REFERENCE ONLY

# UNIVERSITY OF LONDON THESIS

Degree MD Year 2005 Name of Author BANVALLE, D

**COPYRIGHT**
This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

**COPYRIGHT DECLARATION**
I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

**LOAN**
Theses may not be lent to individuals, but the University Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: The Theses Section, University of London Library, Senate House, Malet Street, London WC1E 7HU.

**REPRODUCTION**
University of London theses may not be reproduced without explicit written permission from the University of London Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

A. Before 1962. Permission granted only upon the prior written consent of the author. (The University Library will provide addresses where possible).

B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.

C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.

D. 1989 onwards. Most theses may be copied.

***This thesis comes within category D.***

☑ This copy has been deposited in the Library of UCL

☐ This copy has been deposited in the University of London Library, Senate House, Malet Street, London WC1E 7HU.

C:\Documents and Settings\lproctor.ULL\Local Settings\Temporary Internet Files\OLK36\Copyright - thesis.doc

*Thesis submitted for the degree of Doctor of Medicine (M.D.)*
*University of London*

*Dr. Diana Baralle B.Sc., M.B.B.S. (London), MRCP*

# Molecular Pathology of Neurofibromatosis type 1(NF1)

Supervisors: Professor Charles ffrench-Constant
Professor Robin Winter

2004

UMI Number: U591701

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

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

UMI U591701
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

*To my family*

## Acknowledgments

This work was carried out in the Molecular Genetics laboratory, Addenbrookes Hospital and the Department of Pathology at the University of Cambridge under the supervision of Professor C. ffrench-Constant and Professor R. Winter. I am deeply indebted to Prof. ffrench-Constant for accepting me as a M.D. student in his laboratory and for his excellent guidance and support throughout the project.

Special thanks goes to my collaborators at the Institute of Genetic Engineering and Biotechnology, my father Professor. F.E. Baralle and brother Dr M. Baralle for discussion and assistance throughout this time. Also to Dr. J. Whittaker for her support during the time I spent at the Molecular Genetics lab.

I am grateful to all my colleagues and members of both laboratories for their friendship, support and co-operation, it has been a pleasure to work with them.

The financial assistance of Action Medical Research is gratefully acknowledged.

Last, but not the least, thank you to my family; my parents, children and especially Matthew for their support and patience.

## Contents

**Page**

**Abbreviations** 13

**Abstract** 15

**Introduction**

*1* *Neurofibromatosis type 1* 16

1.1 Clinical features of Neurofibromatosis type 1 (NF1) 16

1.2 Neurofibromatosis type 1 gene and its product 22

1.2.1 Neurofibromin function 23

1.2.2 Expression of neurofibromin 27

1.2.3 Loss of heterozygosity (LOH) for the *NF1* gene 29

1.2.4 RNA processing and clinical variability in NF1 30

1.2.4.1 Alternative transcript expression 31

1.2.4.2 Unequal allelic expression 32

1.2.4.3 *NF1* RNA editing 32

1.3 Mutation analysis and genotype-phenotype correlations 34

1.3.1 Noonan-neurofibromatosis syndrome (NFNS) 36

2 *RNA Splicing* 38

2.0.1 Mutations affecting splicing 38

2.0.2 Fundamentals of RNA splicing 41

2.1 The spliceosome 42

2.1.1 The splicing reaction 42

2.1.2 Sequence elements (cis) on pre-mRNA required for splicing 44

| | | |
|---|---|---|
| 2.2 | Protein components involved in splicing | 45 |
| 2.2.1 | Small Nuclear Ribonucleoprotein Particles (snRNPs) | 45 |
| 2.2.2 | Non-snRNP splicing factors | 47 |
| 2.3 | The spliceosomal Complex | 49 |
| 2.3.1 | General model for spliceosome assembly and spliceosome cycle | 49 |
| 2.3.2 | Early complex formation: Intron and exon definition models | 52 |
| 2.3.3 | Splice site recognition and bridging interactions | 57 |
| 2.4 | Alternative splicing | 58 |
| 2.4.1 | Involvement of cis-acting elements in splice site selection | 61 |
| 2.4.2 | Trans acting factors | 66 |
| **3** | ***Set-up of project*** | 68 |
| **Materials and Methods** | | 72 |
| **4.1** | **Materials** | 72 |
| 4.1.1 | Chemical reagents | 72 |
| 4.1.2 | Enzymes | 72 |
| 4.1.3 | Synthetic oligonucleotides | 72 |
| 4.1.4 | Plasmid constructs | 75 |
| 4.1.4.1 | Commercial plasmids used | 75 |
| 4.1.4.2 | Minigene system construct | 75 |
| 4.1.5 | Bacterial culture | 75 |
| 4.1.6 | Cell culture | 76 |

| | | |
|---|---|---|
| **4.2** | **Methods** | 77 |
| 4.2.1 | Nucleic acids preparation | 77 |
| 4.2.1.1 | Small scale preparation of plasmid DNA from bacterial cultures | 77 |
| 4.2.1.2 | Large scale preparation of plasmid DNA from bacterial cultures | 77 |
| 4.2.1.3 | Preparation of RNA from the cultured cells | 78 |
| 4.2.1.4 | cDNA synthesis from RNA | 78 |
| 4.2.1.5 | Agarose gel electrophoresis | 79 |
| 4.2.2 | Estimation of nucleic acid concentration | 79 |
| 4.2.2.1 | Spectrophotometric | 79 |
| 4.2.3 | Enzymatic modification of DNA | 80 |
| 4.2.3.1 | Restriction enzymes | 80 |
| 4.2.3.2 | Large fragment of E. Coli polymerase I | 80 |
| 4.2.3.3 | Dephosphorylation of DNA 5' termini | 81 |
| 4.2.3.4 | T4 DNA ligase | 81 |
| 4.2.4 | Elution and purification of DNA fragment from agarose gels | 82 |
| 4.2.5 | PCR reactions | 82 |
| 4.2.5.1 | PCR reactions for cloning purposes | 82 |
| 4.2.5.2 | PCR reactions for mutation analysis | 83 |
| 4.2.6 | Sequence analysis | 85 |
| 4.2.6.1 | Sequence analysis for plasmid DNA | 85 |
| 4.2.6.2 | Sequencing for *NF1* mutation analysis | 85 |
| 4.2.6.3 | Sequence analysis | 86 |
| 4.2.7 | Site-directed mutagenesis | 87 |

4.2.8 Pull-down analysis procedure of affinity purification of RNA bound proteins 87

4.2.8.1 Preparation of RNA templates for pull-down analysis 87

4.2.8.2 Purification of RNA-bound protein complexes 88

4.2.8.3 Nanoelectronspray mass spectrometry and database homology search 88

4.2.9 Maintenance and analysis of cells in culture 89

4.2.10 Transfection of recombinant DNA into cells maintained in culture 90

4.2.11 Measurement of transfection efficiency 91

4.3 Clinical methods and recruitment of patients 91

4.4 Fundamentals of the functional minigene splicing assay 92

4.5 Amplification of exon 3 in mutant and WT DNA and cloning into pUC19 95

4.6 Amplification of exon 37 mutant and WT DNA and cloning into pUC19 97

**Results** 99

**Part 1*: Mutation screening of the NF1 gene*** 99

5.1 Sequence analysis of *NF1* 99

5.1.2 Mutation spectrum 104

5.1.3 Recurrent mutations 106

5.1.4 Polymorphisms 110

**Part 2:** ***Mutation screening in Noonan-Neurofibromatosis (NFNS)*** 112

5.2 Noonan-neurofibromatosis (NFNS) 112

**Part 3:** ***Effect of different mutations in NF1 on splicing efficiency*** 117

5.3 Fundamentals of the functional minigene splicing assay 117

5.4 *NF1* intronic mutation exon 3+5 G>C 117

5.4.1 The intronic mutation 3+5 G>C is a splicing mutation 118

5.4.2 The intronic mutation 3+5 G>C excludes exon 3 from the mature transcript by failure of snRNP U1 to recognise the 5' splice site 120

5.4.3 Involvement of a context-specific hnRNP H1 binding site hampering its interaction with U1 snRNA 122

5.5 *NF1* 'nonsense' mutation exon 37 C6782G 127

5.5.1 Mutation exon 37 C6792G causes skipping of this exon in the mature transcript 127

5.5.2 The 'same sense' C6792T mutation also causes skipping of exon 37 130

5.5.3 Skipping of exon 37 is caused by interruption of an exonic splice enhancer? 130

5.6 *NF1* 'silent' mutation exon 28 5172 G>A 136

5.6.1 Exon 28 G5172A does not cause skipping of the exon 136

5.7 *NF1* 'silent' mutation exon 6 846 G>A in a family with familial primitive neuroectodermal tumours (PNETs) 138

5.7.1 Exon 6 846 G>A sequence variant does not cause skipping of this exon 138

5.7.2 *PMS2* gene mutation as a possible cause of a NF1-like phenotype with a highly malignant phenotype and high reccurrence risk 140

**Discussion** 142

**Part 1:** ***Mutation screening of the NF1 gene*** 142

6.1 Automated comparative sequence analysis identifies mutations in 89% of NF1 patients and confirms a mutation cluster in exons 11-17 142

**Part 2:** ***Different mutations in the NF1 gene are associated with NFNS*** 144

**Part 3:** ***Effect of different mutations in NF1 on splicing efficiency*** 147

7.1 The splicing assay and diagnostic usefulness 147

7.2 An intronic mutation affects exon 3 splicing with the involvement of a context-specific hnRNP H1 binding site hampering its interaction with U1 snRNA 148

7.3 A 'nonsense' mutation that causes skipping of exon 37 in NF1 and definition of an exonic splice enhancer 152

7.4 Silent mutations that did not affect splicing and a new form of recessive cancer sydnrome 153

**References** 155

## Figures List                                                                                    Page

### Introduction


| | | |
|---|---|---|
| Table 1 | NIH NF1 Diagnostic criteria | 17 |
| Figure 1 | Café-au-lait macules | 19 |
| Figure 2 | Multiple neurofibromas | 19 |
| Figure 3 | Neurofibromin accelerates inactivation of active GTP bound Ras | 25 |
| Table 2 | Missense and silent mutations associated with altered splicing | 40 |
| Figure 4 | Exon splicing and lariat formation | 43 |
| Figure 5 | Spliceosome assembly pathway | 50 |
| Figure 6 | Intron and exon definition models | 53 |
| Figure 7 | Types of alternative splicing events | 59 |
| Figure 8 | Example of mutation detected by CSA | 69 |
| Figure 9 | Frameshift mutations in CSA | 71 |


### Materials and Methods


| | | |
|---|---|---|
| Table 3 | Primers and conditions for NF1 PCRs | 84 |
| Figure 10 | PTB minigene | 93 |
| Figure 11 | pcDNA3 minigene | 93 |
| Figure 12 | The assay | 94 |
| Figure 13 | Cloning strategy for NF1 exon 3 in PTB and steps in minigene | 96 |
| Figure 14 | Cloning strategy for exon 37 in pcDNA3 and the minigene assay | 98 |

**Results**


| | | |
|---|---|---|
| Table 4 | Summary of NF1 mutations and sequence variations found | 101-103 |
| Figure 15 | Histogram showing site of mutations found in NF1 | 109 |
| Figure 16 | Weighted distribution of missense and single base changes in the NF1 gene | 109 |
| Table 5 | Polymorphisms | 111 |
| Table 6 | NFNS Clinical features | 113 |
| Figure 17 | NFNS Clinical photographs | 114 |
| Figure 18 | Exon 3 amplification and schematic representation of amplicon | 119 |
| Figure 19 | RNA products generated by splicing assay | 121 |
| Figure 20 | Pull-down analysis studies | 124 |
| Figure 21 | *in-vitro* rescue of aberrant splicing | 126 |
| Figure 22 | A) Schematic representation of pcDNA3 with NF1 exon 34-38 as insert B)RT-PCR analysis C) Diagram of the process of aberrant splicing in the presence of C6792G mutation | 129 |
| Figure 23 | Diagram of secondary structure of RNA including the ESE sequence hypothesis | 132 |
| Figure 24 | RT-PCR analysis results for A6791T, A6793T, A6794T | 133 |
| Figure 25 | Minigene analysis of all possible mutations between bases 6788-6799 | 135 |

Figure 26 Results for exon 28 G>A sequence variation 137

Figure 27 Results for exon 6 G>A sequence variation 137

**Discussion**

Figure 28 Model to explain observation of restricted availability of of U1 snRNA base pairing sequences with the 5'SS due to hnRNPH steric hindrance 151

**Abbreviations**

The standard abbreviations used in this dissertation follow IUPAC rules. All the abbreviations are defined also in the text when they are introduced for the first time. The abbreviations mentioned only once are not included in this list.

| | |
|---|---|
| bp | Base pairs |
| nt | Nucleotides |
| aa | Amino acid |
| kb | Kilobase |
| kDa | Kilodalton |
| dNTPs | Deoxynucleoside triphosphate (A, C, G and T) |
| N | Nucleotide (A or C or G or T) |
| DNA | Deoxyribonucleic acid |
| cDNA | Copy DNA |
| ESE | Exonic Splicing Enhancer |
| ISE | Intronic Splicing Enhancer |
| ESS | Exonic Splicing Silencer |
| ISS | Intronic Splicing Silencer |
| snRNA | Small nuclear RNA |
| snRNP | Small nuclear ribonucleoprotein particles |
| hnRNP | Heterogenous ribonuclear protein |
| SR | Arginine-serine rich protein |
| ss | Splice site |
| NMD | Nonsense-mediated decay |
| NAS | Nonsense mutation-altered splicing |
| DTT | Dithiothreitol |
| EDTA | Ethylenediamine tetra-acetic acid |
| TBE | Tris-borate-EDTA (buffer) |
| PBS | Phosphate buffer saline |
| $ddH_20$ | Double-distilled water |
| Pu | Purine |
| ddNTP | Dideoxyriboncleoside triphosphates |
| kDa | Kilodalton |
| RNA | Ribonucleic acid |
| Tris | Tris (hydroxyethyl) amino-ethane |
| NF1 | Neurofibromatosis type 1 |
| NFNS | Noonan-neurofibromatosis |
| LOH | Loss of heterozygosity |
| PNET | Primitive neuroectodermal tumour |
| CSA | Comparative sequence analysis |
| AC | Adenylyl cyclase |

| | |
|---|---|
| NIH | National institutes of Health |
| CNS | Central nervous system |
| GAP | GTPase activating protein |
| GRD | GAP-related domain |
| GTP | guanosine triphosphate |
| PKA | Protein kinase A |
| MPNST | Malignant peripheral nerve sheath tumour |
| HDA | hetroduplex analysis |
| SSCP | Single strand conformation polymorphism |
| TGGE | temperature gradient gel electrophoresis |
| DGGE | denaturing gradient gel electrophoresis |
| PTT | protein truncation test |
| FISH | fluorescent in situ hybridisation |
| ORF | open reading frame |
| PCR | polymerase chain reaction |
| CAL | café-au-lait |
| PTC | premature termination codon |
| WT | wild type |

**Abstract**

Neurofibromatosis type 1 (NF1) is a common autosomal dominant genetic disorder caused by mutations in the *NF1* gene. Mutation detection in *NF1* has been a major challenge due to the large size of the gene and lack of mutational hotspots. This study reports mutation screening of 91 subjects fulfilling NIH NF1 diagnostic criteria in which a mutation detection rate of 89% was achieved using automated comparative sequence analysis (ACSA) and many novel mutations are reported. This detection rate makes this single technique appropriate for routine clinical practice. The data confirms that mutations are evenly distributed along the coding sequence of the *NF1* gene and the presence of a second putative functional domain upstream of the GRD at exons 11-17. A related disorder is also studied: Neurofibromatosis-Noonan syndrome and found to be a subtype of NF1 with mutations in the *NF1* gene.

Abnormalities of pre-mRNA splicing represent an important mechanism by which gene mutations cause disease. Effects on splicing can be predicted from genomic DNA sequence analysis if mutations alter highly conserved canonical splicing signals. However, it is extremely difficult to predict the effects of changes in intronic and exonic sequences not obviously involved in the splicing process. The significance of point mutations, including missense and silent changes are difficult to clarify. This study presents an efficient and simple test using genomic DNA to construct a minigene and analyse the effect on splicing of sequence variations in *NF1*. In particular two mutations are described: an intronic mutation that perturbs NF1 gene splicing and can be rescued by coexpression of an altered U1 snRNA that restores normal base pairing, and a nonsense mutation that interrupts an exonic splice enhancer.

# INTRODUCTION

## 1. Neurofibromatosis type 1

### *1.1 Clinical Features of Neurofibromatosis type 1 (NF1)*

Neurofibromatosis type 1 (NF1), formerly known as Von Recklinghausen Neurofibromatosis is a common genetic disorder affecting approximately 1 per 3000-5000 people. It is a fully penetrant autosomal dominant disorder and is defined by strict diagnostic criteria, as established by the National Institute of Health (NIH) Consensus Conference in 1987 (NIH, 1987). However, clinical descriptions of possible NF1 patients date from as early as the $13^{th}$ century (Mulvihill, 1988). There is wide variability of clinical manifestations, many causing considerable morbidity and even mortality. Approximately half of all NF1 patients represent *de novo* cases (Huson *et al.*, 1989).

The diagnostic criteria for NF1 formulated by the *National Institutes of Health Consensus Development Conference Statement Neurofibromatosis* include the following major disease features which are predominantly of neural crest origin: *café-au-lait* spots, freckling in the axillary or inguinal regions, neurofibromas, Lisch nodules, optic pathway glioma, distinctive osseous lesions (sphenoid dysplasia, thinning of long bone cortex with or without bowing and pseudarthrosis) and a positive family history of NF1. Two or more criteria are necessary for a diagnosis of NF1- Table 1. Most major disease features appear with increasing age, nevertheless, diagnosis should be possible by the age of 5 years when the penetrance of the disease is considered virtually 100% (Huson *et al* 1989).

**Table 1**

*Diagnostic criteria for Neurofibromatosis type 1 (NF1)*

2 or more criteria are necessary for a diagnosis of NF1

- 6 or more *café-au-lait* spots over 5mm in greatest diameter in prepubertal individuals and over 15mm in greatest diameter in postpubertal individuals
- two or more neurofibromas of any type or one or more plexiform neurofibromas
- freckling in the axillary or inguinal region
- two or more Lisch nodules (iris hamartomas)
- optic or chiasma glioma
- a distinctive osseus lesion, such as sphenoid dysplasia or thinning of long bone cortex, with or without bowing or pseudarthrosis
- a first degree relative with NF1 according to the above criteria

*Café-au-lait* spots are generally the first presenting sign of NF1 and are observed in 82% of NF1 affected children before the end of the first year (Huson *et al.*, 1989). In adults *Café-au-lait* spots tend to become less pronounced and may even disappear (**Figure 1**) (Huson *et al.*, 1988; Riccardi, 1992). *Skin fold freckling* of the axillae or groin is often the next presenting major disease feature and is observed in 81% of children before 6 years of age (Obringer *et al.*, 1989). Skin fold freckling may also be encountered in the nape of the neck, under the chin and in the female submammary region (Gutmann *et al.*, 1997).

*Neurofibromas* are divided into two major types: dermal and plexiform neurofibromas. Both types consist of a mixture of Schwann cells, perineural fibroblasts, endothelial cells and mast cells. Dermal neurofibromas are cutaneous or subcutaneous tumours, which originate from terminal nerve branches in the skin- **Figure 2**. They start to appear prior to puberty and mainly develop on the trunk. They are rarely painful but may cause pruritus and can become a major cosmetic burden (Huson, *et al* 1989). Plexiform neurofibromas originate from major nerve plexuses and major peripheral nerves, and as a consequence may cause severe complications. They are present in approximately one third of the NF1 population and are congenital, although they may present later due to growth (Riccardi, 1992). The overlying skin may be abnormal with either signs of hypertrophy, hyperpigmentation or hypertrichosis. Two types of plexiform neurofibromas have been described (Huson, 1994). Diffuse plexiform neurofibromas are soft subcutaneous swellings with ill-defined margins. Nodular plexiform neurofibromas are ovoid or spherically shaped, feel firm and are well circumscribed.

*Lisch nodules* are small pigmented hamartomas of the iris. They are present in 92% of NF1 patients older than 6 years of age (Huson *et al* 1989). As a consequence these iris hamartomas are of great diagnostic importance for NF1.

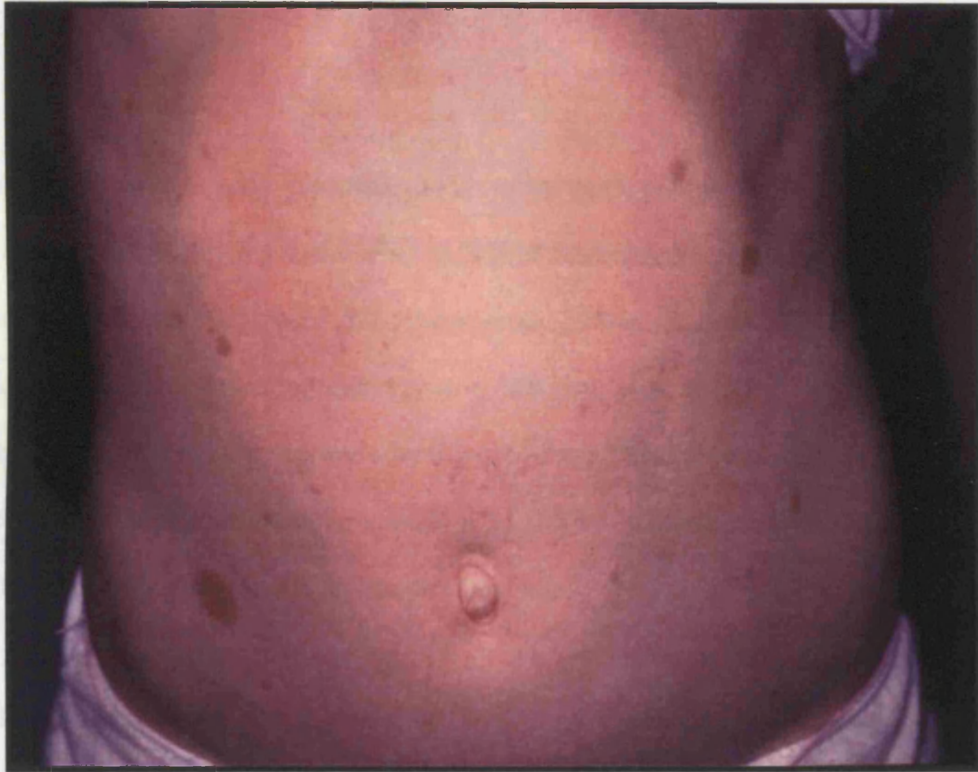


**Figure 1** Café-au-lait macules

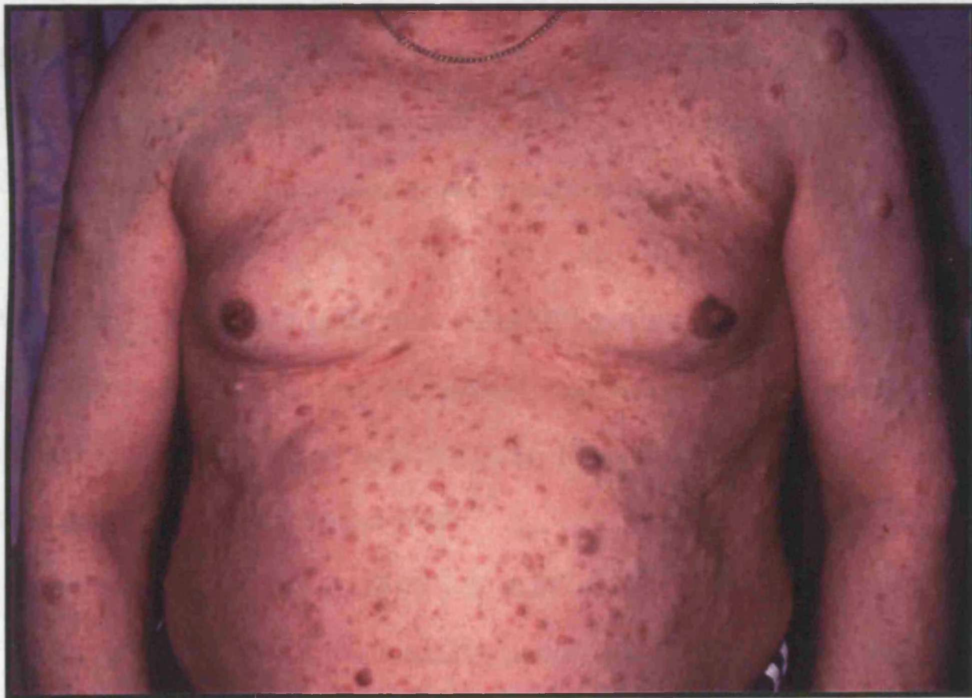


**Figure 2** Multiple neurofibromas

*Optic pathway gliomas* or pilocytic astrocytomas of the optic pathway are an important complication of NF1 in childhood occurring in 19% of cases (median age of presentation of symptomatic tumours: 4.9 years) (Listernick *et al.*, 1994). Generally optic gliomas associated with NF1 are less progressive than in children without NF1. Symptomatic optic pathway gliomas seldomly appear after 6 years of age. Routine screening with MRI imaging of 176 children showed optic glioma in 33 (19%). However only half of these children developed glioma-associated signs or symptoms.

*Osseous lesions* which are particular features of NF1 are sphenoid wing dysplasia and congenital bowing or thinning of long cortical bones with or without pseudarthrosis. Sphenoid wing dysplasia is rare in NF1 patients but occasionally presents with pulsatile exophthalmus. Congenital bowing generally occurs anterolaterally and is located in the tibia or fibula in 3% of NF1 patients (Huson, *et al* 1989). However, NF1 has been implicated in more than half of cases of congenital pseudarthrosis of the tibial bone.

The following signs are also frequently observed in NF1 patients: macrocephaly, short stature, hypertelorism, thorax abnormalities, scoliosis, precocious puberty, hypertension and various learning-, speech-, motor- and behavioural difficulties. These are considered minor features of the disease (Huson, *et al* 1989).
Learning disabilities are observed in 30-45% of NF1 patients (North *et al.*, 1995). Cognitive problems include lower full scale IQ, multidimensional cognitive deficits, reading disabilities and neuromotor deficits. Speech and behavioural problems are frequently seen but are not specific. A less common severe complication of NF1 is mental

retardation with prevalence between 4.8% and 8% (North *et al.*, 1994). The aetiology of cognitive impairment in NF1 has not been established. An early pathological study addressed its anatomical basis and postulated that neurocognitive deficit is due to migrational abnormalities in the developing brain (Rosman and Pearce, 1967). Hyperintense lesions on T2-weighted brain MRI occur in about 60% of children with NF1, but they tend to disappear in adulthood and are seldom observed in patients over 30 years of age. The lesions do not cause overt neurological symptoms. The nature of these lesions is uncertain but it has been postulated that they may represent areas of dysplastic glial proliferation and aberrant myelination (DiMario *et al.*, 1993).

In addition there is a lifelong increased risk of some malignancies. Although some cancers, notably neurosarcomas, CNS tumours and phaeochromocytoma are known to occur with increased frequency in NF1 patients, it is unclear whether there is an increased risk of other cancers.

Other subtypes, or related disorders also exist, these include segmental (or mosaic) NF1, autosomal dominant café-au lait spots alone, Watson syndrome, Neurofibromatosis-Noonan syndrome (discussed in section 1.3.1), familial spinal neurofibromatosis, familial gastrointestinal neurofibromatosis (Carey, *et al* 1999). It is unclear whether these are disorders with mutations in the NF1 gene or distinct clinical entities.

In summary, complications of NF1 may occur in many different organs and may cause severe morbidity or mortality. The course of the disease is unpredictable with intra- and inter- familial variability. In addition there is a significant number of patients who have a clinical picture suspicious of NF1 but do not fulfil the diagnostic criteria (personal

experience). As a consequence, diagnosis can be difficult and the availability of accurate and rapid genetic testing would have a major impact on the management of these patients and families.

### 1.2 *Neurofibromatosis type 1 gene and its product*

The *NF1* gene was mapped to chromosome 17q11.2 and is thought to be a tumour suppressor gene because loss of heterozygosity is associated with the occurrence of benign and malignant tumours in neural crest tissues (Colman *et al.*, 1995; Legius *et al.*, 1992; Serra *et al.*, 1997) as well as myeloid malignancies (Shannon *et al.*, 1994). It spans a region of about 350kb of genomic DNA and contains 60 exons (Cawthon *et al.*, 1990; Viskochil *et al.*, 1993; Wallace *et al.*, 1990). It harbours at least three other embedded genes - *EV12A, EV12B and ONGP*, transcribed from the opposite strand of *NF1* intron 27b. The *NF1* gene transcribes several mRNAs in the size range 11-13kb- the three known alternatively spliced exons are 9a, 23a and 48a; these are expressed ubiquitously but more so in the neurons, oligodendrocytes and non-myelinating Schwann cells (Daston *et al.*, 1992). The most common transcript codes for a polypeptide of 2,818 amino acids called neurofibromin (Danglot *et al.*, 1995; Li *et al.*, 1995; Marchuk *et al.*, 1991). The protein is 280kD (DeClue *et al.*, 1991) and has been found in all tissues and cell lines examined and has been detected in human, rat and mouse tissues (Gutmann *et al.*, 1991).

### 1.2.1 Neurofibromin function

A 360 amino acid region of the predicted protein product, neurofibromin, shows homology to the GTPase activating (GAP) family of proteins in yeast (IRA1 and IRA2) and mammals (Buchberg *et al.*, 1990; Tanaka *et al.*, 1991; Xu *et al.*, 1990). Until recently this was the only known functional domain of the NF1 gene (GRD, GAP-related domain) and spans exons 20-27a (bases 3497-4661). It was found to interact with p21ras and GTP (Martin *et al.*, 1990) and is thought to downregulate the Ras pathway and prevent uncontrolled cell proliferation by converting active Ras GTP (guanosine triphosphatase) to inactive RasGDP. In keeping with this, RasGTP levels may be elevated in human NF1 peripheral nerve tumours in which neurofibromin is reduced or absent (Guha *et al.*, 1996). This region of the gene also represents a 'hot spot' for mutations consistent with an important functional role (Upadhyaya *et al.*, 1997). The primary sequence of the GRD shares homology with other known GAPs, especially with the more closely related NF1 GAPs found in yeasts (IRA1 and IRA2) and *Drosophila.* The most widely characterised signal transduction pathway in mammalian cells that directly involves Ras is the Raf-MAPK kinase pathway (Bernards, 1995; Marshall, 1995). Ras functions through its guanine nucleotide-binding activity and its intrinsic GTPase activity. Upon activation by guanine nucleotide exchange factors, such as GRB2-SOS complex, Ras will exchange GDP for GTP, stimulating Ras to undergo a structural conformational change upon GTP-binding and activating downstream effectors such as Raf-1 kinase, a serine/threonine switch which phosphorylates and activates its downstream target, MAPKK/ERK. These kinases will eventually activate gene expression in the nucleus to stimulate cellular

proliferation or differentiation events. GAPs act as upstream regulators of Ras by stimulating the intrinsic GTPase activity of Ras, which is otherwise a very slow process. Downstream effector functions for GAP have also been reported (Boguski and McCormick, 1993).

