed randomly across the *NF1* gene and no deletion hot spot was found. The deletions are unique and different from those reported previously (Upadhyaya et al., 1994; Shen et al., 1996; Feldkamp et al., 1998; Carey and Viskochil, 1999). No founder effect has been found in these deletions.

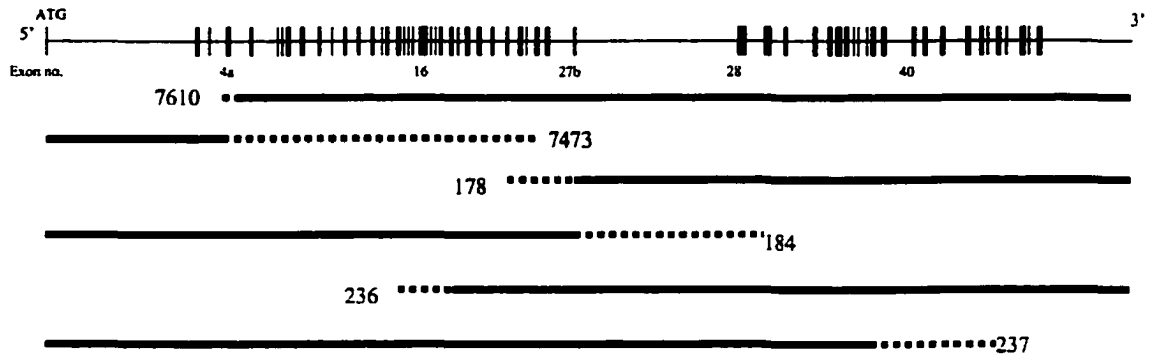
### **3.2.2 Discussion**

The approximate distribution of the six deletions on the *NF1* gene is schematically presented in Figure 39. The six novel deletions are randomly distributed across *NF1* gene and no deletion hot spot exists (Fig. 38).

Approximately 70% *NF1* mutations are either frameshift or nonsense mutations and predicted to produce truncated neurofibromins that are detectable by the recently developed *NF1* protein truncation test. However, large gene deletions would escape detection by protein truncation test as well as most other PCR based mutation identification methods. Loss of heterozygosity analysis (LOH) is an important deletion detection technique (Rasmussen et al., 1998). LOH is able to detect deletions encompassing even a single marker.

Loss of heterozygosity is the loss of alleles on one chromosome detected by assaying for markers for which an individual is constitutively heterozygous. Four





**Figure 38. Approximate extension and distribution of 6 *NF1* deletions.** Dot lines indicate the position of the breakpoints. Black lines indicate the deleted regions. The *NF1* gene is depicted at the top. The exons are indicated as boxes on the thin horizontal line (representing introns). The diagram is not quite to scale.

mechanisms have been proposed to explain the loss of alleles on one chromosome: (1) mitotic nondisjunction leading to loss of a whole chromosome, (2) reduplication to give three copies of one chromosome leading to the loss of another chromosome, (3) mitotic recombination leading to the loss of parts of the chromosome that are distal to the crossover, and (4) *de novo* interstitial deletion leading to a interstitial deletion (Strachan and Read, 1996; Rasmussen et al., 1998). However, LOH depends on the informativeness of the marker(s) and needs for parental DNA. Nevertheless, combination of loss of heterozygosity analysis, southern blotting and southern blot densitometry analysis can be used as a powerful method to detect gross deletions, especially in non-family cases. In our study, six novel deletions were detected, including 2 family cases and 4 non-family cases.

Recent study shows that 4-7% of NF1 patients have large NF1 gene deletions (Lazaro et al., 1996; Ainsworth et al., 1997; Cnossen et al., 1997; Valero et al., 1997). However, our results show that the proportion of large deletion is much higher. About 10.5% of the deletions are family cases and 20% non-family cases. This may indicate that considering family or non-family cases separately, the percentage of deletion is different from overall deletion percentage. These six deletions are unique and different from those reported previously. No founder effect has been found.

In our study, with family deletions, both large deletions are derived from maternal chromosomes. This is in agreement with previous reports that gross deletions are predominantly of maternal origin (Rasmussen et al., 1998; Upadhyaya et al., 1995;

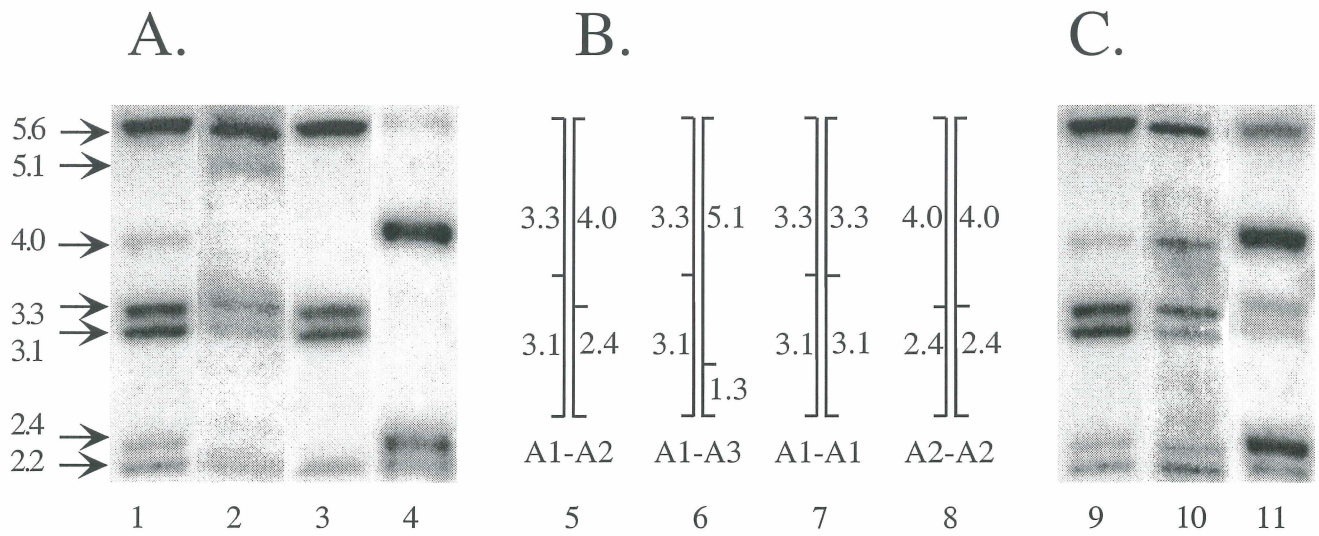
Upadhyaya et al., 1998; Lazaro et al., 1996; Ainsworth et al., 1997). This may suggest that deletions may occur more frequently during oogenesis. It has been suggested that gross deletions are commonly associated with distinct manifestations such as learning disability/mental retardation, dysmorphic features, neurofibroma, and developmental delay (Upadhyaya et al, 1998; Upadhyaya et al, 1996; Cnossen et al, 1997; Tonsgard et al, 1997; Wu et al, 1997; Leppig et al, 1997; Valero et al, 1997; Riva et al, 1996; Kayes et al, 1994; Kayes et al, 1992). However, large *NFI* deletions might not always associate with unusual clinical features (Rasmussen et al, 1998). In our study, the clinical features of deletion 7610 did not show distinct manifestations. No learning disability/mental retardation, dymorphic features and developmental delay have been diagnosed, but neurofibroma were found. Our observations are in agreement with the conclusion that gross deletions might not always associated with unusual clinical manifestations. Unfortunately, the phenotypes for the other deletions were not available.

### **3.3 A novel and very peculiar *HincII* polymorphism in the 5' region of the human Neurofibromatosis type 1 (*NF1*) gene (Appendix 1)**

#### **3.3.1 Results**

During characterization of the deletions, when genomic DNA was digested with *HincII* and analyzed by Southern blotting, a *HincII* polymorphism was detected in the 5' region of the *NF1* gene. The probe used was GE2-400 (nucleotides 2 to 401, exons 1 to 4a, of the *NF1* cDNA) (Marchuk et al., 1991) obtained by digestion with *EcoRI* and *SacII* of GE2 (550 bp) cDNA probe (Wallace et al., 1990). Allelic frequencies were determined after *HincII* digestion of the DNA from 67 unrelated persons: 14 *NF1* patients and 53 unaffected individuals from an ethnically diverse population of Quebec. About 10-15  $\mu$ g of *HincII*-digested genomic DNA was analyzed by Southern blotting with [ $\alpha$ -<sup>32</sup>P]dCTP-labeled GE2 (400 bp) probe (Sambrook et al., 1989).

Figure 39 shows the Southern blot profiles of the polymorphism. Seven fragments of 2.2 kb, 2.4 kb, 3.1 kb, 3.3 kb, 4.0 kb, 5.1 kb and 5.6 kb were detected with the GE2 (400 bp) probe. The 2.2 kb and 5.6 kb fragments were present in all individuals examined. The different alleles are A1 (3.1 and 3.3 kb), A2 (2.4 and 4.0 kb), and A3 (1.3 and 5.1 kb). The heterozygote forms of the polymorphism are either 3.1 kb and 3.3 kb on one chromosome and 2.4 kb and 4.0 kb on the other chromosome (A1-A2) (Fig. 39A, lane 1; Fig. 39B lane 5) or 3.1 kb and 3.3 kb on one chromosome and 1.3 kb and 5.1 kb on the other chromosome (A1-A3) (Figure 39A, lane 2; Figure 39B, lane 6). However, the



**Figure 39. *HincII* polymorphism pattern.** (A) Southern blots of different DNAs. Numbers on the left indicate the approximate fragment lengths in kb as determined by 1 kb molecular weight marker. (B) Schematic drawings representing the possible de visu chromosomal segregation of the polymorphism patterns corresponding to A. The polymorphism fragments are indicated in kb. The *HincII* sites are represented as short horizontal lines. (C) Southern blots of three different heterozygotes illustrating the different ratios of the intensity of the allelic band in the order of 1 (A2) to 3 (A1) (lane 9), 1 (A1) to 1 (A2) (lane 10) and 3 (A2) to 1 (A1) (lane 11).

presumed 1.3 kb was not detected. The 1.3 kb may represent a fragment of an intron, and therefore would be undetectable with a cDNA probe. The homozygotes are either A1-A1 (Figure 39A, lane 3; Figure 39B, lane 7) or A2-A2 (Figure 39A, lane 4; Figure 39B, lane 8). The presumed heterozygote A2-A3 and homozygote A3-A3 were not observed.

From the Southern blots, the A1-A2, A1-A3, A1-A1 and A2-A2 forms were identified. The heterozygosity was 0.776 (52 in 67) for A1-A2 and 0.015 (1 in 67) for A1-A3. The homozygosity was 0.149 (10 in 67) and 0.06 (4 in 67) for A1-A1 and A2-A2 respectively. Thus the allele frequencies are 0.545, 0.448 and 0.007 for A1, A2 and A3 respectively. The allele frequencies are calculated as in appendix 2. If one locus is assumed, the theoretical frequencies given by the Hardy-Weinberg equilibrium law for A1-A1, A1-A2 and A2-A2 are 0.297, 0.488 and 0.201 (For the calculation, see Appendix 2). The theoretical distribution is 19.9, 32.7 and 13.5. It differs significantly from the observed distribution of 10, 52 and 4 ( $p < 0.001$ ). These results are clearly incompatible. They could be the result of a technical error or not a single NF1 locus involved. This discrepancy can however be explained if one assumes that GE2 (400 bp) reveals not one but two similar polymorphisms. For example, the subjects with the 3.1 and 3.3 bands may be homozygotes at two loci A1-A1 and B1-B1.

This is supported by further examination of the blots and reports in the literature. Figure 39C shows the ratios of the intensities of the A1 and A2 allelic bands of the blots are of the order of 1 to 3 (lane 9), 1 to 1 (lane 10) or 3 to 1 (lane 11). This suggests

strongly the segregation of four copies instead of two of A1 and A2. Moreover it has been reported that in addition to the *NF1* gene on chromosome 17, there are at least nine *NF1* related loci on chromosomes 2, 12, 14, 15, 20, 21 and 22 (Cummings et al., 1993; Purandare et al., 1995a; Wang et al., 2000). These loci have been observed in PCR based co-amplification with the *NF1* gene, especially with exons upstream of exon 28. Our results agree with these observations. Thus probe GE2-400 most probably detects the polymorphism in the *NF1* gene and at least in one other related locus.

As defined by the GE2 (400 bp) probe (Li et al., 1995; Marchuk et al., 1991), this *HincII* polymorphism is localized in the 5' region of the *NF1* gene. It must be in an intron since there is no *HincII* site in the cDNA.

### **3.3.2 Discussion**

The restriction fragment length polymorphism described here will be a useful diagnostic and research tool in the analysis of *NF1* both in the indirect diagnosis and in the detection of deletions. This *HincII* polymorphic marker may be especially useful in researches on *NF1* related loci on chromosomes 2, 12, 14, 15, 20, 21 and 22 (Cummings et al., 1993; Purandare et al., 1995a; Wang et al., 2000), as well other potential chromosomes. However, the drawback of this *HincII* polymorphism is the cumbersome Southern blot method which requires relatively large quantity of DNA and has relatively low sensitivity. Most polymorphic markers are PCR-based, which facilitates a sensitive and rapid approach (Ainsworth and Rodenhiser, 1991). Localization of the *HincII*

polymorphism should permit to devise a PCR assay which would provide sensitive and rapid detection.



### **3.4 Linkage disequilibrium analysis between four intragenic polymorphic microsatellites of the NF1 gene in the Québec population**

#### **3.4.1 Results**

With the pedigree and haplotype data of the NF1 families, statistical analysis was performed for linkage and linkage disequilibrium in the Québec population.

Linkage disequilibrium was investigated using four intragenic microsatellites, at introns 26 (IVS26-2.3, Andersen et al., 1993), 27b (IVS27AC28.4 and IVS27AC33.1, Lazaro et al., 1994) and 38 (IVS38GT53.0, Lazaro et al., 1993). Three of them are CA repeats and the IVS26-2.3 is a TAGA/TAGG/TAAGA repeat. These polymorphisms are distributed along an approximately 85 kb genomic DNA region of the *NF1* gene (Li et al., 1995). The IVS26-2.3 polymorphism is the most proximal to the 5' end of this set of markers. It is located at intron 26 and is approximately 40 kb apart from microsatellites at intron 27b. The distance between markers IVS27AC28.4 and IVS27AC33.1 is about 5 kb, and the distance between IVS27AC28.4 and IVS38GT53.0 is 20 kb (Li et al., 1995, Valero et al., 1996).

We haplotyped four microsatellite loci in 19 NF1 families from Québec, 82 individuals and 164 chromosomes (For pedigrees and haplotypes, see Appendix 3). One family (family 7610) was found to carry a deletion in these loci in the *NF1* gene. Thus, one chromosome each from the mother (I-2) and the son (II-2) of this family did not give

information. Therefore, 162 chromosomes are informative for the four microsatellites, IVS26-2.3, IVS27AC28.4, IVS33.1 and IVS38GT53.0. All these microsatellite repeats revealed length polymorphisms with several alleles and high heterozygosity values ranging from 0.58 to 0.62 (Table 8). The allele sizes and frequencies were obtained after analyzing 79 unrelated haplotypes and are summarized in Table 8. No significant difference was observed on frequencies between NF1-bearing chromosomes and normal chromosomes (Table 8). This indicates that the markers and the disease are linked. For the haplotype analysis, many *NF1* chromosomes shared the 108 bp allele for marker IVS27AC33.1 with allele frequency near 0.4 (30/79 for overall founders, 20/52 for normal individuals and 10/27 for NF1 patients) (Table 8).

The result for LD between markers is included in Table 9. Mostly the significance of the test is consistent between the Fisher's test and the  $\chi^2$  test. Since  $\chi^2$  test is valid only in the large sample limit, a condition hardly satisfied in our sparse joint count tables, we trust the Fisher's test result more. With the exception of between markers IVS26-2.3 and IVS27AC33.1, all other marker pairs show significant LD. This observation is expected since our markers are within a single gene, and LD could be observed for markers within 1 Mb (Peterson et al. 1995).

A linkage analysis using the program LINKAGE (result file on Appendix 4) under a dominant model (penetrance values are  $P(DD)=P(D+)=0.9$ ,  $P(++)=0.003$ ) was carried out. There are strong linkage signals for all 4 markers with LOD score of 3.29,

**Table 8. Alleles sizes, frequency distributions (overall, normal and NF1), and estimated heterozygosity for each microsatellite**

Microsatellite	Allele	Size (bp)	Allele number	Overall	Normal (total=52)	NF1 (total=27)	Observed heterozygosity
<b>IVS26-2.3</b>							<b>48/82=0.585</b>
	1	204	1	0.0127	0.0192	0.0000	
	2	206	9	0.1139	0.1346	0.0741	
	3	208	7	0.0886	0.0962	0.0741	
	4	210	20	0.2532	0.2885	0.1852	
	5	212	10	0.1266	0.0962	0.1852	
	6	214	5	0.0633	0.0577	0.0741	
	7	216	4	0.0506	0.0769	0.0000	
	8	218	4	0.0506	0.0385	0.0741	
	9	222	11	0.1392	0.0962	0.2222	
	10	226	8	0.1013	0.0962	0.1111	
<b>IVS27AC28.4</b>							<b>49/82=0.598</b>
	1	215	3	0.0380	0.0577	0.0000	
	2	217	18	0.2278	0.2692	0.1481	
	3	219	24	0.3038	0.3077	0.2963	
	4	221	7	0.0886	0.0577	0.1481	
	5	223	1	0.0127	0.0192	0.0000	
	6	225	12	0.1519	0.1346	0.1852	
	7	227	12	0.1519	0.1346	0.1852	
	8	229	2	0.0253	0.0192	0.0370	
	9	231	0	0.0000	0.0000	0.0000	
<b>IVS27AC33.1</b>							<b>51/82=0.622</b>
	1	104	1	0.0127	0.0192	0.0000	
	2	106	16	0.2025	0.1923	0.2222	
	3	108	30	0.3797	0.3846	0.3704	
	4	110	5	0.0633	0.0385	0.1111	
	5	112	1	0.0127	0.0192	0.0000	
	6	114	8	0.1013	0.1346	0.0370	
	7	116	18	0.2278	0.2115	0.2593	
<b>IVS38GT53.0</b>							<b>51/82=0.622</b>
	1	171	3	0.0380	0.0577	0.0000	
	2	173	7	0.0886	0.0962	0.0741	
	3	175	3	0.0380	0.0385	0.0370	
	4	177	3	0.0380	0.0385	0.0370	
	5	181	10	0.1266	0.1154	0.1481	
	6	183	7	0.0886	0.1154	0.0370	
	7	185	24	0.3038	0.2500	0.4074	
	8	187	17	0.2125	0.2308	0.1852	
	9	189	5	0.0633	0.0577	0.0741	

**Table 9. The association between microsatellite markers**

Marker pairs	P-value from a Fisher's test	99% interval of p-value	p-value from $\chi^2$ test
IVS26-2.3/IVS27AC28.4	0***		0***
IVS26-2.3/IVS27AC33.1	0.18	(0.177-0.182)	0.17
IVS26-2.3/IVS38GT53.0	0.0002***	(0.0001-0.0003)	0.077
IVS27AC28.4/IVS27AC33.1	0.035*	(0.034-0.037)	0.030*
IVS27AC28.4/IVS38GT53.0	0***		0.0078**
IVS27AC33.1/IVS38GT53.0	0***		0.0028**

\*significant  $0.01 < p < 0.05$

\*\*highly significant  $0.001 < p < 0.01$

\*\*\*very highly significant  $p < 0.001$

2.97, 3.64 and 2.76 at recombination fraction  $\theta=0.01$ . Multipoint linkage analysis using GENEHUNTER program leads to a LOD score of 4.35. It is interesting to note that the 3-generation pedigree #11 contributes 32% of the LOD score; the pedigree #12 contributes 20% of the LOD score. Five pedigrees out of 18 combinedly contribute 84% of the LOD score, whereas eight pedigrees each contributes either nothing or less than 2% of the LOD score.

The existence of a founder haplotype is tested by the HAL program (result file on Appendix 5). The data does not provide evidence for the existence of such founder haplotype. When the founder haplotype is assumed to contain only one marker allele, the p-values are between 0.24 and 0.85, depending on the marker allele chosen to be that founder haplotype, and depending on the test statistic. If the founder haplotype is assumed to contain 2, 3, or 4 markers, the p-value ranges are 0.55 - 0.90, 0.64 - 0.82, and 0.32 - 0.82.

Using the allele counts from affected and unaffected pedigree founders, the DISLAMB program (result file on Appendix 6) in the DISEQ package determines the p-value under the null hypothesis of no association between the disease and a marker. For the chi-square test, the p-values for all four markers are between 0.42 and 0.79. For the likelihood ratio test, the p-values are all close to 0.5. In other words, there is no indication of population association between the disease and any marker in the dataset.

Because there is no evidence for association, the multi-point program DISMULT in the DISEQ package was not run.

The family-based association tests obtained by the GASSOC program (results file on Appendix 7) do not show any significant results. The p-values for a generalized TDT statistic (GTDT in GASSOC) are 0.16, 0.11, 0.18, 0.61; the p-values for TDT-like test under a dominant disease model (GDOM in GASSOC) are 0.18, 0.15, 0.25, 0.60; and those for a TDT-like test under a recessive model (GREC in GASSOC) are 0.20, 0.18, 0.29, 0.31. None of these p-values are significant. The lack of a family-based association and the presence of linkage points out that there is a lack of association between the disease and the markers. This conclusion is consistent with our population-based association study.