In addition, studies in *Drosophila melanogaster* and in mice have shown that neurofibromin modulates the activity of the rutabaga-encoded adenylyl cyclase (AC) (Guo *et al.*, 2000; Tong *et al.*, 2002), an enzyme crucial for cAMP signaling. Neurofibromin has also been shown to associate with microtubules (Xu and Gutmann, 1997), suggesting that it is involved in the regulation of multiple signalling pathways in the brain. All of these biochemical pathways have been implicated in synaptic plasticity and in learning and memory- **Figure 3**.

Studies in *D. melanogaster* (60% protein wide homology for neurofibromin) showed that homozygosity for mutation of NF1 causes associative learning deficits and that these deficits are dependent on AC. Furthermore, the cAMP pathway is crucial for memory formation in a variety of vertebrate and invertebrate species (Silva *et al.*, 1998), and AC activity is also impaired in cells from (lethal) homozygous mutant NF1 mice.

The human and mouse forms of neurofibromin are highly homologous (98% sequence similarity) (Bernards *et al.*, 1993), as are the promoter sequences of the gene, suggesting that both the biochemistry of the protein and the transcriptional regulation of the gene are conserved across species (Hajra *et al.*, 1994). Investigation of all the symptoms of NF1 in a single mouse has proven difficult. However the effects of NF1 mutations in humans and mice show interesting parallels. For example, in mice and probably in humans,

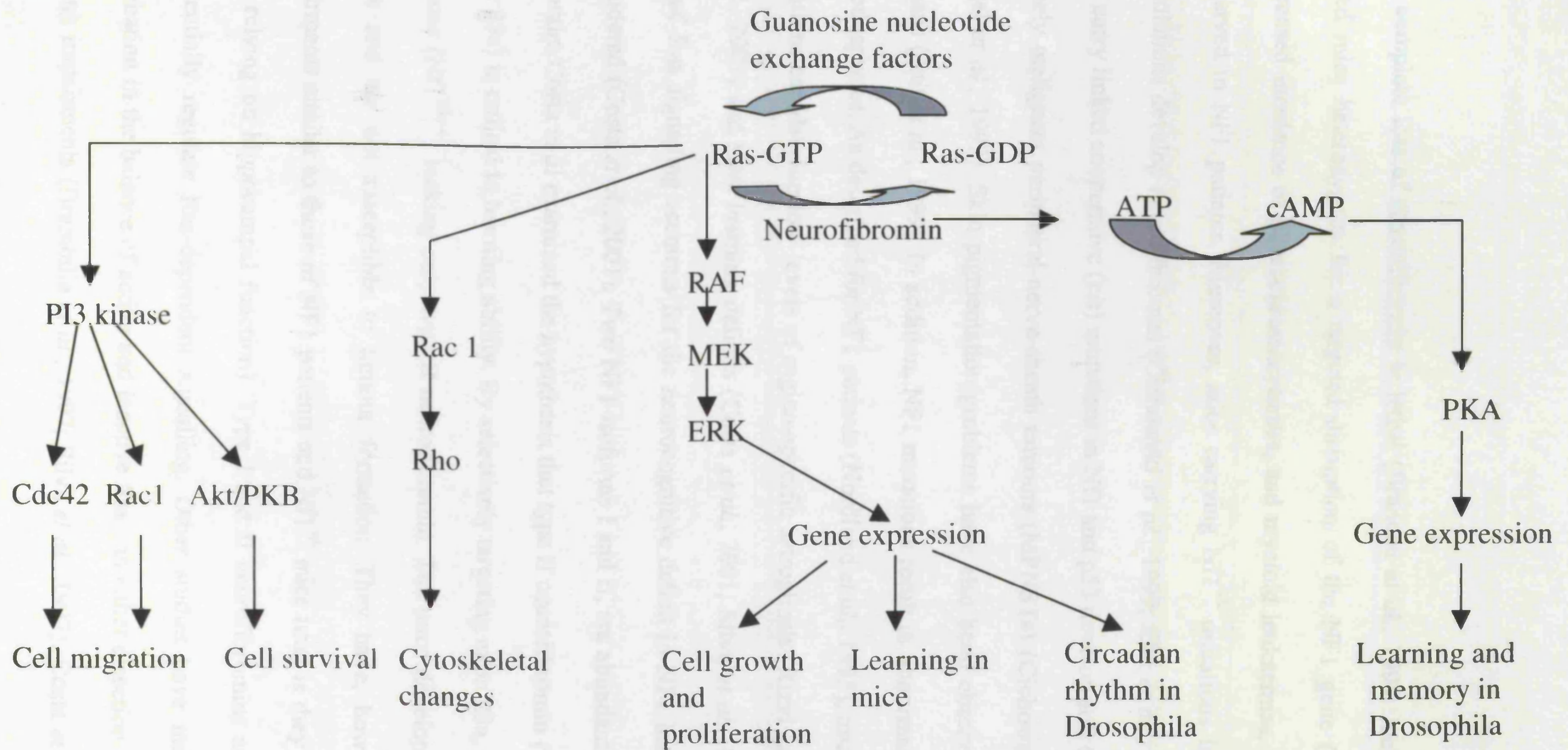



**Figure 3** Neurofibromin accelerates inactivation of active GTP-bound Ras, such that NF1 loss in specific cells results in increased Ras activity and dysregulated cell growth. In addition, neurofibromin may also be required for cAMP- and protein kinase A (PKA)-mediated gene transcription.

the complete loss of neurofibromin is lethal (Brannan *et al.*, 1994; Jacks *et al.*, 1994). Aged mice heterozygous for a targeted disruption of the NF1 gene ($Nf1^{+/-}$) have an increased incidence of phaeochromocytomas, and myeloid leukaemias, two phenotypes observed in NF1 patients. Moreover, mice carrying $Nf1^{+/-}$ mutations in particular cell populations develop neurofibromas (Cichowski *et al.*, 1999; Zhu *et al.*, 2002) and mice that carry linked cooperative (*cis*) mutations in Nf1 and p53 develop soft tissue sarcomas, namely malignant peripheral-nerve-sheath tumours (MPNSTs) (Cichowski *et al.*, 1999; Vogel *et al.*, 1999). Skin pigmentation problems have also been observed in mice and humans (Atit *et al.*, 1999). In addition, NF1 mutations result in abnormal brain function in both species. As described for NF1 patients (Nordlund *et al.*, 1995), mice carrying NF1 mutations can develop low levels of region-specific astrogliosis (Rizvi *et al.*, 1999; Zhu *et al.*, 2001) and have learning deficits (Costa *et al.*, 2001; Silva *et al.*, 1997). Whether altered Ras signalling accounts for the neurocognitive deficit in NF1 has recently been considered (Costa *et al.*, 2001). Two NF1 isoforms I and II, are abundantly expressed in the brain. Costa *et al* examined the hypothesis that type II neurofibromin (which includes exon 23a) is critical to learning ability. By selectively targeting exon 23a, they generated a mouse ($Nf1^{23a-/-}$) lacking only type II neurofibromin. The mice develop normally, are viable and are not susceptible to tumour formation. They have, however, learning impairments similar to those of NF1 patients and $Nf1^{+/-}$ mice (that is they fare poorly in tasks relying on hippocampal function). Type I and II neurofibromins are believed to differentially regulate Ras-dependent signalling. Other studies have indicated that a perturbation in the balance of active and inactive Ras - in either direction - can result in learning impairments (Brambilla *et al.*, 1997; Silva *et al.*, 1997). Costa *et al* later show

that the learning deficits of $Nf1^{+/-}$ mice can be rescued by genetic and pharmacological manipulations that decrease Ras function (Costa *et al.*, 2002).

### 1.2.2 Expression of Neurofibromin

Although the complex distribution pattern of neurofibromin is now known in detail, there is no absolute correlation of protein expression with disease. Thus, patient's neurofibromin is expressed both in tissues affected in NF1 and in apparently unaffected tissues (Upadhyaya, 1998). The complexity of neurofibromin suggests that NF1 mutations may act in more than one cell type to produce the manifestations of disease. For example, the three major cellular components of neurofibromas (neurons, Schwann cells and fibroblasts) all express neurofibromin. It is likely that many of the manifestations are due to abnormal interactions between affected cell types as well as cell-autonomous defects (Sherman *et al* 1998).

The normal distribution of neurofibromin has been investigated in chicken, mouse, rat and some human tissues. In all these species, neurofibromin accounts for a very low proportion of cellular protein, even in tissues with the highest neurofibromin expression, such as the brain (Daston *et al.*, 1992). Neurofibromin distribution is similar across species. It is expressed ubiquitously from the onset of organogenesis to mid-stage embryonic development.

Three main categories of tissues can be distinguished with respect to neurofibromin levels during development and differentiation (Sherman *et al* 1998):

a) spinal motor neurons, ganglionic neurons, some brain neurons, adrenal medullary

cells and oligodendrocytes all increase their expression of neurofibromin concurrent with terminal differentiation and relatively high expression persists in the adult.

b) Skeletal and cardiac muscle show strong expression during development and no detectable expression in adults.

c) Other tissues such as lung, liver and kidney clearly express neurofibromin in development with no obvious peak, and the level in the adult is low or absent.

In discussing neurofibromin distribution, most studies have not discriminated among possible neurofibromin variants generated by alternative splicing. Exons 23a, 48a and 9a of the NF1 transcript can be alternatively spliced (Danglot *et al.*, 1995). There are tissue specific differences in the relative levels of NF1 mRNAs.

The difference in neurofibromin distribution inside cells is demonstrated by analysis of neurons and keratinocytes in tissue sections using antibody staining (Malhotra and Ratner, 1994). Keratinocytes show plasma membrane neurofibromin localisation whereas in cerebellar Purkinje neurons it is localised to smooth endoplasmic reticulum (Nordlund *et al.*, 1993). This shows that there are cell type specific differences in the subcellular localisation of neurofibromin. How this is regulated and the functional significance of differences are unknown. It could, in principle, be determined by specific NF1 splice variants, or the ratios of the variants, but this remains to be demonstrated. Other mechanisms such as differential expression of neurofibromin-binding partners could also influence neurofibromin intracellular localisation.

### 1.2.3 Loss of heterozygosity (LOH) for the NF1 gene

Tumorigenesis in NF1 is believed to follow the 'two-hit' hypothesis postulated for tumour-suppressor genes, with one allele constitutively inactivated and the other somatically mutated (Cavanee *et al* 1983). Somatic loss of heterozygosity (LOH) has been described for NF1-associated malignancies (Xu *et al* 1992) and in benign neurofibromas, but only a few of the latter yielded a positive result (Colman *et al.*, 1995). However, using a systematic approach of searching for somatic inactivation of the Nfl gene in neurofibromas, it has been shown that small subtle mutations occur with similar frequency to that of LOH in benign neurofibromas and that somatic inactivation of the NF1 gene is a general event in these tumours (Side *et al* 1992). Explanations given for the failure to detect LOH in some studies includes the difficulty in analysing a large gene and the availability of informative probes. But it is also possible, as described below, that NF1 tumours may occasionally arise due to inappropriate expression of edited or alternatively spliced isoforms of NF1. Until recently it was unknown whether neurofibroma formation requires *NF1* LOH in a single cell type or in some specific complement of all the cell types commonly present in these tumours. Through the use of a conditional (cre/lox) allele, it has been shown that loss of *NF1* in the Schwann cell lineage is sufficient to generate tumours. In addition, complete NF1-mediated tumorigenicity requires both a loss of *NF1* in cells destined to become neoplastic as well as heterozygosity in non-neoplastic cells (Zhu *et al.*, 2002). Therefore there is a requirement for a permissive haploinsufficient environment to allow tumorigenesis. The notion that tumour formation is a coordinated process in which incipient tumour cells

recruit collaborating cells from the environment has established a firm foothold (Hanahan and Weinberg, 2000). Among the requirements, it has been acknowledged that cell cycle suppressors must be shut down, growth factor requirements must be eluded, blood vessel formation must be induced, and apoptotic signals must be evaded. NF1$^{+/-}$ mast cells invade pre-neoplastic nerves and remain present throughout the development of the tumour. Previous studies have described the enhanced proliferative properties of heterozygous mast cells from NF1 patients and from NF1$^{+/-}$ mice (Ingram *et al.*, 2000). Given the breadth of cytokine expression found in degranulating mast cells, it is tempting to speculate that these cells could play a central role in the initiation of neurofibroma formation.

### 1.2.4 RNA Processing and clinical variability in NF1

One of the intriguing features of NF1 is the great variation in expressivity of disease traits across those affected. To date, the underlying source of this variation remains somewhat unclear. An analysis of variation in expression of NF1 features in families including twins, concluded that this was determined to a large extent by the genotyping at other 'modifying' loci and that these modifying genes were trait specific (Easton *et al.*, 1993). However, evidence suggests that aberrations in normal *NF1* RNA processing may be involved (Skuse *et al* 1997). This evidence includes: (i) differences in the relative ratios of the type I and type II splice variants in NF1 tumours compared with non-tumour tissues; (ii) unequal expression of mutant and normal NF1 alleles in cultured cells derived from NF1 patients; (iii) the existence of NF1 tumours which display *NF1* mRNA editing

levels that are greater than that seen in non-NF1 tumours; and (iv) tissue-specific and developmental stage-specific expression of particular alternative NF1 transcripts (Skuse *et al* 1997).

**1.2.4.1 Alternative transcript expression.**

Expression of type I and type II transcripts appears to respond to extracellular factors, thereby suggesting that modulation of *NF1* gene expression of these two alternative transcripts is regulated by some epigenetic mechanism. Changes in the levels of NF1 gene expression have been observed in response to environmental factors such as cerebral ischaemia in the rat (Giordano *et al.*, 1996). A study of the rat phaeochromocytoma cell line PC12 revealed that the relative levels of the type I transcript compared with type II changed in response to treatment with a variety of factors, including nerve growth factor (NGF) and dexamethasone among others (Metheny and Skuse, 1996). Other transcripts identified are: type III (rodent), which includes exon 23b, type IV (rodent), which includes exons 23b but not 23a. Transcripts have also been identified which differ outside of the GRD. One such transcript contains an additional exon (9br) between exons 9 and 10a that is only expressed in the brain (Danglot *et al.*, 1995). Another involves an alternative splicing event that occurs at the 3' end of the *NF1* transcript to produce a transcript containing exon 48a which is observed in developing and adult skeletal and cardiac muscle (Gutmann and Collins, 1993). Transcripts containing exon 48a but not 23a, are termed type 3 and those with both 48a and 23a are type 4. This observed tissue specificity is, at first glance, unexpected due to the lack of muscle pathology in NF1 patients. However, findings of cardiac abnormalities in transgenic mice homozygously

deficient for *NF1* suggest that *NF1* gene expression is necessary for proper muscle development (Brannan *et al.*, 1994). Finally the N-isoform lacks the region encoding the GRD, and includes an additional four amino acids normally found in the carboxy terminus (Suzuki *et al.*, 1992). Expression of this centrally truncated transcript has been observed in both normal and brain tumours and its function remains unclear.

#### 1.2.4.2 Unequal allelic expression

RNA from cultured fibroblasts derived from 15 NF1 patients and from whole white blood cells from another patient was analysed for *NF1* allele expression (Hoffmeyer *et al.*, 1995). The studies revealed variable levels of allelic expression. The investigators also looked at allelic expression of *NF1* transcripts in nuclear RNA. The ratios were consistently similar. These results suggest that the detected unequal *NF1* allelic expression arises either from different stability or differential transport to the cytoplasm.

#### 1.2.4.3 NF1 mRNA editing

RNA editing is a form of post-transcriptional processing by which the coding sequence of the RNA is changed from that which is prescribed by the encoding DNA. RNA editing thus provides another level by which gene expression can be regulated and protein diversity expanded. Base modification mRNA editing involves the chemical modification of existing nucleotides within the transcript to convert them to different ones. This mechanism acts on the *NF1* mRNA so that a cytidine at position 3916 is deaminated to

become a uracil (Skuse *et al.*, 1996). The consequence of *NF1* nRNA editing is that an in frame stop codon is introduced in the 5' portion of the *NF1* GRD. This may result in an expression of a truncated form of neurofibromin, lacking the complete GRD, or may lead to an unstable mRNA through nonsense mediated decay. Which of these two alternatives applies to the edited *NF1* mRNA remains to be determined. In either case there is potential to inactivate *NF1* tumour suppressor activity without involving mutations to the *NF1* gene itself. It is possible that a shortened form of neurofibromin is expressed which either has a unique function compared with the full-length protein or somehow interacts with the full length protein to regulate its function. Whether or not *NF1* mRNA editing plays a role in the pathogenesis of NF1 or NF1 tumours has not been proven definitely. A study of 23 tumours resected from NF1 and non-NF1 patients demonstrated a trend for higher levels of NF1 mRNA editing in tumours compared with non tumour tissues (Cappione *et al.*, 1997). Interestingly, neither mouse nor rat mRNAs undergo RNA editing (Skuse *et al* 1996). This is most likely due to sequence divergence between the human and murine species at the editing site.

The findings described in the above three sections suggest that the classical 2-hit model for tumour suppressor inactivation used to explain NF1 tumorigenesis can be expanded to include the post-transcriptional mechanisms which regulate *NF1* gene expression. Aberrations in these mechanisms may lead to the pathogenesis of NF1 and may play a role in the observed clinical variability.

## 1.3 *Mutation analysis and genotype-phenotype correlations*

Despite the good diagnostic criteria, many young children and some older patients fail to meet them and are in limbo regarding the future, needing regular follow-up to see whether further features appear that would confirm the diagnosis. For these patients, and for families who desire prenatal diagnosis, a conclusive test would be helpful.

Mutation detection in *NF1* has been made difficult by a number of factors: the large size of the gene, the existence of a number of homologous pseudogene sequences spread throughout the genome, the lack of defined mutational hotspots, and many intronic changes that appear to affect splicing. To overcome these problems, a variety of techniques have been employed for screening the *NF1* gene and a database of *NF1* gene mutations, as reported to the International NF1 Genetic Analysis Consortium, is now available on the Internet (http://www.clam.com/nf/nf1gene). Most studies have been based on single-strand conformation polymorphism (SSCP) (Ainsworth *et al.*, 1993; Cawthon *et al.*, 1990; Upadhyaya *et al.*, 1994; Upadhyaya *et al.*, 1992), heteroduplex analysis (HDA) (Shen *et al.*, 1993; Upadhyaya *et al.*, 1994), temperature gradient gel electrophoresis (TGGE) (Boddrich *et al.*, 1995), denaturing gradient gel electrophoresis (DGGE) (Valero *et al.*, 1994), chemical cleavage of mismatch (Purandare *et al.*, 1995), the protein truncation test (Heim *et al.*, 1995), and long reverse transcriptase-polymerase reaction(RT-PCR) (Martinez *et al.*, 1996). The early studies were focussed on specific regions or groups of exons making it difficult to determine the respective mutation detection efficiencies of these methods in the context of NF1 and the distribution of these mutations within the gene. However no single technique has been 100% effective at

identifying pathological lesions within the NF1 gene. More recently three studies have used combinations of techniques to examine the entire *NF1* gene. In the largest study to date involving 500 patients, Fahsold (Fahsold *et al.*, 2000) used a protein truncation test (PTT), TGGE, and direct genomic sequencing (DGS) to examine all individual exons, finding sequence variants in 301 patients. Within these variants 278 mutations were considered pathogenic. In the smaller studies the methodologies were used sequentially to raise mutation detection rates. In the study of Ars et al (Ars *et al.*, 2000) on 80 patients, using cDNA-SSCP and HDA a detection rate of 70-80% of mutations was achieved. Messiaen et al (Messiaen *et al.*, 2000) used PTT, FISH, southern blot and cytogenetic analysis in 67 patients and reported a detection rate of 95% including a high frequency of unusual splicing defects. The most sensitive single technique for mutation detection reported to date (73%) is the protein truncation test (Park and Pivnick, 1998) but this will not pick up any missense mutations.

Despite the high rates of mutation detection, little evidence for a genotype/phenotype correlation has emerged from this previous work in NF1. This is perhaps not surprising given the considerable intra-familial variability in NF1, which suggests that epigenetic and other factors determine the clinical phenotype. However, a relationship between whole gene deletions (thought to occur in 10% of cases) and a more severe NF1 phenotype has been reported. The definition of severe in this context is unclear, but appears to include facial anomalies, learning disability or mental retardation and large numbers of cutaneous neurofibromas (Cnossen *et al.*, 1997; Kayes *et al.*, 1994; Wu *et al.*, 1995). Some ascertainment bias may play a role as the numbers in these studies are small

and more severely affected individuals are more likely to be investigated for large-scale genomic alterations. Other genes are likely to play minor roles in the NF1 phenotype (Easton *et al.*, 1993) and hints of phenotype/genotype correlations exist in a few scientific publications. For example, a family with only café-au-lait spots appears to show linkage to the NF1 gene region, despite having no other signs of NF1 (Abeliovich *et al.*, 1995). A small study suggested that large NF1 deletions are more common in patients who develop malignancies such as malignant peripheral nerve sheath tumours (Wu *et al.*, 1999a). In another example, all five members of a family with an exon 46 frameshift mutation consistently showed spinal neurofibomas and café-au-lait spots, with other features only variably found. There are also a few reports about *NF1* gene mutations in patients who fit other syndrome diagnoses better, which suggests that different clinical disorders might be allelic: two reports of families/patients with Watson syndrome (which has some overlapping features) having *NF1* gene mutations (Tassabehji *et al.*, 1993; Upadhyaya *et al.*, 1992) and a patient with LEOPARD syndrome (Wu *et al.*, 1996).

### 1.3.1 Noonan-Neurofibromatosis syndrome (NFNS)

An association with Noonan syndrome and NF1 was first reported in 1985 (Allanson *et al.*, 1985). They presented a total of seven patients with NF1 and additional features of Noonan syndrome. Noonan syndrome (NS) is another dominantly inherited disorder with an estimated incidence of 1 in 1,000-2,500 live births. It is characterised by dysmorphic facial features, proportionate short stature and heart malformations (most commonly

pulmonic stenosis and hypertrophic cardiomyopathy). Webbed neck, chest deformity, cryptorchidism, mental retardation and clotting abnormalities are also frequently associated with this disease (Mendez *et al* 1985). A gene causing this syndrome in approximately 50% of cases, a protein tyrosine phosphotase *PTPN11,* has recently been identified (Tartaglia *et al.*, 2001). Subsequent to the first reports, approximately 30 further cases of NF-Noonan syndrome have been reported. It has been suggested that this phenotype of Neurofibromatosis-Noonan syndrome (NFNS) is more common than previously appreciated, as Colley *et al* (Colley *et al.*, 1996) examined 94 sequentially identified patients with NF1 from their genetic register and found Noonan features in 12. There are four possible mechanisms for this association of Noonan syndrome with NF1: (a) a chance occurrence of NS and NF1, (b) NFNS as an unusual variant of NF1 (with mutations in the NF1 neurofibromin gene), (c) NFNS as an unusual variant of NS, and (d) NFNS as a distinct genetic disease. While the original cases in the literature were isolated and reported as a new syndrome entity in which vertical transmission has been reported (Quattrin *et al.*, 1987), the genetic register study suggest an association. Large studies of NS (Sharland *et al.*, 1992) make no reference to patients with CAL spots or neurofibromas suggesting that NF1 features do not occur frequently in classical NS and that NFNS is unlikely to be an unusual variant of NS. However, smaller studies have described CAL macules (Ahlbom *et al.*, 1995). There has been one family reported with evidence for the NF-NS phenotype being additive (Bahuau *et al.*, 1998). In this family, the two autosomal dominant disorders were inherited together by some members of the family. And a further 3 cases were reported by Colley et al (Colley *et al* 1996).

## 2. RNA SPLICING

### 2.0.1 *Mutations affecting splicing*

Splicing signals are a frequent target of mutations in genetic diseases and cancer. In a widely cited survey, Krawczak *et al* (Krawczak *et al.*, 1992) estimated that at least 15% of point mutations that result in a human genetic disease cause RNA splicing defects, a figure that is supported by the annotation of ~16,000 point mutations in the current Human Gene Mutation Database. Most splicing mutations that are considered in these surveys directly affect the standard consensus splicing signals, and typically lead to skipping of the neighbouring exon. Less frequently, the mutations create an ectopic splice site or activate a cryptic splice site, thereby changing the overall splicing pattern of the mutant transcript.

Mutation analysis of a gene will generate sequence variations that can generally be classified into the following groups: nonsense, frameshift, missense, splice site or silent mutations. However these classifications might be misleading when not supported by characterisation at the mRNA level, because mutations that affect sequences that are important for splicing modulation are likely to have a profound effect on the translated product. For example, if a missense causes exon skipping, the mutant protein, instead of having just a single amino-acid difference from the wild-type, will carry a large internal deletion or, if the open reading frame (ORF) is not maintained, an entirely different and probably shorter carboxy-terminal domain.

Indeed, there is growing evidence that misclassification of mutations might commonly occur, and that the general extent of splicing mutations has been underestimated. In a

recent study (Ars *et al.*, 2000) of NF1 where analysis was at the DNA and RNA level ~50% of the patients (26 out of 52) were found to have disease due to mutations that result in aberrant splicing. Of these mutations 13% (7 out of 52) would have been erroneously classified as frameshift, missense or nonsense mutations if the analysis had been limited to genomic sequence. **Table 2** (Cartegni *et al.*, 2002) lists missense and silent mutations associated with altered splicing in the literature. The fact that even mutations that are predicted to be translationally silent cause exon skipping is particularly significant, because these mutations must act at the RNA level (except for possible subtle effects on translation efficiency owing to codon preferences). These mutations probably alter *cis* -elements that are important for correct splicing. Furthermore, it is very likely that such mutations are generally under-reported, because they might be incorrectly assumed to be neutral polymorphisms that do not merit further characterisation. For nonsense mutation-altered splicing (NAS) four possible mechanisms have been proposed (Cartegni *et al.*, 2002): a) In the nuclear scanning model a presumptive translation-like machinery in the nucleus scans the reading frame and surveys its integrity before splicing. b) Indirect nonsense mediated decay (NMD) model. c) Secondary-structure disruption. d) Exonic splice enhancer (ESE) disruption. In the latter theory, the premature stop codon generated by the mutation, fortuitously disrupts the recognition motif for an RNA-binding protein that enhances splicing, such as an SR protein, and exon inclusion is no longer favoured.

| Gene | Mutation | Exon | Gene | Mutation | Exon |
|---|---|---|---|---|---|
| **Missense Mutations** | | | **Silent Mutations** | | |
| ADA | A215T | 7 | APC | R623R | 14 |
| ATM | E2032K | 44 | AR | S888S | 8 |
| ATP7A | G1302R | 4 | ATM | S706S | 16 |
| BRCA1 | E1694K | 18 | | S1135S | 26 |
| CFTR | G58E | 9 | CYP27A1 | G112G | 2 |
| | D565G | 12 | FAH | N232N | |
| F8 | R1997W | 19 | FBN1 | I2118I | 51 |
| FAH | Q279R | 9 | HEXA | L187L | 5 |
| FBN2 | D1114H | 25 | HMBS | R28R | 3 |
| FECH | A155P | 4 | HPRT1 | F199F | 8 |
| HEXB | P404L | 11 | ITGB3 | T420T | 9 |
| HMBS | E29L | 3 | LIPA | Q277Q | |
| HPRT1 | G40V | 2 | MAPT | L284L* | 10 |
| | R48H | 3 | | N296N* | 10 |
| | A161E | 6 | | S305S* | 10 |
| | G180E | 8 | MLH1 | S577S | 16 |
| | G180V | 8 | NF1 | K354K | 8 |
| | E182K | 8 | PAH | V399V | 11 |
| | P184L | 8 | PDHA1 | G185G | 6 |
| | D194Y | 8 | PKLR | A423A | 9 |
| | E197K | 8 | PTPRC | P48P | 4 |
| | E197V | 8 | PTS | E81E | 4 |
| | D201V | 8 | RET | I647I | 11 |
| IL2RG | R285Q | 6 | SMN1 | F280F | 7 |
| IVD | R21C | 2 | TNFRSF5 | T136T | 5 |
| | R21P | 2 | UROD | E314E | 9 |
| | D20N | 2 | | | |
| MAPT | N279K* | 10 | | | |
| | S305N* | | | | |
| MLH1 | R659P | 17 | | | |

**Table 2** Missense and silent mutations associated with altered splicing. * mutations that increase exon inclusion. ADA, adenosine deaminase; APC, adenomatosis polyposis coli; AR, androgen receptor; ATM, ataxia telangectasia mutated; ATP7A, ATPase, Cu transporting, α-polypeptide; BRCA1, breast cancer 1; CFTR, cystic fibrosis transmembrane conductance regulator; CYP27A1, sterol-27-hydroxylase; F8, coagulation factor VIII, procoagulant component; FAH, fumarylacetoacetate hydrolase; FBN1, fibrillin 1; FBN2, fibrillin 2; FECH, ferrochelatase; HEXA, hexosaminidase, α-polypeptide; HEXB, hexosaminidase, β-polypeptide; HMBS, hydroxymethylbilane synthase; HPRT1, hypoxanthine phosphoribosyltransferase1; IL2RG, interleukin2 receptor-γ; ITGB3, integrin-β3; IVD, isovaleryl coenzyme A dehydrogenase; LIPA, lipase A; MAPT, microtubule-associated protein tau; MLH1, mutL homologue; PAH, phenylalanine hydroxylase; PDHA1, pyruvate dehydrogenase; PKLR, pyruvate kinase, liver and red blood cells; PMM2, phosphomannonutase2; PTPRC, protein-tyrosine phosphatase receptor type C; PTS, 6-pyruvoyltetrahydropterin synthase; RET, rearranged during transfection protooncogene; RHAG, Rhesus blood group-associated glycoprotein; SM1, survival of motor neuron 1; TNFRSF5, tumour-necrosis factor receptor superfamily, member 5 (CD40); UROD, uroporphyrinogen decarboxylase.

### 2.0.2 Fundamentals of RNA splicing

The coding sequences of most eukaryotic genes are interrupted by non-coding stretches of DNA known as intervening sequences or introns. The presence of intervening sequences was first described in non-coding region of adenovirus (Berget *et al.*, 1977; Chow *et al.*, 1977), but was quickly shown to be a common feature of cellular genes. In fact in 1977, for the first time, Jeffreys and Flevell described the presence of "large insert" in the coding sequence of rabbit α-globin gene (Jeffreys and Flavell, 1977). Exons must be excised from primary transcripts and the flanking exon joined together before the mature RNA is exported from the nucleus. This process is called splicing and occurs in a macromolecular complex known as a spliceosome, which consists of five small ribonuclear particles (snRNPs) and a large number of non-snRNP protein splicing factors. The primary transcript of protein coding genes - pre-mRNA, or heterologous nuclear RNA (hnRNA), is synthesised by RNA polymerase II within the nuclei of eukaryotic cells. Its average size is much larger than mRNA and because of this it is very unstable. From the time hnRNA emerges from the transcription complex, and throughout the time it is in the nucleus, it is associated with proteins. The physical form of hnRNA is a ribonucleoprotein particle (hnRNP) in which hnRNP proteins (which comprise at least 20 distinct proteins) bind cooperatively to nascent pre-mRNA to form complexes encompassing about 500 nucleotides of RNA (Dreyfuss *et al.*, 1993).