### **3.4.2 Discussion**

The detection of linkage disequilibrium has been used to identify candidate genes or regions and to narrow down the location of disease genes. Recently attention has focused on the use of whole-genome linkage disequilibrium studies to map common disease gene by using single nucleotide polymorphisms (SNPs) (Kruglyak, 1999).

Both linkage and linkage-disequilibrium (LD) measures a co-segregation between a marker and the disease. Linkage analysis examines evidence of co-segregation of a marker and the disease within the families, while linkage disequilibrium analysis

examines association between a marker allele and the disease in the whole population. Linkage results from recombination events in the last 2-3 generations. Linkage disequilibrium results from ancestral recombination events. Such an association is present completely when there is only one or a few disease haplotype at the time the mutation was first created (founder effect), and if the marker is close enough to the disease locus.

In our analysis, we observed a strong linkage between the four microsatellite markers and also between the 4 microsatellites markers and the NF1 disease (LOD from 2.76 to 3.64) (LINKAGE program). This is in agreement with the previous works with the AAAT *Alu* repeats in which five tetra-allelic *Alu* length polymorphisms were observed to lack independence. These five polymorphisms are strongly linked (Messiaen et al., 1993). A high degree of linkage disequilibrium was observed among seven intragenic RFLPs (Jorde et al., 1993). The linkage between the AAAT *Alu* repeats with microsatellite markers ACI27.2 and IVS38GT53.0 were also observed, but not with IVS27AC28.4 (Valero et al., 1996). This linkage was also confirmed by six polymorphism markers, including two exon markers (Purandare et al., 1996).

Gene mapping studies indicates that strong LD between disease alleles and microsatellite marker alleles are at a distance of about 1 Mb (Peterson et al., 1995). At the time of the founder mutation, all polymorphic loci spanning the region will be in complete linkage disequilibrium because there is exactly one haplotype containing the disease gene. So for every marker locus in the haplotype there is a unique allele associated

with the disease. However, with subsequent generations this disequilibrium gradually dissipates. The pace of the dissipation is a function of the recombination frequency. Markers close to the disease gene will manifest higher levels of disequilibrium than those further away. Thus linkage disequilibrium depends on two factors: recombination frequency and the number of generations since the mutation was introduced into the population. This provides the rationale for tracing down the founder. Strong linkage disequilibrium suggests a relatively recent origin for the founder mutation (Martin et al., 2000; Votruba et al., 1998; Jorde et al., 1993).

In our study, we observed the 108 bp alleles for IVS27AC33.1 with a frequency near 0.4 (0.37-0.38). However, the allele frequency of IVS27AC33.1 does not appear to be different between the affected and unaffected individuals, therefore there is no LD between this allele and the disease.

Results from population-based analysis by programs HAL and DISLAMB indicate that linkage disequilibrium between a marker and the NF1 disease is weak (p-values are all too large to reject the null hypothesis). It suggests that the existence of the founder effect is questionable. However, the results are in agreement with the previous described using eight RFLPs in the *NF1* region (Jorde et al., 1993). Virtually no linkage disequilibrium is observed between NF1 and any of the 8 RFLPs. All p values for  $\chi^2$  estimates were greater than 0.05 (Jorde et al., 1993).



The "family-based association" analysis is developed more recently (MacLean et al., 2000). However, family-based association is much more reliable than population-based, because one always have a "case" and a "control" matched. These analysis detect linkage when the linkage disequilibrium between the disease and the marker is present. We used the family-based association analysis program GASSOC in our study. No linkage disequilibrium was observed, in agreement with the results from population-based association analysis. It also suggests no founder effect exists at the *NF1* locus in the Québec population.

The failure to detect founder effect in the *NF1* gene in our Québec population collection may be due to (1) the high sporadic mutation rate (50% in literature, 21% in this study) in the *NF1* locus and (2) the absence of 100% penetrance that is the same haplotype presenting in both affected and unaffected individuals. These two factors make the founder analysis in *NF1* gene delicate. Additionally, in this study from 14 *NF1* disease haplotypes only 4-5 haplotypes clearly showed to be linked to the disease. These haplotypes are 210-221-106-183 in pedigree 11 (Appendix 3) (LOD=1.41), 222-227-106-185 in pedigree 12 (LOD=0.85), 212-219-106-185 in pedigree 19 (LOD=0.56), 208-221-116-175 in pedigree 7 (LOD=0.52) and perhaps 210-217-108-187 in pedigree 16 (LOD=0.30). This data set is too small to clearly infer the existence of a founder.

Because Québec is a young isolated population, the founder events would be necessarily much more recent (Martin et al., 2000). This should give strong linkage disequilibrium score. In addition, the microsatellite markers used in our analysis are very

close (0.085 cM) and not easy for crossover events to break the founder haplotype. However, our results do not support the existence of linkage disequilibrium in the Québec population. In most cases, the key problem in any population-based linkage disequilibrium study is the poorly characterized population. Our results suggest that we may not be able to consider the Québec population as an isolated population anymore due to the rapid and recent mixture with other ethnic immigrants.

No founder haplotype exists in the Québec population. It may suggest that the *NFI* mutations does not come from France, but instead originate from the sporadic new mutations in Québec.

## CHAPTER 4 CONCLUSION

Four parts of work have been achieved in this study. They are (1) a splice site mutation has been identified and characterized, (2) six deletions have been identified and characterized, (3) a *HincII* polymorphism has been detected at the 5' region of the *NF1* gene, and (4) linkage disequilibrium of NF1 has analyzed in the Québec population.

Neurofibromatosis type 1 is one of the most common hereditary disorders affecting the human nervous system and characterized with variable clinical presentations. Since the cloning of the *NF1* gene in 1990, a great deal has been learned about the genetics and biology of this disorder. NF1 has been elevated to a model status (Seizinger, 1993; Carey and Viskochil, 1999)

In this study, from the nineteen Québec NF1 families, one family (5313) has been found to carry a splicing mutation and two families (7610 and 7473) have deletions. From twenty non-family NF1 patients, four patients found to carry deletions. This confirms that the mutation rate is high in the *NF1* gene. The mutations are unique and different from those reported previously. As expected for a disease with so high mutation rate, no linkage disequilibrium and founder effect exist in this Québec population collection. The splicing mutation study suggests that the 11-12a region of *NF1* plays an important role in the activity of neurofibromin.

This study first time showed the splicing mechanism of a splicing mutation. The mutation exerts its effect on splicing process by preventing efficient assembly of complex A across the exon.

The two family deletions both arise from maternal alleles and are not associated with unusual clinical features. The six deletions are distributed randomly across *NF1* gene. This suggests that there is no deletion hot spot in the gene.

This study points out the importance to study mutations at genomic, RNA and protein levels. Special attention should be paid to the aberrations at RNA and protein level, which could be different to that predicted solely at the genomic level. It is reasonable to hypothesize that variations in the RNA splicing mechanism may lead to different levels of the aberrantly spliced mRNA. Splicing mutations could account for part of the clinical variability that is observed in NF1 patients carrying the same mutation.

The study proves that the analysis of NF1 mutation to be complex. But the accumulation of mutational data could reveal functional domain, genotype-phenotype correlation, and eventually tumorigenic mechanisms in the NF1 disorder.

## REFERENCES

- Abernathy C R, Rasmussen S A, Stalker H J, Zori R, Driscoll D J, Williams C A, Kousseff B G, Wallace M R (1997) *NFI* mutation analysis using a combined heteroduplex/SSCP approach. *Hum Mutat*, 9, 548-554.
- Ainsworth P J, Chakraborty P K, Weksberg R (1997) Example of somatic mosaicism in a series of de novo neurofibromatosis type 1 cases due to a maternally derived deletion. *Hum Mutat*, 9, 452-457.
- Ainsworth P J, Rodenhiser D I (1991) Rapid nonradioactive detection by PCR of pHHH202/*RsaI* RFLP linked to Neurofibromatosis type 1. *Am J Hum Genet*, 49, 1098-1099.
- Ainsworth P, Rodenhiser D, Stuart A, Jung J (1994) Characterization of an intron 31 splice junction mutation in the neurofibromatosis type 1 (*NFI*) gene. *Hum Mol Genet*, 3, 1179-1181.
- Andersen L B, Tarlé S A, Marchuk D A, Legius E, Collins F S (1993) A highly informative compound nucleotide repeat in the neurofibromatosis (*NFI*) gene. *Hum Mol Genet*, 2, 1083.
- Ars E, Serra E, Garcia J, Kruyer H, Gaona A, Lazaro C, Estivill X (2000) Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum Mol Genet*, 9, 237-247.
- Asamoah A, North K, Doran S, Wagstaff J, Ogle R, Collins F S, Korf B R (1995) 17q Inversion involving the neurofibromatosis type one locus in a family with neurofibromatosis type one. *Am J Med Genet Neuropsychiatric Genet*, 60, 312-316.

- Bahuau M, Houdayer C, Assouline B, Blanchet-Bardon C, Le Merrer M, Lyonnet S, Giraud S, Recan D, Lakhdar H, Vidaud M, Vidaud D (1998) Novel recurrent nonsense mutation causing neurofibromatosis type 1 (NF1) in a family segregating both NF1 and Noonan syndrome. *Am J Med Genet*, 75, 265-272.
- Bahuau M, Petel A, Vidaud M, Lacombe D, Vidaud D, Lyonnet S (1999) Glial cell line-derived neurotrophic factor gene (*GDNF*) as a candidate type-1 neurofibromatosis (*NF1*) modifier towards intestinal neuronal dysplasia. *Am J Hum Genet*, 65s, A4
- Ballester R, Marchuk D, Boguski M, Saulino A, Letcher R, Wigler M, Collins F S (1990) The *NF1* locus encodes a protein functionally related to mammalian GAP and yeast IRA protein. *Cell*, 63, 851-859.
- Barker D, Wright E, Nguyen K (1987) Gene for Von Recklinghausen, neurofibromatosis is in the pericentromeric region of chromosome 17. *Science*, 236, 1100-1102.
- Bassett J H D, Pannett A A J, Forbes S A, Thakker R V, McCarthy M, Read A P, The B T, Larsson C, Kytola S, Leisti J, Weber G, Giraud S, Zhang C X, Calender A, Hoppener J W M [The European Consortium on MEN1] (1997) Linkage disequilibrium studies in multiple endocrine neoplasia type 1 (*MEN1*). *Hum Genet*, 100, 657-665.
- Basu T N, Gutmann D H, Fletcher J A, Glover T W, Collins F S, Downward J (1992) Aberrant regulation of ras proteins in malignant tumour cells from type 1 neurofibromatosis patients. *Nature*, 356, 713-715
- Berget S (1995) Exon recognition in vertebrate splicing. *J Biol Chem*, 270, 2411-2414.

- Blanchette M, Chabot B (1999) Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J*, 18, 1939-1952.
- Bleyl S, Ainsworth P, Nelson L, Viskochil D, Ward K (1994) An ancient Ta subclass L1 insertion results in an intragenic polymorphism in an intron of the *NF1* gene. *Hum Mol Genet*, 3, 517-518.
- Brannan C I, Perkins A S, Vogel K S (1994) Targeted disruption of the neurofibromatosis type-1 gene leads to developmental abnormalities in heart and various neural crest-derived tissues. *Genes Dev*, 8, 1019-1029.
- Buchberg A M, Cleveland L S, Jenkins N A, Copeland N G (1990) Sequence homology shared by neurofibromatosis type 1 gene and IRA-1 and IRA-2 negative regulators of the RAS cycle AMP pathway. *Nature*, 347, 291-294.
- Cappione A J, French B L, Skuse G R (1997) A potential role for *NF1* mRNA editing in the pathogenesis of NF1 tumors. *Am J Hum Genet*, 60, 305-312.
- Carey J C, Viskochil D H (1999) Neurofibromatosis type 1: a model condition for the study of the molecular basis of variable expressivity in human disorders. *Am J Med Genet*, 89, 7-13.
- Cawthon R M, Weiss R, Xu G, Viskochil D, Culver M, Stevens J, Robertson M, Dunn D, Gesteland R, O'Connell P, White R (1990) A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure and point mutation. *Cell*, 62, 193-201.
- Chabot B (1994) Synthesis and purification of RNA substrates. In Hames, D and Higgins, S (eds), *RNA processing*. Oxford University Press, Oxford, UK. Vol. 1, pp. 1-29.

- Chabot B (1996) Directing alternative splicing: cast and scenarios. *Trends Genet*, 12, 473-478.
- Chabot B, Blanchette M, Lapierre I, LaBranche H (1997) An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol Cell Biol*, 17, 1776-1786.
- Cichowski K, Shih T S, Schmitt E, Santiago S, Reilly D, McLaughlin M E, Bronson R T, Jacks T (1999) Mouse models of tumor development in neurofibromatosis type 1. *Science*, 286, 2172-2176.
- Cnossen M H, van der Est M N, Breuning M H, van Asperen C J, Breslau-Siderius E J, van der Ploeg A T, de Goede-Bolder A, van den Ouweland A M W, Halley D J J, Niermeijer M F (1997) Deletions spanning the neurofibromatosis type 1 gene: implications for genotype-phenotype correlations in neurofibromatosis type 1? *Hum Mutat*, 9, 458-464
- Colman S D, Wallace M R (1994) Neurofibromatosis type 1. *Euro J Cancer*, 30A, 1974-1981.
- Cummings L M, Glatfeller A, Marchuk D A (1993) *NF1* related loci on chromosomes 2, 12, 14, 15, 20, 21, and 22: a potential role for gene conversion in the high spontaneous mutation rate of *NF1*. *Am J Hum Genet*, 53 (suppl. 3), A672.
- Cummings L M, Glatfelter A, Marchuk D A (1993) *NF1*-related loci on chromosomes 2, 12, 14, 15, 20, 21 and 22: a potential role for gene conversion in the high spontaneous mutation rate of *NF1*? *Am J Hum Genet*, 53, 672.



- Danglot G, Regier V, Fauvet D, Vasal G, Kujas M, Bernheim A (1995) Neurofibromatosis 1 (*NF1*) mRNAs expressed in the central nervous system are differently spliced in the 5' part of the gene. *Hum Mol Genet*, 4, 915-920.
- Daston M M, Scrabble H, Nordlund M, Sturbaum A K, Nissen L M, Ratner N (1992) The protein product of the neurofibromatosis type 1 gene is expressed at highest abundance in neurons, Schwann cells and oligodendrocyts. *Neurone*, 8, 415-419
- Dignam J D, Lebovitz R M, Roeder R G (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res*, 11, 1475-1489.
- Doerflinger N, Linder C, Ouahchi K, Gyapay G, Weissenbach J, Paslier D, Rigault P, Belal S, Hamida C B, Hentati F, Hamida M B, Pandolfo M, DiDonato S, Sokol R, Kayden H, Landrieu P, Durr A (1995) Ataxia with vitamin E deficiency: refinement of genetic localization and analysis of linkage disequilibrium by using new markers in 14 families. *Am J Hum Genet*, 56, 1116-1124.
- Dominski Z, Kole R (1991) Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol*, 11, 6075-6083.
- Dominski Z, Kole R (1992) Cooperation of pre-mRNA sequence elements in splice site selection. *Mol Cell Biol*, 12, 2108-2114.
- Easton D F, Ponder M A, Huson S M, Ponder B A J (1993) An analysis of variation in expression on neurofibromatosis type 1 (*NF1*): evidence for modifying genes. *Am J Hum Genet*, 53, 305-313.

- Eisenbarth I, Beyer K, Krone W, Assum G (2000) Toward a survey of somatic mutation of the *NF1* gene in benign neurofibromas of patients with neurofibromatosis type 1. *Am J Hum Genet*, 66, 393-401.
- Elston R C, Johnson W D (1990) *Essentials of biostatistics*. F.A. Davis Company, Philadelphia.
- Fahsold R, Hoffmeyer S, Mischung C, Gille C, Ehlers C, Kucukceylan N, Abdel-Nour M, Gewies A, Peters H, Kaufmann D, Buske A, Tinschert S, Nurnberg P (2000) Minor lesion mutational spectrum of the entire *NF1* gene does not explain its high mutability but points to a functional domain upstream of the GAP-related domain. *Am J Hum Genet*, 66, 790-818.
- Fang L J, Feingold J, Lemieux B, Thirion J P (1999) A novel and very peculiar *HincII* polymorphism in the 5' region of the human neurofibromatosis type 1 (*NF1*) gene. *Ann Génét*, 42, 231-233.
- Feingold J (1991) Le déséquilibre de liaison. *Médecine/sciences*, 7, 161-168.
- Feldkamp M M, Gutmann D and Guha A (1998) Neurofibromatosis type 1: piecing the puzzle together. *Can J Neuro Sci*, 25, 181-191.
- Friedman J M (1999) Epidemiology of Neurofibromatosis type 1. *Am J Med Genet (Semin Med Genet)*, 89, 1-6.
- Gasparini P, D'Agruma L, Pio de Cillis G, Balestrazzi P, Mingarelli R, Zelante L (1996) Scanning the first part of the neurofibromatosis type 1 gene by RNA-SSCP: identification of three novel mutations and of two new polymorphisms. *Hum Genet*, 97, 492-495.

- Gasparini P, Grifa A, Origone P, Coviello D, Antonacci R, Rocci M (1993) Detection of a neurofibromatosis type 1 (*NFI*) homologous sequence by PCR: implications for the diagnosis and screening of genetic diseases. *Mol Cell Probes*, 7, 415-418.
- Goddard K A B, Yu C, Oshima J, Miki T (1996) Toward localization of the Werner syndrome gene by linkage disequilibrium and ancestral haplotyping: lessons learned from analysis of 35 chromosome 8p11.1-21.1 markers. *Am J Hum Genet*, 58, 1286-1302.
- Greer W L, Riddell D C, Girouard G S, Sparrow S M, Tatlidil C, Dobson M J, Neumann P E (1999) Linkage disequilibrium mapping of the Nova Scotia variant of Niemann-Pick disease. *Clin Genet*, 55, 248-253.
- Hall A (1992) Signal transduction through small GTPases – a tale of two GAPs. *Cell*, 69, 389-391.
- Heim R A, Kam-Morgan L N W, Binnie C G, Corns D D, Cayouette M C, Farber R A, Aylsworth A S, Silverman L M, Luce M C (1995) Distribution of 13 truncating mutations in the neurofibromatosis 1 gene. *Hum Mol Gene*, 4, 975-981.
- Higgins J J, Litvan I, Pho L T, Li W, Nee L E (1998) Progressive supranuclear gaze palsy is in linkage disequilibrium with the tau and not the alpha-synuclein gene. *Neurology*, 50, 270-273.
- Hoffmeyer S, Assum G, Kaufmann D, Schwenk K, Krone W (1994) A deletion in the 5'-region of the neurofibromatosis type 1 (*NFI*) gene. *Hum Genet*, 94, 97-100.
- Hoffmeyer S, G Assum (1994) An *RsaI* polymorphism in the transcribed region of the neurofibromatosis (*NFI*) gene. *Hum Genet*, 93, 481-482

- Hoffmeyer S, Nurnberg P, Ritter H, Fahsold R, Leistner W, Kaufmann D, Krone W (1998) Nearby stop codons in exons of the neurofibromatosis type 1 gene are disparate splice effectors. *Am J Hum Genet*, 62, 269-277.
- Horn D, Robinson P N, Boddich A, Buske A, Tinschert S, Nurnberg P (1996) Three novel mutations of the *NF1* gene detected by temperature gradient gel electrophoresis of exons 5 and 8. *Electrophoresis*, 17, 1559-1563.
- Huopaniemi L, Rantala A, Tahvanainen E, Chapelle A, Alitalo T (1997) Linkage disequilibrium and physical mapping of x-linked juvenile retinoschisis. *Am J Hum Genet*, 60, 1139-1149.
- Huson S M, Campston D A S, Harper P S (1989) A genetic study of von Recklinhausen neurofibromatosis in southeast wales. I. Prevalence, fitness, mutation rate and effect of five patients. *Am J Hum Genet*, 54, 424-436.
- Hutter P, Antonarakis S E, Delozier-Blanchet C D, Morris M (1994) Exon skipping associated with A→G transition at +4 of the IVS33 splice donor site of the neurofibromatosis type 1 (*NF1*) gene. *Hum Mol Genet*, 3, 663-665.
- Jacks T, Shih T S, Schmitt E M (1994) Tumour predisposition in mice heterozygous for a targeted mutation in *Nf1*. *Nat Genet*, 7, 353-361.
- Jadayel D, Fain P, Upadhyaya M (1990) Paternal origin of new mutations in von Recklinghausen neurofibromatosis. *Nature*, 343, 558-559.
- Jorde L B, Watkins W S, Viskochil D, O'Connell P, Ward K (1993) Linkage disequilibrium in the neurofibromatosis 1 (*NF1*) region: implications for gene mapping. *Am J Hum Genet*, 53, 1038-1050.

- Kayes L M, Burke W, Riccardi V M, Bennett R, Ehrlich P, Rubenstein A, Stephens K (1994) Deletions spanning the neurofibromatosis 1 gene: identification and phenotype of five patients. *Am J Hum Genet*, 54, 424-436.
- Kayes L M, Riccardi V M, Burke W, Bennett R L, Stephens K (1992) Large *de novo* DNA deletion in a patient with sporadic neurofibromatosis 1, mental retardation, and dysmorphism. *J Med Genet*, 29, 686-690.
- Kendler K S, MacLean C J, Ma Y, O'Neil F A, Walsh D, Straub R E (1999) Marker-to-marker linkage disequilibrium on chromosome 5q, 6p, and 8p in Irish high-density Schizpphrenia pedigrees. *Am J Med Genet*, 88, 29-33.
- Klose A, Robinson N, Gewies A, Kluwe L, Kaufmann D, Buske A, Tinschert S, Perters H (1998) Two novel mutations in exons 19a and 20 and a *BsaI* polymorphism in a newly characterized intron of the neurofibromatosis type 1 gene. *Hum Genet*, 102, 367-371.
- Korf B R (1996) *Human Genetics: a problem-based approach*. Blackwell Sci. Boston, USA. P.102-104.
- Krawczak M, Cooper D N (1997) The human gene mutation database. *Trends Genet*, 13, 121-122.
- Krawczak, M., Reiss, J, Cooper, D. N. (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet*, 90, 41-54.
- Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet*, 22, 139-144.