## 2.1 *The spliceosome*

The removal of introns from pre-mRNA is a critical aspect of gene expression. The inaccurate recognition of exon-intron boundaries or the failure to remove an intron generates aberrant mRNAs that are either unstable or code for defective or deleterious protein isoforms. This reaction takes place in the spliceosome, which is formed by several RNP subunits termed uridine-rich small ribonucleoproteins (UsnRNP), and numerous non-snRNP splicing factors. Each UsnRNP consists of a UsnRNA complexed with a set of eight Sm or Sm-like proteins and several particle-specific proteins (Will, 1997).

The spliceosome acts through a multitude of RNA-RNA, RNA-protein and protein-protein interactions to precisely excise each intron and join exons in the correct order (Madhani and Guthrie, 1994; Nilsen, 1994).

### 2.1.1 The Splicing Reaction

Splicing takes place in two catalytic steps involving two consecutive trans-esterification reactions (Lamond, 1993). During step I, an adenosine residue generally located within 100 nucleotides of the 3' end of the intron, in a sequence element known as the branch point sequence (BPS), carries out a nucleophilic attack on the 5' splice site. This reaction generates the splicing intermediates (free exon 1 and lariat-exon 2). During step II, exon 1 attacks at the 3' splice site to generate splicing products (spliced exon and lariat intron)- **Figure 4**.

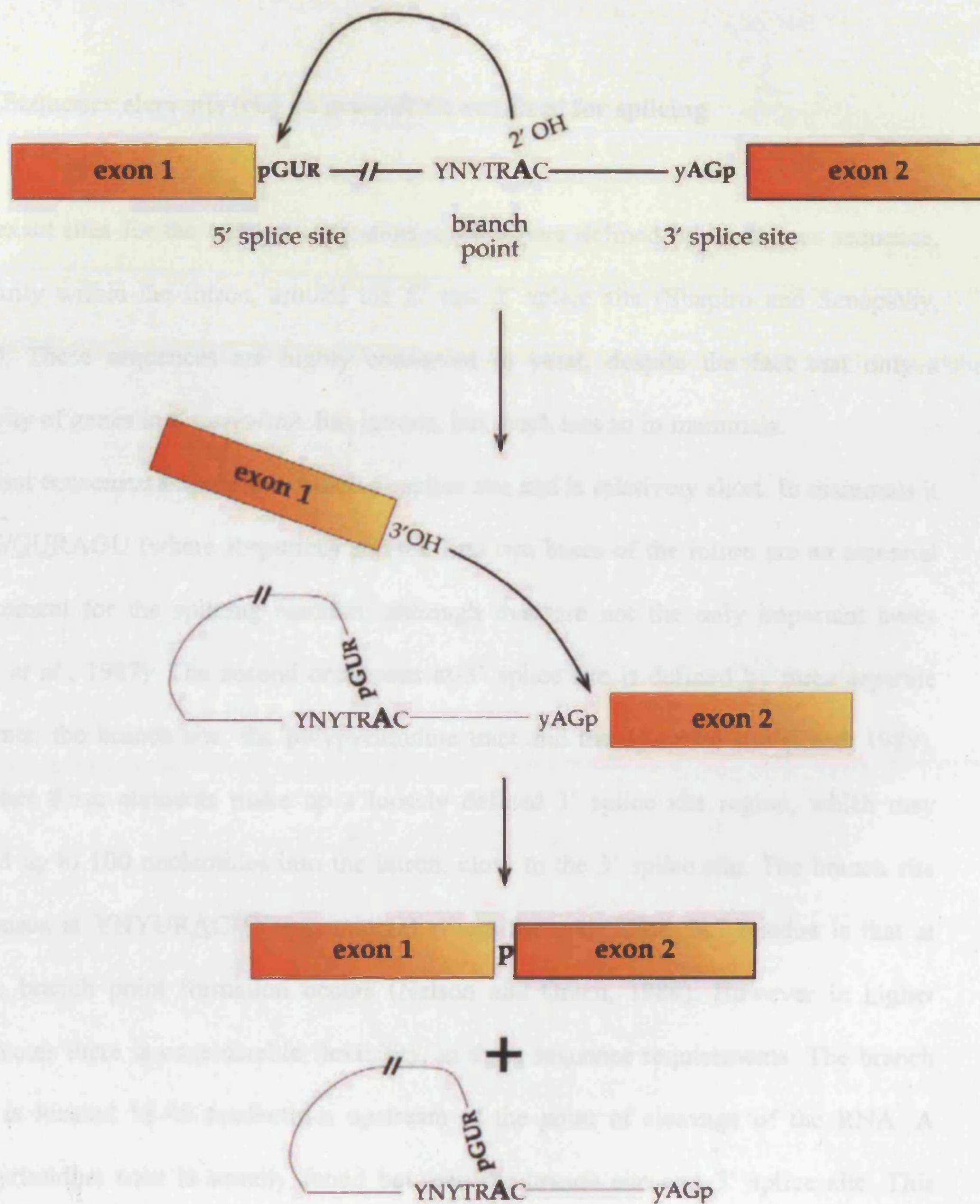



**Figure 4** Exon splicing and lariat formation

### 2.1.2 Sequence elements (cis) on pre-mRNA required for splicing

The exact sites for the trans-esterification reactions are defined by consensus sequence, primarily within the intron, around the 5' and 3' splice site (Shapiro and Senapathy, 1987). These sequences are highly conserved in yeast, despite the fact that only a minority of genes in *S. cerevisae* has introns, but much less so in mammals.

The first consensus sequence is called 5' splice site and is relatively short. In mammals it is AG/GURAGU (where R=purine) and the first two bases of the intron are an essential requirement for the splicing reaction, although they are not the only important bases (Aebi *et al.*, 1987). The second consensus at 3' splice site is defined by three separate elements: the branch site, the polypyrimidine tract and the 3' splice site (Reed, 1989). Together these elements make up a loosely defined 3' splice site region, which may extend up to 100 nucleotides into the intron, close to the 3' splice site. The branch site consensus is YNYURAC (Y=pyrimidine) where the underlined "A" residue is that at which branch point formation occurs (Nelson and Green, 1989). However in higher eukaryotes there is considerable flexibility, in these sequence requirements. The branch point is located 18-40 nucleotides upstream of the point of cleavage of the RNA. A polypyrimidine tract is usually found between the branch site and 3' splice site. This sequence is more pronounced in mammals than in yeast, where the length of this tract is variable and for this reason introns are classified as short and long polypyrimidine-tract introns, which have distinguishable properties in splicing. The 3' splice site junction is defined by YAG/G.

As previously mentioned, consensus sequences in higher eukaryotes are less conserved

than in yeast. However, there are many sequences in mammalian transcripts that match the consensus splice sites but most of them are pseudo-splice sites. To increase the overall fidelity of the splicing reaction additional sequences are present in exons and in introns, these are called enhancers or silencers. Their features and role during the splicing will be described in section 2.4.

## 2.2 *Protein components involved in splicing*

In higher eukaryotes, over 70 spliceosomal proteins have been identified thus far. Although their precise role in some cases remain unknown, spliceosomal proteins have been implicated in a growing number of functions. Most of these proteins share very similar structural features such as RNA binding domains and/or protein binding domains.

### 2.2.1 Small Nuclear Ribonucleoprotein Particles (snRNPs)

snRNP particles form part of the catalytic macromolecular complex of the spliceosome. Each snRNP particle consists of an snRNA molecule complexed with a set of eight Sm or Sm-like proteins and several particle specific proteins (Will and Luhrmann, 1997). The major spliceosomal snRNPs U1, U2, U4, U5 and U6 are responsible for splicing the vast majority of pre-mRNA introns (U2-type introns). A group of less abundant snRNPs, U11, U12, U4atac and U6atac, together with U5, are subunits of the so-called minor spliceosome that splices a rare class of pre-mRNA introns, denoted U12-type (Burge, 1999).

snRNAs U1, U2, U4, U5 and U6 are characterised by their small size, metabolic stability and a high degree of sequence conservation (Kambach *et al.*, 1999). They show high complementarity to the consensus splice sites on the pre-mRNA.

The snRNAs are transcribed by RNA polymerase II, with the only exception of U6 (and presumably U6atac snRNA) that is transcribed by RNA polymerase III and entirely assembled in the nucleus. The other pre-snRNAs must be transported to the cytoplasm where snRNPs assembly with Sm proteins is initiated and then re-imported into the nucleus thanks to the bipartite snRNP nuclear localization signal (NLS) formed by Py3cap and the Sm core domain (Fischer *et al.*, 1993). The U4 and U6 snRNAs are extensively base paired (amounting to >20 base pairs) in the U4/U6 snRNP. They associate in the nucleus forming a larger ribonucleoprotein complex. The U5 snRNA then assembles in an ATP dependent reaction with the U4/U6 snRNP giving the U4/U6•U5 three snRNPs particles (Konarska and Sharp, 1987).

The structural core of snRNPs is formed by eight proteins, called Sm proteins, B', B, D1, D2, D3, E, F and G. This class of common proteins play an essential role in the biogenesis of the snRNPs. The Sm proteins form three distinct heteromeric complexes prior to their interaction with the highly conserved Sm site (PuAU4-6GPu flanked by two stem-loop structures) of the U1, U2, U4 and U5 snRNAs (Raker *et al.*, 1996). Sm-like proteins belonging to the Sm protein family are specifically required for the assembly of U6 snRNA. This subclass of Sm-like proteins shares the conserved structural motif characteristic of all Sm-proteins family members (Mayes *et al.*, 1999), but can be isolated as a heteromeric complex in the absence of U6 snRNA (Achsel *et al.*, 1999).

Besides Sm proteins there are other particle-specific proteins that associate with snRNAs

(Will, 1997). U1-70K and U1-A proteins bind directly to the RNA and are involved in the splice site recognition and selection, while U1-C associates via protein-protein interactions with U1-70K and other Sm proteins. A subset of U2 snRNP proteins also play a critical role in tethering the U2 snRNP to the pre-mRNA. These proteins include the heteromeric splicing factors SF3a and SF3b (Brosi *et al.*, 1993) and bind 20-nucleotide region just upstream of the branch site in a sequence-independent manner (Champion-Arnaud and Reed, 1994; Gozani *et al.*, 1996). At least five different proteins associate with U4/U6 snRNP, including a 15.5 KDa protein, polypeptides of 20, 60 and 90 KDa that form complex with one another (Teigelkamp *et al.*, 1998). U5 snRNP particle presents a complex protein composition. U5 220KDa protein cross-links to both 5' and 3' splice sites, as well as to the exon flanking these two splice sites (Umen and Guthrie, 1995).

### 2.2.2 Non-snRNP splicing factors

**• U2AF**

U2 snRNP auxiliary factor (U2AF) is an essential splicing factor that is required for the binding of U2 snRNP to the pre-mRNA (Ruskin *et al.*, 1985). This factor is a heterodimer composed of two subunits named U2AF65 and U2AF35. U2AF65 contains two functional domains consisting of a sequence specific RNA binding-region composed of three canonical RNA-recognition motif (RRM) and a N-terminal short serine/arginine domain involved in protein-protein interaction (Zamore and Green, 1991). U2AF65 interacts with the polypyrimidine tract. It participates in the recognition of the essential

AG dinucleotide at 3' splice site at the earliest stage of spliceosome assembly (Wu *et al.*, 1999b). These studies have also shown that U2AF35 is essential for splicing 'AG-dependent' introns, which are introns with a weak polypyrimidine tract, but is dispensable for 'AG-independent' introns, which have a strong 3' splice site.

- **SR proteins**

SR proteins are a superfamily of highly conserved proteins that play an important role in splicing control (Fu, 1993; Zahler *et al.*, 1992). The name SR proteins reflect the presence of a characteristic serine/arginine domain present in the carboxy-terminal region of these proteins. Sequence analysis of SR proteins revealed the presence of at least one amino-terminal RNA recognition motif (RRM) required for the interaction with pre-mRNA. Various human SR proteins have been identified using different techniques. Currently the SR protein family contains 10 known members: SRp20, SRp30 (SF2/ASF), SRp30b (SC35/PR264), SRp30c, SRp40, SRp46, SRp54, SRp55, SRp75 (Graveley, 2000). SR proteins are functionally redundant in the splicing of some introns, but exhibit unique functions in the removal of others (Kawano *et al.*, 2000; Longman *et al.*, 2000). They participate in both constitutive and alternative splicing (discussed in section 2.4). It has been demonstrated that the initial binding of SR proteins with pre-mRNA is sufficient to commit the splicing pathway and to facilitate spliceosomal assembly (Fu, 1993). The RRMs mediate RNA binding and determine substrate specificity for individual SR proteins. The hallmark of the RS domain is its significant post-translational phosphorylation of serine residues (Gui *et al.*, 1994; Roth *et al.*, 1991). The phosphorylation step is required for the dissociation of the splicing factors from nuclear

speckles, the nuclear compartments enriched in proteins involved in pre-mRNA splicing, and the consequent recruitment to the sites of transcription (Misteli *et al.*, 1998).

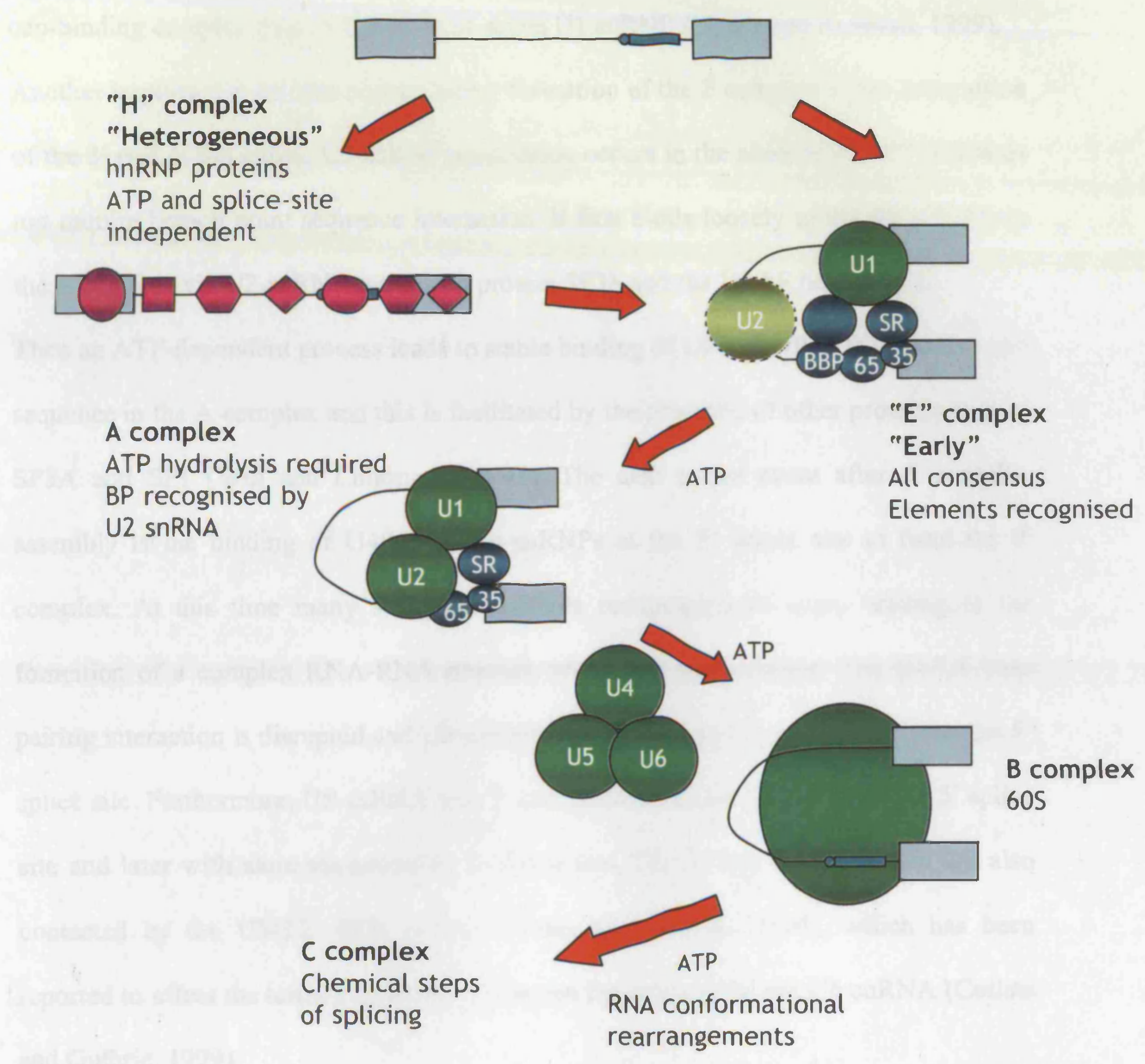
## 2.3 *The Spliceosomal Complex*

Two of the main functions of the spliceosomal snRNPs are to recognise the 5' and 3' intron/exon boundaries and to assemble onto these sites the macromolecular enzyme that catalyse the splicing reaction. In mammals four distinct spliceosomal complexes, which form in the temporal order E, A, B and C, have been detected (**Figure 5**). These general steps in spliceosomal assembly are now well understood, and its consequent recruitment to sites of transcription (Misteli *et al.*, 1998).

### 2.3.1 General model for spliceosome assembly and spliceosome cycle

Assembly of the major spliceosome is initiated by the ATP-dependent recognition of the 5' splice site by the U1 snRNP, which leads to the formation of the spliceosomal complex E. This interaction is mediated by base pairing of the U1 snRNA with the 5' splice site, as well as by protein-protein and protein-pre-mRNA interaction involving U1-70K and U1-C proteins (Will and Luhrmann, 1997). Novel insights into spliceosomal assembly come from studies in *S. cerevisae* which have shown that the first proteins that interact with the 5' splice site during the commitment complex (E complex counterpart) formation are the

**Figure 5. Spliceosome assembly pathway**



cap-binding complex protein CBC80 and seven U1 snRNP (Zhang and Rosbash, 1999). Another important event that occurs during formation of the E complex is the recognition of the 3' end of the intron. U2 snRNP association occurs in the absence of ATP and does not require branch point sequence interaction. It first binds loosely to the pre-mRNA in the E complex via U2-snRNP associated protein SF3b and the U2AF heterodimer.

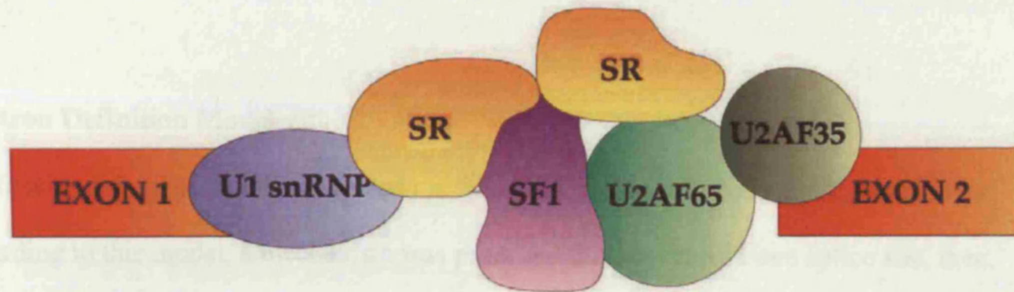
Then an ATP-dependent process leads to stable binding of U2-snRNP to the branch point sequence in the A complex and this is facilitated by the presence of other proteins such as SF3A and SF1 (Will and Luhrmann, 1997). The next major event after A complex assembly is the binding of U4/U5•U6 tri-snRNPs at the 5' splice site to form the B complex. At this time many structural snRNPs rearrangements occur leading to the formation of a complex RNA-RNA network within the spliceosome. The U4/U6 base pairing interaction is disrupted and U6 snRNA base pairs with U2 snRNA and also the 5' splice site. Furthermore, U5 snRNA loop I base pairs with exon sequence at the 5' splice site and later with exon sequences at 3' splice site. The 5' and 3' splice sites are also contacted by the U5-220 KDa protein (Umen and Guthrie, 1995), which has been reported to affect the tertiary interaction between the splice sites and U6 snRNA (Collins and Guthrie, 1999).

An unsolved problem is the catalysis of RNA in pre-mRNA splicing. It is widely believed that this is mediated by RNA. The spliceosome is a metallo-enzyme (Sontheimer *et al.*, 1997; Steitz and Steitz, 1993) and from recent studies has been shown that U6 snRNA coordinates metal ion thus contributing to pre-mRNA splicing (Yean *et al.*, 2000). U6 is a good candidate for catalysis because it is highly conserved through evolution and also forms intramolecular and intermolecular helices that are analogous to autocatalytic group

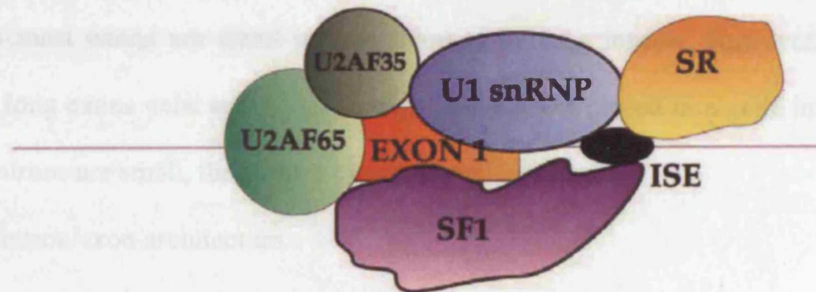
II introns.

### 2.3.2 Early Complex Formation: Intron and Exon Definition Model

Interactions between 5' and 3' splice sites and the factors that recognise them have been observed in the earliest steps of spliceosomal assembly. Two models of interaction have been proposed according to intron size- **Figure 6**.




**Intron definition model**




**Exon definition model**

**Figure 6** Intron and exon definition models

**a) Intron Definition Model**

The first model suggested that the intron is the unit recognised by the splicing machinery. According to this model, a mechanism was proposed that recognised one splice site, then, following scanning through the intron located the second site (Lang and Spritz, 1983). Such interactions have been observed in *in vitro* mRNA splicing with short introns.

• The yeast model.

In yeast, messengers often have unique introns and their length is usually below 100 nucleotides. This provides direct support for the pairing of splice sites model across introns during the first step of spliceosome assembly (Goguel and Rosbash, 1993).

• Vertebrate intron/exon architecture.

In vertebrates most exons are small and are flanked by long introns. However, some exceptionally long exons exist and when an expanded exon is placed in a gene in which the flanking introns are small, the exon is constitutively included.

• *Drosophila* intron/exon architecture.

In *Drosophila* most exons are 100 to 180 nucleotides in length; however 15% are more than 550 nucleotides. There are a number of genes in *Drosophila* that have an inverted intron/exon architecture with respect to the one found in vertebrate genes. In at least three *Drosophila* genes, 5' splice site mutations of small introns cause intron retention instead of exon skipping as observed in vertebrates (Berget *et al.*, 1977; Talerico and Berget, 1994). As for vertebrate short exons, *Drosophila* short introns seem to have size limitations. Expanding the size of the short introns causes repression of splicing or activation of cryptic site. Further evidence for a particular mechanism in the recognition of short introns derives from the observation that spliceosomal complex A formation

requires sequences at both ends of the intron but not a polypyrimidine tract, while in vertebrates it requires only sequence at the 3' splice site including the polypyrimidine tract.

**b) Exon Definition Model**

In 1990 Susan Berget's group proposed a model in which the exon was the unit recognised by the splicing machinery and the identification of splice sites was facilitated by interactions across the exon (Robberson *et al.*, 1990). Basically the exon definition theory proposes that in pre-mRNAs with multiple short exons and long introns the splicing machinery searches for two closely spaced splice sites. This model explains splice site selection of most mammalian pre-mRNAs which contain very long introns and weakly conserved splice sites (Hawkins, 1988). According to this model, an interaction is first established between the 5' and 3' splice sites across the short exon, followed by an interaction across the long intron and the juxtaposition of the neighboring exons.

• Exon size requirement.

The exon length can affect splicing. Internal vertebrate exons have minimum and maximum length requirements (Hawkins, 1988).

Simultaneous recognition of splice sites bordering an exon suggests that a minimal separation between the sites might be required to prevent steric hindrance between the factors that recognise individual sites. This is demonstrated, when a constitutively recognised internal exon was internally deleted below 50 nucleotides and was skipped by *in vivo* splicing machinery (Dominski and Kole, 1991). In addition, Black *et al* (Black, 1991) have shown that when the length of the N1 (neuron specific) exon is extended to

109 nucleotides, the exon is constitutively included, implying that the exon is normally skipped because it is too short to allow spliceosomes to assemble at both ends simultaneously. Instead, the expansion *in vitro* of internal exons to lengths above 300 nucleotides determines the activation of cryptic splice site inside the exon or exon skipping (Berget *et al.*, 1977) demonstrating the exon length limitation for efficient splicing. In fact, less than 1% of known internal exons in vertebrates are longer than 400 nucleotides.

On the other hand, expansion of internal exons in vertebrate genes with moderate to large introns have two phenotypes: activation of internal cryptic sites within the expanded exons to create small exons or skipping of the entire exon (Robberson *et al.*, 1990).

• Effect of splice site mutations.

*In vitro* (Dominski and Kole, 1991; Talerico and Berget, 1994) and *in vivo* (Dominski and Kole, 1994; Xu *et al.*, 1993) experiments revealed that mutations within splice sites influence the splicing of both introns flanking the exon and not only of the intron bearing the mutated splice site, as predicted from intron recognition oriented theories. Furthermore, mutations at the 5' splice site are suppressed by mutations that improve the consensus of the upstream 3' splice site (Carothers *et al.*, 1993).

• A network of interactions span the exon.

Experiments carried out on the preprotachykinin pre-mRNA showed evidence for exon bridging interactions that occur between U1 snRNP bound at the 5' splice site and U2AF65 bound at the 3' splice site. U2AF65 is indeed recruited to the polypyrimidine tract by interactions with the downstream 5' splice site and the U1 snRNP (Hoffman and Grabowski, 1992). SR proteins have been implicated in these interactions in different

experimental models (see section 2.4.2.).

• Exon enhancer sequences.Alternatively spliced exons with suboptimal splice sites can contain particular sequences that work as enhancers of splicing, helping in the recruitment of splicing factors to an exon otherwise not recognised. These sequences will be explained in section 2.4.

• First and last exon recognition.

The lack of one functional splice site at the first and the last exon indicate special mechanisms for their recognition. To have efficient removal of the first intron, the capping and the proteins that bind the cap are essential (Izaurralde *et al.*, 1994). Therefore both factors that recognise the cap and the 5' splice site, are necessary to define the first exon. The last exon is usually longer than internal exons (average length 600 nucleotides versus 137). Removal of the last intron involves RNA splicing and polyadenylation factors (Gunderson and Kopito, 1994).

### 2.3.3 Splice site recognition and bridging interactions

One of the critical steps of the splicing reaction is initial recognition of the splice site that leads to formation of complex E. It has been proposed that splice site recognition and pairing across introns is promoted by a network of interactions involving the SR proteins. SF2/ASF and SC35 associate via their RS domain with another RS-domain containing protein the U1 snRNP component U1-70K at 5' splice site (Eperon *et al.*, 1993; Kohtz *et al.*, 1994; Zuo and Manley, 1994).

SR proteins are also involved in the subsequent bridging interaction. Different models

have been proposed for bridging interaction according to exon and intron size. In the intron bridging model U2ASF and SF1 cooperatively interact to recognize the branch point and the polypyrimidine tract (Berglund *et al.*, 1997). The RS domain of U2AF35 contacts SC35 that simultaneously interacts with the RS domain of U1 snRNP 70K protein and U1 snRNP is base paired to the 5' splice site (Wu and Maniatis, 1993). These U2AF65-U2AF35-SR-U1 70K interactions were also proposed to function as cross-exon contacts in constitutive and regulated mammalian splicing (Reed, 1996; Zuo and Maniatis, 1996).

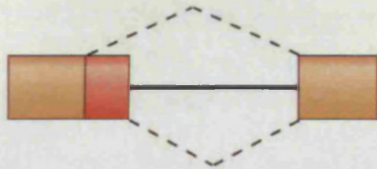
In Exonic Splicing Enhancer (ESE)-dependent splicing a different class of proteins are involved. These are SRm160/300 proteins (SR-related Matrix protein) that are SR repeats containing proteins that lack the RNA recognition motif. In this case a splicing co-activator model for the function of SRm160/300 has been proposed. SRm160/300 promotes splice-site pairing and splicing through multiple cooperative interactions with factors bound to pre-mRNA, including SR proteins, U1 and U2 snRNPs (Blencowe *et al.*, 1998; Eldridge *et al.*, 1999).

### 2.4 *Alternative splicing*

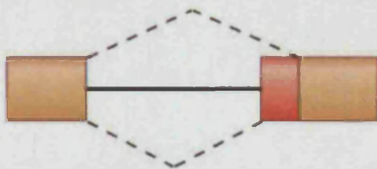
Alternative RNA splicing is the process that allows the selection of different combination of splice sites within precursor mRNA- **Figure 7**. This process is seen in nearly all metazoan organisms as a means for producing functionally diverse polypeptides from a single gene (Lopez, 1998).

(a) Retained intron

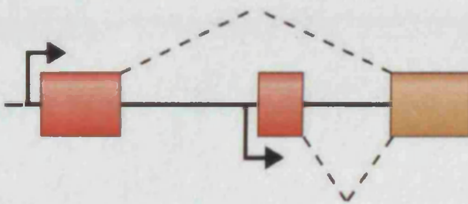
(b) Competing 5′ splice sites



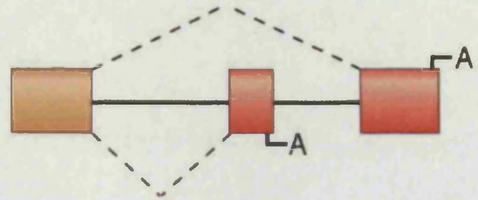
(c) Competing 3′ splice sites



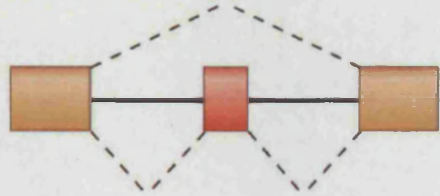
(d) Multiple promoters



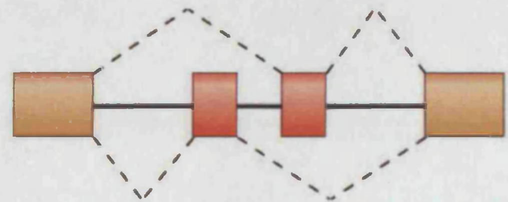
(e) Multiple poly(A) sites




(f) Cassette exons



(g) Mutually exclusive exons



Current Opinion in Chemical Biology

**Figure 7** Types of alternative splicing events.

Alternative splicing is especially common in vertebrates. Alignment of EST sequences and mapping the resulting mRNA families to the human genome provided a minimum estimate that 35% of human genes show variably spliced products (Croft *et al.*, 2000). However, since ESTs derive from a limited number of tissues or developmental stages, and cover a small portion of each mRNA, the true percentage is likely much higher and is responsible for much of the complexity of the proteome. This partly explains the unexpected finding that the human genome might consist of only 31,000-39,000 genes (McPherson *et al.*, 2001; Venter *et al.*, 2001). Because a single primary transcript can have several regions that each undergo alternative splicing, the resulting combinatorial effects of selecting different splice sites can be very pronounced, and genes that code for tens to hundreds of different isoforms are common (Graveley, 2001). There are also remarkable examples of hundreds and even thousands of functionally divergent mRNAs and protein being produced from a single gene. A recent discovery in *Drosophila* is a fascinating example of the extremely high number of proteins that can be produced from one gene. The DSCAM gene presents different exons used in a mutually exclusive manner. For each of these exons there are different numbers of alternative forms. If all the combinations of these exons were used, the single DSCAM gene would produce 38,016 different DSCAM proteins (Schmucker *et al.*, 2000).