- Kruglyak L, Daly M J, Reeve-Daly M P, Lander E S (1996) Parametric and non-parametric linkage analysis: a unified multipoint approach. *Am J Hum Genet*, 58, 1347-1363.
- Lavigueur A, LaBranche H, Kornbliht A R, Chabot B (1993) A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Gene Dev*, 7, 2405-2417.
- Lazaro C, Antonia G, Xu G, Weiss R and Xavier Estivill (1993) A highly informative CA/GT repeat polymorphism in intron 38 of the human neurofibromatosis type 1 (*NF1*) gene. *Hum Genet*, 92, 429-430.
- Lazaro C, Gaona A, Ainsworth P (1996) Sex differences in mutational rate and mutational mechanism in the *NF1* gene in neurofibromatosis type 1 patients. *Hum Genet*, 98, 696-699.
- Lázaro C, Gaona A, Estivill X (1994) Two CA/GT repeat polymorphisms in intron 27 of the human neurofibromatosis type 1 (*NF1*) gene. *Hum Genet*, 93, 351-351.
- Lázaro C, Gaona A, Xu G, Weiss R, Estivill X (1993) A highly informative CA/GT repeat polymorphism in intron 38 of the human neurofibromatosis type 1 (*NF1*) gene. *Hum Genet*, 92, 429-430.
- Ledbetter D M, Rich D C, O'Connell P, Leppert M, Carey J (1989) Precise localization of *NF1* to 17q11-2 by balanced translocations. *Am J Hum Genet*, 44, 20-24.
- Legius E, Marchuk D A, Collins F S, Glover T W (1993) Somatic deletion of the neurofibromatosis type 1 gene in a neurofibrosarcoma supports a tumour suppressor gene hypothesis. *Nat Genet*, 3, 122-126.

- Legius E, Marchuk D A, Hall B K (1992) NF1-related locus on chromosome 15. *Genomics*, 13, 1316-1318.
- Leppig K A, Kaplan P, Viskochil D, Weaver M, Ortenberg J, Stephens K (1997) Familial neurofibromatosis 1 microdeletions: cosegregation with distinct facial phenotype and early onset of cutaneous neurofibromata. *Am J Hum Genet*, 73, 197-204.
- Leppig K A, Viskochil D, Neil S, Rubenstein A, Johnson V P, Zhu X L, Brothman A R, Stephens K (1996) The detection of contiguous gene deletions at the neurofibromatosis 1 locus with fluorescence in situ hybridization. *Cytogenet Cell Genet*, 72, 95-98.
- Li W (1998) Comments on 'simplicity and complexity in gene evolution'. *Complexity*. 3, 10.
- Li Y, O'Connell P, Breidebach H H, Cawthon R, Stevens J, Xu G, Neil S, Robertson M, White R, Viskochil D (1995) Genomic organization of the neurofibromatosis 1 gene (*NF1*). *Genomics*, 25, 9-18.
- Luhrmann R, Kastner B, Bach M (1990) Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochem Biophys Acta*, 1087, 265-292.
- Luijten M, Wang Y, Smith B T, Westerveld A, Smink L J, Dunham I, Roe B A, Hulsebos T J (2000) Mechanism of spreading of the highly related neurofibromatosis type 1 (*NF1*) pseudogenes on chromosomes 2, 14 and 22. *Eur J Hum Genet*, 8, 209-214.
- Luijten M, Wang Y, Smith B T, Westerveld A, Smink L J, Dunham I, Roe B A, Hulsebos T J (2000) Mechanism of spreading of the highly related neurofibromatosis type 1 (*NF1*) pseudogenes on chromosomes 2, 14 and 22. *Eur J Hum Genet*, 8, 209-214.

- MacLean C J, Martin R B, Sham P C, Straub R E, Kendler K S (2000) The trimmed-haplotype test for linkage disequilibrium. *Am J Hum Genet*, 66, 1062-1075.
- Marcela V, Moore A T, Shomi S B (1998) Demonstration of a founder effect and fine mapping of dominant optic atrophy locus on 3q28-qter by linkage disequilibrium method. *Hum Genet*, 102, 79-86
- Marchuk D A, Saulino A M, Tavakkol R, Swaroop M, Wallace M R, Anderson L B, Mitchell A I, Gutmann D H, Boguski M, Collins E S. (1991) cDNA cloning of the type 1 neurofibromatosis gene: complete sequence of the *NF1* gene product. *Genomics*, 11, 931-940.
- Martin G A, Viskochill D, Bollag G, McCabe P C, Crosier W J, Haubruck H, Conroy L, Clark R, O'Connell P, Cawthon R M, Innis M A, McCormick F (1990) The GAP-related domain of the neurofibromatosis type 1 gene product interacts with ras p21. *Cell*, 63, 843-849.
- Martin R B, MacLean C J, Sham P C, Straub R E, Kendler K S (2000) Tests for linkage disequilibrium: haplotypes, multiplex pedigees, and complex traits. *Hum Hered*, appearing.
- Martinsson T, Sjoberg R-M, Hedborg F, Kogner P (1997) Homozygous deletion of the neurofibromatosis-1 gene in the tumor of a patient with neuroblastoma. *Cancer Genet Cytogenet*, 95, 183-189.
- Maynard J, Krawczak M, Upadhyaya M (1997) Characterisation and significance of nine novel mutations in exon 16 of the neurofibromatosis type 1 (*NF1*) gene. *Hum Genet*, 99, 674-676.



- Messiaen L, Callens T, De Paepe A, Craen M, Mortier G (1997) Characterisation of two different nonsense mutations, C6792A and C6792G, causing skipping of exon 37 in the *NF1* gene. *Hum Genet*, 101, 75-80.
- Messiaen L, De Bie S, Moens T, Van den Enden A, Leroy J (1993) Lack of independence between five DNA polymorphisms in the *NF1* gene. *Hum Mol Genet*, 2, 485.
- Mitchison H M, O'Rawe A M, Lerner T J, Taschner P E M et al. (1995) Refined localization of the Batten disease gene (*CLN3*) by haplotype and linkage disequilibrium mapping to D16S288-D16S388 and exclusion from this region of a variant form of Batten disease with granular osmiophilic deposits. *Am J Med Genet*, 57, 312-315.
- Modafferi E F, Black D L (1997) A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon. *Mol Cell Biol*, 17, 6537-6545.
- Moore M J, Query C C, Sharp P A (1993) Splicing of precursors to mRNA by the spliceosome. In *The RNA world*. Cold Spring Harbor Laboratory Press.
- Natacci F, Colapietro P, Riva P, Corrado L, Rossi L N, Maninetti M M, Casciati M C, Zambrino C A, Lanzi A, Larizza L (1999) Distribution and high frequency of novel alleles at *NF1* polymorphic markers in the Italian population. *Mol Cell Probes*, 13, 415-420.
- Park V M, Kenwright K A, Sturtevant D B, Pivnick E K (1998) Alternative splicing of exons 29 and 30 in the neurofibromatosis type 1 gene. *Hum Genet*, 103, 382-385.
- Park V M, Pivnick E K (1998) Neurofibromatosis type 1 (*NF1*): a protein truncation assay yielding identification of mutations in 73% of patients. *J Med Genet*, 35, 813-620.

- Perrin G, Morris M A, Antonarakis S E, Boltshauser E, Hutter P (1996) Two novel mutations affecting mRNA splicing of the neurofibromatosis type 1 (*NF1*) gene. *Hum Mutat*, 7, 172-175.
- Peterson A C, Rienzo A D, Lehesjoki A, Chapelle A, Slatkin M, Freimer N B (1995) The distribution of linkage disequilibrium over anonymous genome regions. *Hum Mol Genet*, 4, 887-894.
- Purandare S M, Cawthon R, Nelson L M, Sawada S, Watkins W S, Ward K, Jorde L B, Viskochil D H (1996) Genotyping of PCR-based polymorphisms and linkage-disequilibrium analysis at the *NF1* locus. *Am J Hum Genet*, 59, 159-166.
- Purandare S M, Huntsman B H, Li Y, Zhu X L, Sawada S, Brothman A, White R, Cawthon R, Viskochil D (1995a) Identification of neurofibromatosis type 1 (*NF1*) gene homologous loci by direct sequencing *in situ* hybridization and PCR amplification of somatic cell hybrids. *Genomics*, 30, 476-485.
- Purandare S M, Lanyon W G, Arngrimsson R, Connor J M (1995b) Characterisation of a novel splice donor mutation affecting position +1 in intron 18 of the *NF-1* gene. *Hum Mol Genet* 4, 767-768.
- Purandare S M, Lanyon W G, Connor J M (1994) Characterisation of inherited and sporadic mutations in neurofibromatosis type-1. *Hum Mol Genet*, 3, 1109-1115.
- Rasmussen S A, Colman S D, Ho V T, Abernathy C R, Arn P H, Weiss L, Schwartz C, Saul R A, Wallace M R (1998) Constitutional and mosaic large *NF1* gene deletions in neurofibromatosis type 1. *J Med Genet*, 35, 468-471.
- Reyniers E, de Boule K, Marchuk D A, Andersen L B, Collins F S, Willems P J (1993) An *EcoRI* RFLP in the 5' region of the human *NF1* gene. *Hum Genet*, 92, 631.

- Riccardi V M (1992) Neurofibromatosis: Phenotype, Natural history and pathogenesis. Johns Hopkins University Press, Baltimore.
- Riccardi V M (1993) Genotype, malleotype, phenotype, and randomness: lessons from neurofibromatosis-1 (NF1). *Am J Hum Genet*, 53, 301-304.
- Riva P, Castorina P, Manoukian S, Dalpra L, Doneda L, Marini G, den Dunnen J, Larizza L (1996) Characterization of a cytogenetic 17q11.2 deletion in an NF1 patient with a contiguous gene syndrome. *Hum Genet*, 98, 646-650.
- Robinson P N, Boddreich A, Peters H, Tinschert S, Buske A, Kaufmann D, Numberg P (1995) Two recurrent nonsense mutations and a 4 bp deletion in a quasi-symmetric element in exon 37 of the *NF1* gene. *Hum Genet*, 96, 95-98.
- Rodenhiser D, Howland K (1995) A noval *RsaI* polymorphism within intron 39 of the neurofibromatosis type 1 (NF1) gene. *Hum Genet*, 95, 241-242.
- Rubenstein A E, Korf B R (eds) (1990) Neurofibromatosis. A handbook for patients, families, and health-care professionals. Thieme Medical Publishers, Inc., New York.
- Sambrook J, Fritsch E F, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor. (2nd ed.), 9.17-9.19.
- Schaid D J (1996) General score tests for association of genetic markers with disease using cases and their parents. *Genet Epidemiol*, 13, 423-449.
- Scheurlen W G, Senf L (1995) Analysis of the GAP-related domain of the neurofibromatosis type 1 (*NF1*) gene in childhood brain tumors. *Int J Cancer*, 64, 234-238.

- Schleutker J, Laine A-P, Haataja L, Reulund M, Weissenbach J, Aula P, Peltonen L (1995) Linkage disequilibrium utilized to establish a refined genetic position of the Salla disease locus on 6q14-q15. *Genomics*, 27, 286-292.
- Seizinger B R (1993) NF1: a prevalent cause of tumorigenesis in human cancers? *Nat Genet*, 3, 97-99.
- Seizinger B R, Rouleau G A, Ozelius L J (1987) Genetic linkage of Von Recklinghausen neurofibromatosis to the nerve growth factor receptor gene. *Cell*, 49, 589-594.
- Serra E, Puig S, Otero D, Gaona A, Kruyer H, Ars E, Estivill X, Lazaro C (1997) Confirmation of a double-hit model for the NF1 gene in benign neurofibromas. *Am J Hum Genet*, 61, 512-519.
- Sham P (1998) *Statistics in Human Genetics*. Arnold, London.
- Shen M H, Harper P S, Upadhyaya M (1996) Molecular genetics of neurofibromatosis type 1 (NF1). *J Med Genet*, 33, 2-17.
- Side L, Taylor B, Cayouette M, Conner E, Thompson P, Luce M, Shannon K (1997) Homozygous inactivation of the NF1 gene in bone marrow cells from children with neurofibromatosis type 1 and malignant myeloid disorders. *New England J Med*, 336, 1713-1720.
- Skuse G R, Cappione A J (1997) RNA processing and clinical variability in neurofibromatosis type 1 (*NF1*). *Hum Mol Genet*, 6, 1707-1712.
- Skuse G R, Cappione A J, Sowden M, Metheny L J, Smith H C (1996) The neurofibromatosis type 1 messenger RNA undergoes base-modification RNA editing. *Nucleic Acid Res*, 24, 478-486.

- Spielman R S, McGinnis R E, Ewens W J (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet*, 52, 506-516.
- Stephens K, Kayes L, Riccardi V M, Rising M, Sybert V P, Pagon R A (1992) Preferential mutation of the neurofibromatosis type 1 gene in paternally derived chromosomes. *Hum Genet*, 88, 279-282.
- Strachan T, Read A P (1996) *Human Molecular Genetics*, BIOS Scientific Publishers Ltd.
- Strachan T, Read A P (1996) *Human molecular genetics*. BIOS Scientific Publishers Ltd, Oxford. pp. 466-471
- Stumpf S, Alksne J F, Annegers J F (1988) Neurofibromatosis Conference statement. National Institutes of Health Consensus Development Conference. *Arch Neurol*, 45, 575-578.
- Suzuki H, Takahashi K, Kubota Y, Shibahara S (1992) Molecular cloning of a cDNA coding for neurofibromatosis type 1 protein isoform lacking the domain related to ras GTPase-activating protein. *Biochem Biophys Res Commun*, 187, 984-990.
- Suzuki Y, Suzuki H, Kayama T, Yoshimoto T, Shibahara S (1991) Brain tumors predominantly express the neurofibromatosis type 1 gene transcripts containing the 63 base insert in the region coding for GTPase activating protein-related domain. *Biochem Biophys Res Commun*, 181, 955-961.
- Takahashi K, Suzuki H, Kayama T, Suzuki Y, Yoshimoto T, Sasano H, Shibahara S (1994) Multiple transcripts of the neurofibromatosis type 1 gene in human brain and in the brain tumors. *Clin Sci*, 87, 481-485.

- Terwilliger J D (1995) A powerful likelihood method for the analysis of linkage disequilibrium between trait loci one or more polymorphic marker loci. *Am J Hum Genet*, 56, 777-787.
- Terwilliger J D, Ott J (1994) *Handbook of Human Genetic Linkage* (The John Hopkins University Press).
- Tongard J H, Yelavarthi K K, Cushner S, Short M P, Lindgren V (1997) Do *NF1* gene deletions result in a characteristic phenotype? *Am J Hum Genet*, 73, 80-86.
- Treisman R, Orkin S H, Maniatis T (1983) Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. *Nature*, 302, 591-596.
- Upadhaya M, Shaw D J, Harper P S (1994) Molecular basis of neurofibromatosis type 1 (NF1): mutation analysis and polymorphisms in the *NF1* gene. *Hum Mutat*, 4, 83-101.
- Upadhyaya M, Maynard J, Osborn M, Huson S M, Ponder M, Ponder B A J, Harper P S (1995) Characterisation of germline mutations in the neurofibromatosis type 1 (*NF1*) gene. *J Med Genet*, 32, 706-710
- Upadhyaya M, Osborn M J, Maynard J, Kim M R, Tamanoi F, Cooper D N (1997) Mutational and functional analysis of the neurofibromatosis type 1 (*NF1*) gene. *Hum Genet*, 99, 88-92
- Upadhyaya M, Robert S H, Maynard J, Sorour E, Thompson P W, Vaughan M, Wilkie A O M, Hughes H E (1996) A cytogenetic deletion, del(17)(q11.22q21.1) in a patient with sporadic neurofibromatosis type 1 (NF1) associated with dysmorphism and developmental delay. *J Med Genet*, 33, 148-152.

- Upadhyaya M, Ruggieri M, Maynard J, Osborn M, Hartog C, Mudd S, Penttinen M, Cordeiro I, Ponder M, Ponder B A, Krawczak M, Cooper D N (1998) Gross deletions of the neurofibromatosis type 1 (*NF1*) gene are predominantly of maternal origin and commonly associated with a learning disability, dysmorphic features and developmental delay. *Hum Genet*, 102, 591-597.
- Valero M C, Pascual-Castroviejo I, Velasco E, Moreno F, Hernandez-Chico C (1997) Identification of *de novo* deletions at the *NF1* gene: no preferential paternal origin and phenotypic analysis of patients. *Hum Genet*, 99, 720-726.
- Valero M C, Velasco E, Valero A, Moreno F, Hernandez-Chico C (1996) Linkage disequilibrium between four intragenic polymorphic microsatellites of the *NF1* gene and its implications for genetic counselling. *J Med Genet*, 33, 590-593
- Viskochil D (1999) Neurofibromatosis 1. *Am J. med Genet* 89: v-viii.
- Viskochil D, Buchberg A M, Xu G, Cawthon R M, Stevens J, Wolff R K, Culver M, Carey J C, Copeland N G, Jenkins N A, White R, O'Connell P (1990) Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell*, 62, 187-192.
- Vogel K S, Klesse L J, Velasco-Miguel S, Meyers K, Rushing E J, Parada L F (1999) Mouse tumor model for neurofibromatosis type 1. *Science*, 286, 2176-2179.
- Vortuba M, Moore A T, Bhattacharya S S (1998) Demonstration of a founder effect and fine mapping of dominant optic atrophy locus on 3q28qter by linkage disequilibrium method. *Hum Genet*, 102, 79-86.
- Wallace M R, Marchuk D A, Anderson L B, Letcher R, Odeh H M, Saulino A M, Fountain J W, Brereton A, Nicholson J, Mitchell A L, Browstein B H, Collin E S

- (1990) Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science*, 249, 181-186.
- Ward K, O'Connell P, Carey C, Leppert M, Jolley S, Plaetke R, Ogden B, White R (1990) Diagnosis of neurofibromatosis 1 by using tightly linked flanking DNA markers. *Am J Hum Genet*, 46, 943-949.
- Wei N, Lin C Q, Modafferi E F, Gomes W A, Black D L (1997) A unique intronic splicing enhancer controls the inclusion of the agrin Y exon. *RNA*, 3, 1275-1288.
- Weiss B, Bollag G, Shannon K (1999) Hyperactive Ras as a therapeutic target in Neurofibromatosis type 1. *Am J Med Genet (Semin Med Genet)*, 89, 14-22.
- White M B, Carvalho M, Derse D, O'Brien S J, Dean M (1992) Detecting single base substitutions as heteroduplex polymorphisms. *Genomics*, 12, 301-306.
- Wu B L, Austin M A, Schneider G H, Boles R G, Korf B R (1995) Deletion of the entire NF1 gene detected by the FISH: four deletion patients associated with severe manifestations. *Am J Med Genet*, 59, 528-535.
- Wu B L, Schneider G H, Korf B R (1997) Deletion of the entire NF1 gene causing distinct manifestations in a family. *Am J Hum Genet*, 69, 98-101.
- Xu G F, Nelson L, O'Connell P, White R (1991) An *AluI* polymorphism intragenic to the neurofibromatosis type 1 (*NF1*) gene. *Nucleic Acid Res*, 19, 3764-3765.
- Xu G, Lin B, Tanaka K, Dunn D, Wood D, Gesteland R, White R, Weiss R, Tamanoi F. (1990) The catalytic domain of the neurofibromatosis type 1 gene product stimulates ras GTPase and complements ira mutants of *S. cerevisiae*. *Cell*, 63, 835-841.
- Xu G, O'Connell P, Viskochil D (1990) The neurofibromatosis type 1 gene encodes a protein related to GAP. *Cell*, 62, 599-608



Xu G, O'Connell P, Viskochill D, Cawthon R, Robertson M, Culver M, Dunn D, Stevens J, Gesteland R, White R, Weiss R. (1990) The neurofibromatosis type 1 gene encodes a protein related to GAP. *Cell*, 62, 599-608.

Xu W, Liu L, Ponder M, Ponder B A(1991) A *TaqI* polymorphism in the human NF1 gene. *Nucleic Acid Res*, 19, 4570.

Zubay G (1987) *Genetics*. Benjamin/Cummings Pub. California, USA.

## **Acknowledgments**

I would like to express my sincere gratitude to my supervisor, Dr. Jean-Paul Thirion, for useful advises, constant encouragement and guidance throughout the period of this study. I gratefully acknowledge his suggestions on this dissertation.

I would like to thank the members of my advisory committee, Dr. Benoit Chabot, Dr. Joseph Weber, Dr. Majambu M'Bikay and Dr. Claude Déry for their valuable advices throughout the period of this study. I greatly appreciate the kindness of Dr. Dominique Vidaud, Dr. Michelle Vidaud, and Dr. Josué Feingold.

I would like to give my special thanks to Mr. Nader Chalhoub for his kind help when I started this project. I would like to thank Dr. Bernard Lemieux and Robert Giguère for the blood samples of the NF1 families; Dr. G. Rouleau and Ms Louise LeBuis for the lymphoblastoid cell lines. I would like to thank Dr. Wentian Li and Mr. Martin J. Simard for their nice co-operation. I am grateful to my colleagues, and all past and present members of Department of Microbiology for their cooperation and help.