Variation in mRNA structures takes many different forms (Lopez, 1998). Exons can be spliced into the mRNA or skipped. Introns that are normally excised can be retained in the mRNA. The position of either 5' or 3' splice sites can shift to make exons longer or shorter. In addition to these changes in splicing, alterations in transcriptional start site or polyadenylation site also allow production of multiple mRNAs. All of these changes in

mRNA structure can be regulated in diverse ways, depending on sexual genotype, cellular differentiation, or the activation of particular cell signaling pathways.

The mechanisms that determine which splice sites are utilized and how this selection is regulated in different cell types or developmental stages have still not been precisely defined. However, much progress has been made in the identifying the *cis*-acting elements and the *trans*-acting factors involved in alternative splicing.

### 2.4.1 Involvement of *Cis*-Acting Elements in Splice Site Selection

The question of splice-site choice is intimately connected to the problem of normal recognition of constitutive splice sites. A feature shared by both regulatory sequences and splice-site signals is that they are usually short and often degenerate. *Cis*-acting determinants that influence competing splicing pathways include relative strength of 5' splice site, the branch point sequence, the polypyrimidine tract, the proximity between 5' splice site and branch point, sequences between the branch point-polypyrimidine tract and 3' splice site. In some cases, the formation of pre-mRNA secondary structure participates in the regulation of splice site selection by modifying the physical distance within introns, or by being involved in the definition of the exon. Moreover, additional exonic or intronic sequences are often involved in the correct recognition of splice sites, when suboptimal splice sites are present. These sequences can act by increasing the recognition or decreasing it and are respectively named enhancers or silencers.

• **Exon/intron architectural requirements**

As described in the previous section, the average vertebrate genes consists of multiple small exons separated by introns that are considerably larger. The average size of exons is 137 nucleotides and exons longer than 300 nucleotides or shorter than 50 nucleotides are not favoured by splicing machinery.

A compensatory relationship between exon and intron size has been proposed. Sterner *et al* observed that large internal exons are problematic for recognition if they are flanked by large introns, suggesting that naturally occurring large vertebrates exons might be flanked by small introns (Sterner *et al.*, 1996).

• **The 5' and 3' splice site requirements**

The strength of splice sites is a determinant in their selection. As previously described, at 3' splice site exon/intron junction the consensus sequence is determined by 4 bases, while at the 5' exon/intron junction this consensus sequence is extended to 9 bases (Shapiro and Senapathy, 1987). The GU and AG dinucleotides at 5' and 3' intron boundaries are nearly invariant in higher eukaryotes and mutations in one of these conserved bases completely abolishes splicing (Langford *et al.*, 1984). It has been reported in different studies that mutations of a weak splice site that take it closer to the consensus sequence can lead to constitutive recognition (Huh and Hynes, 1993; Muro *et al.*, 1999).

Mutations affecting splice sites causing aberrant splicing are observed more frequently in the 5' splice site region than in 3' splice site region (Nakai and Sakamoto, 1994). Furthermore, alternative-exon splice sites deviate more from the consensus: their 5' splice sites deviate from consensus sequence mostly at the +4 and +5 position, while

adenosine is more frequently used at the –3 position of the 3' splice site (Stamm *et al.*, 2000).

**• The polypyrimidine tract**

Recognition of branch point sequences in the metazoan can be affected by the adjacent polypyrimidine tract (Mullen *et al.*, 1991; Reed, 1989). It has been shown that progressive deletion of the polypyrimidine tract abolishes lariat formation, spliceosome assembly and splicing (Mullen *et al.*, 1991; Roscigno *et al.*, 1993).

Despite the important role of polypyrimidine tract in splicing, there appears to be great flexibility in the specific sequence of a given tract, for example the number of uridines involved (Norton, 1994; Roscigno *et al.*, 1993).

The length of the polypyrimidine tract is also important. In fact, as the number of consecutive uridines decreases, the location of pyrimidine tract becomes increasingly important showing a position dependent effect. Short tracts immediately adjacent to the 3' splice site are utilised more than twice as much as short tracts adjacent to the branch point (Coolidge *et al.*, 1997).

**• The branch point sequence requirement**

While in yeast there is an invariant branch point sequence (UACUA<u>A</u>C) which base pairs with U2 snRNA, in metazoan this sequence is not highly conserved. However, several lines of evidence suggest that the mammalian branch point is specified primarily by its proximity to the 3' splice site. Most branch points have been mapped within 18-40 nucleotides of the 3' splice site (Reed and Maniatis, 1985; Ruskin *et al.*, 1985). There are

cases in which the branch point is located more than 40 nucleotides from 3' splice site, is functional and is essential for the regulation of alternative splicing (Smith *et al.*, 1989). Generally, the mutation of the adenine residue involved in the lariat formation strongly reduces splicing efficiency of the downstream exon (Reed and Maniatis, 1988).

• **RNA secondary structure**

The *cis*-acting elements regulating constitutive and alternative splicing interact with *trans*-acting factors. The secondary structure that pre-mRNA creates might directly influence this kind of interaction.

Many studies have proposed RNA secondary structure as a regulator of alternative splicing. In the case of rat calcitonin/CGRP pre-mRNA splicing, the 3' splice acceptor of exon 4 forms a stable stem-loop structure. Mutations that destabilize the stem abolish the use of this splice acceptor site *in vitro* (Coleman and Roesser, 1998). In the fibronectin EDA exon there is both an ESS and an ESE. Functional studies coupled to secondary structure analysis suggest that the role of the ESS element may be exclusively to ensure proper RNA conformation and raise the possibility that the display of the ESE in a loop position may represent a significant feature of the exon splicing regulatory region (Muro *et al.*, 1999). An example of 5' splice site mutation that affects a stem-loop structure can be found in the tau gene. In this case the disruption of secondary structure increases exon 10 inclusion and this results in the production of an aberrant protein isoform that leads to neurodegeneration (Grover *et al.*, 1999).

• **Enhancers and silencers**

Additional intronic and exonic sequences are often necessary for efficient and accurate choice of the correct splice site. These sequences can either promote (enhancers) or repress (silencers) splice site usage.

SR proteins bound to splicing enhancers are thought to function by recruiting components of the splicing machinery to the nearby intron (Reed, 1996). Most splicing enhancers are located within 100 nucleotides of the 3' splice site and are not active when located further away (Tian and Maniatis, 1994). However in *Drosophila melanogaster* doublesex pre-mRNA, the *dsx* enhancer element functions when located 300 nucleotides downstream of the regulated 3' splice site (Tian and Maniatis, 1993). Recent studies show that the activity of splicing enhancers decreases as a function of distance from the 3' splice site. Most of the enhancer elements obtained contained extended purine-rich sequences (more than 65% purine rich), but a second, novel class of sequences lacking stretches of purines was also identified (Tian and Kole, 1995). In fact, Schaal and Maniatis recently identified pyrimidine-rich enhancers that are more than 67% pyrimidine-rich and function such as strong enhancers (Schaal and Maniatis, 1999).

In addition to sequences that promote exon inclusion, there are sequences that inhibit splicing; exonic or intronic splicing silencers. The silencers are less well characterised; they can be purine or pyrimidine-rich and bind a diverse array of proteins (Fairbrother and Chasin, 2000). A negative element (CAGG) involved in the down regulation of fibronectin EDA exon inclusion was identified within EDA exon (Caputi *et al.*, 1994), where it seems to be a determinant for RNA conformation (Muro *et al.*, 1999). An example of an intronic splicing silencer is described for CFTR exon 9 alternative

splicing. This extended region in intron 9 acts as a silencer by recruiting SR proteins (Pagani *et al.*, 2000).

### 2.4.2 *Trans*-Acting Factors

The *trans*-acting factors that regulate alternative splicing are members of at least two protein families. The most important one is the SR protein family and the other is the hnRNP protein family. This last group of proteins is associated with newly synthesised pre-mRNA as well as mature mRNA during nucleo-cytoplasm transport. They can act on splicing by affecting correct spliceosomal assembly.

**• SR proteins as alternative splicing regulator**

Besides their role in constitutive splicing, SR proteins can also regulate alternative splicing either by altering the 5' or 3' splice site choice in pre-mRNA containing competiting sites (Bai *et al.*, 1999; Krainer *et al.*, 1990). Enhancer-bound SR proteins can promote the utilisation of a downstream 5' splice site and this activity may involve the recruitment of U1 snRNP to the splice site (Eperon *et al.*, 1993). In the presence of competing 5' splice sites, different SR proteins show distinct preferences for promoting the use of proximal or distant splice sites. High concentrations of ASF/SF2 and SC35 proteins promote the use of 5' splice site proximal to the 3' splice site, while SRp40 and SRp55 promote the use of distal sites (Zahler *et al.*, 1992).

There are different components that determine the effect of SR proteins in the regulation

of splite site choice; one of these is the level of their expression (Hofmann and Wirth, 2002).

**• hnRNP as a splicing regulator**

hnRNP were first described as a major group of chromatin-associated RNA-binding proteins. These proteins contain RNA-binding motifs that possess different RNA sequence binding preferences (Dreyfuss *et al.*, 1993). In human hnRNP complexes more than 30 proteins have been identified and traditionally their principal function was described as packaging of nascent pre-mRNAs in order to protect and organise them in a nucleosome-like structure. In recent years a more dynamic role of hRNP proteins has been suggested. These proteins may be viewed as a subset of the trans-acting factors involved in pre-mRNA processing and some of them also act in nucleus-cytoplasm mRNA export (Krecic and Swanson, 1999; Weighardt *et al.*, 1996).

Several intronic splicing enhancers and repressors associate with hnRNPs. One of the best-analysed hnRNP proteins is hnRNP A1. It was observed that in splicing extracts the relative ratio between A1 and the SR protein SF2/ASF determines the usage of duplicated 5' splice sites in a β-globin construct (Mayeda *et al.*, 1994).

Another well-studied protein of this family is hnRNP I, best known as PTB (Polypyrimidine Tract Binding Protein). This protein is involved in the regulation of alternative splicing of several genes. PTB recognises as its optimal binding site, the RNA sequence UCUU within a polypyrimidine rich context (Singh *et al.,* 1995) that sometimes overlaps with U2ASF binding site and simple competition could account for the inhibitory action of PTB.

## Part 3. Set-up of Project

From this review it is clear that a series of problems exist relating to NF1 and need to be addressed.

1. Molecular Diagnostics: Although the application of the use of a cascade of techniques allows a useful mutation detection rate, the extremely time-consuming nature of the work makes them impractical for routine clinical or scientific work. An alternative strategy based on automated sequencing techniques would clearly be a more sensitive and informative method for mutation testing. However these techniques are not widely used as a first screen since analysis of the results can be both cumbersome and time-consuming, and with current methodologies raw sequence data often requires extensive manipulation before analysis. The Cambridge Molecular Genetics laboratory have previously described a novel analytical approach, comparative sequence analysis (CSA) to overcome this problem (Mattocks *et al.*, 2000). Analysis software associated with automated DNA sequence generates a single electrophoretogram comprising four different coloured traces, each representing a different base. Interpretation of this data then requires manual checking of each individual base and comparison with the equivalent base in a known 'normal' sequence. CSA splits the sequence data into its four component traces, each representing the distribution of a single base within the sequences fragment. Each of these four traces are then overlaid on the corresponding 'normal' trace and mutations clearly stand out as peaks with no analogous partner in the 'normal' trace-
**Figure 8**

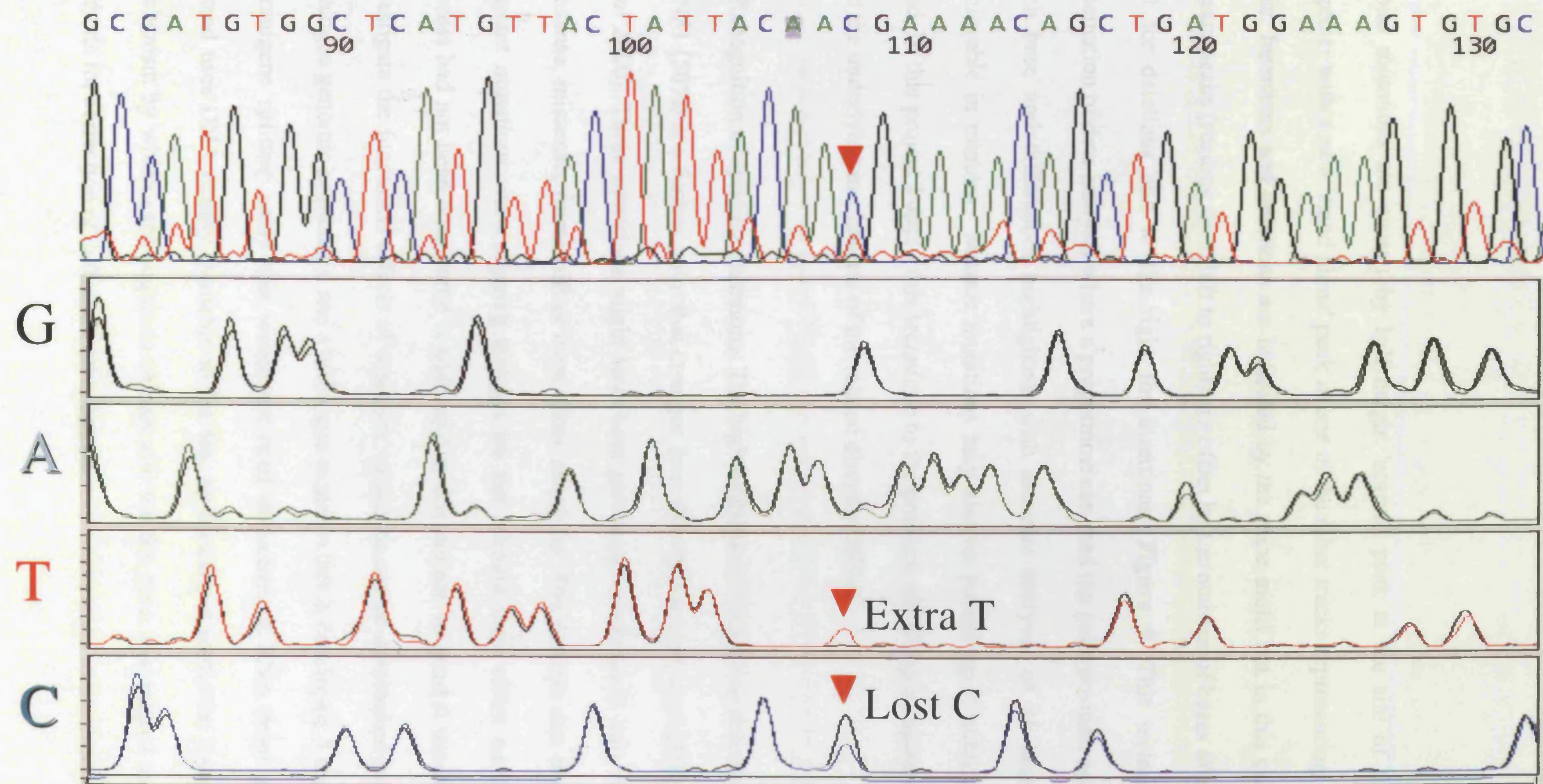



**Figure 8** Example of mutation detected by CSA. Normal sequence indicated in black, test sample indicated in colour. The base represented by each track is indicated in the top left of each panel. A heterozygote point mutation is seen C>T. The half-height peak in the C trace and the extra peak in the T trace can be easily seen. These differences can be read by the software program ACSA.

Point mutations are shown by half height 'normal' peak at the site of the mutation together with a new 'stand alone' peak in one of the other tracks representing the mutated base. Insertions and deletions are indicated by the same motif but in this case all down stream peaks (reading from left to right) are offset by the number of bases affected; to the left for deletions and to the right for insertions- **Figure 9**. This technique allows automation of data analysis where a programme can read the peak position and height for each base and differences highlighted, with accurate analyses of 96 lanes therefore achievable in minutes. Mosaic mutations may also be picked up if automation is not used. In this project I apply this technique to the problem of NF1 gene mutation analysis and the underlying mechanism of the related disorder, NFNS.

2. Recognition of splicing mutations: The high incidence of mutations that affect splicing in NF1 (50%) is a further point that emerges from this review (Ars, *et al* 2000; Messiaen, *et al* 2000). These mutations might have been previously erroneously only classified as nonsense, missense, frameshift or even silent mutations. The concern that diagnostically relevant mutations were slipping through the net because their effect on the splicing process had not been considered is addressed in this project. In general terms, studies to investigate the functional effects of sequence variations can be overlooked in the rush to catalogue genomic sequences, and a technique to assess this is developed. I test the use of a minigene splicing assay that would not need extraction of RNA from patients but instead uses DNA readily available to the lab, by looking at particular mutations. The mechanism by which each sequence change affected the gene was studied and potential methods for correction of defect are investigated.

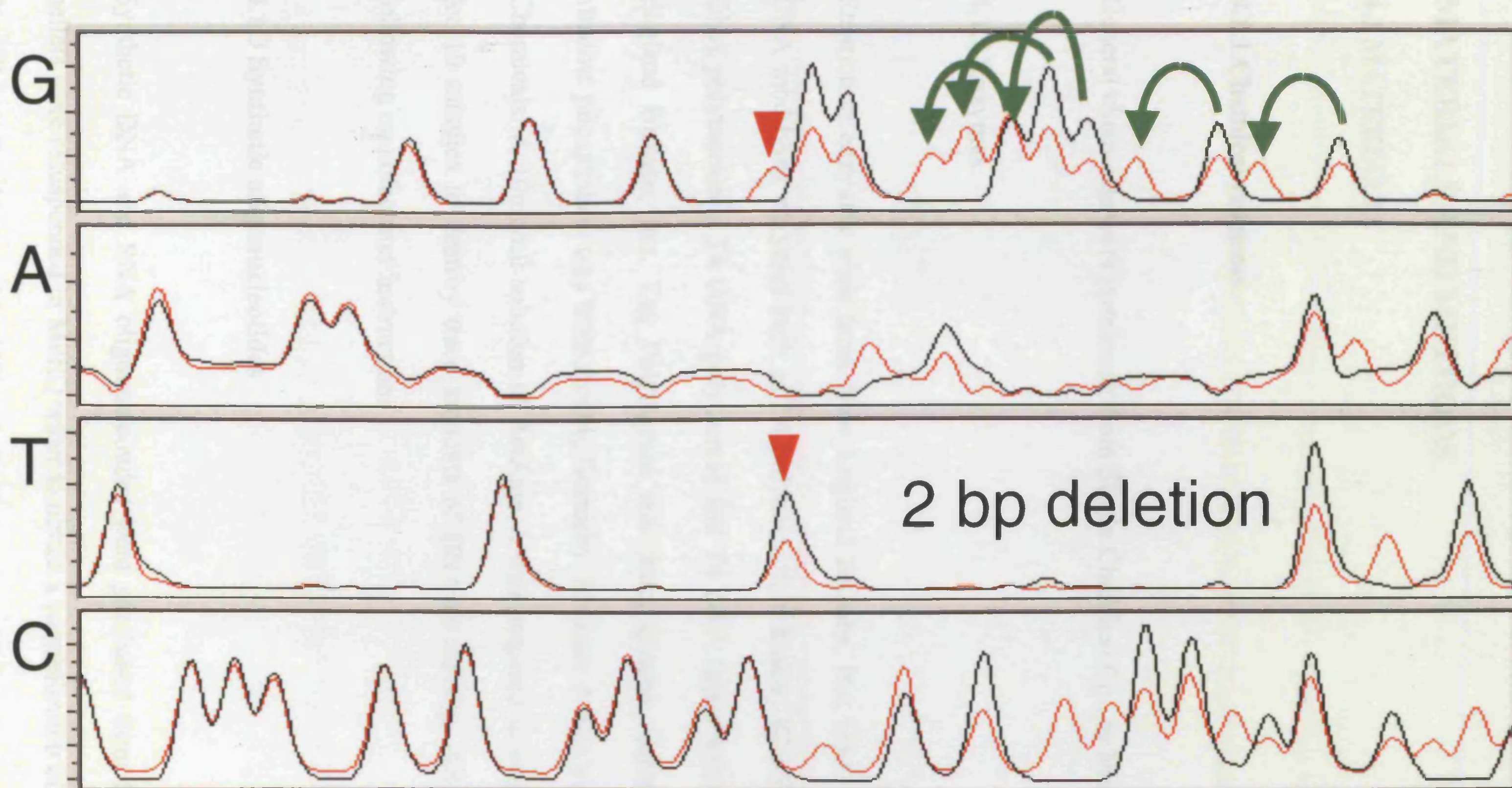



**Figure 9**. Frame shift mutations in CSA. The same two indicators are seen at point of mutation, and downstream bases are offset by number of bases involved.

# MATERIALS AND METHODS

## 4.1 MATERIALS

### 4.1.1 Chemical reagents

General chemicals were purchased from Sigma Chemical Co., or Merck.

### 4.1.2 Enzymes

Restriction enzymes were from New England Biolabs, Inc, USA or Roche, Germany. DNA modifying enzymes such as T4 polynucleotide kinase, Klenow fragment of *E.coli* DNA polymerase I, T4 DNA polymerase and T4 DNA ligase were obtained from New England Biolabs, Inc. Taq Polymerase was from Qiagen, Germany. Calf intestinal alkaline phosphatase was from Roche, Germany. RNAase A was purchased from Sigma Chemicals. A 10mg/ml solution of RNAase A was prepared in sterile water and boiled for 10 minutes to destroy trace amounts of DNAase activity. All enzymes were used following manufacturer instructions.

### 4.1.3 Synthetic oligonucleotides

Synthetic DNA and RNA oligonucleotides were purchased from Sigma-Genosys. The pellets were resuspended in MilliQ water to obtain a concentration of 1µg/µl.

The oligonucleotides used in this work are (bases in bold represent the relevant restriction

enzyme sites or the base mutated to achieve complementarity/site directed mutagenesis):

- NF1/exon3(NdeI): sense 5'GGAATTC**CATATG**TCAAGATTCTGGTACAGGTC3'

  antisense 5'GGAATT**CATATG**TCTCAAGGTAACATCTATCC3'

- Oligo pUC19: sense 5'TAAAACGACGGCCAGT3'

  antisense 5'ACAGCTATGACCAT3'

- Ori insert: sense 5'AGCGTTATGGCCATGATGTTGTC3'

  antisense 5'GATCAACAACAATAGTCTTCCAT3'

- Oligo PTB: sense 5'CAACTTCAACTCCTAAGCCACTGC3'

  antisense 5' TAGGATCCGTCACCAGGAAGTTGGTTAAATCA3'

- Oligo snRNP U1: sense 5'GATCTCATAGTTACCTGGCAGGGGAGATACCAT3'

  antisense 5'GATCATGGTATCTCCCCTGCCAGGTAACTATGA3'

- Oligo NF1/exon 3 wild-type -2G>A: sense 5'GTCTTGCTGAGGTAAGTAAATTG3'

  antisense 5' ATTTACTTACCTCAGCAAGAC3'

- Oligo NF1/exon 3 G>C,-2G>A: sense 5'GTCTTGCTGAGGTAACTAAATT3'

  antisense 5' ATTTAGTTACCTCAGCAAGAC3'

- Oligo NF1/exon 3 wild-type -2G>T: sense 5' GTCTTGCTGTGGTAAGTAAATTG3'

  antisense 5' ATTTACTTACCACAGCAAGAC3'

- Oligo NF1/exon 3 G>C,-2G>T: sense 5' GTCTTGCTGTGGTAACTAAATT3'

  antisense 5'ATTTAGTTACCACAGCAAGAC3'

- Nf1/exon37 (Bam HI/Xho I):

  sense 5'ACGT**GGATCC**ATGGAAGAGACCAAGCAAGTTTTGA3'

  antisense 5' TGCA**CTCGAG**TCACTTGTCATTGAATATACGGAGA3'

- NF1 exon 37 sequencing: sense 5'TGTATTAGCAAACGAGTGTC3'

5'GTGCAAGTGTCTGGACCAGTGG3'

5'CCAGTGGACAGAACTAGCTC3'

5'CCTTGGAAACAGTCACAGAAGC3'

5'GCTAGCTACCAAGATCACCA3'

antisense 5'TCCTTAATGAAGTAATCAGAG3'

- Oligo external NF1/exon 37: sense 5'GCTAGCTACCAAGATCACCA3'

  antisense 5'ACAGTACTTGGCAATAGCAGATAA3'

- Oligo A6791T: sense 5'CCTGACACTTTCAACAGTCAAGTTC3'

  antisense 5'GAACTTGACTGTTGAAAGTGTCAGG3'

- Oligo A6793T: sense 5'CCTGACACTTACTACAGTCAAGTTC3'

  antisense 5'GAACTTGACTGTAGTAAGTGTCAGG3'

- Oligo A6794T: sense 5'CCTGACACTTACATCAGTCAAGTTC3'

  antisense 5'GAACTTGACTGATGTAAGTGTCAGG3'

- Oligo C6795G: sense 5'CCTGACACTTACAAGAGTCAAGTTC3'

  antisense 5'GAACTTGACTCTTGTAAGTGTCAGG3'

- Oligo C6792T/A6793T: sense 5'CCTGACACTTATTACAGTCAAGTTC3'

  antisense 5'GAACTTGACTGTAATAAGTGTCAGG3'

- Oligo C6792T/A6794T: sense 5'CCTGACACTTATATCAGTCAAGTTC3'

  antisense 5'GAACTTGACTGATATAAGTGTCAGG3'

- Oligo pcDNA3: sense 5'ACAGAAGCTTTGTTGGAGATC3'

  antisense 5'ATTTAGGTGACACTATAGAATAG3'

- Oligo exon 28-Nde: sense 5'**GGGAATTCCATATG**GAGATTTGTCTCTTCTC3'

  antisense 5' **GGGAATTCCATATG**CCTTACGTGACATTTTATAC3'

- Oligo exon 6-Nde: sense 5'**GGGAATTCCATATG**GGATAGATCAAGATAGCTC3'

antisense 5'GGGAATTCCATATGCCTCTGCCAAATAAAATTATG3'

### 4.1.4 Plasmid constructs

#### 4.1.4.1 Commercial plasmids used

The plasmid used in this work was pUC19 (New England Biolabs), the cloning vector which consists of 2686 basepairs. It contains a polylinker sequence of 54 bp within a *lac Z* gene. It contains an ampicillin resistance gene.

#### 4.1.4.2 Minigene system construct

This system allows the pre mRNA splicing reaction to be studied. In all cases, except the exon 37 studies, a modified version of the EDB α-globin-fibronectin minigene was used (Muro *et al.*, 1999). The EDB exon subject to alternative splicing is removed to create an insertion site for the sequence of interest and is represented by a unique Nde I cleavage site. For the exon 37 studies pcDNA3 (Invitrogen) was used linearised with BamHI and Xho I enzymes.

### 4.1.5 Bacterial culture

The K12, *E. coli* family strain DH5α was used for transformation by plasmid and growth of the plasmid.

They were maintained in the short term as single colonies on agar plates at 4°C and for longer term storage were kept as glycerol stocks, made by adding sterile glycerol to a final 15% v/v. Glycerol stocks were stored at -20°C. An overnight culture of bacteria was grown in LB medium.

LB: Luria-Bertani medium per litre:

Difco Bactotryptone 10 g, Oxoid yeast extract 5g, NaCl 10g, pH 7.5.

Bacterial growth media were sterilised before using by autoclaving. The appropriate antibiotic, ampicillin, was added to media to a final concentration of 200 μg/ml.

### 4.1.6 Cell culture

The following cell lines were used:

Hep3B; human, hepatocellular carcinoma, fibroblast-like.

HeLa; human, cervix, epithelial, adenocarcinoma.

Cos 7; cercopitheous aethiops (Monkey, African green), kidney; SV40 transformed, fibroblast-like.

## 4.2 METHODS

### 4.2.1 Nucleic Acid Preparations

#### 4.2.1.1 Small-scale preparation of plasmid DNA from bacterial cultures

Purification of small amounts of recombinant plasmid DNA was performed using the Qiagen miniprep kit, following the manufacturer's instructions. The procedure is based on alkaline lysis of bacterial cells followed by adsorption of DNA onto silica in the presence of high salt concentrations. After washing and elution of plasmid DNA the final pellet was resuspended in 50 μl of $dH_2O$ and 5 μl of such preparations were routinely taken for analysis by restriction enzyme digests.

#### 4.2.1.2 Large-scale preparations of plasmid DNA from bacterial cultures

For large-scale preparations of plasmid DNA, the Qiagen Midi kit was used following manufacturers instructions. This was achieved with overnight inoculum of 200mls using LB medium. The protocol is based on a modified alkaline lysis procedure, followed by binding of plasmid DNA to a Qiagen Anion-Exchange resin under appropriate low-salt and pH conditions. RNA, proteins, and low-molecular weight impurities were removed by a medium-salt wash. Plasmid DNA was eluted in a high-salt buffer and then concentrated and desalted by isopropanol precipitation. The concentration of DNA produced was quantified using mass spectrometry and stored at -20°C.

#### 4.2.1.3 Preparation of RNA from the cultured cells

RNeasy kits (Qiagen) were used to isolate total RNA from cultured cells. This technology uses a silica-gel-based membrane to bind RNA. The cells are first lysed and homogenised in the presence of a highly denaturing guanidine isothiocyanate (GITC)-containing buffer. Ethanol is added to provide appropriate binding conditions, and the sample is applied to an RNeasy mini column where total RNA binds to the membrane. The manufacturers protocol was followed throughout. The RNA is eluted in water and stored at -80°C.

The RNA quality was checked by electrophoresis on 0.8% agarose gels and the RNA concentration quantified by mass spectrometry.

#### 4.2.1.4 cDNA synthesis from RNA

First strand cDNA was synthesised using a Moloney Murine Leukemia virus (M-MulV) reverse transcriptase kit from Amersham. Preassembled bulk first strand cDNA reaction mixes, DTT and pd(N)$_6$ primer were used. The manufacturer's instructions were followed: the RNA solution was heated to 65°C for 10 minutes and chilled on ice, and the appropriate volume of reaction mix, DTT, primer was added. The mixture was incubated at 37°C for 1 hour. 1.5-3 μl of the cDNA reaction mix was used for the PCR analysis.

#### 4.2.1.5 Agarose gel electrophoresis

Gels for DNA were prepared by dissolving agarose (Sigma) in hot TBE buffer at concentrations of 0.8-2% (w/v) and adding 0.5μg/ml ethidium bromide (Sigma). The resulting solution was left to set (30 min, rt). DNA samples were mixed with one-quarter volume gel loading solution (Sigma) before being loaded into the gel and electrophoresed. DNA was visualized under UV transillumination

### 4.2.2 Estimation of Nucleic Acid Concentration

#### 4.2.2.1 Spectrophotometric

An optical density of 1.0 at 260 nm is roughly equivalent to a concentration of 50 μg/μl for double stranded DNA, 33 μg/μl for single stranded DNA and 40 μg/μl for RNA samples. The ratio of values for optical densities measured at 260 nm and 280 nm is 1.8 for pure sample of DNA and 2 for RNA. These are reduced by protein contaminants and therefore the values were used to assess both the concentration and the purity of the samples.

### 4.2.3 Enzymatic Modification of DNA

#### 4.2.3.1. Restriction enzymes

Restriction endonucleases recognise and cut within specific sequences of double stranded DNA leaving blunt ends, or 5' and 3' protruding ends. These were used in the construction and analysis of recombinant plasmids. Each restriction enzyme functions optimally in a buffer of specific ionic strength. All buffers were supplied by the same company that supplied the enzymes and were used according with the manufacturer's instructions.

For analytical digests 100-500 ng DNA were digested in a volume of 10-20 μl containing 5 U of the appropriate restriction enzyme per μg DNA. The digest was incubated for 3-6 hours at the optimal temperature required by the enzyme used.

Preparative digests were made of 1-20 μg DNA using the above conditions but a larger reaction volume. Enzymatic activity was then removed either by incubation at 75°C for 20 minutes or phenol-chloroform extraction.

#### 4.2.3.2. Large fragment of *E. Coli* Polymerase I

The large fragment of DNA Polymerase I (Klenow) lacks the 5' to 3' exonuclease activity of the intact enzyme, but retains 3' to 5' exonuclease activities. It is used to catalyse the polymerisation of deoxyribonucleotide triphosphates in a 5' to 3' direction on a template of double stranded DNA with a recessed 3' hydroxyl and protruding 5'

phosphoryl terminus to generate a flush-ended DNA molecule. This was useful for creating compatible ends for ligation during construction of recombinant plasmids and to digest specific A residues added by Taq DNA polymerase at the 5' terminus.

Klenow fragment was used with the proper buffer supplied by New England Biolabs Inc. at a final concentration of 5 U per μg DNA. When a "fill-in" was required (DNA fragments with protruding 3' ends) dNTP with a final concentration of 0.5 mM were added. The mixture was incubated at 37°C for 20 minutes.

#### 4.2.3.3 Dephosphorylation of DNA 5' termini

Calf intestinal phosphatase catalyses the removal of 5' terminal phosphate groups from linear DNA molecules. This is used to reduce self-ligation of vector DNA during generation of recombinant clones, thus increasing the proportion of resulting recombinants containing the required DNA inserts.

This reaction was carried out in a final volume of 50-100 μl using 1U of enzyme per 0.5 μg DNA incubating for 1 hour at 37°C. The enzyme was then inactivated by adding 1 mM EDTA and incubating for 20 minutes at 75°C.

#### 4.2.3.4 T4 DNA ligase

T4 DNA ligase catalyses the formation of a phosphodiester bond between adjacent 3' hydroxyl and 5' phosphoryl termini in DNA, requiring ATP as a cofactor in this reaction. This enzyme was used to join double stranded DNA fragments with compatible sticky or

blunt ends, during generation of recombinant plasmid DNA.

20 ng of linearised vector were ligated with a 5-10 fold molar excess of insert in a total volume of 20 μl containing 1x ligase buffer and 1 U T4 DNA ligase. The reaction was carried out at room temperature for 3-4 hours in most cases and at 16°C for blunt- ended ligations.

#### 4.2.4 Elution and purification of DNA fragments from agarose gels

This protocol was used to purify small amounts (less than 1 μg) of DNA for subcloning. The DNA samples were loaded onto an agarose minigel and electrophoresed as described 10.2.1.5.

The DNA was visualised with UV light and the required DNA fragment band was excised from the gel. A Gel Extraction kit was used (Qiagen, Germany). This kit uses a silica-membrane to adsorb DNA from agarose gel in the presence of high salt concentrations while contaminants pass through the column. The DNA is eluted with Tris buffer or water. The manufacturer's protocol was followed.

#### 4.2.5 PCR reactions

##### 4.2.5.1 PCR reactions for cloning purposes.

The polymerase chain reaction was performed on genomic or plasmid DNA following the basic protocols of the Boehringer and Promega Taq DNA Polymerases. The volume of the reaction was 50-100 μl with 1x Taq buffer, dNTP mix 200 μM each, oligonucleotide

primers 100 00nM each, Taq DNA Polymerase 2.5-5 U. 0.1 ng for plasmid and 100 ng of genomic DNA were used for amplification. The amplification conditions varied for each particular PCR but normally comprised a hot start (94°C, 3 min) followed by the denaturing step (94°C, 1 min), annealing (45-65°C, 1 min) and extension (72°C, 1 min) The denaturing-annealing-extension cycle was repeated 30-35 times. The amplifications were performed on a ThermoHybaid PCR Express or PTC-100 programmable thermal controller.

#### 4.2.5.2 PCR reactions for mutation analysis

Genomic DNA was extracted from peripheral blood using the Nucleon BACC2 kit and re suspended in TE buffer to a concentration of approximately 500μg/ml. The entire reading frame of the NF1 gene was PCR amplified exon by exon, in a total of 58 amplicons. In general primers were designed so that each amplicon comprised an exon flanked by at least 50 base pairs of intron on either side. One primer in each pair was tailed with an M13 primer sequence (M13F: TGTAAAACGACGGCCAGT or M13R: CAGGAAACA GCTATGACC) to provide a binding site for M13 universal sequencing primer. Care was taken to ensure that any repetitive sequences flanking the exon lay at the opposite end of the amplicon with respect to the M13 sequencing primer sequence. In general the PCR temperature profile was: Initial denaturation at 95°C for 2 minutes, followed by 30 cycles with annealing for 30 seconds, elongation at 72°C for 1 minute, and denaturation at 95°C for 30 seconds. Primer sequences, annealing temperatures and buffers used are given in **Table 3.**

## Table 3: Primers and conditions for NF1 PCRs.

| Forward primer | | Reverse primer | | Length | Ta | Buffer | Seq Primer | Orientation |
|---|---|---|---|---|---|---|---|---|
| 1F.A | AGACCCTCTCCTTGCCTCTT | 1R.0 | M13RX - ATGGAGGGTCGGAGGCTG | 458 | 55.0 | D | M13R | Reverse |
| 2F.1 | TGGCAAGTAAGTTATTTATGGTC | 2R.1 | M13F - AATCAAAAAGAAAAGAAAGCAA | 288 | 55.0 | B | M13F | Reverse |
| 3F.1 | GTTTGCCTTAGACTTTAGTTTTT | 3R.1 | M13F - CATCTGTACTTTGGGACATAA | 368 | 55.0 | B | M13F | Reverse |
| 4aF.0 | TTTGAAAATTTTCATAATAGAAAATGT | 4aR.0 | M13RX - GGTCAAAGCTGCTGTGAG | 435 | 55.0 | C | M13R | Reverse |
| 4bF.0 | M13FX - TCCTGGCCTCAAGTGGTC | 4bR.0 | M13RX - TTATAAAANCCAGATTGGTGTTC | 386 | 57.0 | B | M13F | Forward |
| 4cF.0 | M13FX - TCCTAGCAGACAACTATCGA | 4cR.0 | AAAAAAAAATCAATCGTATCCTTA | 582 | 55.0 | C | M13F | Forward |
| 6F.0 | M13FX - CATGTTTATCTTTTAAAAATGTTGCC | 6R.0 | ATAATGGAAATAATTTTGCCCTCC | 323 | 55.0 | C | M13F | Forward |
| 7F.0 | M13FX - TGCTATAATATTAGCTACATCTGG | 7R.0 | CCTATGAACTTATCAACGAAGAG | 395 | 55.0 | B | M13F | Forward |
| 8F.0 | M13FX - TTGTGTGGGTAATGTGTTGA | 8R.0 | AAATATAGTTAGATAAAAACCAATG | 270 | 55.0 | B | M13F | Forward |
| 9F.1 | TATTTGCTGTTCTTTTTGG | 9R.1 | M13F - AATTTAGCAATACCTTTTGG | 204 | 55.0 | B | M13F | Reverse |
| 9aF.0 | TCCGCTGTGGCTCAGAACAC | 9aR.0 | M13RX - AGTAGAAGAGGATGCACAGCC | 315 | 60.0 | B | M13R | Reverse |
| 10aF.0 | M13FX - GATAAACAGAGCATACAACTCA | 10aR.0 | AAATGCAATAGAAAGGAGGTGA | 270 | 55.0 | B | M13F | Forward |
| 10bF.A | AACTATTGAGTGTTCTACTATACC | 10bR.0 | M13RX - TTGGCGATTCAGCTAAACC | 374 | 55.0 | G | M13R | Reverse |
| 10cF.1 | M13F - CGTCCAGCCTAGTTCTAGAA | 10cR.1 | ACCCCTTCTTTCTCCAT | 444 | 55.0 | C | M13F | Forward |
| 11F.0 | M13FX - ATAAGTACTCCAGTGTTATGT | 11R.0 | TAAAGTTGAAATTTAAAAAATTAAAGTAC | 179 | 52.0 | C | M13F | Forward |
| 12aF.0 | M13FX - AAACCTTACAAGAAAAACTAAGCT | 12aR.0 | ATTACCATTCCAAATATTCTTCCA | 325 | 55.0 | C | M13F | Forward |
| 12bF.0 | M13FX - TTTCTAGTAAATCTCCTTCAAGT | 12bR.0 | ATGAAATTTACCAAATTTCATTCAG | 404 | 55.0 | C | M13F | Forward |
| 13F.1 | AAGATATTTTGGGGTTTGAA | 13R.1 | M13R - GCCATGTGCTTTGAGG | 508 | 55.0 | B | M13R | Reverse |
| 14F.0 | M13FX - AGCTTATCAGGTTCTCCATTGG | 15R.0 | M13R - AGTTAACAGACAAAAGTCAACTTTACAG | 535 | 55.0 | D | M13F | Forward |
| 14F.0 | M13FX - AGCTTATCAGGTTCTCCATTGG | 15R.0 | M13R - AGTTAACAGACAAAAGTCAACTTTACAG | 535 | 55.0 | D | M13R | Reverse |
| 16F.0 | M13FX - TGGATAAAGCATAATTTGTCAAGT | 16R.0 | TAGAGAAAGGTGAAAAATAAGAG | 571 | 55.0 | B | M13F | Forward |
| 17F.0 | CTGTGTGTTTAGATCAGTCA | 17R.0 | M13RX - TTTATCAATTACTACCAGTATCAG | 339 | 55.0 | B | M13R | Reverse |
| 18F.0 | AGAAGTTGTGTACGTTCTTTTCT | 18R.0 | M13RX - TCCTTTCTACCAATAACCGC | 388 | 55.0 | I | M13R | Reverse |
| 19aF.0 | M13FX - ATGTCACTTAGGTTATCTGG | 19aR.0 | TGTAATTAAGTAGTTATAACTCTC | 290 | 60.0 | B | M13F | Forward |
| 19bF.1 | M13F - ATTTGAGGGGAAGTGAA | 19bR.2 | M13R - ATGGGTATTATTTCTTTGCT | 312 | 55.0 | E | M13R | Reverse |
| 20F.0 | M13FX - CATTCACACCATGCACATATGATTG | 21R.0 | M13R - GATTTGCTATGTGCCAGGCAC | 703 | 55.0 | E | M13F | Forward |
| 20F.0 | M13FX - CATTCACACCATGCACATATGATTG | 21R.0 | M13R - GATTTGCTATGTGCCAGGCAC | 703 | 55.0 | E | M13R | Reverse |
| 22F.0 | M13FX - TGCTACTCTTTAGCTTCCTAC | 22R.0 | ACAGCGGTTCTATGTGAAAAG | 537 | 55.0 | B | M13F | Forward |
| 23-2F.0 | M13FX - CTTAATGTCTGTATAAGAGTCTC | 23-2R.0 | ACTTTAGATTAATAATGGTAATCTC | 290 | 55.0 | C | M13F | Forward |
| 23aF.0 | CAGAAATAGTATACATGATTGGGT | 23aR.0 | M13RX - CTATTTTCTGCCAGAATTAGTAGA | 465 | 55.0 | C | M13R | Reverse |
| 24F.0 | M13FX - TTGAACTCTTTTGTTTTCATGTCTT | 24R.0 | GGAATTTAAGATAGCTAGATTATC | 289 | 55.0 | C | M13F | Forward |
| 25F.0 | M13FX - AATATAATAATTATATTTGGGAAGGT | 25R.0 | GAAAATATTTGATTCAAACAGAGC | 360 | 55.0 | B | M13F | Forward |
| 26F.1 | M13F - CCCTCCATATTTGTAATCTT | 26R.1 | TTAAACGGAGAGTGTTCAC | 324 | 58.0 | B | M13F | Forward |
| 27aF.1 | AGGGCATTTTAATCTTTTAT | 27aR.1 | M13F - GCAAACTCTCCTTCTCAAC | 406 | 55.0 | F | M13F | Reverse |
| 27bF.1 | M13R - GCCAAATTACCCTTTAGAATG | 27bR.1 | AACCAAACTTGCCATCTCT | 348 | 55.0 | C | M13R | Forward |
| 28F.1 | M13F - GACTTTGAAGAATTGTTTTATA | 28R.0 | M13RX - CTATAGGGGTAGGACACC | 693 | 55.0 | E | M13F | Forward |
| 28F.1 | M13F - GACTTTGAAGAATTGTTTTATA | 28R.0 | M13RX - CTATAGGGGTAGGACACC | 693 | 55.0 | E | M13R | Reverse |
| 29F.1 | M13F - TACAATGGTGGGAACTCTTCCTTA | 29R.0 | M13RX - ACCAACACTGCATACCTTCCAATT | 635 | 55.0 | E | M13F | Forward |
| 29F.1 | M13F - TACAATGGTGGGAACTCTTCCTTA | 29R.0 | M13RX - ACCAACACTGCATACCTTCCAATT | 635 | 55.0 | E | M13R | Reverse |
| 30F.0 | GCCTCACAGTGCTCTTATGG | 30R.0 | M13RX - CCTTCAGAGTTCCTTCAGTTAC | 628 | 55.0 | C | M13R | Reverse |
| 31F.0 | TGTTGATGTGATTTTCATTGACCA | 31R.0 | M13RX - CCAATGTGGCACCAGATAAATAT | 337 | 55.0 | B | M13R | Reverse |
| 32F.0 | M13FX - ATCTAGTATTTTTGAGGCCTCAG | 32R.0 | CAGATATGCTATAGTACAGAAGG | 334 | 55.0 | C | M13F | Forward |
| 33F.0 | M13FX - TATCTGTTTTATCATCAGGAGG | 33R.1 | M13R - TAAAATGGAGAAAGGAACTGG | 497 | 55.0 | E | M13R | Reverse |
| 34F.1 | M13F - AAAATGAAACATGGAACTTTAG | 34R.0 | M13RX - TAAGCATTAAGTACAAAATAGCAC | 439 | 55.0 | E | M13F | Forward |
| 35F.0 | ATGTATTCAGAGTATCCCCTT | 35R.0 | M13RX - ATTTAAACGTCTTTTAGAAAAATG | 458 | 55.0 | C | M13R | Reverse |
| 37F.0 | M13FX - ATTCCGAGATTCAGTTTAGGAG | 37R.0 | AGTAACATTCAACACTGATACCC | 256 | 55.0 | H | M13F | Forward |
| 38F.1 | ACATGGGTAATTTAGGAAGATAAG | 38R.1 | M13F - CAACAAGAAAAGATGGAAGA | 335 | 55.0 | C | M13F | Reverse |
| 39F.0 | CTACTGTGTGAACCTCATCAACC | 39R.0 | M13RX - GTAAGACATAAGGGCTAACTTACTTC | 306 | 55.0 | I | M13R | Reverse |
| 40F.0 | GGGAAGAAGACCTCAGCAGAT | 40R.0 | M13RX - AACTTTCTGCTCTGCCACGCA | 345 | 55.0 | I | M13R | Reverse |
| 41F.0 | TTCATCCTGTTTTAAGTCACACTTG | 41R.0 | M13RX - TTGCCTCCATTAGTTGGAAAATTG | 295 | 55.0 | H | M13R | Reverse |
| 42F.0 | GAAGGAGCAAACGATGGTTG | 42R.0 | M13RX - AAACTTTGCTACACTGACATGG | 371 | 55.0 | C | M13R | Reverse |
| 43F.0 | GTATTAGAGCTTTCTTTGAGTC | 43R.0 | M13RX - GAATGGTAGAGTTTATCTTTTG | 421 | 55.0 | C | M13R | Reverse |
| 44F.2 | M13R - GGTGAAGTGATTATCCAGGTG | 44R.2 | TGTTATCTGTTTGGGTGAGAAG | 328 | 55.0 | E | M13R | Forward |
| 45F.0 | M13R - AGATAACAATTCAGCCACAAAG | 45R.0 | TAAAGACAGGCACGAAGGT | 225 | 55.0 | F | M13R | Forward |
| 46F.1 | M13R - ATTTTGGCACATTATTCTGG | 46R.1 | GCGCATGTTAGCAAGTT | 316 | 55.0 | C | M13R | Forward |
| 47F.0 | CTGTTACAATTAAAAGATACCTTGC | 47R.0 | M13RX - GTGTTCTTAAAGCAGGCATAC | 204 | 55.0 | H | M13R | Reverse |
| 48F.0 | TTTTGGCTTCAGATGGGGATTTAC | 48R.0 | M13RX - AAGGGAATTCCTAATGTTGGTGTC | 373 | 55.0 | B | M13R | Reverse |
| 48aF.0 | ATCTAGTATCTAATTGTATTTCACCC | 48aR.0 | M13RX - AGACTGAGCTTACAGGGAC | 265 | 55.0 | A | M13R | Reverse |
| 49F.0 | M13FX - CTGGGAGAAACAGGCTATAC | 49R.0 | CACTTTCTTTGCAGTTGTTCTG | 639 | 58.0 | E | M13F | Forward |

### 4.2.6 Sequence analysis

#### 4.2.6.1 Sequence analysis for plasmid DNA

Sequence analysis of plasmid DNA and inserts was performed by the sequencing facility of the Cambridge University Biochemistry Department, using ABIprism sequencing machines (Perkin Elmer).

#### 4.2.6.2 Sequencing for NF1 mutation analysis

PCR products were purified by adding 1 U shrimp alkaline phosphatase (Amersham) and 10U exonuclease I (NEB) to each reaction and incubating at 37°C for 30 min followed by 95°C for 10 min to denature the enzymes. Purified PCR products were then sequenced in a single orientation dependent on the location of the M13 tail on the PCR primers. Reactions comprised 2 µl of PCR product, 0.7 µl BigDye V sequencing reaction mix (Applied Biosystems), 1.05 µl 5x sequencing diluent (Applied Biosystems), 0.35 µl primer at 10 pmol/µl, 5% (v/v) betaine (Sigma) and dd$H_2O$ to 7µl. Initial denaturation was at 95°C for 5 min, followed by 25 cycles with annealing at 50°C for 15 sec, elongation at 60°C for 4 min, and denaturation at 95°C for 30 sec. For batches D02 and D04 completed sequencing reactions were purified using Dynapure™ Dye Terminator Removal Ver.2 (Dynal A.S, Oslo, Norway) according to the manufacturer's protocol. Products were resuspended in 1.5 µl formamide and 0.8 µl of this was dotted on 96 well RapidLoad™ membrane combs (Web Scientific Crewe, UK). The ABI 377 was loaded

using the 'water protocol' as described by Web Scientific. For batch D05 four plates of sequencing reactions were transferred to a single Multiscreen 384 well filter plate (Millipore, Bedford, UK) and purified according to the manufacturer's protocol. The samples were then re suspended in 15 µl Hi-Di formamide (Applied Biosystems) with 0.05 µl GeneScan™ 500 LIZ size standard (Applied Biosystems) per reaction and transferred to a 384 MicroAmp optical plate (Applied Biosystems) for loading on the ABI 3100.

#### 4.2.6.3 Sequence analysis

Analysis was carried out using an automated comparative sequence analysis (CSA) programme AutoCSA12. This is described in (Mattocks, *et al* 2004). In brief, the technique involves horizontal alignment of each of the four base traces separately with a control trace. The programme then calculates peak position and height for each base and will highlight all differences. No electrophoretograms were examined directly unless variations were called by the programme. Called variations were examined in GeneScan V3 (Applied Biosystems) by non-automated CSA (Mattocks *et al.*, 2000). For batches D02 and D04 horizontal alignment was achieved by generating an artificial size standard applying arbitrary numbers to the peaks in one of the four traces (usually red C trace). As a fifth color became available batch D05 was horizontally aligned using the GeneScan™ 500 LIZ size standard. Sequence variations were confirmed by re-sequencing using a new aliquot of DNA.

### 4.2.7 Site directed mutagenesis

Site directed mutagenesis involved the substitution of a specific nucleotide by another. Synthetic mutagenic oligonucleotides are used for this method. Details of the primers used are listed in section 4.1.3.

### 4.2.8 Pull-down analysis-procedure of affinity purification of RNA-bound proteins

This was performed by Dr E Buratti at International Centre of Genetic Engineering and Biotechnology (ICGEB), Italy. His methods are described below.

#### 4.2.8.1 Preparation of RNA templates for pull-down analysis

RNA templates were obtained by amplifying the respective exon/intron sequences using a forward primer carrying a T7 polymerase target sequence (5'TACgTAATACgACTCACTATAg3') with 12 nucleotides complementary for the specific exon and a reverse primer carrying 18 nucleotides of the target sequence. The amplified products were then purified and approximately 1μg of DNA was transcribed using T7 RNA polymerase (Stratagene) as described elsewhere (Buratti and Baralle, 2001).

#### 4.2.8.2 Purification of RNA-bound protein complexes

500 pmole of RNA (approx. 15μg of T7 RNA polymerase-synthesized 100-mer RNA) are placed in a 400μl reaction mixture containing 100mM NaOAC pH=5.0 and 5mM sodium m-periodate (Sigma), incubated for 1 hour in the dark room at room temperature, ethanol precipitated, and resuspended in 100μl of 1.0 M NaOAC, pH 5.0. The RNA was mixed with 300μl of adipic acid dehydrazide agarose bead 50% slurry (Sigma) equilibrated in 100mM NaOAC pH 5.0 (with four 10ml washes) and the mix was incubated for 12 hours at 4°C on a rotator. The beads with the bound RNA were pelleted, washed 2 times with 1ml 2M NaCl and then 3 times with 1ml washing buffer (5mM HEPES pH 7.9, 1mM $MgCl_2$, 0.8 mM magnesium acetate). A protein mix (in binding buffer: 20mM HEPES pH 7.9, 72mM KCl, 1.5 mM $MgCl_2$, 0.78mM magnesium acetate, 0.52mM DTT, 3.8% glycerol, 0.75mM ATP and 1mM GTP, 0.5 mg/ml heparin) containing approximately 1-2mg of HeLa cell nuclear extract (C4, Biotech) was then added to the beads and incubated in a rotator for 30 minutes at room temperature. After incubation the beads were pelleted by centrifugation at 300rpm for 3 minutes and washed 4 times with 1.5 ml of washing buffer before addition of SDS sample buffer and loading on a 10% SDS-PAGE gel.

#### 4.2.8.3 Nanoelectrospray mass spectrometry and database homology search

Internal sequence analysis from the Coomassie blue-stained bands excised from the SDS-PAGE gel was performed using an electrospray ionization mass spectrometer 9LCQ

DECA XP, Thermo Finnigan). The bands were digested by trypsin, and the resulting peptides were extracted with water and 60% acetonitrile/1% trifluoroacetic acid. The fragments were then analysed by mass spectrometry, and the proteins were identified by analysis of the peptide MS/MS data with Turbo SEQUEST (Thermo Finnigan) and MASCOT (Matrix science).

### 4.2.9 Maintenance and analysis of cells in culture

Hep3B, Hela and Cos 7 cell lines were maintained in 100x20 mm Falcon tissue culture plates, incubated at 37°C and with 5% carbon dioxide. These cell lines were grown in Dulbecco's modified Eagle's medium supplemented with 4.5g/l glucose, 10% foetal calf serum, 50 µg/ml penicillin/streptomycin and 4 mM glutamine.

Plates containing a confluent monolayer of cells were passaged 1 in 3-5 with 1% trypsin as follows. Cells were washed with 5 ml PBS solution, then were incubated at 37 °C with 1 ml PBS/EDTA/trypsin solution (PBS containing 0.02% w/v EDTA and 2% v/v trypsin solution) for 2 minutes or until cells were dislodged. After adding 10 ml of media cells were pelleted by centrifugation in polypropylene tubes and resuspended in 5ml pre-warmed medium. 1-2 ml of this cell suspension was added to 10ml medium in fresh plate and was gently mixed before incubation. This procedure was required on average once every three days.

### 4.2.10 Transfection of recombinant DNA into cells maintained in culture

Cells were passaged as above into 6-well tissue culture plates and grown to a confluency of 40-70%.

The transfection was performed by Effectene non-liposomal transfection reagent (Qiagen). The Effectene reagent was used in conjunction with the enhancer and the DNA-condensation buffer (Buffer EC) to achieve transfection. In the first step of effectene-DNA complex formation, the DNA is condensed by interaction with the enhancer in a defined buffer system. The reagent is added to the condensed DNA to produce condensed effectene-DNA complexes. These are mixed with medium and directly added to the cells: 1 μg of expression plasmid was mixed with 0.1μg/μl of buffer EC to a total volume of 150μl. 8μl of enhancer was added and mixed. The mixture was incubated at room temperature for 2-5 min. The mixture was 25μl effectene transfection reagent was added to the DNA-enhancer mixture and mixed. The samples were incubated at room temperature for 5-10 minutes to allow transfection-complex formation. 1ml of growth medium was then added to the transfection complexes and added drop-wise onto the cells in the 60mm dishes. These were then swirled and incubated at 37° for up to 48 hours. The medium was then collected and the cells were washed with PBS. RNA isolation followed as described in section 4.2.1.3.

### 4.2.11 Measurement of transfection efficiency

48 hours after transfection, the medium was collected and a colorimetric enzyme assay for the quantitative determination of secreted human growth hormone was performed using hGH ELISA reagent kit (Roche Diagnostics GmbH, Manneim, Germany). The 200µl of diluted medium was loaded onto the microtitre plate and incubated for 1 hour at 37°C to allow binding of the protein to the plate. The unbound proteins were then removed by five washings with washing buffer. A 200µl sample of anti-hGH-DIG was added in each well following incubation of 1h at 37°. After five washings, 200µl of anti-DIG-POD antibody solution was added and incubated for 1h at 37°C. After repeated washings, POD substrate was added and photometric measurements were performed at 405nm.

### 4.3 Clinical methods and recruitment of patients

131 unrelated subjects were referred for NF1 mutation analysis mainly from the Addenbrooke's Hospital, Cambridge but also from other centres within East Anglia. All patients were examined for signs and symptoms of NF1 either by personal consultation or by careful examination of the medical notes. 91 patients fulfilled the NIH Consensus development conference criteria ((NIH), 1987). Other patients without the diagnosis of NF1 were included in the mutation analysis to act as controls. The NFNS patients were recruited either from Cambridge or GOS Genetics departments and examined by myself (Case 6) , Dr Baraitser or Dr Elmslie (all patients).

### 4.4 Fundamentals of the functional minigene splicing assay

The functional minigene splicing assay is a technique using hybrid minigenes with the exon/s of interest inserted into it. This system allows further study of the sequence in question and assessment of splicing. The minigene constructs are transiently transfected into eukaryotic cells and the splicing products obtained by RNA extraction and amplified by RT-PCR using oligonucleotides specific for the region of interest. The size of the cDNA bands obtained when analysed by agarose gel electrophoresis will be the sum of bases for each exon included in the transcript. Thus a calculation is made and this confirms exon skipping or inclusion. The minigene sytems used in this study are: PTB (a modified version of the α-globin-fibronectin EDB (Muro *et al.*, 1999; Niksic *et al.*, 1999) and pcDNA3 (Invitrogen), for small and large inserts respectively. With regards to PTB, part of a flanking EDB exon from an intronic region (from the intronic Nde I site -1 to intronic Xba I site +1) has been removed. The sequence of interest is then cloned into this unique Nde I site created (**Figure 10**). Transcription of this minigene is by use of the α-globin promoter and the SV40 enhancer. In pcDNA3, the sequence of interest is cloned into the polylinker site. Transcription is through a CMV promoter (**Figure 11**). The minigene assay is shown in **Figure 12.**

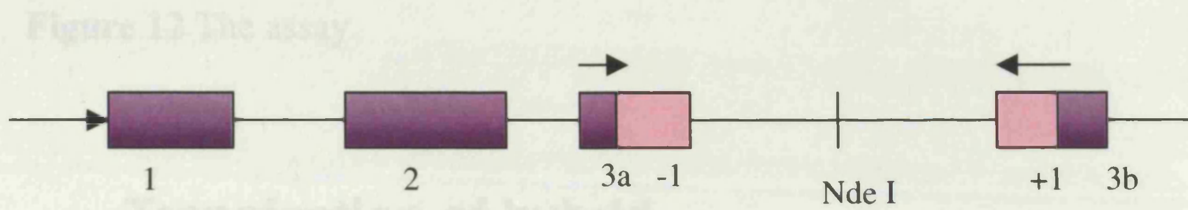



**Figure 10** The PTB minigene. The alpha globin and fibronectin EDB exons are represented by purple and pink boxes respectively. Transcription is guided by an interrupted alpha globin promoter and SV40 enhancer (5' arrow). The Nde I restriction site used in this project is shown.

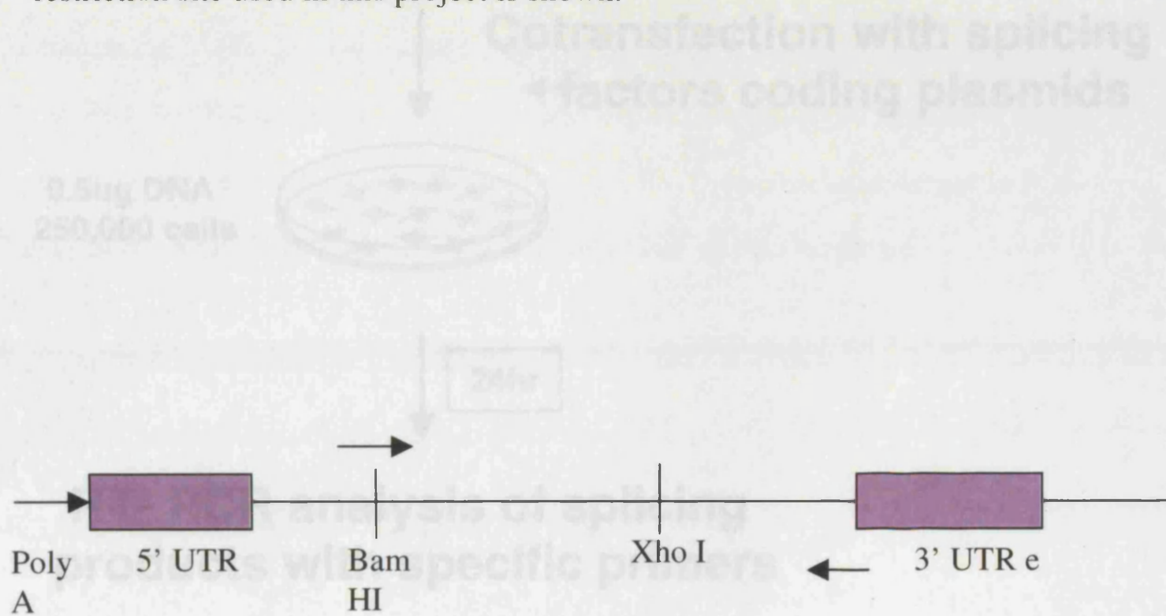
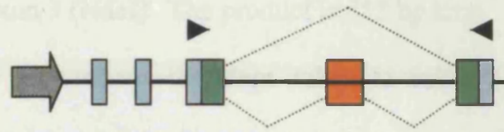



**Figure 11** pcDNA 3 minigene. The sequence of interest is cloned into the area between restriction sites Bam HI and Xho I. Transcription is guided by a CMV promoter.