## **Appendix 1**

**A novel and very peculiar *Hind*III polymorphism in the 5' region of the human Neurofibromatosis type 1 (*NF1*) gene, *Annales de génétique*, 42, 231-233, 1999**

BRIEF REPORT  
NOTE BRÈVE

A NOVEL AND VERY PECULIAR *HincII*  
POLYMORPHISM IN THE 5' REGION OF THE HUMAN  
NEUROFIBROMATOSIS TYPE 1 (NF1) GENE

L.J. FANG<sup>1</sup>, J. FEINGOLD<sup>2</sup>,  
B. LEMIEUX<sup>3</sup>, J.P. THIRION<sup>1</sup>

FANG L.J., FEINGOLD J., LEMIEUX B., THIRION J.P. – A novel and very peculiar *HincII* polymorphism in the 5' region of the human neurofibromatosis type 1 (NF1) gene.

Ann Génét, 1999, 42, n° 4, 231-233.

**SUMMARY:** We report a *HincII* polymorphism in the 5' end of the neurofibromatosis type 1 gene (NF1) as detected with a probe made of exons 1 to 4a (nucleotides 2 to 401 of the cDNA). This *HincII* site is most probably in an intron. Evidence presented suggests the probe reveals not one but two similar polymorphisms.

**KEY-WORDS:** Neurofibromatosis type 1. – Polymorphism. – NF1. – Human genome – *HincII* restriction site.

Neurofibromatosis type 1 (NF1), an autosomal dominant disorder, affects about 1 in 3 500 individuals and is characterised by multiple neurofibromas, café-au-lait spots, axillary freckling and Lisch iris nodules [1, 2]. It results from the mutation of the NF1 gene located at 17q11.2, spanning 350 kb with 60 exons [3-7]. Southern blotting to screen for deletions indicated a *HincII* polymorphism in the 5' region of the gene. The probe used was GE2-400 (nucleotides 2 to 401, exons 1 to 4a, of the NF1 cDNA) [5] obtained by digestion with *EcoRI* and

FANG L.J., FEINGOLD J., LEMIEUX B., THIRION J.P. – Un nouveau polymorphisme très particulier dans la région 5' du gène humain de la neurofibromatose de type 1 (NF1). (En Anglais).

Ann Génét, 1999, 42, n° 4, 231-233.

**RÉSUMÉ:** Un polymorphisme *HincII*, situé en position 5' du gène NF1 de la neurofibromatose de type 1 a été mis en évidence. Il a été détecté grâce à une sonde faite des exons 1 à 4a (nucléotides 2 à 401 du cDNA). Ce site est fort probablement dans un intron. Nos résultats suggèrent aussi que la sonde révèle non pas un mais deux polymorphismes similaires.

**MOTS-CLÉS:** Neurofibromatose de type 1. – Polymorphisme. – NF1. – Génome humain – Site de restriction *HincII*.

*SacII* of GE2-550bp cDNA probe [8]. Allelic frequencies were determined after *HincII* digestion of the DNA from 67 unrelated persons: 14 NF1 patients and 53 unaffected individuals from an ethnically diverse population of Québec. Genomic DNA was extracted either from mononuclear cells isolated from whole blood by ficoll-hypaque gradient centrifugation or from lymphoblastoid cell lines. About 10-15 µg of genomic DNA was analyzed by Southern blotting [9]. Briefly, *HincII*-digested genomic DNA was separated by gel electropho-

Correspondance: Dr J.-P. Thirion (address opposite).

Submitted for publication September 9, 1999.  
Accepted October 27, 1999.

1. Département de Microbiologie, Faculté de Médecine, Université de Sherbrooke, SHERBROOKE, Québec, Canada J1H 5N4.

2. INSERM-155, Université Paris 7, 2, place Jussieu, 75251 PARIS (France).

3. Département de Pédiatrie, Faculté de Médecine, Université de Sherbrooke, SHERBROOKE, Québec, Canada J1H 5N4.

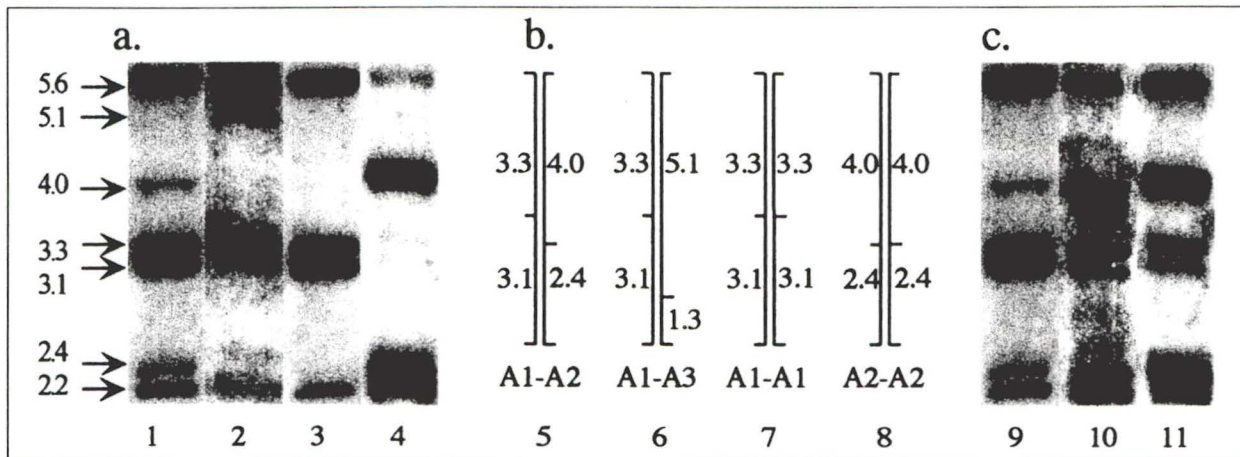


Fig. 1. - *HincII* polymorphism pattern.

a) Southern blots of different DNAs. Numbers on the left indicate the approximate fragment lengths in kb as determined by 1 kb molecular weight marker.

b) Schematic drawings representing the possible *de visu* chromosomal segregation of the polymorphism patterns corresponding to Fig. 1a. The polymorphism fragments are indicated in kb. The *HincII* sites are represented as short horizontal lines.

c) Southern blots of three different heterozygotes illustrating the different ratios of the intensity of the allelic band in the order of 1 (A2) to 3 (A1) (lane 9), 1 (A1) to 1 (A2) (lane 10) and 3 (A2) to 1 (A1) (lane 11).

resis and alkali transferred to a Hybon<sup>TM</sup>-N<sup>+</sup> nylon membrane (Amersham). Membranes were prehybridized for at least 6 hours at 65°C in a solution composed of 5 × SSC, 5 × Denhardt's solution, 0.5% SDS and 20 µg/ml of denatured salmon sperm DNA. Hybridization was carried out overnight at 65°C in the same solution, with denatured [ $\alpha$ -<sup>32</sup>P] dCTP-labeled GE2-400 probe. Membranes were washed first with 2 × SSC at room temperature for 20 mins, then with a solution of 0.1 × SSC and 0.1% SDS at 65°C for 30 mins. Autoradiography was carried out at -70°C using double intensifying screens.

Figure 1 shows the Southern blot profiles of the polymorphism. Seven fragments of 2.2 kb, 2.4 kb, 3.1 kb, 3.3 kb, 4.0 kb, 5.1 kb and 5.6 kb were detected with the GE2-400 probe. The 2.2 kb and 5.6 kb fragments were present in all individuals examined. The different alleles are A1 (3.1 and 3.3 kb), A2 (2.4 and 4.0 kb), and A3 (1.3 and 5.1 kb). The heterozygote forms of the polymorphism are either 3.1 kb and 3.3 kb on one chromosome and 2.4 kb and 4.0 kb on the other chromosome (A1-A2) (fig. 1a, lane 1; fig. 1b lane 5) or 3.1 kb and 3.3 kb on one chromosome and 1.3 kb and 5.1 kb on the other chromosome (A1-A3) (fig. 1a, lane 2; fig. 1b, lane 6). However, the presumed 1.3 kb was not detected. The 1.3 kb may represent a fragment of an intron, and therefore would be undetectable with a cDNA probe. The homozygote forms are either A1-A1 (fig. 1a, lane 3; fig. 1b, lane 7) or A2-A2 (fig. 1a, lane 4; fig. 1b lane 8). The presumed

heterozygote form A2-A3 and homozygote form A3-A3 were not observed.

From the Southern blots, A1-A2, A1-A3, A1-A1 and 4 A2-A2 were identified. The heterozygosity was 0.776 (52 in 67) for A1-A2 and 0.015 (1 in 67) for A1-A3. The homozygosity was 0.149 (10 in 67) and 0.06 (4 in 67) for A1-A1 and A2-A2 respectively. Thus the allele frequencies are 0.545, 0.448 and 0.007 for A1, A2 and A3 respectively. If one locus is assumed, the theoretical frequencies given by the Hardy-Weinberg equilibrium law for A1-A1, A1-A2 and A2-A2 are 0.297, 0.488 and 0.201. The theoretical distribution is 19.9, 32.7 and 13.5. It differs significantly from the observed distribution of 10, 52 and 4 ( $p < 0.001$ ). These results are clearly incompatible. They can however be explained if one assumes that GE2-400 reveals not one but at least two similar polymorphisms. For example, the subjects with the 3.1 and 3.3 bands may be homozygotes at two loci A1-A1 and B1-B1.

This is supported by reports in the literature and further examination of the blots. Figure 1c shows the ratio of the intensities of the A1 and A2 allelic bands of the blots are of the order of 1 to 3 (lane 9), 1 to 1 (lane 10) or 3 to 1 (lane 11). This suggests strongly the segregation of four copies instead of two of A1 and A2. Indeed it has been reported that in addition to the *NFI* gene on chromosome 17, there are at least nine *NFI* related loci on chromosomes 2, 12, 14, 15, 20, 21 and 22 [10, 11]. These loci have been observed in PCR based co-amplification with the *NFI* gene, especially with exons

upstream of exon 28. Our results agree with these observations. Thus probe GE2-400 most probably detects the polymorphism in *NF1* gene and at least in one other related locus.

As defined by the GE2-400 probe [4, 5]. This *HincII* polymorphism is localized in the 5' region of the *NF1* gene. It must be in an intron since there is no *HincII* site in the cDNA.

#### ACKNOWLEDGMENTS

We thank Dr M.R. Wallace for the cDNA probe and Drs D. Vidaud and M. Vidaud for discussions.

#### REFERENCES

1. FELDKAMP M.M., GUTMANN D., GUHA A. - Neurofibromatosis type 1: piecing the puzzle together. *Can. J. Neuro. Sci.*, 1998, 25, 181-191.
2. UPADHAYA M., SHAW D.J., HARPER P.S. - Molecular basis of neurofibromatosis type 1 (*NF1*): mutation analysis and polymorphisms in the *NF1* gene. *Hum. Mut.*, 1994, 4, 83-101.
3. CAWTHON R.M., WEISS R., XU G., VISKOCHIL D., CULVER M., STEVENS J., ROBERTSON M., DUNN D., GESTELAND R., O'CONNELL P., WHITE R. - A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure and point mutation. *Cell*, 1990, 62, 193-201.
4. LI Y., O'CONNELL P., BREIDEBACH H.H., CAWTHON R., STEVENS J., XU G., NEIL S., ROBERTSON M., WHITE R., VISKOCHIL D. - Genomic organization of the neurofibromatosis 1 gene (*NF1*). *Genomics*, 1995, 25, 9-18.
5. MARCHUK D.A., SAULINO A.M., TAVAKKOL R., SWAROOP M., WALLACE M.R., ANDERSON L.B., MITCHELL A.L., GUTMANN D.H., COLLINS F.S. - cDNA cloning of the type 1 neurofibromatosis gene: complete sequence of the *NF1* gene product. *Genomics*, 1991, 11, 931-940.
6. VALERO M.C., VELASCO E., VALERO A., MORENO F., HERNANDEZ-CHICO C. - Linkage disequilibrium between four intragenic polymorphic microsatellites of the *NF1* gene and its implications for genetic counselling. *J. Med. Genet.*, 1996, 33, 590-593.
7. WARD K., O'CONNELL P., CAREY C., LEPPERT M., JOLLEY S., PLAETKE R., OGDEN B., WHITE R. - Diagnosis of neurofibromatosis 1 by using tightly linked flanking DNA markers. *Am. J. Hum. Genet.*, 1990, 46, 943-949.
8. WALLACE M.R., MARCHUK D.A., ANDERSON L.B., LETCHER R., ODEH H.M., FOUTAIN J.W., BREETON A., NICHOLSON J., MITCHELL A.L., BROWNSTEIN B.H., COLLINS F.S. - Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three *NF1* patients. *Science*, 1990, 249, 181-186.
9. SAMBROOK J., FRITSCH E.F., MANIATIS T. - Molecular cloning: a laboratory manual (2nd ed.). *Cold Spring Harbour Laboratory Press*, 1989, 2, 9.31-9.62.
10. CUMMINGS L.M., GLATEFELLER A., MARCHUK D.A. - *NF1*-related loci on chromosomes 2, 12, 14, 15, 20, 21 and 22: A potential role for gene conversion in the high spontaneous mutation rate of *NF1*. *Am. J. Hum. Genet.*, 1993, 53 (suppl.3), A672.
11. PURANDARE S.D., HUNTSMAN B.H., LI Y., ZHU X.L., SAWADA S., BROTHMAN A., WHITE R., CAWTHON R., VISKOCHIL D. - Identification of Neurofibromatosis type 1 (*NF1*) gene homologous loci by direct sequencing in situ hybridization and PCR amplification of somatic cell hybrids. *Genomics*, 1995, 30, 476-485.



## Appendix 2

### Calculation of allele frequency for *HincII* polymorphism

In 67 individuals, 10 A1-A1, 52 A1-A2, 1 A1-A3 and 4 A2-A2 were observed. Therefore the frequency of

$$A1 \quad p = (10 + 52/2 + 1/2)/67 = 0.545$$

$$A2 \quad q = (4 + 52/2)/67 = 0.448$$

$$A3 = (0 + 1/2)/67 = 0.0007$$

Theoretical frequencies according to Hardy-Weinberg should be

$$p^2 = 0.545 \times 0.545 = 0.297$$

$$q^2 = 0.448 \times 0.448 = 0.20$$

$$2pq = 2 \times 0.545 \times 0.448 = 0.488$$

The theoretical distribution is

$$A1-A1 = 67 \times p^2 = 19.9 (\neq 10)$$

$$A2-A2 = 67 \times q^2 = 13.5 (\neq 4)$$








$$A1-A2 = 67 \times 2pq = 32.7 (\neq 52)$$

Therefore, it differs significantly from the observed distribution.

### Appendix 3

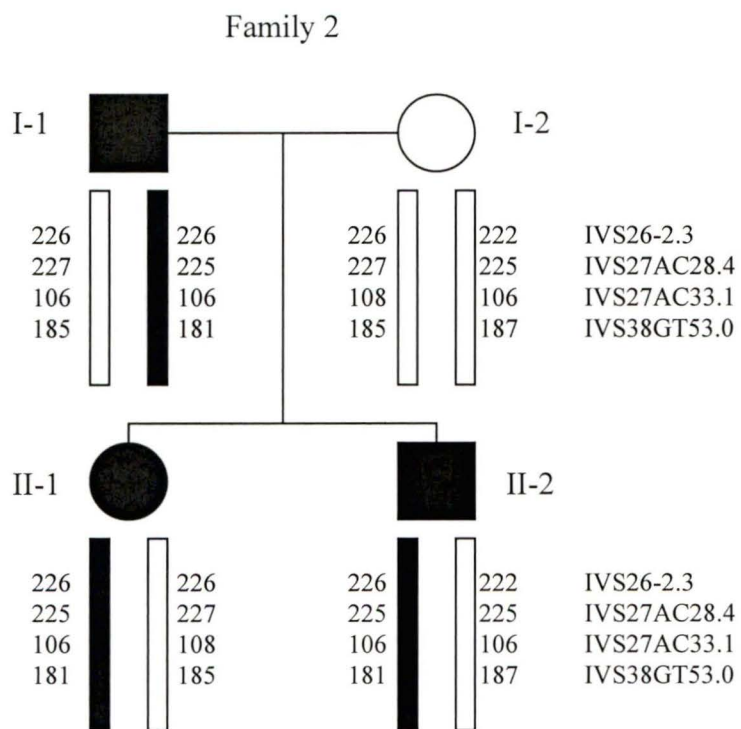
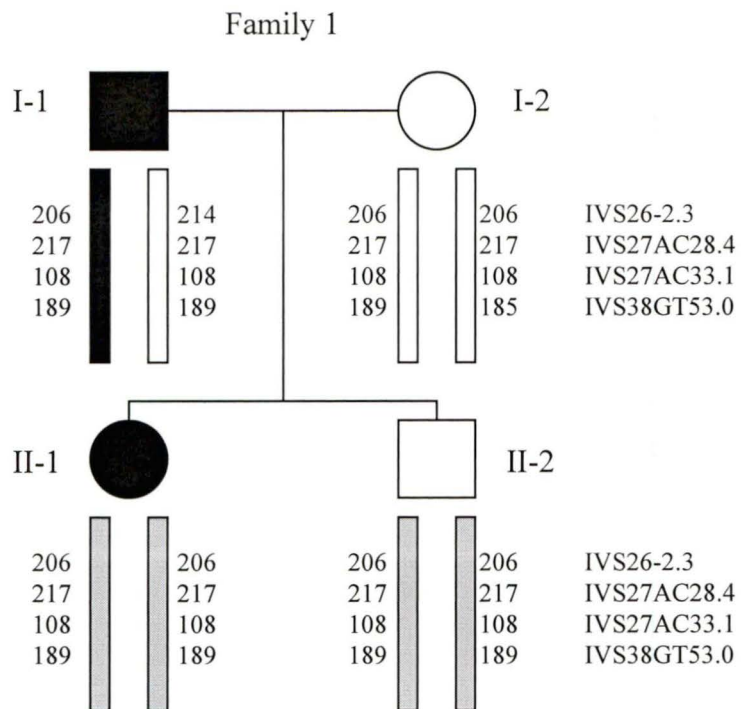
#### Pedigrees and haplotypes of the 19 NF1 families from Québec

Notes:

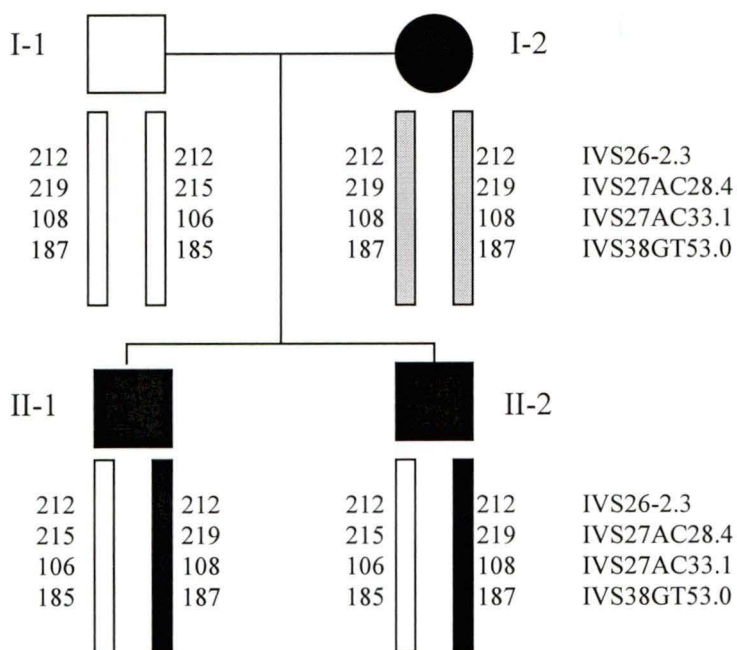
	Normal male individual, wild type
	Normal female individual, wild type
	Male NF1 patient
	Female NF1 patient
	Normal chromosome, non-NF1 chromosome
	Affected chromosome, NF1-bearing chromosome
	Affection status unknown, unable to decide the affecting status

\*The alleles are indicated by bp along each locus. The size (bp) of the allele was determined by pUC18 sequencing ladder.