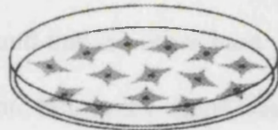
**Figure 12** The assay.

Transfection of hybrid minigenes in different cell lines



Cotransfection with splicing factors coding plasmids

0.5ug DNA
250,000 cells



24hr

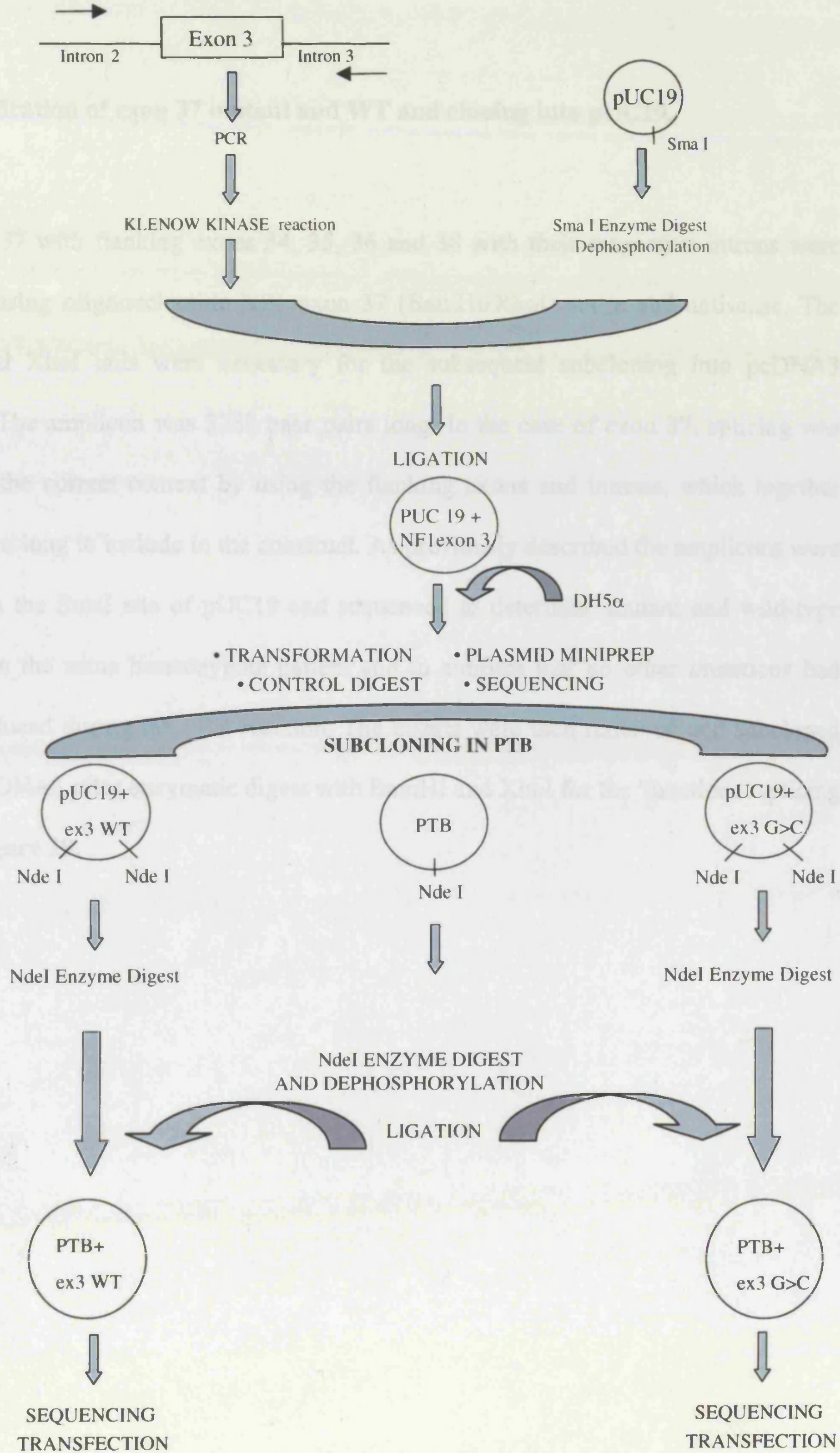
RT- PCR analysis of splicing products with specific primers

### 4.5 Amplification of exon 3 in mutant and WT DNA and cloning into pUC19

*NF1* Exon 3 with flanking 5' and 3' intronic regions was amplified using oligonucleotides 'NF1/exon 3 (NdeI)'. The product is 855 bp long

**Figure 16** outlines the steps taken to achieve cloning of the amplicon into the PTB minigene system. As the patient is heterozygote for the mutation, I was able to determine by sequencing which clones were wild type (these were used as controls) and which had the exon 3 +5 G>C mutation. Sequencing also allowed checking that the entire insert sequence was correct and that no other base changes had been introduced during the PCR Taq polymerase reaction, before the insert was cloned into the PTB minigene.

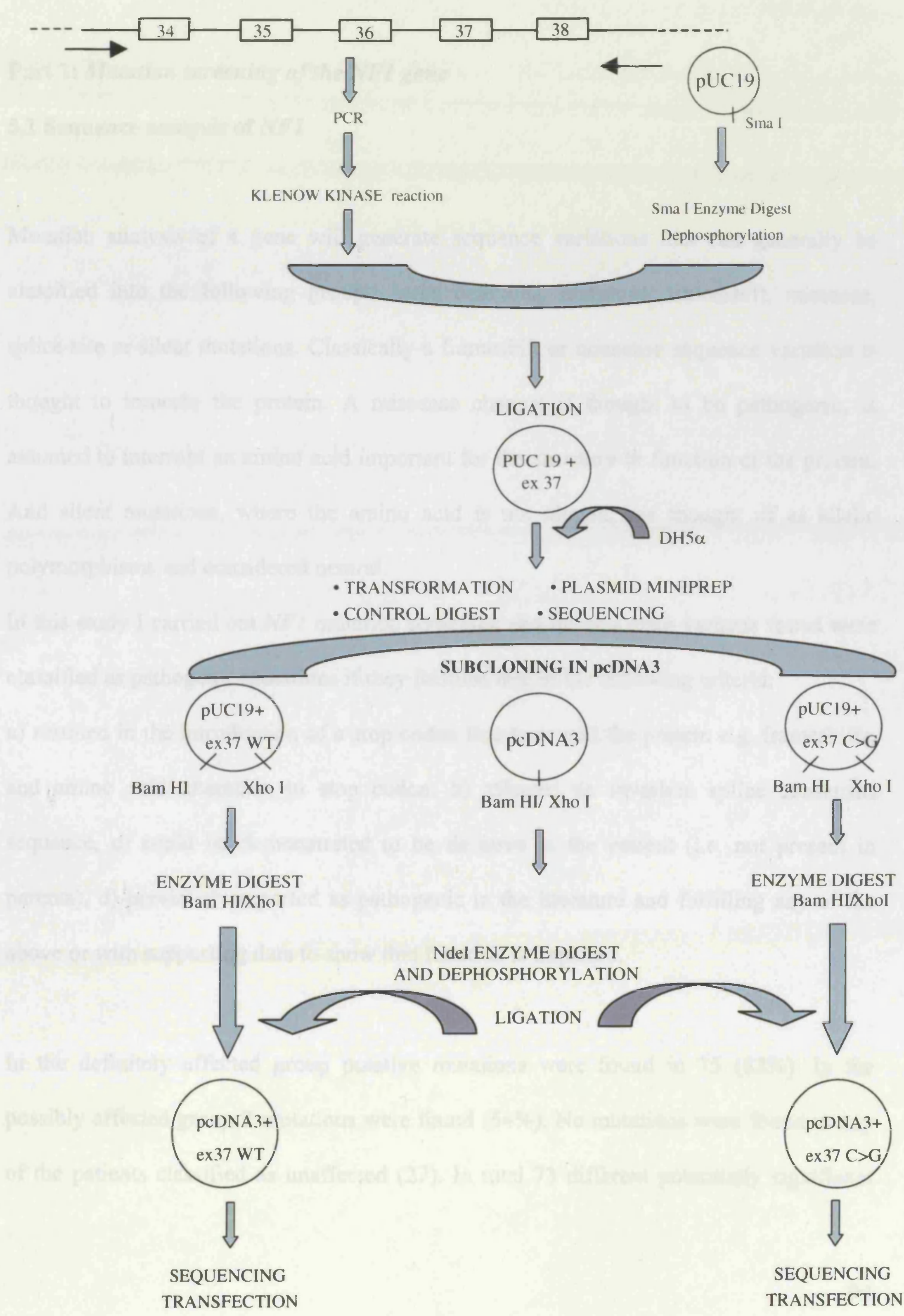
**Figure 13** Cloning strategy for NF1 exon 3 in PTB and steps in the minigene assay.



### 4.6 Amplification of exon 37 mutant and WT and cloning into pUC19.

*NF1* exon 37 with flanking exons 34, 35, 36 and 38 with their respective introns were amplified using oligonucleotide NF1 exon 37 (BamHI/XhoI) sense and antisense. The BamHI and XhoI tails were necessary for the subsequent subcloning into pcDNA3 minigene. The amplicon was 3280 base pairs long. In the case of exon 37, splicing was studied in the correct context by using the flanking exons and introns, which together were not too long to include in the construct. As previously described the amplicons were cloned into the SmaI site of pUC19 and sequenced to determine mutant and wild-type alleles from the same heterozygote patient and to confirm that no other mutations had been introduced during the PCR reaction. The inserts were then removed and subcloned into the pcDNA3 after enzymatic digest with BamHI and XhoI for the 'functional splicing assay' - **Figure 19.**

**Figure 14** Cloning strategy for exon 37 in pcDNA3 and minigene assay



# Results

## Part 1: *Mutation screening of the NF1 gene*

### 5.1 Sequence analysis of *NF1*

Mutation analysis of a gene will generate sequence variations that can generally be classified into the following groups: large deletions, nonsense, frameshift, missense, splice site or silent mutations. Classically a frameshift or nonsense sequence variation is thought to truncate the protein. A missense change, if thought to be pathogenic, is assumed to interrupt an amino acid important for the structure or function of the protein. And silent mutations, where the amino acid is not altered, are thought of as allelic polymorphisms and considered neutral.

In this study I carried out *NF1* mutation screening and the sequence variants found were classified as pathogenic mutations if they fulfilled one of the following criteria:

a) resulted in the introduction of a stop codon that truncated the protein e.g. frameshifts, and amino acid alteration to stop codon, b) affected an invariant splice consensus sequence, c) could be demonstrated to be de novo in the patient (i.e. not present in parents), d) previously reported as pathogenic in the literature and fulfilling any of the above or with supporting data to show that function is impaired.

In the definitely affected group putative mutations were found in 75 (82%). In the possibly affected group 7 mutations were found (54%). No mutations were found in any of the patients classified as unaffected (27). In total 73 different potentially significant

sequence variations were found of which 57 (78%) were novel and previously unreported. **Table 4** provides a summary of the *NF1* mutations and sequence variations found.

**Table 4.** Summary of NF1 mutations and sequence variations found

| NIH | Exon | Mutation | Type | Predicted affect | Accession | CpG |
|---|---|---|---|---|---|---|
| NF | 1 | 1-14 to 7 del 21bp | Frameshift | Truncated protein | | |
| NF | 2 | 127 del CT V65X | Frameshift | Truncated protein | | |
| NF | 3 | 271 G>A E91X | Nonsense | Truncated protein | | |
| NF | 4a | 311 T>A L104X | Nonsense | Truncated protein | | |
| NF | 4b | 499 del TGTT D176X | Frameshift | Truncated protein | CD995532 | |
| NF | 4c | 597/8 del 23 V207X | Frameshift | Truncated protein | | |
| NF | 6 | 787 A>T K263X | Nonsense | Truncated protein | | |
| NF | 7 | 916 del G L316X | Frameshift | Truncated protein | | |
| NF | 7 | 889-2 A>G | Splice acceptor | Truncated protein | CS991467 | |
| NF | 8 | 1070 T>C L357P | Missense | Amino acid substitution | | |
| NF | 9 | 1246 C>T R416X | Nonsense | Truncated protein | CM992366 | Y |
| NF | 10a | 1274 G>A W425X | Nonsense | Truncated protein | CM000779 | |
| NF | 10a | 1318 C>T R440X | Nonsense | Truncated protein | CM950845 | Y |
| NF | 10a | 1318 C>T R440X | Nonsense | Truncated protein | CM950845 | Y |
| NF | 10a | 1318 C>T R440X | Nonsense | Truncated protein | CM950845 | Y |
| NF | 10b | 1527+1 G>C | Splice donor | Truncated protein | | |
| NF | 11 | 1642-2 A>G | Splice acceptor | Truncated protein | | |
| NF | 12a | 1756 del ACTA I603X | Frameshift | Truncated protein | CD982825 | |
| NF | 13 | 2124 del CT E715X | Frameshift | Truncated protein | | |
| NF | 13 | 2034 ins C D699X | Frameshift | Truncated protein | | |
| NF | 13 | 2180 C>G S727X | Nonsense | Truncated protein | | |
| NF | 15 | 2352 G>C W784L | Missense | Amino acid substiitution | | |
| NF | 16 | 2851-2 A>G | Splice acceptor | Truncated protein | | |
| NF | 18 | 2991-4 del ATA | Splice acceptor | Truncated protein | | |
| NF | 19a | 3163 C>T Q1055X | Nonsense | Truncated protein | | |
| NF | 21 | 3528 del A L1183X | Frameshift | Truncated protein | | |
| NF | 21 | 3546 del T L1183X | Frameshift | Truncated protein | | |
| NF | 22 | 3758 to 3762 del TCTAC | Frameshift | Truncated protein | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| NF | 24 | 4265 C>A S1422X | Nonsense | Truncated protein | | |
| NF | 27a | 4537 C>T R1513X | Nonsense | Truncated protein | CM941093 | Y |
| NF | 27a | 4537 C>T R1513X | Nonsense | Truncated protein | CM941093 | Y |
| NF | 27a | 4515-1 G>A | Splice acceptor | Truncated protein | | |
| NF | 28 | 4913 del TCTCT C1661X | Frameshift | Truncated protein | CD972357 | |
| NF | 28 | 4950 ins A Y1650X | Frameshift | Truncated protein | | |
| NF | 31 | 5898 del GA D1987X | Frameshift | Truncated protein | | |
| NF | 31 | 5943+1 G>T | Splice donor | Truncated protein | | |
| NF | 32 | 5944-1 G>C | Splice acceptor | Truncated protein | | |
| NF | 33 | 6181del 8bp D2074X | Frameshift | Truncated protein | | |
| NF | 34 | 6483 C>G Y2161X | Nonsense | Truncated protein | | |
| NF | 35 | 6641+1 del G | Splice donor | Truncated protein | | |
| NF | 36 | 6709 C>T R2237X | Nonsense | Truncated protein | CM000815 | Y |
| NF | 36 | 6709 C>T R2237X | Nonsense | Truncated protein | CM000815 | Y |
| NF | 37 | 6792 C>A Y2264X | Nonsense | Truncated protein | CM981382 | |
| NF | 37 | 6792 C>G Y2264X | Nonsense | Truncated protein | CM981382 | |
| NF | 39 | 7096 ins A N2400X | Frameshift | Truncated protein | | |
| NF | 40 | 7206 del CA C2405X | Frameshift | Truncated protein | | |
| NF | 40 | 7201 A>T K2401X | Nonsense | Truncated protein | | |
| NF | 41 | 7285 C>T R2429X | Nonsense | Truncated protein | CM000818 | Y |
| NF | 42 | 7427 ins TC S2502X | Frameshift | Truncated protein | | |
| NF | 42 | 7486 C>T R2496X | Nonsense | Truncated protein | CM941096 | Y |
| NF | 45 | 7807 -2 A>G | Splice | Truncated protein | | |
| Q | 12b | 1994 C>T S665F | Missense | Amino acid substitution | CM000785 | |
| NF | 12b | 1885 G>A G629R | Missense | Amino acid substitution | CM961026 | Y |
| NF | 12b | 1885 G>A G629R | Missense | Amino acid substitution | CM961026 | Y |
| Q | 16 | 2530 C>T L844F | Missense | Amino acid substitution | CM002379 | |
| Q | 17 | 2970 del AAT ΔM991 | Inframe deletion | Amino acid deletion | CD931025 | |
| Q | 26 | 4402 A>G | Missense | Amino acid substitution | CM971049 | |

| | | | | | |
|---|---|---|---|---|---|
| NF | 7 | 1010 A>T E337V | Missense | Amino acid substitution | |
| NF | 7 | 970 T>C C324R | Missense | Amino acid substitution | |
| NF | 10b | 1466 A>G Y489C | Missense | Amino acid substitution | |
| NF | 10b | 1466 A>G Y489C | Missense | Amino acid substitution | |
| NF | 10b | 1466 A>G Y489C | Missense | Amino acid substitution | |
| NF | 10c | 1595 T>C L532P | Missense | Amino acid substitution | |
| NF | 12a | 1722 C>A S574R | Missense | Truncated protein | |
| NF | 13 | 2040 C>T C680C | Samesense | Truncated protein | |
| NF | 16 | 2514 C>G I838M, 2531 T>C L844P | Missense | Amino acid substitution | |
| NF | 16 | 2531 T>C L844P | Missense | Amino acid substitution | |
| NF | 16 | 2617 C>T R873C | Missense | Amino acid substitution | Y |
| NF | 18 | 3113+3 ins A | Intronic | Truncated protein | |
| Q | 19b | 3217 A>G M1073V | Missense | Amino acid substitution | |
| Q | 21 | 3587 T>G L1196R | Missense | Amino acid substitution | |
| NF | 22 | 3826 C>G R1276G | Missense | Amino acid substitution | Y |
| NF | 22 | 3826 C>G R1276G | Missense | Amino acid substitution | Y |
| NF | 22 | 3827 G>A R1276Q | Missense | Amino acid substitution | Y |
| NF | 22 | 3827 G>A R1276Q | Missense | Amino acid substitution | Y |
| NF | 24 | 4111-8 del GTT | Intronic | Truncated protein | |
| Q | 24 | 4255 A>G K1419E | Missense | Amino acid substitution | |
| NF | 25 | 4312 del GAAΔE1438 | Inframe deletion | Amino acid deletion | |
| NF | 28 | 4973 del 6 ΔIY1658-9 | Inframe deletion | Amino acid deletion | |
| NF | 28 | 5172 G>A K1724K | Samesense | Truncated protein | |
| NF | 29 | 5427 C>T R1808R | Samesense | Truncated protein | |
| NF | 39 | 7096 del 6 ΔNF2366-7 | Inframe deletion | Amino acid deletion | |

Key to abbreviations used in table: NF- fulfils NF1 diagnostic criteria, Q- did not fulfil diagnostic criteria. Missense mutations have been included in this table, only if previously reported as pathogenic or in the cases of: H31R, L145P, Y489C, P533L, and M1073V where non-affected parental samples did not have the same mutation.

### 5.1.2 Mutation spectrum

By far the largest group of variations were single base pair substitutions, which were detected in 56 cases (68%). Of these 38 (68%) were transitions and 18 (32%) were transversions, proportions corresponding closely with those previously observed (Ars *et al.*, 2000). Mutation at CpG dinucleotides account for 20/60 (33%) of the single base pair substitutions. Of these single base pair substitutions the most frequently detected single group of sequence variants were missense substitutions, which were found in 24 subjects (28%). Of these 18 were found in patients who fulfilled the NIH criteria whilst 6 were found in possibly affected patients. Only five of these have been previously described but a further 11 have closely related variations (affecting the same codon) in the literature. Two mutations of note were first, the variant of codon 844 (2531T>C L844P), which was found twice in this study and is one of at least four known variations of this codon. Second, the missense variant 1722 C>A S574R that affects the first base of exon 12a and may exert an effect by amino acid substitution or by interfering with normal splicing.

I also detected 21 nonsense mutations within the single base pair substitutions, of which 16 (76%) have been described by previous investigators. In 13/21 (62%) of the nonsense mutations the mutation occurs at a CpG dinucleotide and alters the primary base of the codon to a T (CGA>TGA) forming a termination signal.

10 of the missense mutations affected bases potentially involved in splicing. These included five substitutions altering splice acceptors and one altering a splice donor. In

addition we found 2 intronic deletions, which directly alter invariant splice site dinucleotides. 2991-4 del ATA changes the intron 17 acceptor site AG>TG, whilst 6641+1 del G appears to shift the intron 35 donor site 1 bp upstream. Three other potentially significant intronic sequence variations were found; 3113+3 ins A, 4111-8 del GTT and 288+5 G>C. None of these has been previously reported

A second class of mutations, frameshifts, were detected in 18 patients. 14 were small deletions ranging from 1 bp to 23 bp in length. In this group only 499 del TGTT and 1756 del ACTA have been previously reported by other investigators. However, four others deletions had closely related mutations described in the literature, such as 3758 to 3762 del TCTAC related to 3759 to 3763 del CTACC was reported by (Ars *et al.*, 2000). All deletions were entirely exonic except 1-14 to 7 del 21bp which deletes 14 bp of non-coding sequence upstream of exon 1 together with 7 bp of exon 1 thereby removing the start codon.

Together, frameshift and nonsense mutations that would be expected to truncate the reading frame constituted 39 of the 82 mutations detected (497.6%) in patients who fulfilled the NIH criteria. The distribution of these mutations appears to be evenly spread within the gene, with 27/44 (61%) predicted to terminate the protein in or upstream of the GRD.

A third class of mutations. insertions and deletions, were detected in 11 patients. Seven insertions were detected comprising five 1 bp insertions and two 2 bp insertions. None of

these insertions has been previously described. Deletions detected included four in-frame deletions; two removing a single amino acid and two removing two consecutive amino acids. Only 2970 del AAT ΔM991 and 7096 del 6 ΔNF2366-7 have been previously reported.

Four different conservative base substitutions were found in five unrelated, definitely affected patients. No other variations were detected in any of these patients and none of these conservative base substitutions were found in 260 normal chromosomes. Sequence variations 2040 C>T C680C and 5427 C>T R1808R were found in two patients in whom no other family members were available for testing at the time of writing. Another patient was found to have 5172 G>A K1724K and, on testing this patients parents, the affected mother was found to carry the same variation while the father was found to be normal at this position. Finally, the variation 846 G>A Q282Q was found in two unrelated patients of eastern descent.

### 5.1.3 Recurrent mutations.

Twenty of the 77 novel variations were detected in more than one unrelated individual or have been previously reported by other investigators. A total of 8 variants were observed more than once in this study. All of these variations were single base pair substitutions and seven of them occurred at hypermutable CpG dinucleotides. In this study 6 sequence variations were each found in two unrelated cases; 1885 G>A G629R, 4537 C>T R1513X, 6709 C>T R2237X, 3826C>G R1276G, and 3827 G>A R1276Q. Base 6792

was mutated differently in two different patients, C>A and C>G both of which give Y2264X. Two sequence variations were detected in three unrelated patients; the nonsense mutation 1318 C>T R440X has been previously reported while the missense variation 1466 A>G Y489C is novel.

Two pairs of mutations were found to be very closely related, affecting a different base in the same codon. Codon 844 was alternatively mutated CTT>TTT (L844F), and CTT>CCT (L844P). The latter sequence change was found once on its own and also in combination with another missense change, 2514 C>G I838M. It is unclear whether these occurred in *cis* or *trans* and which is the pathogenic mutation. As the L844P change was also found alone in a definitely affected patient and is one of at least four known variations of this codon, it is suspected that this change is the causative mutation, although more work to clarify this will be needed. Neither 3826C>G R1276G, 3827 G>A R1276Q have been previously described and both were found in two unrelated cases in this study. However two other mutations of this codon, 3826 C>T R1276X, 3827 G>C R1276Q have been previously reported (Heim *et al.*, 1995). Twenty other mutations found in this study were also found to have closely related mutations in the literature. Of particular note is 1466 A>G Y489C, which was found in 3 cases in this study but has not been previously reported although (Fahsold *et al.*, 2000) reported an identical mutation of an identical codon 2 downstream of Y489C (1472 A>G Y491C).

The positions of all the mutations seen are shown on a histogram- **Figure 15**. Although there seems to be a general spread of mutations found throughout the gene, we found no

sequence changes in exons 1, 5, 9a,15, 20, 23a, 27b, 29, 30, 34, 35, 43, 44, 45, 47, 48, 48a and 49.

While I find that mutations are evenly distributed along the NF1 coding region, the scale of the study allows further analysis of possible mutation hotspots and possible functional domains. 22% of mutations fall within the GAP-related domain, but when the mutations are analysed by type and a weighted distribution calculated for each exon, a number of missense and single amino acid mutations were found to be clustered in the region; exons 11-17. This therefore confirms the findings of (Fahsold *et al.*, 2000) that this represents an important area of the protein. **Figure 11** shows the weighted distribution of missense mutations and single base changes in the NF1 gene. The relative mutation frequency was calculated by dividing the number of missense or single base changes by the base pairs per exon. The value is then shown as a ratio between the exon-specific mutation densities and the average mutation density for the whole gene (28/8457bp). A clustering can be seen between exons 10-17.

**Figure 15.** Histogram showing site of mutations found in NF1

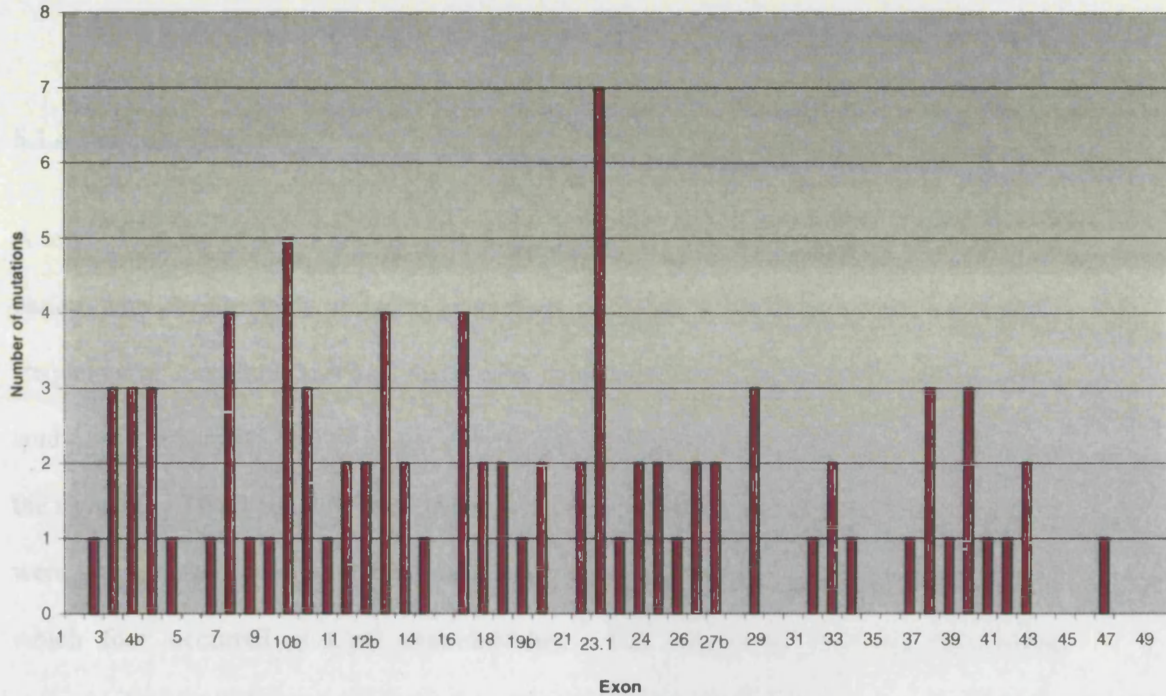
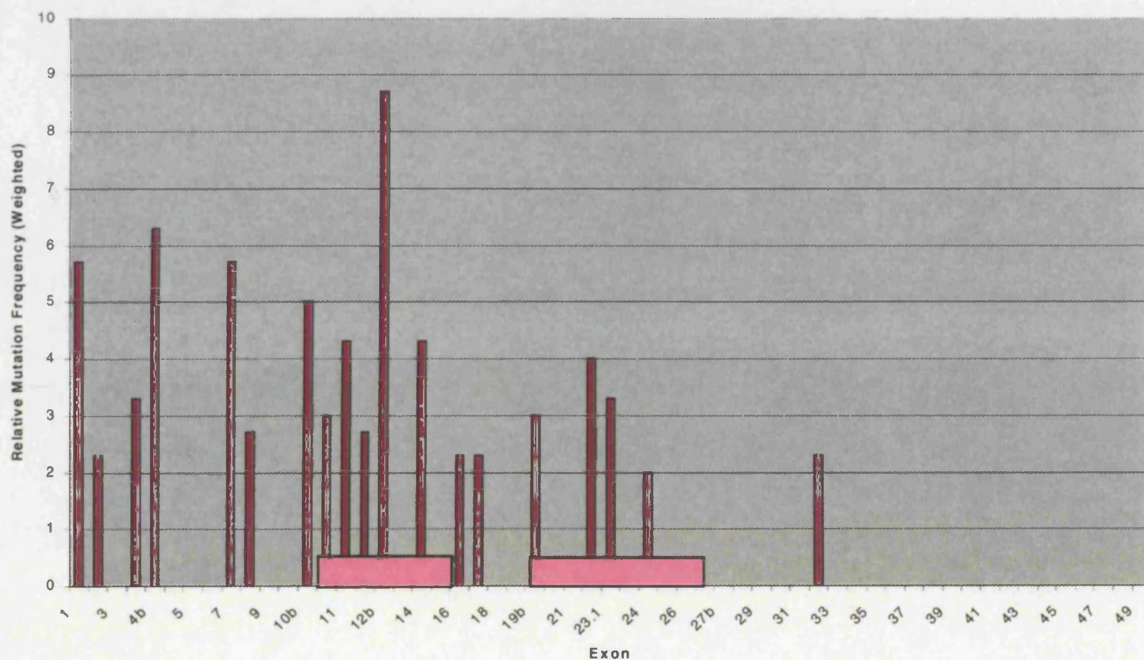



**Figure 16** Weighted distribution of missense mutations and single base changes in the NF1 gene. The relative mutation frequency was calculated the dividing the number of missense or single base changes by the base pairs per exon. The value is then shown as a ratio between the exon-specific mutation densities and the average mutation density for the whole gene (30/8473bp). A clustering can be seen between exons 10-17. Between exon group and number of mutations present $\chi^2=20.85$, df 3, p=0.0001. Mutation rate is greatest between exons 1-20: $\chi^2=11.74$, df 2, p=<0.005, compared to exons 21-49: $\chi^2=9.1$, df 2, p=<0.01. The pink boxes represent cAMP/PKA domain and the GRD.



### 5.1.4 Polymorphisms

A total of 19 sequence variations were found either in a normal subject or in an affected patient who already had a mutation characterised- **Table 5**. Of these 14 had a rare allele frequency of less than 0.05 and eight were only found in a single subject in the entire study (167 subjects). Seven of the polymorphisms found were in coding sequence and the remaining 12 all fell within 40 bases of an exon. All but three of the polymorphisms were single base pair substitutions, six being transversions and 10 being transitions of which four occurred at CpG dinucleotides. The remaining three polymorphisms comprised two deletions involving a single T and one deletion of 5 bp.

| Exon | | Base | P | Q | Aa Change | *p* | *q* | Found with | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Unaff | Definite | Others |
| 1 | | 1-22 G>C | G | C | | 0.984 | 0.016 | Y | 7486 C>T R2496X | Y |
| 2 | | 61-4 del T | 9T | 8T | | 0.996 | 0.004 | | 1642-2 A>G | |
| 2 | | 168 C>T | C | T | S56S | 0.959 | 0.041 | Y | 4045 ins T S1373X | Y |
| 4 | b | 528 T>A | T | A | D176E | 0.996 | 0.004 | Y | | |
| 5 | | 702 A>G | A | G | L234L | 0.728 | 0.364 | Y | many | Y |
| 7 | | 889-31 del ATTAT | | | | 0.996 | 0.004 | | 1885 G>A G629R | |
| 10 | b | 1393-32 C>T | C | T | | 0.707 | 0.403 | Y | many | Y |
| 10 | c | 1528-29 del T | 8T | 7T | | 0.694 | 0.320 | Y | many | Y |
| 11 | | 1642-25 T>C | T | C | | 0.973 | 0.027 | Y | 7096 del 6 | Y |
| 12 | a | 1810 T>C | T | C | L604L | 0.993 | 0.007 | | 4255 A>G K1419E | |
| 16 | | 2617 C>T | C | T | R873C | 0.996 | 0.004 | | 1-14 to 7 del 21bp | |
| 17 | | 2851-16 T>C | T | C | | 0.977 | 0.023 | Y | 3528 del A L1183X | |
| 22 | | 3867C>T | C | T | F1289F | 0.996 | 0.004 | | 1885 G>A G629R | |
| 29 | | 5546-19 T>A | T | A | | 0.598 | 0.471 | Y | many | Y |
| 32 | | 6084+8 C>G | C | G | | 0.996 | 0.004 | | 1318 C>T R440X | |
| 33 | | 6173 C>A | C | A | A2058D | 0.995 | 0.005 | | 6181del 8bp D2074X | |
| 39 | | 7126+37 G>C | G | C | | 0.937 | 0.063 | Y | many | Y |
| 42 | | 7395-29 A>G | A | G | | 0.666 | 0.356 | Y | many | Y |
| 46 | | 8050+20 G>A | G | A | | 0.989 | 0.011 | Y | | |

**Table 5:** Polymorphisms

**Part 2:** ***Mutation screening in Noonan-Neurofibromatosis syndrome (NFNS)***

## 5.2 Noonan-Neurofibromatosis syndrome (NFNS)

Six cases of NFNS were studied. In all cases the presence or otherwise of NF1 features, complying with the NIH NF1 diagnostic criteria were documented. They were also examined specifically for features of Noonan syndrome with the criteria used by Colley et al (Colley *et al.*, 1996) in their clinical study either by myself or Dr Elmslie from GOS. These were Noonan facial appearance with hypertelorism; downslanting palpebral fissures; ptosis; malar hypoplasia and epicanthic folds; low posterior hair line; low-set posteriorly rotated ears; short, broad and/or webbed neck; pectus abnormality; cubitus valgus; short stature; cryptorchidism/genital anomaly and cardiac defect. The clinical features of all six cases are summarised in **Table 6**. Clinical photographs of cases 1, 3, 4 and 6 are shown in **Figure 17.**

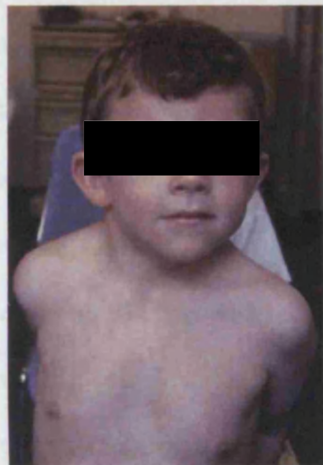
**Table 6**

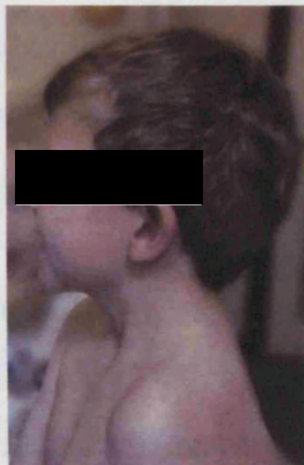
| Case | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **NF1 Features** | | | | | | |
| • >6 CAL macules | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| • >2 neurofibromas | ✓ | | | | | |
| • plexiform neurofibromas | | | | | | |
| • axillary/inguinal freckling | ✓ | ✓ | ✓ | | | ✓ |
| • >2 lisch nodules | ✓ | x | x | x | | x |
| • osseous lesion | | | | | | |
| • Family history | | ✓ | | ✓ | | |
| | | | | | | |
| **Noonan syndrome features** | | | | | | |
| • hypertelorism | | | ✓ | | | |
| • downslanting palpebral fissures | ✓ | | | ✓ | | |
| • ptosis | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| • malar hypoplasia | | | | | | ✓ |
| • epicanthic folds | | | | | ✓ | |
| • low posterior hairline | | | | | ✓ | |
| • low set posteriorly rotated ears | | | ✓ | ✓ | ✓ | |
| • short/broad/webbed neck | ✓ | | ✓ | ✓ | | |
| • pectus abnormality | | | | | | ✓ |
| • cubitus valgus | | | | | | |
| • short stature | | ✓ | | | | |
| • cryptorchidism/genital abnormality | | | | | | |
| • cardiac defect | ✓ | ✓ | x | x | ✓ | x |

All cases except case 5 fulfilled NF1 diagnostic criteria and also had Noonan facies with one or other features of Noonan syndrome. The cardiac defects seen were atrial septal defect, coarctation of aorta and pulmonary stenosis. Four cases had feeding problems in the neonatal period. Other clinical features found were: case 1 also has widely spaced nipples and scoliosis, case 6 has keratosis pilaris. Ages of patients: Case 1- 20yrs, Case 2- 18yrs, Case 3- 7yrs, Case 4- 7yrs, Case 5- 6yrs, Case 6- 9yrs.

**Figure 17** NFNS cases

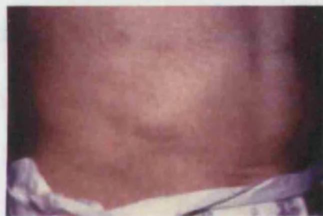
(a) Case 1

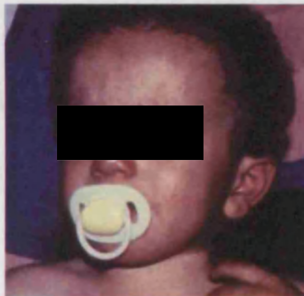


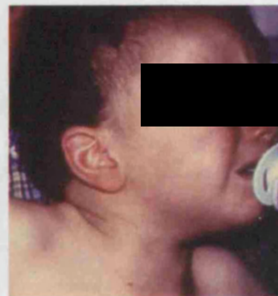


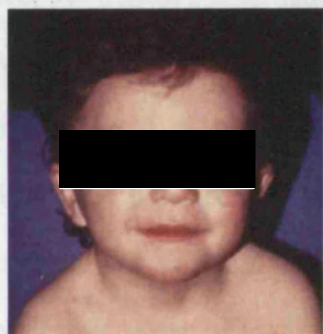
(b) To the left Case 3.

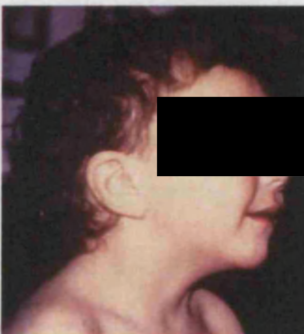
(c) Case 4 below.











(d) Case 6

**Case 1:** A 20 year old man with 7 CAL macules, axillary freckling, 10 neurofibromas, Lisch nodules, and a scoliosis with a structural cervical vertebral abnormality. He had downslanting palpebral fissures, ptosis, a short/broad neck, widely spaced nipples, and an atrial septal defect. He was of short stature and needed extra help at mainstream school. There was no family history of similar findings.

**Case 2:** An 18 year-old man with more than six CAL macules and axillary freckling. He had severe feeding difficulties as n infant, requiring nasogastric tube feeding for 18 months. He had ptosis, low set posteriorly rotated ears, and coarctation of the aorta. Height was on the 10$^{th}$ centile. Intellectual development was normal. His father also had ptosis and one CAL macule.

**Case 3:** A 7 year old boy with more than six CAL macules and axillary freckling. He had hypertelorism, ptosis, low-set posteriorly rotated ears, and a webbed neck. There was no cardiac defect by echocardiography, and intellectual development was normal. There was no significant family history.

**Case 4:** A 7 year-old boy with more than 12 CAL macules. There were no other features of NF1, but his father had 6 CAL macules and axillary freckling. He had downslanting palpebral fissures, ptosis, low set posteriorly rotated ears and a webbed neck. Intellectual development was normal and he had a normal echocardiogram.

**Case 5:** A 6 year-old boy with more than 6 CAL macules. There were no other features of NF1, but his mother had a single, large CAL macule and Lisch nodules, low hairline, and a short neck. He did not therefore fulfil the NIH NF1 diagnostic criteria. He had ptosis, epicanthic folds, low posterior hairline, and low set ears. On echocardiogram he had pulmonary stenosis. As an infant he had had severe feeding problems, which improved at 2

years of age.

**Case 6:** A 9 year-old boy with more than sic CAL macules and axillary and inguinal freckling. He had feeding problems as an infant. He had pectus excavatum, ptosis, malar hypoplasia, scoliosis, keratosis pilaris and developmental delay. Echocardiogram and MRI of the brain were normal. His mother had one large CAL macule on the right thigh.

I examined the possible mechanism for the relationship between NF1 and NS by undertaking mutation analysis of the entire coding region of the NF1 gene in these 6 cases. Mutations in the *NF1* gene were identified in cases 1 and case 5. Case 1 had a 2-bp insertion in exon 23-2, 4095 ins TG. Case 5 was found to have an in-frame 3-bp deletion in exon 25, 4312 del GAA. In the four remaining cases, heterozygous polymorphisms were detected at at least one site within the gene indicating that there was not a whole gene deletion. Parental samples were available only for the mothers of cases 1 and 5, these were negative for the above mutations.

In collaboration with St Georges Hospital Medical School Clinical Genetics Department, and Professor M. Patton and Dr K. Kalidas, the four cases that were not found to have a mutation in the *NF1* gene were screened for sequence variations in the Noonan syndrome *PTPN11* gene. Unfortunately, no more DNA was available on cases 1 and 5. No mutations of the *PTPN11* gene were identified on samples 2, 3, 4 and 6.

**Part 3:** ***Effect of different mutations in NF1 on splicing efficiency.***

## 5.3 The functional minigene splicing assay

The sequence variations generated by the first part of this project were studied and a number were chosen for further analysis using the minigene system. They were chosen because they were either a known mutation that affects splicing, a sequence variation where the mode of causing disease was unclear, or where there was a clinical need to know if this was a disease causing mutation. The assay is described in methods section 4.4 and figure 15.

## 5.4 NF1 intronic mutation exon 3+5 G>C

The sequence variation, exon 3+5 G>C was found in a three generation family that had been tested for *NF1* gene mutations using CSA in Part 1 of this project. The significance of this intronic mutation and whether this was the disease causing mutation was unclear. The index case, at age 82, has classical features of NF1 including multiple café-au-lait macules, neurofibromas and axillary and inguinal freckling. Her son was similarly affected and died in a road traffic accident. Her granddaughter additionally had macrocephaly and died aged 31 from a malignant nerve sheath tumour affecting the coeliac axis.

### 5.4.1 The intronic mutation 3+5 G>C is a splicing mutation

Amplification of exon 3 in mutant and WT DNA and cloning into pUC19 is decribed in methods section 4.5 and **Figure 13.** The functional splicing assay is a technique, as outlined in methods section 4.4, that involves the insertion of exon/exons into a minigene system, in this case PTB ($\alpha$-globin-fibronectin EDB illustrated in the introduction) giving rise to a hybrid minigene. This will allow the inclusion or exclusion of the exon in the transcript to be assessed.

**Figure 18A**, shows the PCR product after exon 3 amplification. **Figure 18B** is a schematic representation of the amplicon incorporating exon 3 and flanking intronic sequences and how this fits in the PTB minigene.

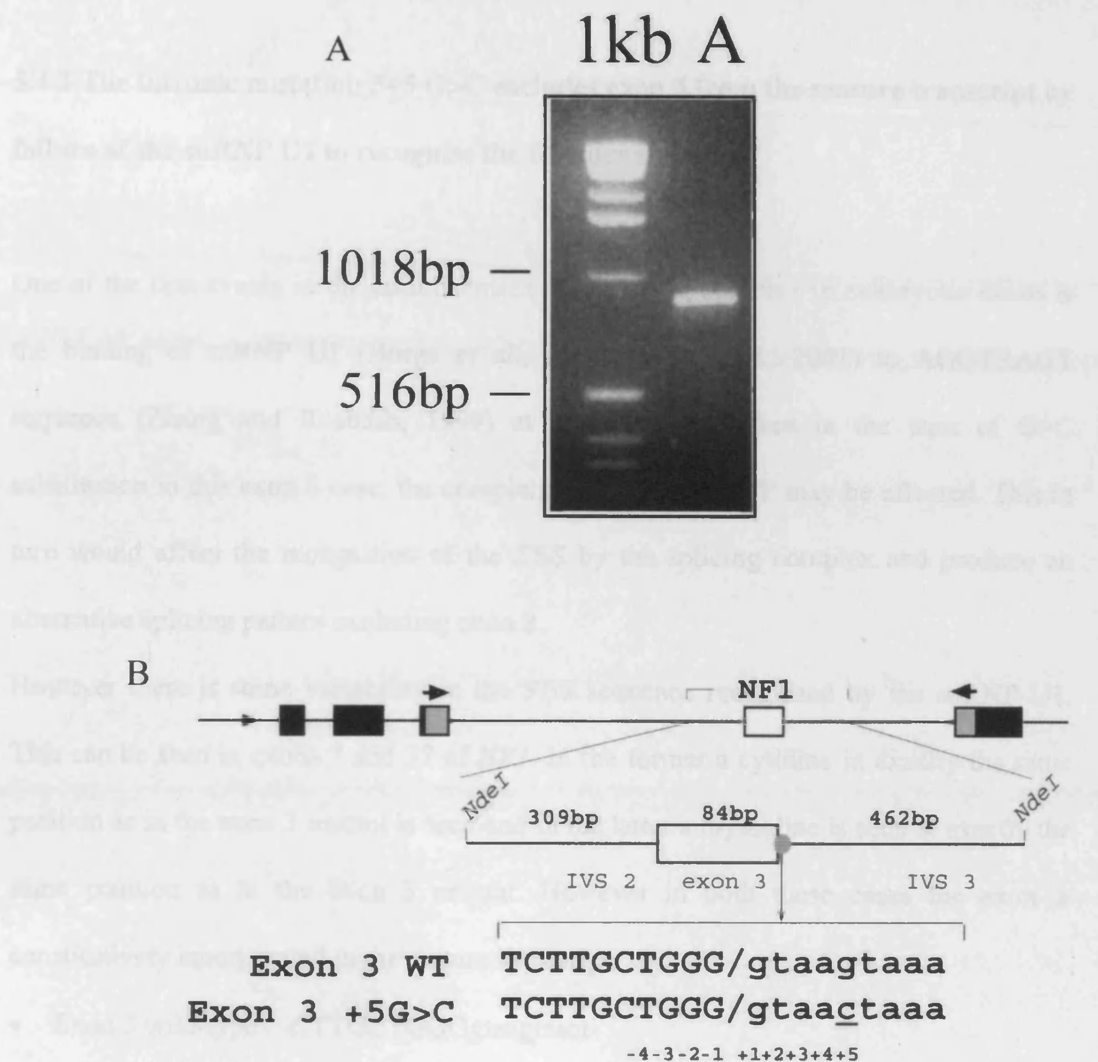



**Figure 18** (A) Exon 3 amplification on 0.8% agarose gel. The product is 855bp long. (B) Schematic representation of amplicon: exon 3 (84bp) plus 309bp and 462bp of flanking intronic sequence at the 5' and 3' ends respectively.

**5.4.2 The intronic mutation 3+5 G>C excludes exon 3 from the mature transcript by failure of the snRNP U1 to recognise the 5' splice site.**

One of the first events in the establishment of a splicing complex in eukaryotic exons is the binding of snRNP U1 (Burge *et al.*, 1998; Stark *et al.*, 2001) to AGGTRAGT sequence (Zhang and Rosbash, 1999) at the 5'SS. Therefore in the case of G>C substitution in this exon 3 case, the complementarity for snRNP may be affected. This in turn would affect the recognition of the 5'SS by the splicing complex and produce an alternative splicing pattern excluding exon 3.

However there is some variability in the 5'SS sequence recognised by the snRNP U1. This can be seen in exons 7 and 37 of *NF1*. In the former a cytidine in exactly the same position as in the exon 3 mutant is seen and in the latter a thymidine is seen at exactly the same position as in the exon 3 mutant. However in both these cases the exon is constitutively incorporated in the mature transcript.

- Exon 3 wild-type -CTTGCTGGGgtaagtaaat-
- Exon 3 G>C - CTTGCTGGGgtaactaaat-
- Exon 7 wild-type -CTTGCTGGGgtaactaaat-
- Exon 37 wild-type- CTTGCTGGGgtaattaaat-

To test the hypothesis that the defective splicing pattern seen in exon 3 is due to reduced affinity for the 5'SS by snRNP U1 a U1 complimentary to the G>C mutation was constructed (U1 C>G), using oligo snRNP U1, that should have the same affinity for the 5'SS as in the wild-type sequence and factor, -**Figure 19**.

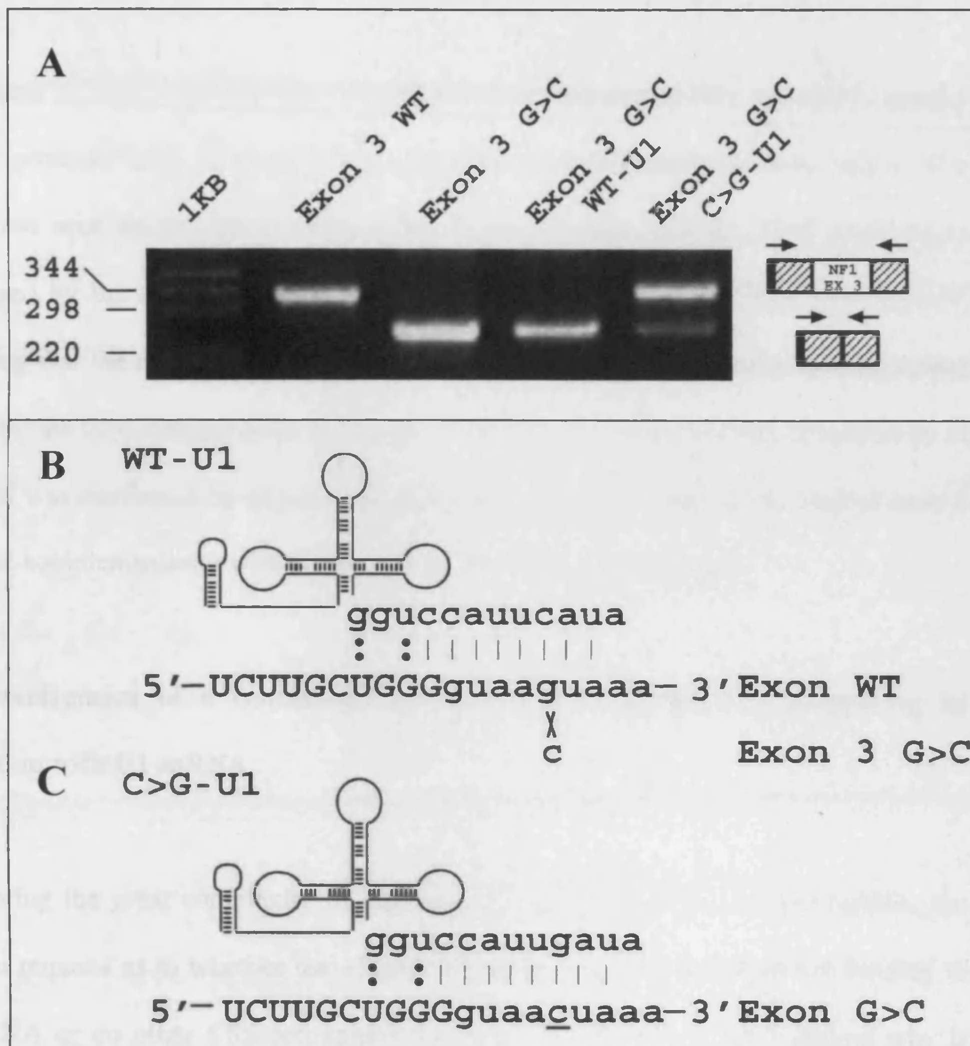



**Figure 19** (A) RNA products generated by the splicing assay. Two products are seen on agarose gel electrophoresis, with the 239 band representing an RNA product lacking exon 3 and the 323bp band product including exon 3, as indicated schematically on the right of the gel. Note the C>G minigene excludes exon 3, while the WT gene includes this exon. Splicing is rescued (that is, exon 3 is included) by the G>C/C>G-U1) but not by expression of the WT-U1 snRNA (exon 3 G>C/WT-U1). (B) Normal base pairing between the 3'end of exon 3 and WT-U1 snRNA. :Indicates wobble association. The nucleotide change observed in the patient (G>C at the +5 position) reduces base pairing between the RNA and the U1 snRNA as shown. (C) The variant snRNA (C>G-U1) changed to complement the nucleotide change seen in the family; note the restoration of appropriate base pairing

The altered U1 was co-transfected with NF1 exon wild-type and NF1 exon G>C and the splicing products analysed by RT-PCR using the same oligonucleotides as before. The results are seen on the gel in **Figure 19.** It can be seen that the G>C mutation is recognised by the modified complimentary U1 where a band of 323bps now appears, indicating that the modified U1 has attached to the 5'SS and rescued the splicing defect caused by the G>C change in Exon 3 (Lane NF1/Exon 3 G>C/U1 C>G). The identity of this band was confirmed by sequencing. This result also shows that in the case of exon 3 complete complementarity with U1 is needed for inclusion of this exon.

### 5.4.3 Involvement of a context-specific hnRNP H1 binding site hampering its interaction with U1 snRNA.

Considering the great complexity of interactions between snRNPs and pre-mRNA, the question remains as to whether the +5G>C substitution simply abolishes U1 binding to pre mRNA or do other 5'SS recognition mechanisms also play a role? Indeed why is complete complementarity not needed for U1 binding in the cases of exons 7 and 37 in the same gene? Therefore further studies on the snRNP U1 and 5'SS interactions were undertaken by pull-down analysis using RNA extracted from cells with mutant and wild-type exon 3. This work was performed by Dr. E. Burratti, ICGEB, Italy (Buratti *et al* 2004).

The first steps of splicesome assembly at the 5' SS can be studied by pull-down analysis where bands representing the different proteins associated with the pre-mRNA are

visualised. To maximise the efficiency of the pull-down analysis 21 nucleotides upstream and 34 nucleotides downstream of the GT consensus sequence of exon 3 were included in the RNA samples. These non-conserved intronic sequences flanking the 5'SS have been previously shown to influence the binding of snRNP U1 (Puig *et al.*, 1999). The results of this study are shown in **Figure 20**. In the exon 3/wild-type lane 4 bands are seen approximately 70Kda, 32.5Kda and two at 25Kda that disappear in the exon 3 mutant (+5 G>C) lane. To determine the identity of these proteins they were sequenced by the Proteomic facility at the ICGEB and compared to proteins present in the HumanGenBankDatabase.

As seen in the figure the proteins were identified as: U1-70K and U1-A, which are components of snRNP U1, and snRNP B/B1 that is a common component of all snRNP U (Stark *et al.*, 2001). Other components of snRNP U: U1-C, E, F, G, D1-D3 were not seen probably because of their low molecular weight (17.5Kda- 8.5 Kda). This result confirms the hypothesis that the +5 G>C mutation inhibits snRNP U1 binding at the 5'SS and thus causes skipping of exon 3.

As mentioned previously, some variability exists in the sequences recognised by snRNP at the 5'SS, as seen in exons 7 and 37 of *NF1*. See sequences below, the blue bases are those that base-pair with snRNP U1.

SnRNP U1: --------gguccauucaua

Exon 3 wild-type: CUUGCUGGGguaaguaaa

Exon 3 G>C: CUUGCUGGGguaacuaaa

Exon 7 wild-type: GAUCUUAAGguaacaugcu

Exon 37 wild-type: CUUAAUAAGguaauuacug

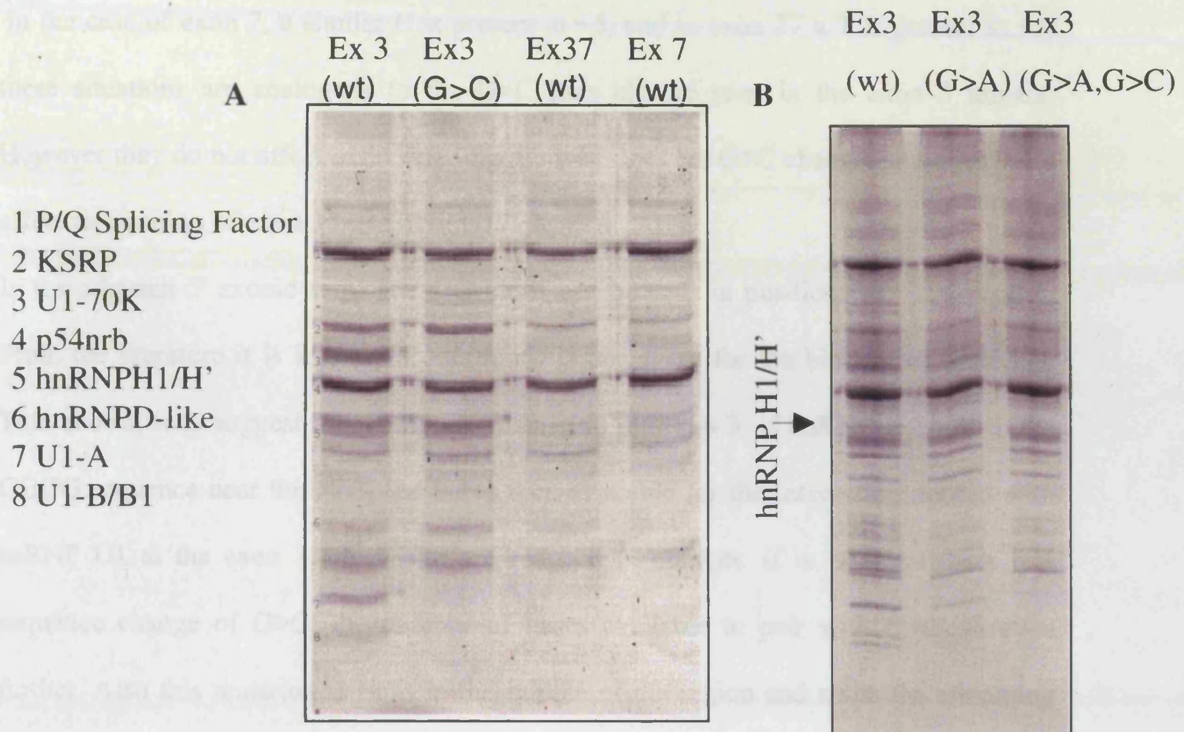



**Figure 20** (A) To gain insight into why NF1 +5 G>C exon 3 is skipped while exons 1,7 and 37 are constitutively included the protein complexes binding in correspondence to the exon/intron boundaries were analysed by RNA pull down analyses and mass spec identification of the bands. It can be seen that the +5G>C mutation causes the loss of the binding of U1 related proteins. Furthermore we observe the presence of hnRNP H1/H (previously shown to play key roles in the splicing process both as a component of splicing enhancers or silencers) in both WT and mutant exon 3 but not in the other exon/intron boundaries. (B)A binding motif for hnRNP H1/H' (GGGG) was observed in exon 3 5'ss. The introduction of a -2G>A substitution destroyed this run both in the wt and mutant minigenes. These two constructs: exon 3 (G>A) and exon 3 (G>A, G>C) were then analysed. Pull down analysis shows the disappearance of a 57kDa protein band, identified as hnRNP H1/H'

In the case of exon 7, a similar C is present at +5, and in exon 37 a T is present at +5, these situations are analogous to the G>C base change seen in the exon 3 mutant. However they do not affect exon skipping. So why does the G>C change so dramatically affect recognition of exon 3 by snRNP U1?

In the adjacent 5' exonic sequence a G motif can be seen in positions -3, -2, -1 and 1. From the literature it is know that this motif is necessary for the binding of hnRNPH. This allows us to suggest a new model of splicing for exon 3: if hnRNP H binds to the GGGG sequence near the 5'SS, the bases then available for the interaction needed with snRNP U1 at the exon 3 splice site are reduced in number. If in addition there is a sequence change of G>C, the number of bases available to pair with diminish even further. Also this mutation is right in the middle of the region and splits the remaining binding region into two sections. The combination of these events could therefore be what prevents snRNP U1 interaction and thus exclusion of exon 3.

To study this model further, the putative hnRNP H binding motif was removed by site directed mutagenesis (see materials and methods), to see if correct splicing could be achieved. In the exon 3 wild-type and G>C mutant a -2G>A sequence change was introduced, to make a double mutant: NF1/exon 3 G>C, G>A. An NF1/exon 3 wild-type, G>A was also constructed. As described previously this was subcloned into the PTB minigene and transfected into Hela cells. At RT-PCR analysis- **Figure 21.** It can be seen that whilst the -2G>A mutation has no effect on inclusion of exon 3 in the wild-type mature transcript, it does rescue the splicing defect that had been seen in +5 G>C mutant confirming the hypothesis.