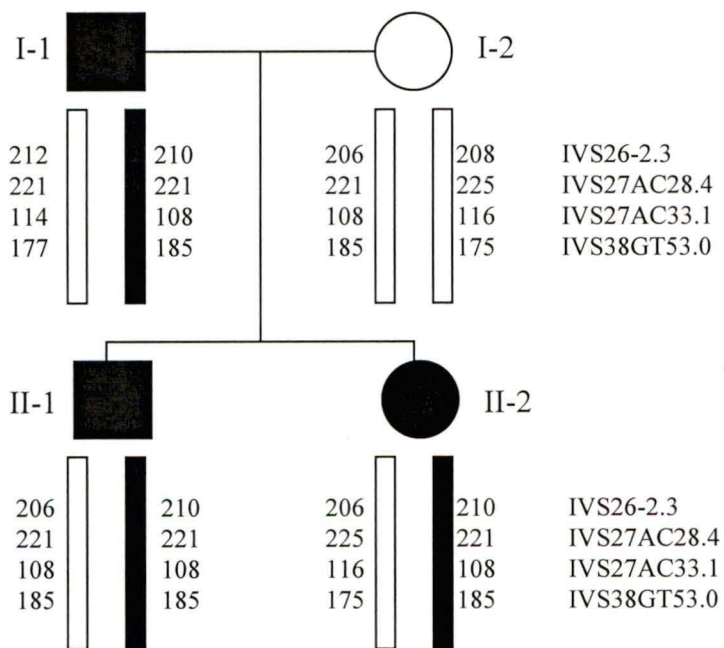


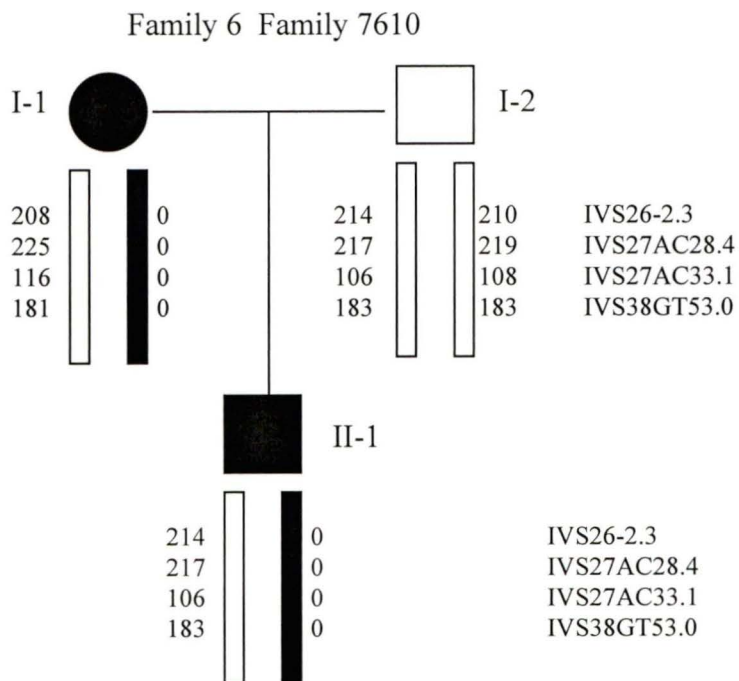
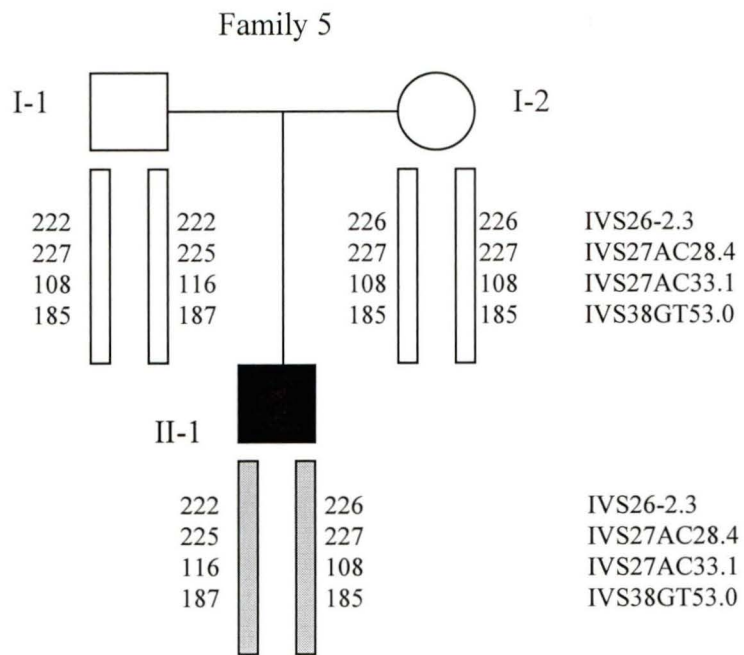


Family 3 Family 5313

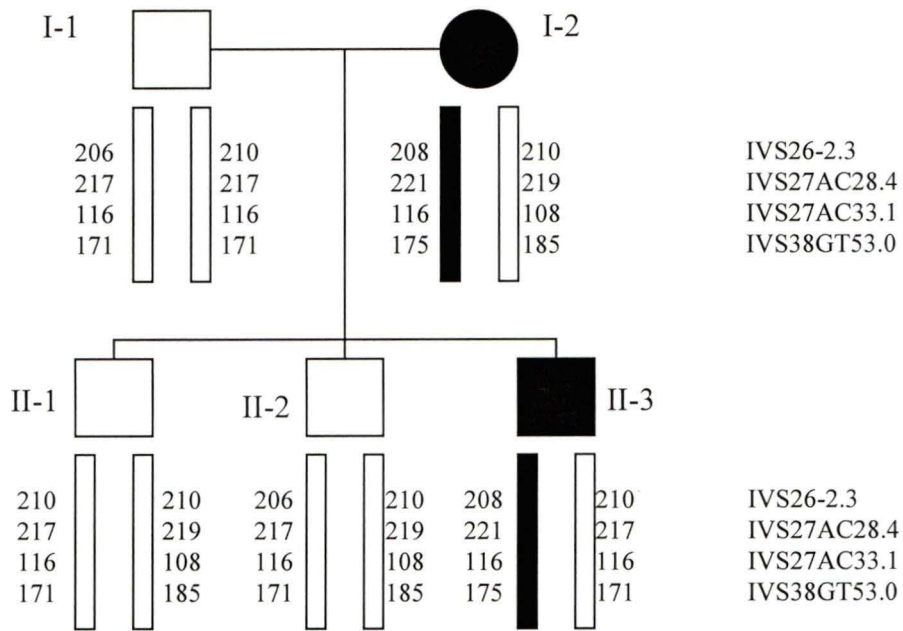


Family 4

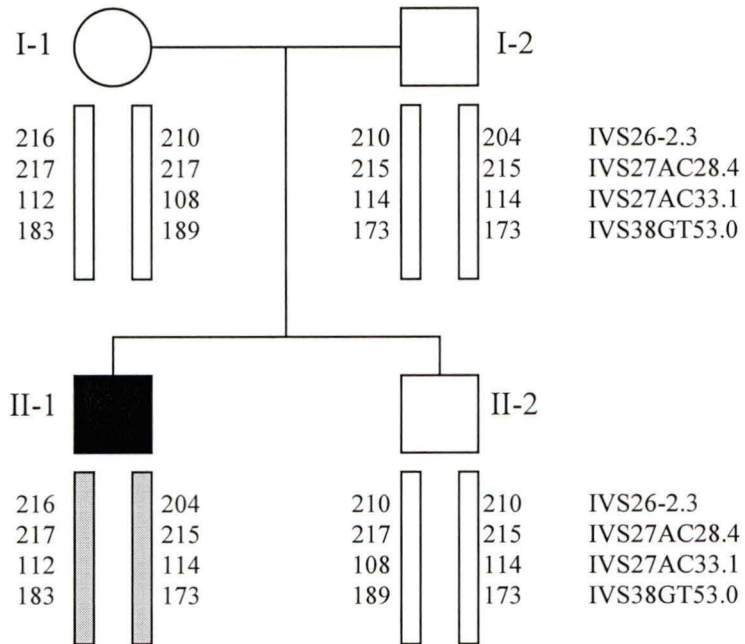




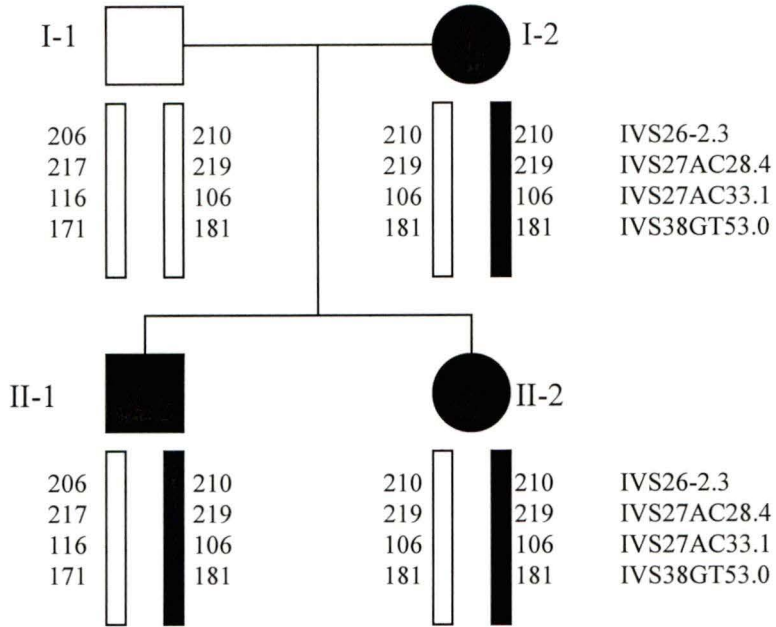
### Family 7



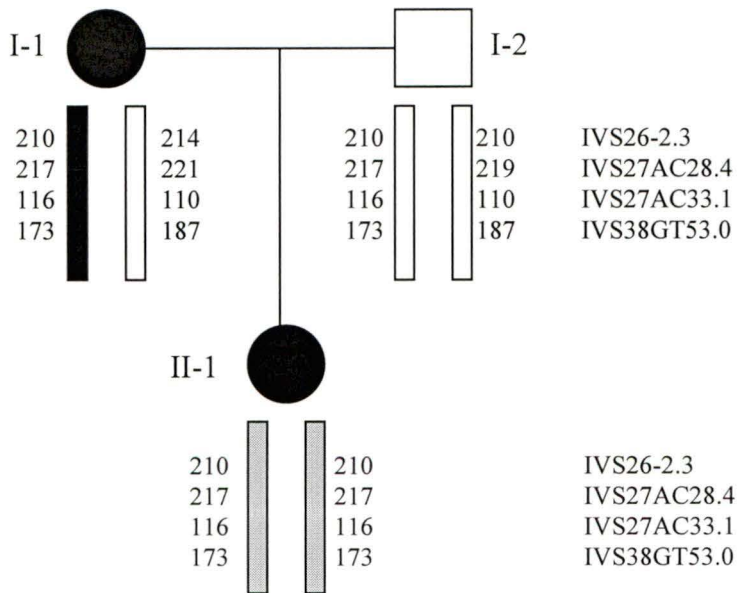
### Family 8



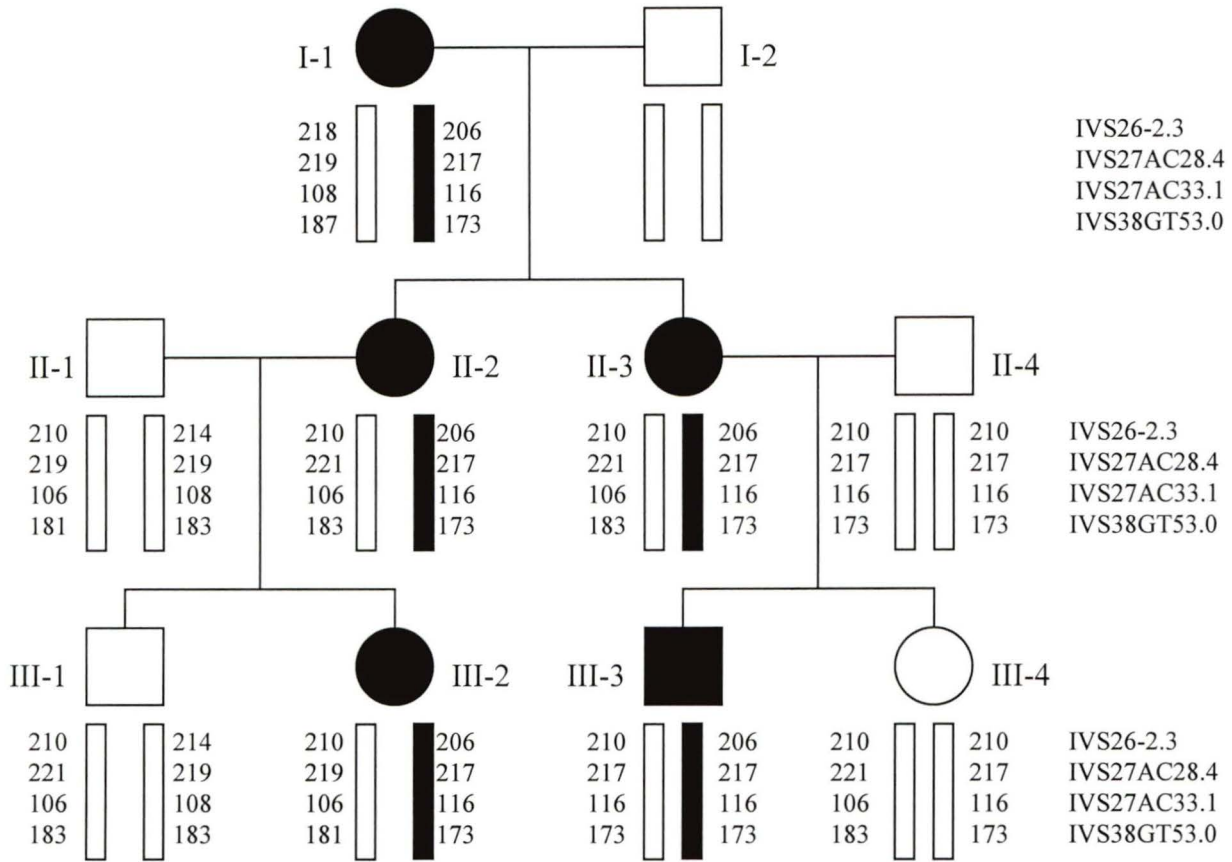
Family 9



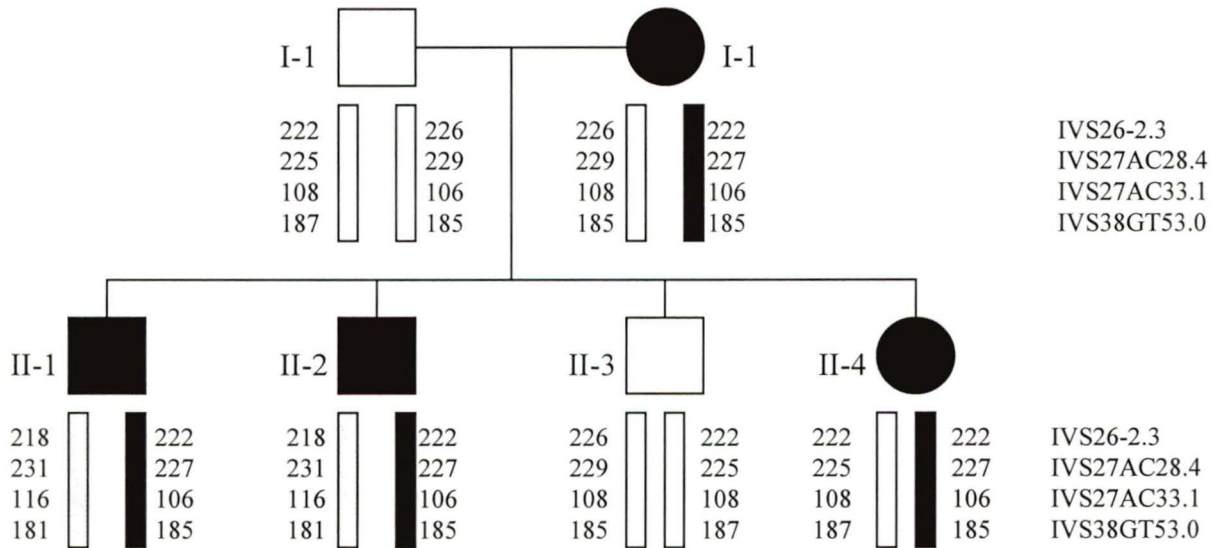
Family 10



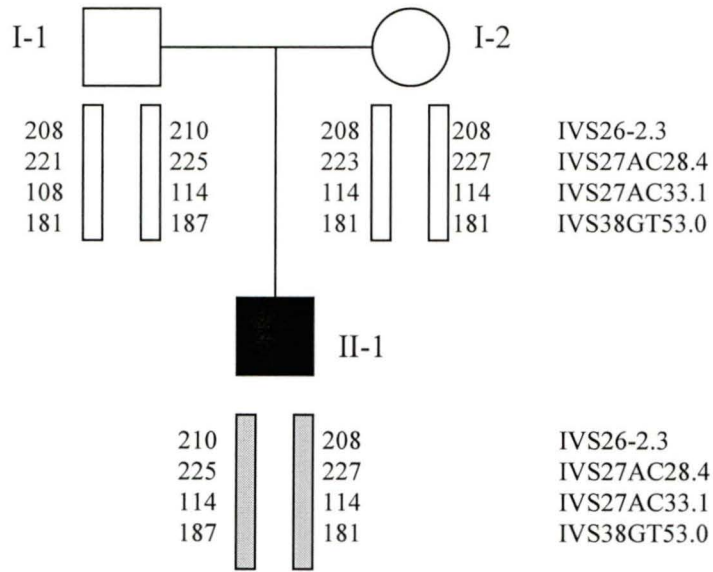
### Family 11



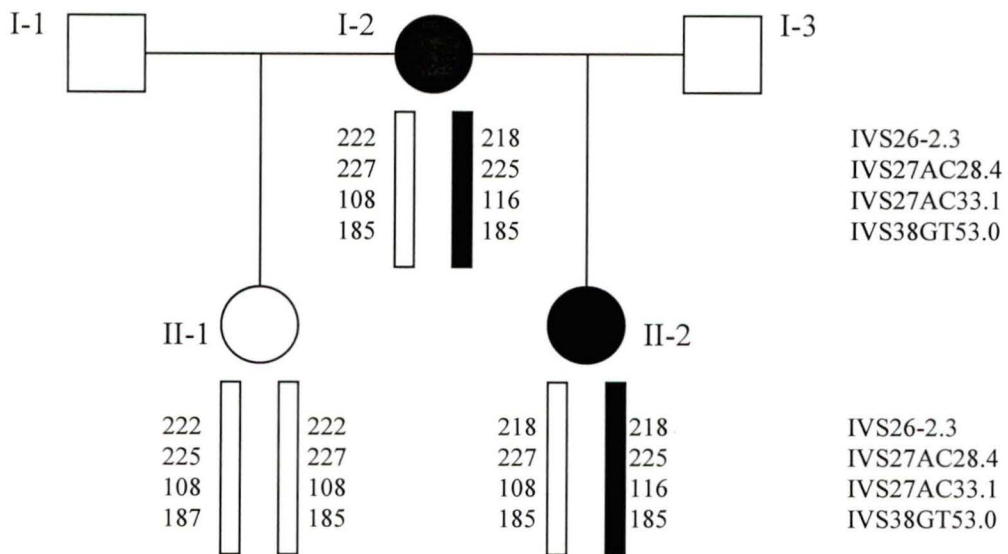
### Family 12



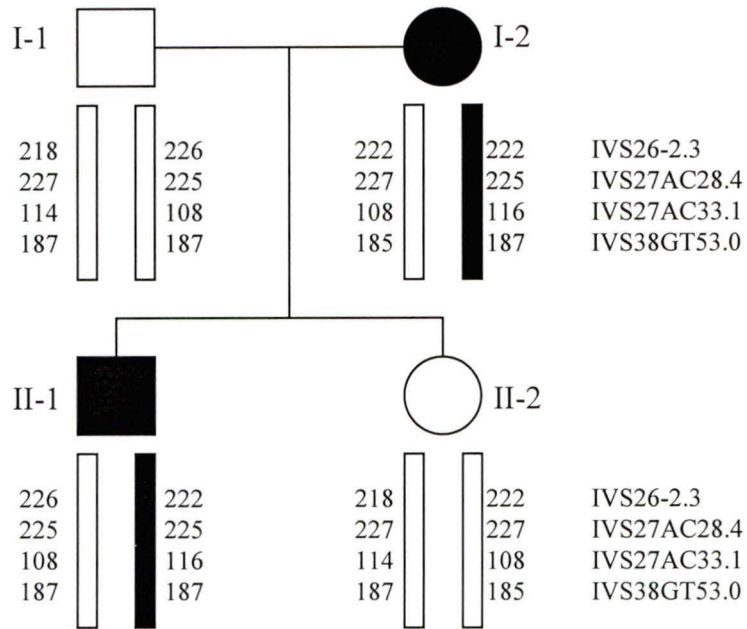
Family 13



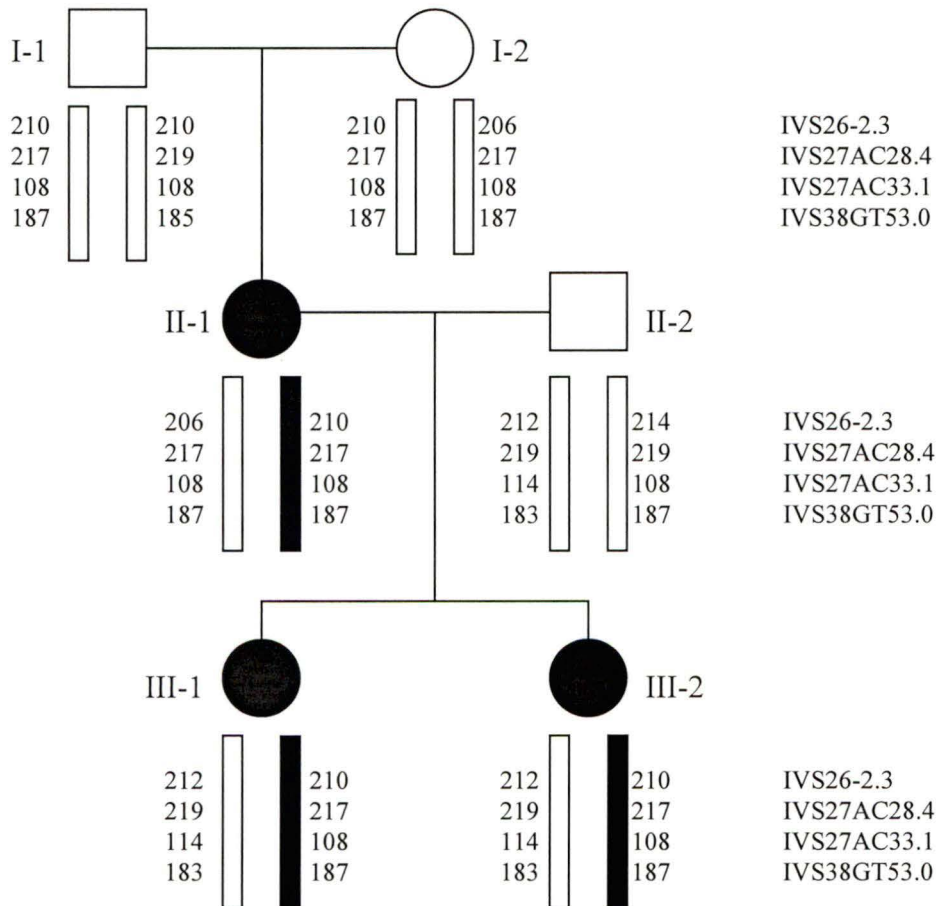
Family 14



Family 15

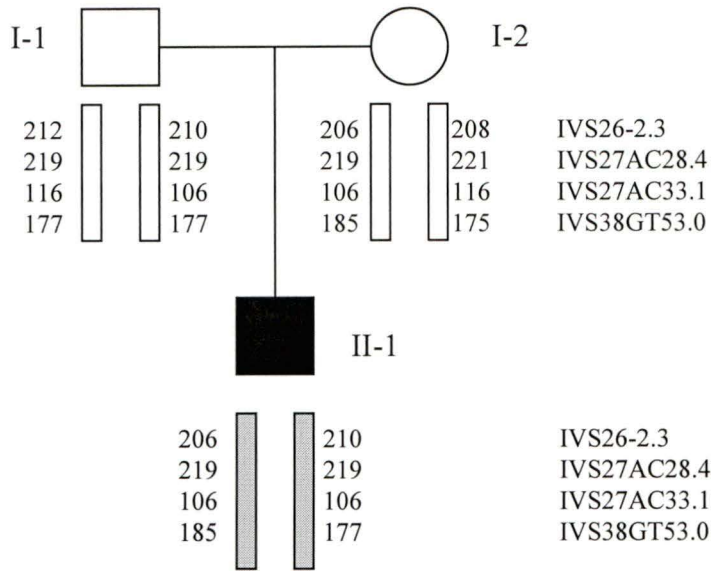


Family 16 Family 7473

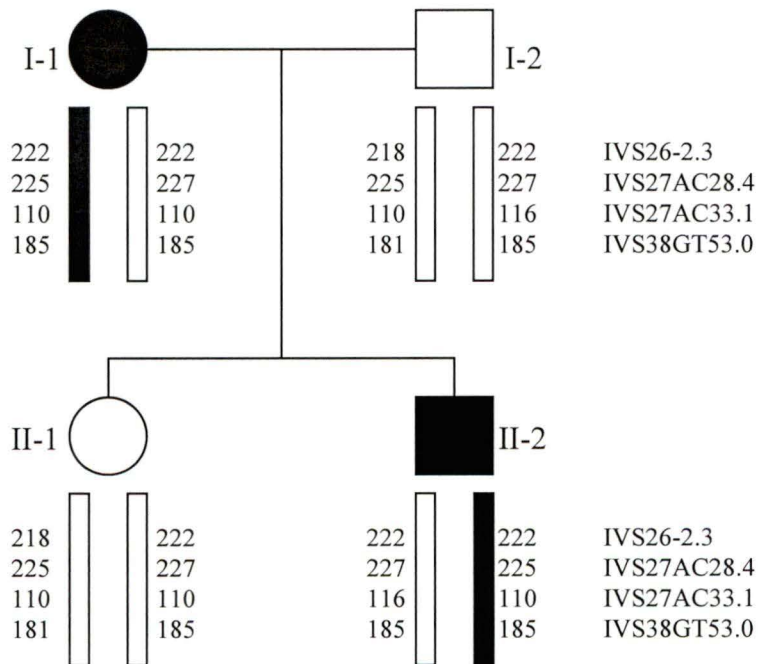




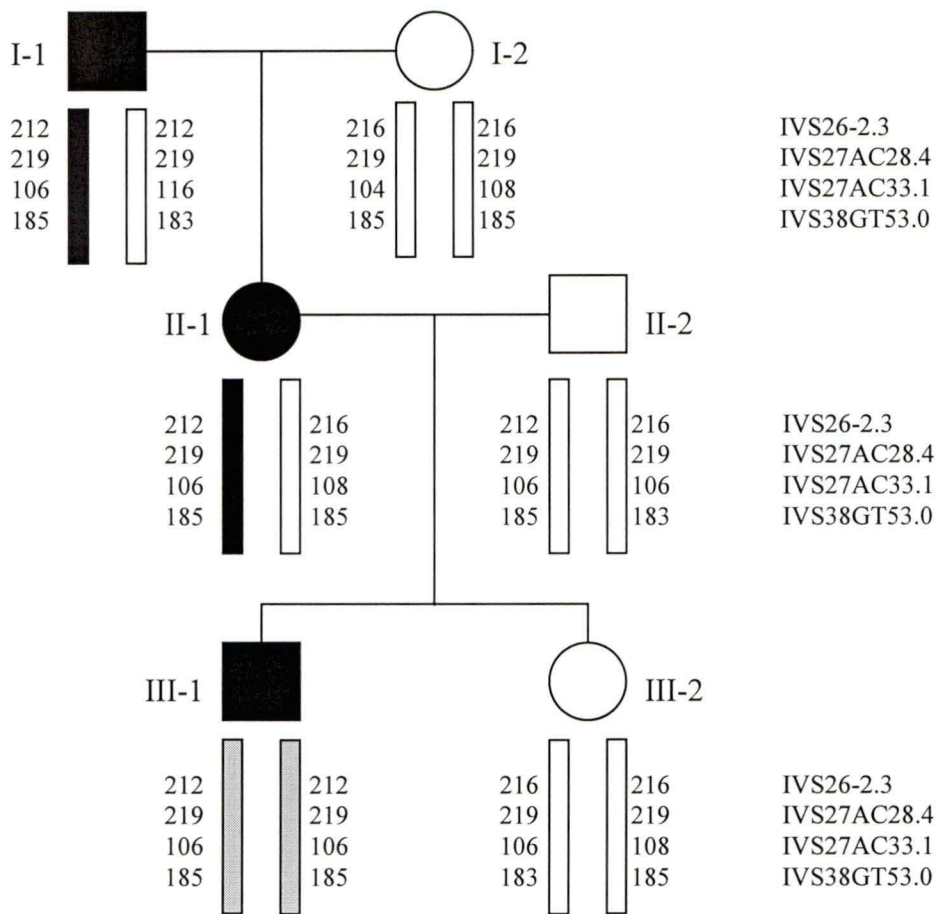
Family 17



Family 18



Family 19



## **Appendix 4**

### **Result file of LINKAGE program**

**LOD TABLE REPORT**

log10(likelihood at theta/likelihood at theta=0.5)

Order	0.005	0.01	0.05	0.1	0.2	0.3	0.4
-----							
m1	3.29	3.29	3.13	2.82	2.09	1.30	0.57
1	-0.65	-0.65	-0.46	-0.32	-0.15	-0.06	-0.01
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0.28	0.28	0.25	0.21	0.13	0.06	0.02
5	0	0	0	0	0	0	0
7	0.50	0.50	0.45	0.39	0.26	0.14	0.04
8	0.10	0.10	0.09	0.07	0.04	0.02	0
9	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0
11	1.39	1.39	1.28	1.15	0.88	0.60	0.30
12	0.58	0.58	0.53	0.46	0.31	0.17	0.05
13	0	0	0	0	0	0	0
14	0.25	0.25	0.22	0.18	0.11	0.05	0.01
15	0	0	0	0	0	0	0
16	0.29	0.29	0.25	0.21	0.13	0.06	0.02
17	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0
19	0.55	0.55	0.52	0.47	0.38	0.27	0.14
m2	2.97	2.97	2.72	2.40	1.73	1.06	0.44
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
7	0.50	0.50	0.45	0.39	0.26	0.14	0.04
8	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0
11	1.39	1.39	1.28	1.15	0.88	0.60	0.30
12	0.83	0.83	0.76	0.67	0.48	0.27	0.08
13	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0
15	0.25	0.25	0.22	0.18	0.11	0.05	0.01
16	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0
m3	3.64	3.64	3.32	2.91	2.09	1.29	0.56
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0.28	0.28	0.25	0.21	0.13	0.06	0.02
5	0	0	0	0	0	0	0
7	0.50	0.50	0.45	0.39	0.26	0.14	0.04
8	0.05	0.05	0.05	0.04	0.02	0.01	0
9	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0

11	1.39	1.39	1.28	1.15	0.88	0.60	0.30
12	0.54	0.54	0.49	0.42	0.28	0.15	0.04
13	0	0	0	0	0	0	
14	0.08	0.08	0.07	0.05	0.03	0.01	0
15	0.25	0.25	0.22	0.18	0.11	0.05	0.01
16	0	0	0	0	0	0	
17	0	0	0	0	0	0	
18	0	0	0	0	0	0	
19	0.55	0.55	0.52	0.47	0.38	0.27	0.14

m4	2.76	2.76	2.50	2.18	1.53	0.92	0.39
1	0	0	0	0	0	0	
2	0.28	0.28	0.25	0.21	0.13	0.06	0.02
3	0	0	0	0	0	0	
4	0.28	0.28	0.25	0.21	0.13	0.06	0.02
5	0	0	0	0	0	0	
7	0.50	0.50	0.45	0.39	0.26	0.14	0.04
8	0.05	0.05	0.05	0.04	0.02	0.01	0
9	0	0	0	0	0	0	
10	0	0	0	0	0	0	
11	1.39	1.39	1.28	1.15	0.88	0.60	0.30
12	0	0	0	0	0	0	
13	0	0	0	0	0	0	
14	0	0	0	0	0	0	
15	0.25	0.25	0.22	0.18	0.11	0.05	0.01
16	0	0	0	0	0	0	
17	0	0	0	0	0	0	
18	0	0	0	0	0	0	
19	0	0	0	0	0	0	

l=5	2.76	2.76	2.50	2.18	1.53	0.92	0.39
1	0	0	0	0	0	0	
2	0.28	0.28	0.25	0.21	0.13	0.06	0.02
3	0	0	0	0	0	0	
4	0.28	0.28	0.25	0.21	0.13	0.06	0.02
5	0	0	0	0	0	0	
7	0.50	0.50	0.45	0.39	0.26	0.14	0.04
8	0.05	0.05	0.05	0.04	0.02	0.01	0
9	0	0	0	0	0	0	
10	0	0	0	0	0	0	
11	1.39	1.39	1.28	1.15	0.88	0.60	0.30
12	0	0	0	0	0	0	
13	0	0	0	0	0	0	
14	0	0	0	0	0	0	
15	0.25	0.25	0.22	0.18	0.11	0.05	0.01
16	0	0	0	0	0	0	
17	0	0	0	0	0	0	
18	0	0	0	0	0	0	
19	0	0	0	0	0	0	

l=5	2.76	2.76	2.50	2.18	1.53	0.92	0.39
1	0	0	0	0	0	0	
2	0.28	0.28	0.25	0.21	0.13	0.06	0.02
3	0	0	0	0	0	0	
4	0.28	0.28	0.25	0.21	0.13	0.06	0.02
5	0	0	0	0	0	0	

7	0.50	0.50	0.45	0.39	0.26	0.14	0.04
8	0.05	0.05	0.05	0.04	0.02	0.01	0
9	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0
11	1.39	1.39	1.28	1.15	0.88	0.60	0.30
12	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0
15	0.25	0.25	0.22	0.18	0.11	0.05	0.01
16	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0

## **Appendix 5**

### **Result file of HAL program**

# File: results.out  
 # Created on Fri Mar 24 15:20:45 2000

>>> Welcome to HAL version 1.1: <<<  
 >>> Software for linkage disequilibrium mapping of <<<  
 >>> complex traits using ancestral haplotypes. <<<  
 >>> Written by Rory Martin & Charles MacLean, 20 May 1999 <<<

Haplotype-wise analysis: number of feasible configurations for:

	Disease position	Ancestral loci	Ancestral haps				
Scan to date:	1	1	1				
Current disease position:		1	1				
Current ancestral loci:			1				
marker loci (X=Disease, "="=fixed)							
	1	2	3	4			
Ancestral:	----X-----						
Loci	1						
Hap	2						
Allele freqs	.11						
	(cM)	Max	# Trans to				
	(cM)	Min	Span	span	affecteds:	# Gens	
	Disease	allele	of anc	of anc	(min)	(max)	since anc
	position	freq	loci	loci	test control	mutation	alpha
	gamma						
	2.5000E-02	1.00E-01	0.00E+00	8.00E+00	2	1	200 1.00E+00
	2.00E-01						

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

-----permutation replicates-----										
--admissible--		---locally-admiss---		-----t+c-----						
somewhere	local	test control	t+c	min	mean	max	min	mean	max	
80	80	16	64	80	80	8.00E+01	80	80	8.00E+01	80

S\_reg: analytical significance of t-value = 3.2059E-01

permutation significance level = 3.1600E-01

regression coefficient = 1.1812E-01

std error of coefficient = 2.5346E-01

regression constant = 1.8047E-01

std error of constant = 4.2507E-02

t-value for regression coefficient = 4.6604E-01

S\_est: permutation significance level = 6.1200E-01

anc hap stat: raw, normed = 8.4579E-02, -5.2506E-01

replicate samples: mean, std dev = 5.5106E-01, 8.8844E-01

S\_trim: permutation significance level = 3.7700E-01

anc hap stat: raw, normed = -2.7581E-01, 4.6694E-01

replicate samples: mean, std dev = -9.2696E-01, 1.3945E+00

ovlap	--#haps--	--#peds--	----- prob_() -----							
share	lh	rh	t	c	t	c	trim	t	c	p_t_hs2

0.5	1	0	2	7	2	6	9.6E-01	1.2E-01	1.1E-01	2.8E-01
-----	---	---	---	---	---	---	---------	---------	---------	---------



0.0 0 0 14 57 12 18 4.3E-02 8.8E-01 8.9E-01 7.2E-01

Haplotype-wise analysis: number of feasible configurations for:  
Disease position Ancestral loci Ancestral haps

Scan to date: 1 1 2  
Current disease position: 1 2  
Current ancestral loci: 2  
marker loci (X=Disease, "="=fixed)  
1 2 3 4

Ancestral: ----X-----  
Loci 1  
Hap 4  
Allele freqs .25

(cM) Max # Trans to  
(cM) Min Span span affected: # Gens  
Disease allele of anc of anc (min) (max) since anc  
position freq loci loci test control mutation alpha  
gamma  
2.50000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00  
2.00E-01

Number of replicates (feas/ total) = 1000 / 1000  
Founder haplotypes:

-----permutation replicates-----  
--admissible-- --locally-admiss-- -----t+c-----  
somewhere local test control t+c min mean max min mean max  
80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

S\_reg: analytical significance of t-value = 3.5801E-01  
permutation significance level = 3.6400E-01  
regression coefficient = 6.7319E-02  
std error of coefficient = 1.8506E-01  
regression constant = 1.7870E-01  
std error of constant = 4.6264E-02  
t-value for regression coefficient = 3.6378E-01

S\_est: permutation significance level = 8.3600E-01  
anc hap stat: raw, normed = 0.0000E+00, -5.8570E-01  
replicate samples: mean, std dev = 4.8958E-01, 8.3589E-01

S\_trim: permutation significance level = 5.2500E-01  
anc hap stat: raw, normed = -5.4286E-01, -1.0257E-01  
replicate samples: mean, std dev = -4.3410E-01, 1.0604E+00

overlap --#haps-- --#peds-- ----- prob\_() -----  
share lh rh t c t c trim t c p\_t\_hs2  
-----  
0.5 1 0 4 16 4 8 9.6E-01 2.5E-01 2.5E-01 3.9E-01  
0.0 0 0 12 48 11 18 3.6E-02 7.5E-01 7.5E-01 6.1E-01

Haplotype-wise analysis: number of feasible configurations for:

```

Disease position  Ancestral loci  Ancestral haps
Scan to date:      1          1          3
Current disease position:      1          3
Current ancestral loci:      3
marker loci (X=Disease, "="=fixed)
  1  2  3  4
Ancestral:  ---X-----
Loci        1
Hap         5
Allele freqs .12
(cM) Max # Trans to
(cM) Min Span span affecteds: # Gens
Disease allele of anc of anc (min) (max) since anc
position freq loci loci test control mutation alpha
gamma
2.50000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00
2.00E-01

```

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

```

-----permutation replicates-----
--admissible--      ---locally-admiss--- -----t+c-----
somewhere local test control t+c min mean max min mean max
80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

```

S\_reg: analytical significance of t-value = 4.8973E-01

permutation significance level = 5.0000E-01

regression coefficient = 6.2444E-03

std error of coefficient = 2.4250E-01

regression constant = 1.8673E-01

std error of constant = 4.2869E-02

t-value for regression coefficient = 2.5750E-02

S\_est: permutation significance level = 9.7600E-01

anc hap stat: raw, normed = -4.6157E-01, -1.1482E+00

replicate samples: mean, std dev = 5.4462E-01, 8.7634E-01

S\_trim: permutation significance level = 4.8100E-01

anc hap stat: raw, normed = -1.3673E+00, -4.3284E-01

replicate samples: mean, std dev = -7.7096E-01, 1.3777E+00

```

ovlap --#haps-- --#peds-- ----- prob_() -----
share lh rh t c t c trim t c p_t_hs2
-----
0.5 1 0 3 7 3 4 9.6E-01 1.9E-01 1.1E-01 2.8E-01
0.0 0 0 13 57 11 17 4.3E-02 8.1E-01 8.9E-01 7.2E-01

```

Haplotype-wise analysis: number of feasible configurations for:

```

Disease position  Ancestral loci  Ancestral haps
Scan to date:      1          1          4
Current disease position:      1          4
Current ancestral loci:      4
marker loci (X=Disease, "="=fixed)

```

```

      1 2 3 4
Ancestral: ---X-----
Loci      1
Hap       9
Allele freqs .14
      (cM) Max # Trans to
      (cM) Min Span span affecteds: # Gens
      Disease allele of anc of anc (min) (max) since anc
      position freq loci loci test control mutation alpha
gamma
2.50000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00
2.00E-01

```

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

```

-----permutation replicates-----
--admissible-- ---locally-admiss--- -----t+c-----
somewhere local test control t+c min mean max min mean max
      80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

```

S\_reg: analytical significance of t-value = 2.5294E-01  
 permutation significance level = 2.4900E-01  
 regression coefficient = 1.4900E-01  
 std error of coefficient = 2.2397E-01  
 regression constant = 1.7594E-01  
 std error of constant = 4.3372E-02  
 t-value for regression coefficient = 6.6526E-01

S\_est: permutation significance level = 2.9400E-01  
 anc hap stat: raw, normed = 3.6023E-01, -1.5705E-01  
 replicate samples: mean, std dev = 4.8592E-01, 8.0036E-01

S\_trim: permutation significance level = 1.6700E-01  
 anc hap stat: raw, normed = 4.2324E-01, 9.5178E-01  
 replicate samples: mean, std dev = -7.5381E-01, 1.2367E+00

```

      overlap --#haps-- --#peds-- ----- prob_() -----
share lh rh t c t c trim t c p_t_hs2
-----
0.5 1 0 3 9 3 6 9.6E-01 1.9E-01 1.4E-01 3.0E-01
0.0 0 0 13 55 11 18 4.2E-02 8.1E-01 8.6E-01 7.0E-01

```

Haplotype-wise analysis: number of feasible configurations for:

```

      Disease position Ancestral loci Ancestral haps
Scan to date:          1          2          5
Current disease position:          2          5
Current ancestral loci:          1
      marker loci (X=Disease, "="=fixed)
      1 2 3 4
Ancestral: ---X-----
Loci      1
Hap       2
Allele freqs .22

```

```

(cM) Max # Trans to
(cM) Min Span span affecteds: # Gens
Disease allele of anc of anc (min) (max) since anc
position freq loci loci test control mutation alpha
gamma
2.5000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00
2.00E-01

```

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

```

-----permutation replicates-----
--admissible-- ---locally-admiss--- -----t+c-----
somewhere local test control t+c min mean max min mean max
80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

```

S\_reg: analytical significance of t-value = 2.2007E-01  
permutation significance level = 2.0300E-01  
regression coefficient = 1.5077E-01  
std error of coefficient = 1.9531E-01  
regression constant = 1.7110E-01  
std error of constant = 4.5016E-02  
t-value for regression coefficient = 7.7195E-01

S\_est: permutation significance level = 3.8500E-01  
anc hap stat: raw, normed = 3.3483E-01, -2.0648E-01  
replicate samples: mean, std dev = 5.2557E-01, 9.2372E-01

S\_trim: permutation significance level = 2.1500E-01  
anc hap stat: raw, normed = 5.1312E-01, 9.2321E-01  
replicate samples: mean, std dev = -5.6656E-01, 1.1695E+00

```

overlap --#haps-- --#peds-- ----- prob_() -----
share lh rh t c t c trim t c p_t_hs2
-----
0.5 0 1 4 13 4 7 9.6E-01 2.5E-01 2.0E-01 3.5E-01
0.0 0 0 12 51 11 17 3.8E-02 7.5E-01 8.0E-01 6.5E-01

```

Haplotype-wise analysis: number of feasible configurations for:

```

Disease position Ancestral loci Ancestral haps
Scan to date: 1 2 6
Current disease position: 2 6
Current ancestral loci: 2
marker loci (X=Disease, "="=fixed)
1 2 3 4
Ancestral: ----X-----
Loci 1
Hap 3
Allele freqs .30

```

```

(cM) Max # Trans to
(cM) Min Span span affecteds: # Gens
Disease allele of anc of anc (min) (max) since anc
position freq loci loci test control mutation alpha
gamma

```

2.50000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00  
2.00E-01

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

```
-----permutation replicates-----
--admissible--      --locally-admiss--  -----t+c-----
somewhere local test control t+c min mean max min mean max
  80  80  16  64  80  80 8.00E+01  80  80 8.00E+01  80
```

S\_reg: analytical significance of t-value = 7.8702E-01  
permutation significance level = 7.9800E-01  
regression coefficient = -1.4050E-01  
std error of coefficient = 1.7648E-01  
regression constant = 2.0732E-01  
std error of constant = 4.7315E-02  
t-value for regression coefficient = -7.9613E-01

S\_est: permutation significance level = 8.5000E-02  
anc hap stat: raw, normed = 1.9187E+00, 1.6816E+00  
replicate samples: mean, std dev = 5.0555E-01, 8.4035E-01

S\_trim: permutation significance level = 9.5500E-01  
anc hap stat: raw, normed = -1.9716E+00, -1.5830E+00  
replicate samples: mean, std dev = -3.5177E-01, 1.0233E+00

```
overlap --#haps-- --#peds-- ----- prob_() -----
share lh rh t c t c trim t c p_t_hs2
-----
0.5 0 1 2 21 2 8 9.7E-01 1.2E-01 3.3E-01 4.6E-01
0.0 0 0 14 43 12 16 3.4E-02 8.8E-01 6.7E-01 5.4E-01
```

Haplotype-wise analysis: number of feasible configurations for:

```
      Disease position  Ancestral loci  Ancestral haps
Scan to date:          1          2          7
Current disease position:          2          7
Current ancestral loci:          3
      marker loci (X=Disease, "="=fixed)
      1 2 3 4
Ancestral:  ---X-----
Loci        1
Hap         6
Allele freqs .15
      (cM)  Max  # Trans to
      (cM)  Min  Span  span  affecteds:  # Gens
      Disease allele of anc of anc (min) (max) since anc
      position freq loci loci test control mutation alpha
gamma
2.50000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00
2.00E-01
```

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

```

-----permutation replicates-----
--admissible--      --locally-admiss-- -----t+c-----
somewhere local test control t+c min mean max min mean max
      80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

```

```

S_reg: analytical significance of t-value = 2.0275E-01
      permutation significance level = 1.8400E-01
      regression coefficient = 1.8601E-01
      std error of coefficient = 2.2362E-01
      regression constant = 1.7317E-01
      std error of constant = 4.3303E-02
      t-value for regression coefficient = 8.3183E-01

```

```

S_est: permutation significance level = 2.8700E-01
      anc hap stat: raw, normed = 3.6022E-01, -1.2648E-01
      replicate samples: mean, std dev = 4.6548E-01, 8.3225E-01

```

```

S_trim: permutation significance level = 1.5800E-01
      anc hap stat: raw, normed = 4.2311E-01, 9.7368E-01
      replicate samples: mean, std dev = -7.7103E-01, 1.2264E+00

```

```

      overlap --#haps-- --#peds-- ----- prob_() -----
share lh rh t c t c trim t c p_t_hs2
-----
0.5 0 1 3 9 3 8 9.6E-01 1.9E-01 1.4E-01 3.0E-01
0.0 0 0 13 55 10 18 4.2E-02 8.1E-01 8.6E-01 7.0E-01

```

Haplotype-wise analysis: number of feasible configurations for:

```

      Disease position Ancestral loci Ancestral haps
Scan to date:          1          2          8
Current disease position:          2          8
Current ancestral loci:          4
      marker loci (X=Disease, "="=fixed)
      1 2 3 4
Ancestral: ----X-----
Loci          1
Hap           7
Allele freqs .15
      (cM) Max # Trans to
      (cM) Min Span span affecteds: # Gens
      Disease allele of anc of anc (min) (max) since anc
      position freq loci loci test control mutation alpha
gamma
2.5000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00
2.00E-01

```

Number of replicates (feas/total) = 1000 / 1000

Founder haplotypes:

```

-----permutation replicates-----
--admissible--      --locally-admiss-- -----t+c-----
somewhere local test control t+c min mean max min mean max
      80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

```

S\_reg: analytical significance of t-value = 5.7559E-01  
 permutation significance level = 5.4800E-01  
 regression coefficient = -4.1431E-02  
 std error of coefficient = 2.1735E-01  
 regression constant = 1.9048E-01  
 std error of constant = 4.3808E-02  
 t-value for regression coefficient = -1.9062E-01

S\_est: permutation significance level = 8.9600E-01  
 anc hap stat: raw, normed = -8.8819E-02, -7.5212E-01  
 replicate samples: mean, std dev = 5.0649E-01, 7.9152E-01

S\_trim: permutation significance level = 4.7800E-01  
 anc hap stat: raw, normed = -7.9639E-01, -5.2742E-02  
 replicate samples: mean, std dev = -7.3167E-01, 1.2271E+00

```

ovlap --#haps-- --#peds-- ----- prob_() -----
share lh rh t c t c trim t c p_t_hs2
-----
0.5 0 1 2 11 2 6 9.6E-01 1.2E-01 1.7E-01 3.3E-01
0.0 0 0 14 53 12 18 4.2E-02 8.8E-01 8.3E-01 6.7E-01

```

Haplotype-wise analysis: number of feasible configurations for:  
 Disease position Ancestral loci Ancestral haps

Scan to date: 2 3 9  
 Current disease position: 1 1  
 Current ancestral loci: 1

marker loci (X=Disease, "="=fixed)  
 1 2 3 4

Ancestral: -----X-----  
 Loci 1  
 Hap 2  
 Allele freqs .22

(cM)	Min	Span	Max	# Trans	affecteds:	# Gens
Disease position	allele freq	of anc loci	of anc loci	(min) test control	(max) mutation	since anc alpha
7.5000E-02	1.00E-01	0.00E+00	8.00E+00	2	1	200
2.00E-01						1.00E+00

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

```

-----permutation replicates-----
--admissible-- ---locally-admiss--- -----t+c-----
somewhere local test control t+c min mean max min mean max
80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

```

S\_reg: analytical significance of t-value = 2.2021E-01  
 permutation significance level = 2.0300E-01  
 regression coefficient = 1.5076E-01  
 std error of coefficient = 1.9542E-01  
 regression constant = 1.7110E-01

std error of constant = 4.5041E-02  
 t-value for regression coefficient = 7.7149E-01

S\_est: permutation significance level = 3.8600E-01  
 anc hap stat: raw, normed = 3.3449E-01, -2.0707E-01  
 replicate samples: mean, std dev = 5.2584E-01, 9.2408E-01

S\_trim: permutation significance level = 2.1600E-01  
 anc hap stat: raw, normed = 5.1192E-01, 9.2205E-01  
 replicate samples: mean, std dev = -5.6690E-01, 1.1700E+00

		ovlap		--#haps--		--#peds--		----- prob_()		-----	
share	lh rh	t	c	t	c	trim	t	c	p	t	hs2
0.5	1 0	4	13	4	7	9.6E-01	2.5E-01	2.0E-01	3.5E-01		
0.0	0 0	12	51	11	17	3.8E-02	7.5E-01	8.0E-01	6.5E-01		

Haplotype-wise analysis: number of feasible configurations for:  
 Disease position    Ancestral loci    Ancestral haps

Scan to date:                    2                    3                    10

Current disease position:                    1                    2

Current ancestral loci:                    2

marker loci (X=Disease, "="=fixed)

1 2 3 4

Ancestral: -----X-----

Loci                    1

Hap                    3

Allele freqs                    .30

(cM)    Max    # Trans to

(cM)    Min    Span    span    affecteds:    # Gens

Disease allele    of anc    of anc    (min)    (max)    since anc

position    freq    loci    loci    test control    mutation    alpha

gamma

7.5000E-02 1.00E-01 0.00E+00 8.00E+00    2    1    200 1.00E+00

2.00E-01

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

-----permutation replicates-----

--admissible--

---locally-admiss--- -----t+c-----

somewhere local test control t+c min mean max min mean max

80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

S\_reg: analytical significance of t-value = 7.8690E-01

permutation significance level = 7.9800E-01

regression coefficient = -1.4051E-01

std error of coefficient = 1.7658E-01

regression constant = 2.0732E-01

std error of constant = 4.7341E-02

t-value for regression coefficient = -7.9571E-01

S\_est: permutation significance level = 8.5000E-02

anc hap stat: raw, normed = 1.9202E+00, 1.6824E+00



replicate samples: mean, std dev = 5.0575E-01, 8.4075E-01

S\_trim: permutation significance level = 9.5500E-01

anc hap stat: raw, normed = -1.9728E+00, -1.5833E+00

replicate samples: mean, std dev = -3.5189E-01, 1.0237E+00

```
ovlap --#haps-- --#peds-- ----- prob_() -----
share lh rh t c t c trim t c p_t_hs2
-----
0.5 1 0 2 21 2 8 9.7E-01 1.2E-01 3.3E-01 4.6E-01
0.0 0 0 14 43 12 16 3.4E-02 8.8E-01 6.7E-01 5.4E-01
```

Haplotype-wise analysis: number of feasible configurations for:

```
          Disease position  Ancestral loci  Ancestral haps
Scan to date:                2            3            11
Current disease position:                1            3
Current ancestral loci:                3
marker loci (X=Disease, "="=fixed)
  1 2 3 4
Ancestral: -----X-----
Loci      1
Hap       6
Allele freqs .15
          (cM)  Max  # Trans to
          (cM)  Min  Span  span  affecteds:  # Gens
          Disease allele of anc of anc (min) (max) since anc
          position freq loci loci test control mutation alpha
gamma
7.50000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00
2.00E-01
```

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

```
-----permutation replicates-----
--admissible-- ---locally-admiss--- -----t+c-----
somewhere local test control t+c min mean max min mean max
80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80
```

S\_reg: analytical significance of t-value = 2.0289E-01

permutation significance level = 1.8600E-01

regression coefficient = 1.8601E-01

std error of coefficient = 2.2374E-01

regression constant = 1.7317E-01

std error of constant = 4.3327E-02

t-value for regression coefficient = 8.3135E-01

S\_est: permutation significance level = 3.0600E-01

anc hap stat: raw, normed = 3.5990E-01, -1.2701E-01

replicate samples: mean, std dev = 4.6567E-01, 8.3282E-01

S\_trim: permutation significance level = 1.7700E-01

anc hap stat: raw, normed = 4.2192E-01, 9.7248E-01

replicate samples: mean, std dev = -7.7140E-01, 1.2271E+00

```

ovlap --#haps-- --#peds-- ----- prob_() -----
share lh rh t c t c trim t c p_t_hs2
-----
0.5 1 0 3 9 3 8 9.6E-01 1.9E-01 1.4E-01 3.0E-01
0.0 0 0 13 55 10 18 4.2E-02 8.1E-01 8.6E-01 7.0E-01

```

Haplotype-wise analysis: number of feasible configurations for:

```

Disease position Ancestral loci Ancestral haps
Scan to date: 2 3 12
Current disease position: 1 4
Current ancestral loci: 4
marker loci (X=Disease, "="=fixed)
 1 2 3 4
Ancestral: -----X-----
Loci 1
Hap 7
Allele freqs .15
(cM) Max # Trans to
(cM) Min Span span affecteds: # Gens
Disease allele of anc of anc (min) (max) since anc
position freq loci loci test control mutation alpha
gamma
7.50000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00
2.00E-01

```

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

```

-----permutation replicates-----
--admissible-- ---locally-admiss--- -----t+c-----
somewhere local test control t+c min mean max min mean max
80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

```

S\_reg: analytical significance of t-value = 5.7556E-01

permutation significance level = 5.4900E-01

regression coefficient = -4.1435E-02

std error of coefficient = 2.1747E-01

regression constant = 1.9049E-01

std error of constant = 4.3832E-02

t-value for regression coefficient = -1.9054E-01

S\_est: permutation significance level = 8.8500E-01

anc hap stat: raw, normed = -8.8507E-02, -7.5163E-01

replicate samples: mean, std dev = 5.0675E-01, 7.9196E-01

S\_trim: permutation significance level = 4.8900E-01

anc hap stat: raw, normed = -7.9758E-01, -5.3449E-02

replicate samples: mean, std dev = -7.3196E-01, 1.2277E+00

```

ovlap --#haps-- --#peds-- ----- prob_() -----
share lh rh t c t c trim t c p_t_hs2
-----

```

0.5 1 0 2 11 2 6 9.6E-01 1.2E-01 1.7E-01 3.3E-01  
 0.0 0 0 14 53 12 18 4.2E-02 8.8E-01 8.3E-01 6.7E-01

Haplotype-wise analysis: number of feasible configurations for:

	Disease position	Ancestral loci	Ancestral haps
Scan to date:	2	4	13
Current disease position:		2	5
Current ancestral loci:			1

marker loci (X=Disease, "="=fixed)  
 1 2 3 4

Ancestral: -----X-----  
 Loci 1  
 Hap 2  
 Allele freqs .21  
 (cM) Max # Trans to  
 (cM) Min Span span affecteds: # Gens  
 Disease allele of anc of anc (min) (max) since anc  
 position freq loci loci test control mutation alpha  
 gamma  
 7.5000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00  
 2.00E-01

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

-----permutation replicates-----  
 --admissible-- ---locally-admiss--- -----t+c-----  
 somewhere local test control t+c min mean max min mean max  
 80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

S\_reg: analytical significance of t-value = 1.6052E-01

permutation significance level = 1.7300E-01

regression coefficient = 1.9782E-01

std error of coefficient = 1.9936E-01

regression constant = 1.6734E-01

std error of constant = 4.4578E-02

t-value for regression coefficient = 9.9230E-01

S\_est: permutation significance level = 5.3100E-01

anc hap stat: raw, normed = 1.2921E-01, -4.3915E-01

replicate samples: mean, std dev = 4.7711E-01, 7.9222E-01

S\_trim: permutation significance level = 2.1800E-01

anc hap stat: raw, normed = 6.3041E-02, 5.0733E-01

replicate samples: mean, std dev = -5.2036E-01, 1.1499E+00

share	lh	rh	t	c	t	c	trim	t	c	p_t_hs2	prob_()
0.5	0	1	5	11	5	6	9.6E-01	3.1E-01	1.7E-01	3.3E-01	
0.0	0	0	11	53	10	18	3.9E-02	6.9E-01	8.3E-01	6.7E-01	

Haplotype-wise analysis: number of feasible configurations for:  
 Disease position Ancestral loci Ancestral haps  
 Scan to date: 2 4 14  
 Current disease position: 2 6  
 Current ancestral loci: 2  
 marker loci (X=Disease, "="=fixed)  
 1 2 3 4  
 Ancestral: -----X-----  
 Loci 1  
 Hap 3  
 Allele freqs .38  
 (cM) Max # Trans to  
 (cM) Min Span span affecteds: # Gens  
 Disease allele of anc of anc (min) (max) since anc  
 position freq loci loci test control mutation alpha  
 gamma  
 7.5000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00  
 2.00E-01

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

-----permutation replicates-----  
 --admissible-- ---locally-admiss--- -----t+c-----  
 somewhere local test control t+c min mean max min mean max  
 80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

S\_reg: analytical significance of t-value = 8.8996E-01

permutation significance level = 8.6900E-01

regression coefficient = -2.0007E-01

std error of coefficient = 1.6315E-01

regression constant = 2.2588E-01

std error of constant = 5.0779E-02

t-value for regression coefficient = -1.2263E+00

S\_est: permutation significance level = 1.7900E-01

anc hap stat: raw, normed = 8.0600E-01, 4.0491E-01

replicate samples: mean, std dev = 4.8314E-01, 7.9734E-01

S\_trim: permutation significance level = 9.1200E-01

anc hap stat: raw, normed = -1.2438E+00, -1.1054E+00

replicate samples: mean, std dev = -3.0372E-01, 8.5041E-01

ovlap --#haps-- --#peds-- ----- prob\_() -----  
 share lh rh t c t c trim t c p\_t\_hs2  
 -----  
 0.5 0 1 3 28 3 14 9.7E-01 1.9E-01 4.4E-01 5.4E-01  
 0.0 0 0 13 36 11 14 3.0E-02 8.1E-01 5.6E-01 4.6E-01

Haplotype-wise analysis: number of feasible configurations for:

Disease position Ancestral loci Ancestral haps  
 Scan to date: 2 4 15  
 Current disease position: 2 7  
 Current ancestral loci: 3

```

marker loci (X=Disease, "="=fixed)
  1 2 3 4
Ancestral: -----X-----
Loci      1
Hap       7
Allele freqs .23
           (cM) Max # Trans to
           (cM) Min Span span affecteds: # Gens
           Disease allele of anc of anc (min) (max) since anc
           position freq loci loci test control mutation alpha
gamma
7.50000E-02 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00
2.00E-01

```

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

```

-----permutation replicates-----
--admissible-- ---locally-admiss--- -----t+c-----
somewhere local test control t+c min mean max min mean max
 80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

```

S\_reg: analytical significance of t-value = 1.0737E-01  
permutation significance level = 1.0700E-01  
regression coefficient = 2.3609E-01  
std error of coefficient = 1.9030E-01  
regression constant = 1.6056E-01  
std error of constant = 4.5133E-02  
t-value for regression coefficient = 1.2407E+00

S\_est: permutation significance level = 7.9000E-02  
anc hap stat: raw, normed = 1.6338E+00, 1.2633E+00  
replicate samples: mean, std dev = 4.9906E-01, 8.9823E-01

S\_trim: permutation significance level = 4.4000E-02  
anc hap stat: raw, normed = 1.5341E+00, 1.8426E+00  
replicate samples: mean, std dev = -5.5082E-01, 1.1315E+00

```

overlap --#haps-- --#peds-- ----- prob_() -----
share lh rh t c t c trim t c p_t_hs2
-----
0.5 0 1 6 12 6 9 9.6E-01 3.8E-01 1.9E-01 3.4E-01
0.0 0 0 10 52 9 18 3.7E-02 6.2E-01 8.1E-01 6.6E-01

```

Haplotype-wise analysis: number of feasible configurations for:

```

Disease position Ancestral loci Ancestral haps
Scan to date: 3 5 16
Current disease position: 1 1
Current ancestral loci: 1
marker loci (X=Disease, "="=fixed)
  1 2 3 4
Ancestral: -----X---
Loci      1
Hap       2

```

Allele freqs .21  
(cM) Max # Trans to  
(cM) Min Span span affecteds: # Gens  
Disease allele of anc of anc (min) (max) since anc  
position freq loci loci test control mutation alpha  
gamma  
1.25000E-01 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00  
2.00E-01

Number of replicates (feas/ total) = 1000 / 1000  
Founder haplotypes:

-----permutation replicates-----  
--admissible-- --locally-admiss-- -----t+c-----  
somewhere local test control t+c min mean max min mean max  
80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

S\_reg: analytical significance of t-value = 1.6052E-01  
permutation significance level = 1.7300E-01  
regression coefficient = 1.9782E-01  
std error of coefficient = 1.9936E-01  
regression constant = 1.6734E-01  
std error of constant = 4.4578E-02  
t-value for regression coefficient = 9.9230E-01

S\_est: permutation significance level = 5.3100E-01  
anc hap stat: raw, normed = 1.2921E-01, -4.3915E-01  
replicate samples: mean, std dev = 4.7711E-01, 7.9222E-01

S\_trim: permutation significance level = 2.1800E-01  
anc hap stat: raw, normed = 6.3041E-02, 5.0733E-01  
replicate samples: mean, std dev = -5.2036E-01, 1.1499E+00

ovlap --#haps-- --#peds-- ----- prob\_() -----  
share lh rh t c t c trim t c p\_t\_hs2  
-----  
0.5 1 0 5 11 5 6 9.6E-01 3.1E-01 1.7E-01 3.3E-01  
0.0 0 0 11 53 10 18 3.9E-02 6.9E-01 8.3E-01 6.7E-01

Haplotype-wise analysis: number of feasible configurations for:  
Disease position Ancestral loci Ancestral haps  
Scan to date: 3 5 17  
Current disease position: 1 2  
Current ancestral loci: 2  
marker loci (X=Disease, "="=fixed)  
1 2 3 4  
Ancestral: -----X---  
Loci 1  
Hap 3  
Allele freqs .38  
(cM) Max # Trans to  
(cM) Min Span span affecteds: # Gens  
Disease allele of anc of anc (min) (max) since anc  
position freq loci loci test control mutation alpha

gamma  
 1.25000E-01 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00  
 2.00E-01

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

-----permutation replicates-----  
 --admissible-- --locally-admiss-- -----t+c-----  
 somewhere local test control t+c min mean max min mean max  
 80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

S\_reg: analytical significance of t-value = 8.8996E-01  
 permutation significance level = 8.6900E-01  
 regression coefficient = -2.0007E-01  
 std error of coefficient = 1.6315E-01  
 regression constant = 2.2588E-01  
 std error of constant = 5.0779E-02  
 t-value for regression coefficient = -1.2263E+00

S\_est: permutation significance level = 1.7900E-01  
 anc hap stat: raw, normed = 8.0600E-01, 4.0491E-01  
 replicate samples: mean, std dev = 4.8314E-01, 7.9734E-01

S\_trim: permutation significance level = 9.1200E-01  
 anc hap stat: raw, normed = -1.2438E+00, -1.1054E+00  
 replicate samples: mean, std dev = -3.0372E-01, 8.5041E-01

ovlap	--#haps--	--#peds--	----- prob_() -----								
share	lh	rh	t	c	t	c	trim	t	c	p_t	hs2
0.5	1	0	3	28	3	14	9.7E-01	1.9E-01	4.4E-01	5.4E-01	
0.0	0	0	13	36	11	14	3.0E-02	8.1E-01	5.6E-01	4.6E-01	

Haplotype-wise analysis: number of feasible configurations for:

	Disease position	Ancestral loci	Ancestral haps
Scan to date:	3	5	18
Current disease position:		1	3
Current ancestral loci:			3

marker loci (X=Disease, "="=fixed)

1 2 3 4

Ancestral: -----X---

Loci 1

Hap 7

Allele freqs .23

(cM) Max # Trans to

(cM) Min Span span affecteds: # Gens

Disease allele of anc of anc (min) (max) since anc

position freq loci loci test control mutation alpha

gamma

1.25000E-01 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00

2.00E-01

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

```

-----permutation replicates-----
--admissible--      ---locally-admiss--- -----t+c-----
somewhere local test control t+c min mean max min mean max
  80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

```

S\_reg: analytical significance of t-value = 1.0737E-01  
 permutation significance level = 1.0700E-01  
 regression coefficient = 2.3609E-01  
 std error of coefficient = 1.9030E-01  
 regression constant = 1.6056E-01  
 std error of constant = 4.5133E-02  
 t-value for regression coefficient = 1.2407E+00

S\_est: permutation significance level = 7.9000E-02  
 anc hap stat: raw, normed = 1.6338E+00, 1.2633E+00  
 replicate samples: mean, std dev = 4.9906E-01, 8.9823E-01

S\_trim: permutation significance level = 4.4000E-02  
 anc hap stat: raw, normed = 1.5341E+00, 1.8426E+00  
 replicate samples: mean, std dev = -5.5082E-01, 1.1315E+00

```

ovlap --#haps-- --#peds-- ----- prob_() -----
share lh rh t c t c trim t c p_t_hs2
-----
0.5 1 0 6 12 6 9 9.6E-01 3.8E-01 1.9E-01 3.4E-01
0.0 0 0 10 52 9 18 3.7E-02 6.2E-01 8.1E-01 6.6E-01

```

Haplotype-wise analysis: number of feasible configurations for:

```

Disease position Ancestral loci Ancestral haps
Scan to date:      3          6          19
Current disease position:      2          4
Current ancestral loci:      1
marker loci (X=Disease, "="=fixed)
  1 2 3 4
Ancestral: -----X---
Loci      1
Hap      5
Allele freqs .13
(cM) Max # Trans to
(cM) Min Span span affecteds: # Gens
Disease allele of anc of anc (min) (max) since anc
position freq loci loci test control mutation alpha
gamma
1.25000E-01 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00
2.00E-01

```

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

```

-----permutation replicates-----
--admissible--      ---locally-admiss--- -----t+c-----
somewhere local test control t+c min mean max min mean max
  80 80 16 64 80 80 8.00E+01 80 80 8.00E+01 80

```



S\_reg: analytical significance of t-value = 4.2864E-01  
 permutation significance level = 4.2700E-01  
 regression coefficient = 4.3627E-02  
 std error of coefficient = 2.4259E-01  
 regression constant = 1.8439E-01  
 std error of constant = 4.2884E-02  
 t-value for regression coefficient = 1.7984E-01

S\_est: permutation significance level = 9.6100E-01  
 anc hap stat: raw, normed = 0.0000E+00, -6.1107E-01  
 replicate samples: mean, std dev = 5.3115E-01, 8.6922E-01

S\_trim: permutation significance level = 5.7200E-01  
 anc hap stat: raw, normed = -1.4634E+00, -5.1441E-01  
 replicate samples: mean, std dev = -7.6697E-01, 1.3539E+00

share	lh	rh	t	c	t	c	trim	t	c	p_t	hs2
0.5	0	1	2	8	2	4	9.6E-01	1.2E-01	1.2E-01	2.9E-01	
0.0	0	0	14	56	12	18	4.2E-02	8.8E-01	8.8E-01	7.1E-01	

Haplotype-wise analysis: number of feasible configurations for:

	Disease position	Ancestral loci	Ancestral haps	
Scan to date:	3	6	20	
Current disease position:		2	5	
Current ancestral loci:			2	
marker loci (X=Disease, "="=fixed)				
	1	2	3	4
Ancestral:	-----X---			
Loci	1			
Hap	7			
Allele freqs	.30			
	(cM)	Max	# Trans to	
	(cM)	Min	Span	span affecteds: # Gens
Disease position	allele freq	of anc loci	of anc loci	(min) (max) since anc mutation alpha
gamma	1.2500E-01	1.00E-01	0.00E+00	8.00E+00 2 1 200 1.00E+00
	2.00E-01			

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

	-----permutation replicates-----									
--admissible--	---locally-admiss---					-----t+c-----				
somewhere	local	test	control	t+c	min	mean	max	min	mean	max
80	80	16	64	80	80	8.00E+01	80	80	8.00E+01	80

S\_reg: analytical significance of t-value = 4.0537E-01  
 permutation significance level = 3.9000E-01  
 regression coefficient = 4.1011E-02  
 std error of coefficient = 1.7126E-01

regression constant = 1.8046E-01  
 std error of constant = 4.8817E-02  
 t-value for regression coefficient = 2.3946E-01

S\_est: permutation significance level = 3.8000E-01  
 anc hap stat: raw, normed = 2.4868E-01, -2.3926E-01  
 replicate samples: mean, std dev = 4.3217E-01, 7.6691E-01

S\_trim: permutation significance level = 2.0400E-01  
 anc hap stat: raw, normed = 2.8065E-01, 7.3042E-01  
 replicate samples: mean, std dev = -3.7764E-01, 9.0125E-01

share	lh	rh	t	c	t	c	trim	t	c	p_t	hs2	prob_()
0.5	0	1	6	20	6	12	9.7E-01	3.8E-01	3.1E-01	4.4E-01		
0.0	0	0	10	44	8	18	3.4E-02	6.2E-01	6.9E-01	5.6E-01		

Haplotype-wise analysis: number of feasible configurations for:

	Disease position	Ancestral loci	Ancestral haps
Scan to date:	3	6	21
Current disease position:		2	6
Current ancestral loci:			3

marker loci (X=Disease, "="=fixed)  
 1 2 3 4  
 Ancestral: -----X---  
 Loci 1  
 Hap 8  
 Allele freqs .22  
 (cM) Max # Trans to  
 (cM) Min Span span affecteds: # Gens  
 Disease allele of anc of anc (min) (max) since anc  
 position freq loci loci test control mutation alpha  
 gamma  
 1.25000E-01 1.00E-01 0.00E+00 8.00E+00 2 1 200 1.00E+00  
 2.00E-01

Number of replicates (feas/ total) = 1000 / 1000

Founder haplotypes:

	-----permutation replicates-----									
--admissible--	---locally-admiss---				-----t+c-----					
somewhere	local	test	control	t+c	min	mean	max	min	mean	max
80	80	16	64	80	80	8.00E+01	80	80	8.00E+01	80

S\_reg: analytical significance of t-value = 5.8476E-01  
 permutation significance level = 5.5700E-01  
 regression coefficient = -4.1126E-02  
 std error of coefficient = 1.9211E-01  
 regression constant = 1.9175E-01  
 std error of constant = 4.5562E-02  
 t-value for regression coefficient = -2.1408E-01

S\_est: permutation significance level = 4.8700E-01

anc hap stat: raw, normed = 1.5238E-01, -4.1297E-01  
replicate samples: mean, std dev = 4.9510E-01, 8.2989E-01

S\_trim: permutation significance level = 7.8200E-01  
anc hap stat: raw, normed = -1.1965E+00, -5.8811E-01  
replicate samples: mean, std dev = -5.3554E-01, 1.1239E+00

share	lh	rh	t	c	t	c	trim	t	c	p_t_hs2
0.5	0	1	2	16	2	10	9.6E-01	1.2E-01	2.5E-01	3.9E-01
0.0	0	0	14	48	12	17	3.8E-02	8.8E-01	7.5E-01	6.1E-01

Summary for HAL scan-wise statistics:

S\_reg: permutation significance level = 8.4600E-01  
minimum significance of t-value = 1.0737E-01  
S\_est: permutation significance level = 6.1900E-01  
maximum S\_est value = 1.6824E+00  
S\_trim: permutation significance level = 4.8900E-01  
maximum S\_trim value = 1.8426E+00

Number of replicates (feasible / total) = 1000 / 1000  
# Created on Fri Mar 24 15:21:12 2000

## **Appendix 6**

### **Result file of DISLAMB program**

\*\*\*\*\* marker 1 (ignore 1 alleles)

Disease allele frequency = 0.00100000

---



---

CASE		2		2		5		5		2		0		2		6		3	
CONTROL		7		5		15		5		3		4		2		5		5	

---



---

Estimated parameters for likelihood ratio test:  
Allele frequencies:

Allele	H0:	H1
1	0.11538462	0.11538462
2	0.08974359	0.08974359
3	0.25641026	0.25641026
4	0.12820513	0.12820513
5	0.06410256	0.06410256
6	0.05128205	0.05128205
7	0.05128205	0.05128205
8	0.14102564	0.14102564
9	0.10256410	0.10256410
Lambda	0.00000000	0.000000

LRT Chi-Square = 0.00000 p-value = 0.499999810123862 Lambda = 0.000000

NO SIGNIFICANT EVIDENCE OF LINKAGE DISEQUILIBRIUM BY LRT TEST

2 x n table Chi-square = 7.14636 P-value = 0.520926633189796

NO SIGNIFICANT EVIDENCE OF LINKAGE DISEQUILIBRIUM BY 2 x N TABLE CHI-SQUARE TEST

\*\*\*\*\* marker 2 (ignore 2 alleles)

Disease allele frequency = 0.00100000

---



---

CASE		0		4		8		4		5		5	
CONTROL		3		14		16		3		7		7	

---



---

Estimated parameters for likelihood ratio test:  
Allele frequencies:

Allele	H0:	H1
1	0.03947368	0.03947368
2	0.23684211	0.23684211
3	0.31578947	0.31578947
4	0.09210526	0.09210526
5	0.15789474	0.15789474
6	0.15789474	0.15789474
Lambda	0.00000000	0.000000

LRT Chi-Square = 0.00000 p-value = 0.5000000000000000 Lambda = 0.000000

NO SIGNIFICANT EVIDENCE OF LINKAGE DISEQUILIBRIUM BY LRT TEST

2 x n table Chi-square = 4.94603 P-value = 0.422523771730510

NO SIGNIFICANT EVIDENCE OF LINKAGE DISEQUILIBRIUM BY 2 x N TABLE CHI-SQUARE TEST

\*\*\*\*\* marker 3 (ignore 2 alleles)

```
=====
CASE   | 6| 10| 3| 1| 7|
CONTROL | 10| 20| 2| 7| 11|
=====
```

Estimated parameters for likelihood ratio test:  
Allele frequencies:

Allele	H0:	H1
1	0.20779221	0.20779221
2	0.38961039	0.38961039
3	0.06493506	0.06493507
4	0.10389610	0.10389611
5	0.23376623	0.23376623
Lambda	0.00000000	0.000000

LRT Chi-Square = 0.00000 p-value = 0.5000000000000000 Lambda = 0.000000

NO SIGNIFICANT EVIDENCE OF LINKAGE DISEQUILIBRIUM BY LRT TEST

2 x n table Chi-square = 3.35108 P-value = 0.500879294159147

NO SIGNIFICANT EVIDENCE OF LINKAGE DISEQUILIBRIUM BY 2 x N TABLE CHI-SQUARE TEST

\*\*\*\*\* marker 4

```
=====
CASE   | 0| 2| 1| 1| 4| 1| 11| 5| 2|
CONTROL | 3| 5| 2| 2| 6| 6| 13| 12| 3|
=====
```

Estimated parameters for likelihood ratio test:  
Allele frequencies:

Allele	H0:	H1
1	0.03797468	0.03797468
2	0.08860759	0.08860760
3	0.03797468	0.03797468
4	0.03797468	0.03797468
5	0.12658228	0.12658228
6	0.08860759	0.08860760
7	0.30379747	0.30379746
8	0.21518987	0.21518987
9	0.06329114	0.06329114
Lambda	0.00000000	0.000000

LRT Chi-Square = 0.00000 p-value = 0.499999810123862 Lambda = 0.000000

NO SIGNIFICANT EVIDENCE OF LINKAGE DISEQUILIBRIUM BY LRT TEST

2 x n table Chi-square = 4.73569 P-value = 0.785416372092084

**NO SIGNIFICANT EVIDENCE OF LINKAGE DISEQUILIBRIUM BY 2 x N TABLE CHI-SQUARE TEST**

## **Appendix 7**

### **Result file of GASSOC program**



\*\*\*\*\*  
 \* gassoc Version 1.05  
 \*\*\*\*\*

---



---

ANALYSIS FOR MARKER: 2

---



---

Summary Info:

# of valid lines in input file: 83  
 # of affected cases: 43  
 # of affected cases used in analysis: 25  
 # of affected cases not used: 18  
 # not used due to missing parent or missing parent alleles: 18  
 # not used due to case missing alleles: 0  
 # not used due to inconsistent parent/case alleles: 0

Conditional Logistic:

Final estimates of Beta:

	Allele	Beta	exp(Beta)	SE(Beta)	Z	P(2-sided)
	1	16.2029	10884982.3001	5439.5270	0.0030	0.99762332
	2	0.2750	1.3165	0.6805	0.4040	0.68617994
	3	-1.2268	0.2932	1.1829	-1.0371	0.29968361
	5	-1.0064	0.3655	1.1892	-0.8463	0.39740465
	6	-16.6553	0.0000	2377.2499	-0.0070	0.99440997
	7	-1.4055	0.2452	1.5382	-0.9137	0.36086625
	8	-15.7446	0.0000	3746.0806	-0.0042	0.99664654
	9	1.0986	3.0000	1.1547	0.9514	0.34138810
10	***Unable to compute***					

LR Statistic: 15.3742, df=8, p=0.052265076

Covariance/Correlation Matrix (\*=Corr(Bi,Bj)):

29588454.3824	0.0000*	0.0000*	0.0000*
0.0000*	0.0000*	0.0000*	0.0000*
0.0000	0.4631	0.3224*	0.0000*
0.0000*	0.0000*	-0.0000*	-0.0000*
0.0000	0.2595	1.3992	0.0000*

0.0000*	0.0000*	-0.0000*	-0.0000*	NaN*
0.0000	0.0000	0.0000	1.4143	
0.0003*	0.5817*	-0.0000*	-0.0000*	NaN*
0.0000	0.0427	0.0239	0.9401	
5651317.3153	0.0002*	-0.0000*	-0.0000*	
NaN*				
0.0000	0.0000	0.0000	1.0641	
0.7073	2.3661	-0.0000*	-0.0000*	NaN*
0.0000	-0.0000	-0.0000	-0.0000	
-0.0000	-0.0000	14033120.1943	0.0001*	NaN*
0.0000	-0.0000	-0.0000	-0.0000	
-0.0000	-0.0000	0.3333	1.3333	NaN*
0.0000	-0.0000	-0.0000	-0.0000	
-0.0000	-0.0000	0.0000	0.0000	0.0000

Score Statistics:

	Score	df	P-value	Sim P-value(Simulations=2)
GTDT:	11.7064	8	0.164793530	0.000000000
GDOM:	12.5603	9	0.183530602	0.000000000
GREC:	8.5789	6	0.198678847	0.500000000

Note: Seeds used for random# generation were 942, 18223, 8374

---



---

ANALYSIS FOR MARKER: 3

---



---

Summary Info:

# of valid lines in input file: 83

# of affected cases: 43

# of affected cases used in analysis: 25

# of affected cases not used: 18

# not used due to missing parent or missing parent alleles: 18

# not used due to case missing alleles: 0

# not used due to inconsistent parent/case alleles: 0

Conditional Logistic:

Final estimates of Beta:

Rel. Risk

Allele	Beta	exp(Beta)	SE(Beta)	Z	P(2-sided)
1	15.2029	4004360.2052	2332.9149	0.0065	0.99480047
2	1.5569	4.7440	1.1581	1.3443	0.17884762
4	-0.6270	0.5342	1.2510	-0.5012	0.61622551
5	-16.3895	0.0000	3299.2405	-0.0050	0.99603639
6	0.0661	1.0684	1.7507	0.0378	0.96986349
7	-1.1866	0.3053	1.9256	-0.6162	0.53773424
8	-15.8773	0.0000	2252.2533	-0.0070	0.99437534

LR Statistic: 14.4397, df=7, p=0.043891853

Covariance/Correlation Matrix (\*=Corr(Bi,Bj)):

5442492.1500	0.0000*	0.0000*	0.0000*
0.0000*	0.0000*	0.0000*	
0.0000	1.3413	0.3476*	0.0001*
0.2484*	0.2258*	0.0002*	
0.0000	0.5036	1.5650	0.0004*
0.7146*	0.6497*	0.0006*	
0.0000	0.5036	1.5650	10884988.0060
0.0005*	0.0006*	0.0000*	
0.0000	0.5036	1.5650	3.0650
3.0650	0.9092*	0.0008*	
0.0000	0.5036	1.5650	3.7078
3.0650	3.7078	0.0008*	
0.0000	0.5036	1.5650	3.5650
3.0650	3.5650	5072644.8277	

Score Statistics:

	Score	df	P-value	Sim P-value(Simulations=2)
GTDT:	11.7341	7	0.109647704	0.000000000
GDOM:	10.8301	7	0.146202442	0.000000000
GREC:	7.6061	5	0.179324534	0.000000000

Note: Seeds used for random# generation were 942, 18223, 8374

---



---

#### ANALYSIS FOR MARKER: 4

---



---

Summary Info:

# of valid lines in input file: 83

# of affected cases: 43  
 # of affected cases used in analysis: 25  
 # of affected cases not used: 18  
 # not used due to missing parent or missing parent alleles: 18  
 # not used due to case missing alleles: 0  
 # not used due to inconsistent parent/case alleles: 0

Conditional Logistic:

Final estimates of Beta:

Allele	Beta	exp(Beta)	SE(Beta)	Z	P(2-sided)
1	-15.2029	0.0000	3299.2400	-0.0046	0.99632336
2	1.2634	3.5373	0.7105	1.7782	0.07537569
4	-14.0902	0.0000	1904.8172	-0.0074	0.99409799
5	15.2029	4004360.2052	3299.2400	0.0046	0.99632336
6	0.0000	1.0000	0.8165	0.0000	1.00000000
7	1.1127	3.0426	0.7672	1.4503	0.14697893

LR Statistic: 10.9601, df=6, p=0.089616297

Covariance/Correlation Matrix (\*=Corr(Bi,Bj)):

10884984.2982	0.0000*	0.0000*	0.0000*
-0.0000*	0.0000*		
0.0000	0.5048	0.0002*	0.0000*
-0.0000*	0.6033*		
0.0000	0.3289	3628328.6874	0.0000*
-0.0000*	0.0004*		
0.0000	0.0000	0.0000	10884984.3000
-0.0000*	0.0000*		
-0.0000	-0.0000	-0.0000	-0.0000
0.6667	-0.0000*		
0.0000	0.3289	0.5886	0.0000
-0.0000	0.5886		

Score Statistics:

	Score	df	P-value	Sim P-value(Simulations=2)
GTDT:	8.8168	6	0.184146637	0.500000000
GDOM:	7.7796	6	0.254700460	0.000000000
GREC:	4.9259	4	0.294988290	0.500000000

Note: Seeds used for random# generation were 942, 18223, 8374

---



---

**ANALYSIS FOR MARKER: 5**

---



---

**Summary Info:**

# of valid lines in input file: 83

# of affected cases: 43

# of affected cases used in analysis: 25

# of affected cases not used: 18

# not used due to missing parent or missing parent alleles: 18

# not used due to case missing alleles: 0

# not used due to inconsistent parent/case alleles: 0

**Conditional Logistic:**

**Final estimates of Beta:**

	Rel. Risk				
Allele	Beta	exp(Beta)	SE(Beta)	Z	P(2-sided)
1	0.4960	1.6421	1.7037	0.2911	0.77095833
2	15.4266	5008405.0965	1649.3944	0.0094	0.99253757
3	0.0000	1.0000	1.0000	0.0000	1.00000000
4	-15.2029	0.0000	2332.9149	-0.0065	0.99480047
5	0.4960	1.6421	0.9500	0.5221	0.60161178
6	0.1850	1.2032	0.9127	0.2027	0.83935270
8	0.2615	1.2989	0.6413	0.4078	0.68341738
9	0.0925	1.0969	1.4875	0.0622	0.95040920

LR Statistic: 8.6941, df=8, p=0.368750702

**Covariance/Correlation Matrix (\*=Corr(Bi,Bj)):**

	2.9025	0.0001*	0.0000*	0.0000*
0.5576*	0.1321*	0.1097*	0.0405*	
	0.1610	2720501.9341	0.0000*	0.0000*
0.0001*	0.0003*	0.0003*	0.0001*	
	0.0000	0.0000	1.0000	-0.0000*
0.0000*	0.0000*	0.0000*	0.0000*	
	0.0000	0.0000	-0.0000	5442492.1491
0.0000*	0.0000*	0.0000*	0.0000*	
	0.9025	0.1610	0.0000	0.0000
0.9025	0.2369*	0.1967*	0.0727*	
	0.2054	0.4899	0.0000	0.0000
0.2054	0.8330	0.2939*	0.3068*	

	0.1199	0.2962	0.0000	0.0000
0.1199	0.1720	0.4113	0.0902*	
	0.1027	0.2449	0.0000	0.0000
0.1027	0.4165	0.0860	2.2125	

**Score Statistics:**

	Score	df	P-value	Sim P-value(Simulations=2)
GTDT:	6.3729	8	0.605542445	0.500000000
GDOM:	6.3844	8	0.604251878	1.000000000
GREC:	5.9220	5	0.313884220	0.000000000

Note: Seeds used for random# generation were 942, 18223, 8374