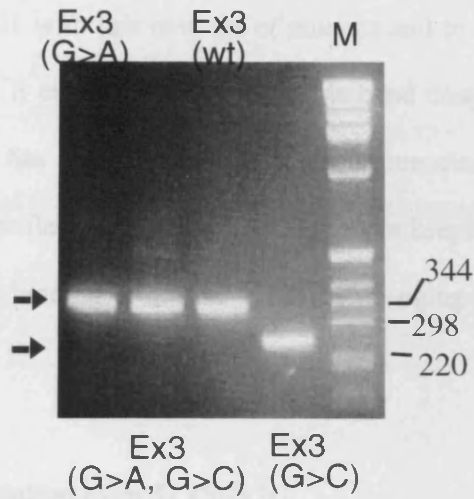



**Figure 21.** -2G>A mutation successfully completely rescues the +5G>C mutation.

A further pull-down analysis was then performed to verify at molecular level the interaction of snRNP U1 with this new set of mutants and to see if hnRNP H was also involved. In **Figure 21** it can be seen that a 57Kda band disappears in the lanes where the -2G>A mutation has been introduced. On sequencing this band after mass spectrometry it was identified as hnRNP H1/H'. This is in keeping with the theory that the -2G>A change prevents binding of hnRNP H1/H' by changing the sequence from gggGT to gAgGT.

### 5.5 NF1 'nonsense' mutation exon 37 C6792G

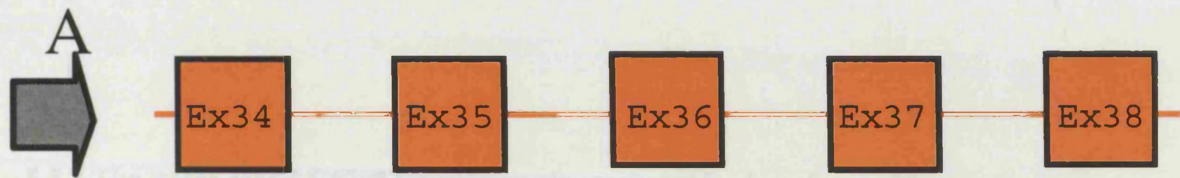
This sample was obtained from a female patient with classical features of NF1, including café-au-lait macules, axillary freckling and multiple neurofibromas. Her son is also affected. This mutation had been previously reported in the literature and known to affect splicing from studies at RNA level (Messiaen *et al.*, 1997).

#### 5.5.1 Mutation exon 37 C6792G causes skipping of this exon in the mature transcript.

Amplification of exon 37 mutant and WT and cloning into pUC19 is described full in methods section 4.6 and **Figure 14**. .Exon 37 and 20 bp of flanking intronic sequence was originally amplified like exon 3 into the PTB minigene. When transfected this gave very poor exon inclusion in both the mutant and wild-type. This indicated that long distance sequences surrounding this exon may be important for inclusion of it. So a large

fragment was cloned into pcDNA3 that is able to allow insertion of longer sequences into the plasmid. It is rare, but not unheard of, that long distance sequences are involved in recognition of an exon (Romano et al *NAR* 2001,29;4: 886-894).

After subcloning of the insert (consisting of exons and respective introns 34, 35, 36, 37, 38) into pcDNA3 the 'functional splicing assay' was performed. The plasmid constructs of both wild-type and mutant NF1/exon 37 C6793G were transfected into HeLa cells and the splicing products analysed by RT-PCR using the oligonucleotides 'Oligo pcDNA 3' sense and antisense. **Figure 22A** shows a schematic representation of the insert and **22B** RT-PCR analysis gel. It can be seen that the C>G substitution at position 6792 which produces a STOP codon, causes skipping or exclusion of not only exon 37 but also of exon 36.



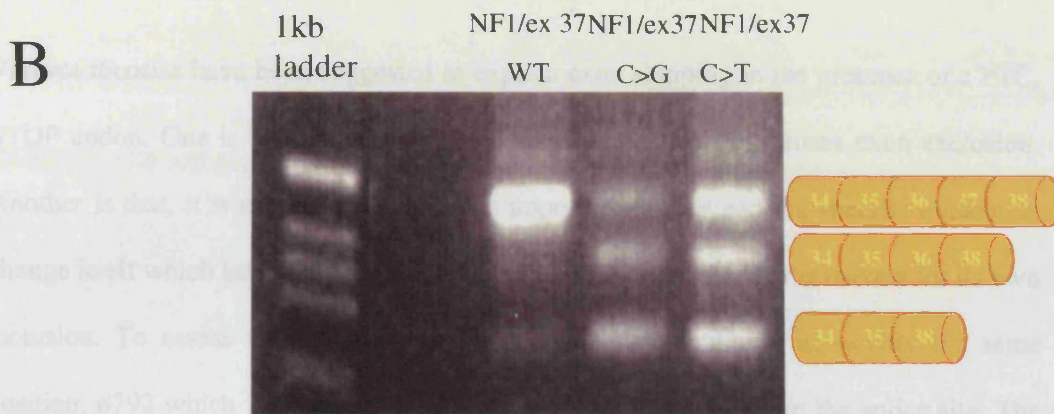
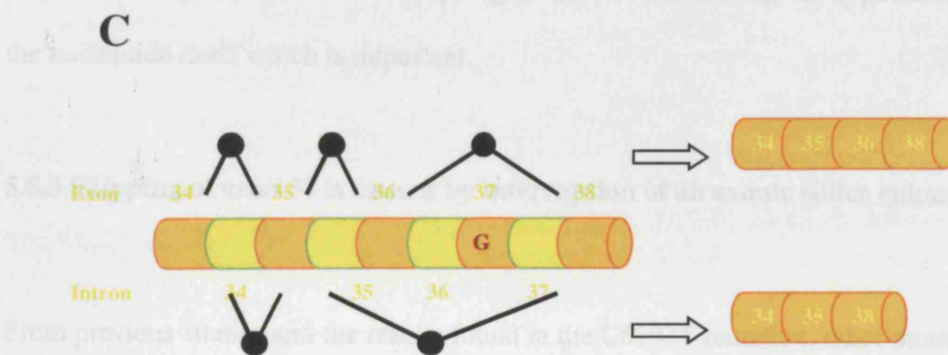




**Figure 22** A) Schematic representation of pcDNA3 with NF1 exons 34-38 as the insert. B) RT-PCR analysis on 2% agarose gel. Two products can be seen; a larger band in lane 1 which includes exon 37 and a smaller band in lane 2 excluding exon 37. Therefore the C>G substitution causes skipping of exon 37. The results also show that as hypothesized, even a translationaly silent mutation in nucleotide 6792 (C6792T) causes skipping. This observation indicates that the nucleotide change itself rather than the creation of a stop codon produces the aberrant splicing. It can also be seen that the cis-acting element affected by the nucleotide change at position 6792 is also necessary for the correct splicing of exon 36. C) Schematic diagram of the process of aberrant splicing in the presence of C6792G mutation.



### 5.5.2 The 'same sense' C6792T mutation also causes skipping of exon 37

Various theories have been suggested to explain exon skipping in the presence of a PTC, STOP codon. One is that the STOP codon is read and that this causes exon exclusion. Another is that, it is not the PTC which is important but instead the specific nucleotide change itself which interrupts an essential sequence in that exon that is needed for its own inclusion. To assess these hypotheses I created a new mutation at exactly the same position, 6792 which is in the middle of exon 37 and 36 bases from the splice site. The change was a *same sense* or 'silent' sequence variation where the C became T (NF1/exon 37 C6792T). This was done by site-directed mutagenesis using oligonucleotides 'Oligo C6792T'. This new amplicon was processed in the same way as the C>G substitution by first cloning into the Sma I site of pUC19, sequencing to confirm the single base change, cut out and subcloned into the BamHI/Xho I site of pcDNA3, transfected in HeLA cells, RNA extracted and RT-PCR analysis performed. As can be seen in **Figure 22** the sequence change C6792T causes skipping of exon 37 confirming the hypothesis that it is the nucleotide itself which is important.

### 5.5.3 Skipping of exon 37 is caused by interruption of an exonic splice enhancer?

From previous studies and the results found in the C6792T mutation, other mutations that may interrupt an ESE sequence in exon 37 were identified and analysed. The single nucleotide changes at a specific position could be altering the sequence of an ESE and thus its function. Indeed an ACAAC motif is seen to be conserved in all cases where

correct splicing occurs. Studies of the secondary structure of RNA in regions with this sequence show reduction of free energy (Turner and Sugimoto, 1988; Zuker and Stiegler, 1981), and a loop motif accessible to splicing factors- **Figure 23**

Therefore this is a putative ESE sequence. ESE's are known to occur in constitutive and alternative exons and are a necessary requirement for the inclusion of that exon (Cooper and Mattox, 1997; Gersappe and Pintel, 1999; Shiga *et al.*, 1997; Vuillaumier-Barrot *et al.*, 1999; Watakabe *et al.*, 1993). To verify whether in the case of exon 37 this particular region does indeed represent an ESE site or not, further studies using site directed mutagenesis were undertaken. Single point mutations were introduced by PCR using specific oligonucleotides. The mutations created are shown in listed below:

- A6791T oligonucleotides used: Oligo A6791T
- A6793T oligonucleotides used: Oligo A6793T
- A6794T oligonucleotides used: Oligo A6794T
- C6795G oligonucleotides used: Oligo C6795G

The DNA fragment studied in all cases is 3280 base pairs long. Sequencing of this to check for the absence of mutations that may have been introduced by the Taq polymerase reaction would be laborious. However the use of two restriction enzyme sites; EcoRI (position 1259) and Afe I (position 1642), respectively upstream and downstream of exon 37 allow a reduction in the number of base pairs in the DNA fragment to be sequenced (386 bp).

The new amplicons were processed as described in previous sections and analysed after RT-PCR analysis. The results can be seen in **Figure 24.** In mutations A6791T, A6793T, A6794T normal splicing was observed. Mutations C6795G showed the triple band

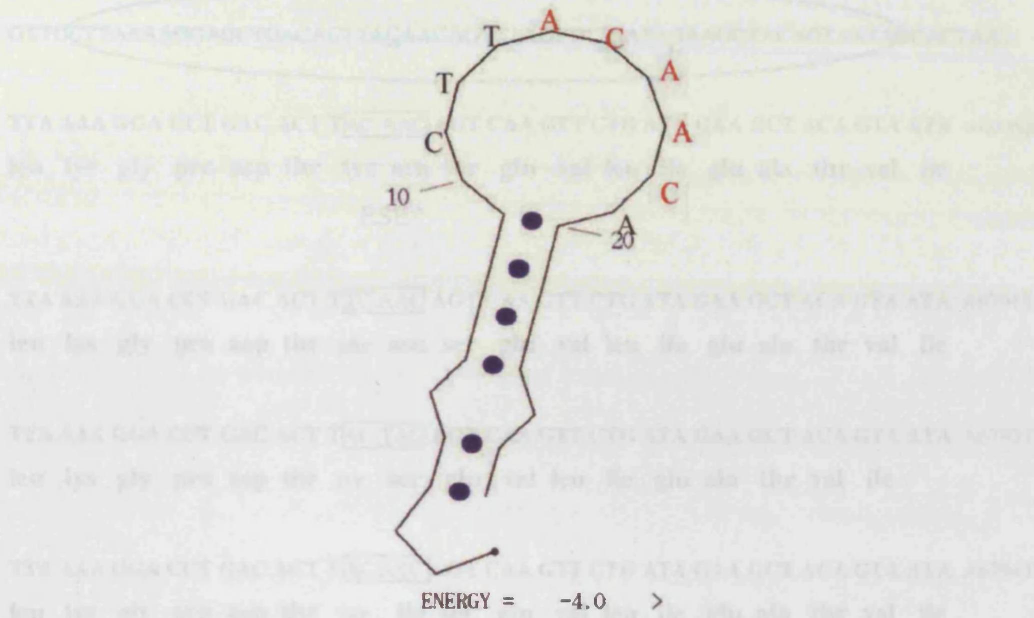



**Figure 23** Diagram of the secondary structure of RNA which incorporates the ESE sequence hypothesis and the corresponding free energy reduction.The red bases represent the conserved sequence of the exposed loop of RNA in which correct splicing occurs.

**Position 6792**

GTTCCTTAAAAGGACCTGACACTTACAACAGTCAAGTTCTGATAGAAGCTACAGTAATAGCACTAA...

**TTA AAA GGA CCT GAC ACT TAC AAC AGT CAA GTT CTG ATA GAA GCT ACA GTA ATA** wild-type
**leu lys gly pro asp thr tyr asn ser gln val leu ile glu ala thr val ile**

ESE?

**TTA AAA GGA CCT GAC ACT TTC AAC AGT CAA GTT CTG ATA GAA GCT ACA GTA ATA** A6791T
**leu lys gly pro asp thr** phe **asn ser gln val leu ile glu ala thr val ile**

**TTA AAA GGA CCT GAC ACT TAC TAC AGT CAA GTT CTG ATA GAA GCT ACA GTA ATA** A6793T
**leu lys gly pro asp thr** tyr **ser gln val leu ile glu ala thr val ile**

**TTA AAA GGA CCT GAC ACT TAC ATC AGT CAA GTT CTG ATA GAA GCT ACA GTA ATA** A6794T
**leu lys gly pro asp thr** tyr **ile ser gln val leu ile glu ala thr val ile**

**TTA AAA GGA CCT GAC ACT TAC AAG AGT CAA GTT CTG ATA GAA GCT ACA GTA ATA** C6795G
**leu lys gly pro asp thr** tyr **lys ser gln val leu ile glu ala thr val ile**

**A6791T A6793T A6794T C6795G**

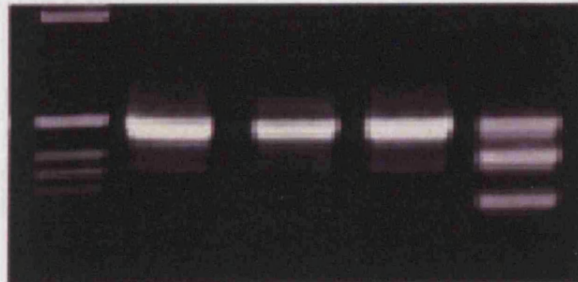
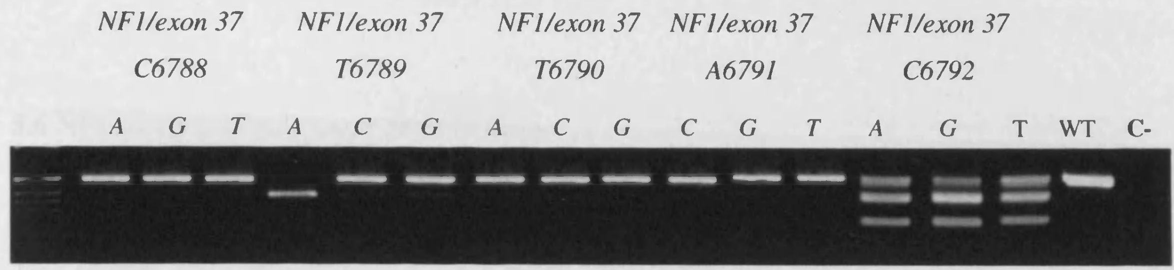
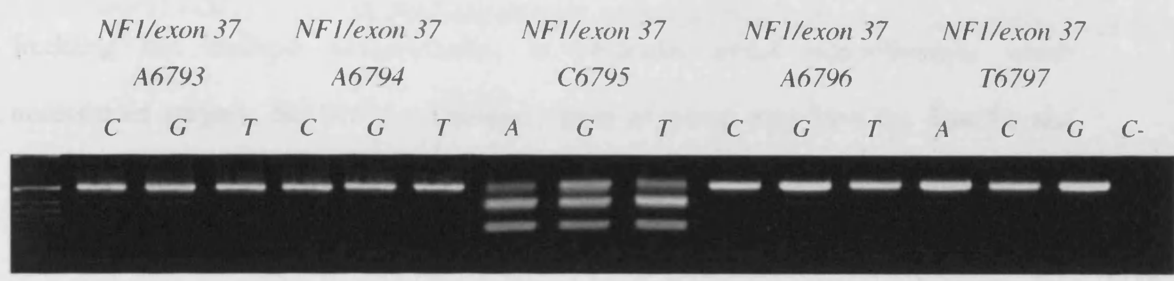


**Figure 24** In mutations A6791T, A6793T, A6794T normal splicing was observed. Mutations C6795G showed a triple band pattern for splicing.

pattern for splicing confirming that the ACAAC region is necessary for splicing regulation. For a fuller picture extensive site-directed mutagenesis was undertaken of all the bases between positions 6788 and 6799 with all three possible substitutions. **Figure 25** show these results. Mutations at positions 6790, 6791, 6793, 6794, 6796, 6797 do not influence splicing, so that exons 34-38 are correctly included. Any mutation at positions 6792 and 6795 lead to aberrant splicing. A three band profile is seen for all of these, where the upper band represents correct assembly of the 5 exons, the intermediate band is composed of exons 34, 35, 36 and 38 with missing exon 37. The lower band consists of exons 34, 35, and 38, in this case both exons 36 and 37 have been skipped. Positions 6789 and 6798 are interesting (both have T here in the WT). If the T is substituted with either C or G normal splicing is observed, whereas if it is substituted with its complementary base, an A, exon 36 skipping (confirmed by sequencing) is seen. Therefore a new, complicated, regulatory element, that does not function through the binding of a SR protein and that is capable of influencing correct exon inclusion is demonstrated.



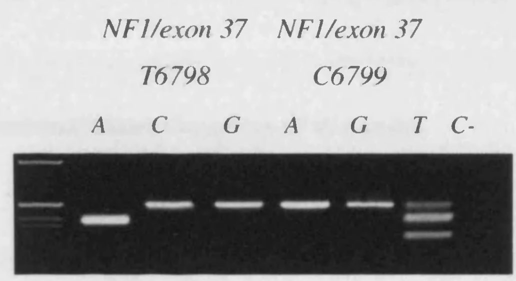




**Figure 25** Minigene analysis on 2% agarose gel of all possible mutations between bases 6788 and 6799. As can be seen, any mutation at position 6792 and 6795 and C6799T generates a splice 3 band pattern with exon 37 skipping (intermediate band) and skipping of both exons 36 and 37 (lower band). A>T substitution at positions 6798 and 6789 skips only exon 36. All other mutations do not alter splicing as can be seen by the single band on the gel of exactly the same size as the WT sequence.

### 5.6 NF1 silent mutation exon 28 5172 G>A

This sample was obtained from a large family with many NF1 affected members. The female proband has classical features of NF1 including café-au-lait macules, axillary freckling and multiple neurofibomas, in particular spinal neurofibromas which necessitated surgery. She has four children, three of whom also have the disorder and additional features including learning difficulties and scoliosis. The entire *NF1* gene was sequenced as described in materials and methods and only a same sense or silent sequence variation was found: exon 28 G5172A (leu-leu). A whole gene deletion was excluded by the checking for the presence of known polymorphisms.

#### 5.6.1 Exon 28 G5172A does not cause skipping of the exon.

Exon 28 with some flanking intronic region was amplified using oligonucleotides exon 28/Nde and processed through the splicing assay as previously described in section 4.2 using initial subcloning into pUC19 and then the PTB minigene system. Exon 28 wild-type and mutant G>A constructs were obtained and checked by sequencing. After transfection, RNA extraction and RT-PCR analysis, it was found that this sequence variation does not affect splicing of exon 28, see **Figure 26.**

This case provides a negative control for this assay, and demonstrates its potential usefulness as a diagnostic splicing assay particularly in the case of assessment of sequence variations found in the high-throughput sequencing of large genes.

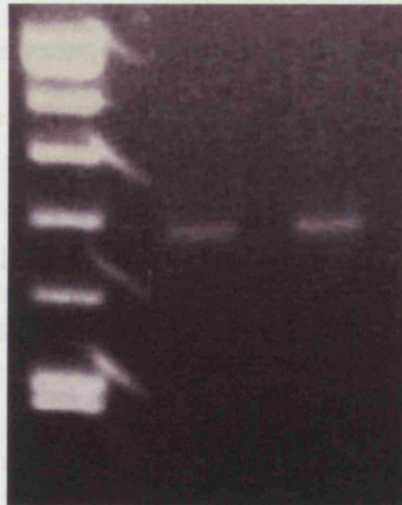



**Figure 26.** Both WT and exon 28 G>A sequence variation show bands of size 431. Exon 28 is included in both cases, therefore there is no aberrant splicing.

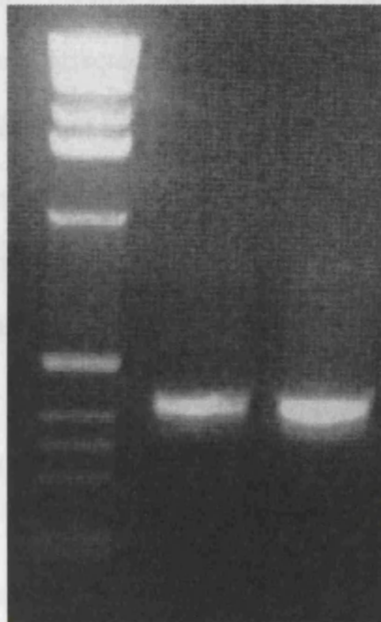



**Figure 27** Both WT and exon 6 846 G>A sequence variation show bands of size 398bp which includes the exon. Therefore there is no aberrant splicing.

## 5.7 NF1 silent mutation exon 6 846 G>A in a family with familial Primitive Neuroectodermal tumours (PNETs) and café-au-lait patches.

These samples were obtained from two sisters who had had supratentorial primitive neuroectodermal tumours (PNETs) diagnosed in childhood, the elder sister subsequently died. PNETs are a highly lethal paediatric malignancy that is usually sporadic and of unknown aetiology. PNETs are not a common brain tumour found in NF1 although case reports do exist in the literature (Scheurlen and Senf, 1995; Thiel *et al.*, 1995). On examination both girls were found to have numerous café-au-lait patches and axillary freckling, such as that seen in neurofibromatosis type 1. They were born from Pakistani consanguineous (first cousin) parents who were not clinically examined for features of NF1 but who claimed to be unaffected. They have a younger brother who at the age of 5 years also has multiple café-au-lait patches.

After sequencing the entire coding region of the NF1 gene by CSA a single base change was found in exon 6 in both affected sisters, G846A. This was a same sense, or silent change that did not affect the amino acid, Gln-Gln.

### 5.7.1 Exon 6 846 G>A sequence variation does not cause skipping of this exon.

Exon 6 with some flanking intronic region was amplified using oligonucleotides exon 6/Nde and processed through the splicing assay as previously described in section 4.2 using the PTB minigene system. Exon 6 wild-type and mutant G>A constructs were

obtained and checked by sequencing. After transfection, RNA extraction and RT-PCR analysis it was found that this sequence variation does not affect splicing, see **Figure 27.**

**5.7.2 PMS2 gene mutation as a possible cause of an NF1-like phenotype with a highly malignant phenotype and high recurrence risk.**

There are at least two reports of PNET in NF1. One of these tumours was peripheral developing in a pre-existing plexiform neurofibroma (Chan *et al* 1996) The other was a large sPNET that developed after chemo- and radiotherapy for a brain stem astrocytoma (Raffel et al 1989 *Neurosurgery* 25:959-964).

A similar sibship to the one described in this thesis was subsequently reported at the European Society of Human Genetics conference 2003 (De Vos et al, St James University Hospital, Leeds). This family consisted of three siblings with brain tumours in childhood from consanguineous parents. The eldest child died from cerebral high grade small-round-cell non-Hodgkin lymphoma, the other two had sPNETs. All three children had multiple CAL macules but did not fulfil NF1 NIH diagnostic criteria. In this family whole genome autozygosity mapping had been undertaken and linkage found to 7p22 including the *PMS2* gene. Subsequently a novel homozygous nonsense mutation in *PMS2* exon 14 was found (De Vos et al 2004 *Am J Hum Genet* 74:954-964). *PMS2* is one of the mammalian genes similar to the mutL DNA mismatch-repair (MMR) gene of *Escherichia coli* (Marti et al 2002 *J Cell Physiol* 191:28-41). The hMutLα heterodimer of *PMS2* with *MLH1* (another mutL homologue) is the major species providing mutL-like MMR activity in human cells (Li and Modrich 1995 *PNAS* 92:1950-1954). Although *MLH1* mutations are the most common cause of hereditary non-polyposis colon cancer (HNPCC) (Peltomaki 2001 *Hum Mol Genet* 10:735-740) *PMS2* mutations are, in contrast, very rare, having been reported in only five families with cancer (Horii *et al.*,

1994; Nicolaides *et al.*, 1995)(De Rosa et al 2000 *Oncogene* 19:1719-1723, Trimbath et al 2001 *Fam Cancer* 1:101-105). In addition review of *PMS2* mutations suggests that they behave as recessive traits, with little or no evidence of cancer predisposition in heterozygotes. This mutation finding significantly widens the spectrum of disorders associated with mismatch repair gene defects, and suggests a new form of recessive cancer syndrome.

CALs and early-onset neoplasia in combination has also been reported, in particular with mutations in *MLH1* (Ricciardone et al 1999 *Cancer Res* 59:290-293).

Subsequently the samples from our family were sent to Leeds and analysed. A *PMS2* gene exon 6 codon 181 homozygous mutation tyr-stop, was found in both the sisters, confirming this as the cause of the signs and symptoms of PNETs with a Neurofibromatosis-like phenotype in this family.

# DISCUSSION

## Part 1: *Mutation screening of the NF1 gene*

### 6.1 Automated comparative sequence analysis identifies mutations in 89% of NF1 patients and confirms a mutation cluster in exons 11-17 distinct from the GAP related domain.

In this study direct sequencing and automated CSA was used to study 167 subjects including 91 unrelated definitely affected NF1 patients. Putative mutations were found in 81 of these affected patients (82%) achieving the highest recorded mutation detection rate using a single technique for this gene. Given that the current estimate of whole gene deletions in this disorder is approximately 10%, this technique may therefore pick up all the remaining classes of mutations and potentially replaces all existing technologies, used either alone or together, with advantages of increased specificity and sensitivity associated with decreased cost and considerably reduced analysis times. A class of mutations that would not be detected with this assay are large gene deletions as the test is not dosage sensitive. We plan to set up an MLPA (multiplex ligation dependent probe amplification) assay to look for whole exon and whole gene deletions in the patients in whom we have not found pathogenic mutations by sequence analysis.

Many new mutations have been identified in our study, with 74% being novel. This

confirms that there are few recurrent mutations in *NF1*. While I find that mutations are evenly distributed along the *NF1* coding region, the scale of this study allows further analysis of possible mutation hotspots and functional domains. The RasGAP activity of the central GAP-related domain (GRD) and the structure of the GRD from neurofibromin have already been well characterised (Izawa *et al.*, 1996; Upadhyaya, 1998). Indeed clustering of mutations has been reported in this region, and much attention has since been concentrated on this area. The GRD is not shown to be a significant mutation hotspot in this study, with only 22% of mutations falling within this region. A second possible functional domain upstream of GRD has also been reported around exons 11-17 (Fahsold *et al.*, 2000). When our mutations were analysed by type and a weighted distribution calculated for each exon, a number of missense and single amino acid deletions were found to be clustered in this region as assessed visually and shown in Figure 11 confirming that this represents an important area of the NF1 protein.

A similar cluster, which was termed a hotspot, was seen in the study of (Fahsold *et al.*, 2000) although I detected more mutations in upstream exons. This region is known to be a cysteine/ serine-rich domain (CSRD) defined by (Izawa *et al.*, 1996). It has three cysteine pairs suggestive of ATP binding as well as three potential cAMP-dependent protein kinase (PKA) recognition sites, that are subject to phosphorylation by PKA. As abnormalities of cAMP signalling have been demonstrated in *Drosophila* neurofibromin mutations, it will be important to clarify further the link between cAMP and the Ras signalling pathway in NF1.

In addition to clinical utility, the availability of a powerful mutation detection technique for the *NF1* gene will allow us to address two longstanding questions in NF1. First, do phenotype-genotype correlations exist? This could be studied in a large number of patients where the whole gene is sequenced and attention paid to the type of mutation and clinical sign/symptom. Second, what is the contribution of the *NF1* gene to some of the rare related disorders, such as segmental NF, gastrointestinal NF, familial spinal NF and familial café au lait spots. The first example to be studied was the rare syndrome Noonan-Neurofibromatosis (NFNS) where patients exhibit features of both disorders.

## Part 2: *Different mutations in the NF1 gene are associated with Neurofibromatosis-Noonan syndrome (NFNS).*

The possible mechanisms for the association of Noonan syndrome with NF1 were examined by undertaking mutation analysis of the *NF1* gene and the *PTPN11* gene (found in 40% of Noonan syndrome cases) in a cohort of 6 NFNS patients. These mechanisms included a) a chance occurrence for the association of Noonan syndrome with NF1, b) NFNS as an unusual variant of NS, c) NFNS as an unusual variant of NF1, and d) NFNS as a distinct genetic disease.

Two mutations were found in the *NF1* gene in this group, and in those without mutations whole gene deletions were excluded by confirming the presence of polymorphic sequence variations in the gene. In this series Case 1 was found to have a 2bp insertion in exon 23-2 and case 5 a 3bp in frame deletion in exon 25. These mutations have not been previously reported and were not found in testing 100 other NF1 patients (Mattocks *et al* 2004).

However, both mutations lie within the GAP-related domain of the gene (exons 20-27a), a region previously shown to contain multiple NF1-causing mutations. I did not find the previously described 3bp deletion in exon 17 in our NFNS patients. However, I have observed this mutation in our series of NF1 patients without NFNS (Results section 3.1 table 3a), showing definitely that this mutation can cause classical NF1 and is not specific for NFNS. Interestingly, case 5 did not fulfill the NIH diagnostic criteria for NF1 and yet was found to have a mutation in this gene. This further emphasises the variability of this disorder.

A disorder that has similar features and has been found to have mutations in the *NF1* gene (Wu *et al.*, 1996) and more recently in the *PTPN11* gene (Digilio *et al.*, 2002) is LEOPARD syndrome (multiple lentigines, electrocardiographic-conduction abnormalities, ocular hypertelorism, pulmonary stenosis, abnormal genitalia, retardation of growth, and sensorineural deafness). However, the hallmark of this disorder are the lentigines, which appear quite different from café au lait macules on clinical examination and therefore does not appear to be a possible diagnosis in these cases. In four cases of this cohort, mutations were not found in either of the two genes; *NF1* or *PTPN11*, suggesting that the disorder is a heterogenous group and may be either a separate syndrome or caused by mutations in further NS genes still to be elucidated. There is no known association or link with the signalling pathways of *NF1* and *PTPN11* to date.

I conclude that NFNS can therefore, in some, be a subgroup of NF1, and importantly, that there does not appear to be a specific *NF1* gene change associated with NFNS. The NFNS syndrome therefore represents further evidence for the wide genotype/phenotype variation seen in *NF1* mutations and the likely diagnosis in these cases. While I have found

mutations in the *NF1* gene in some NFNS patients, our results do not exclude the possibility that mutations in other genes can also cause NFNS. One potential candidate is the *PTPN11* gene responsible for many cases of Noonan syndrome. In all four of the cases in which I did not find NF1 mutations, we were able to sequence this gene (in collaboration), and did not find any mutations. This result would be consistent with a hypothesis that most cases of NFNS represent a form of NF1 rather then Noonan syndrome. However, the inability with current technology to detect mutations in as many as 50% of patients with Noonan syndrome makes it impossible for this hypothesis to be tested, as we cannot exclude the possibility that other cases of NFNS will result from presently undetectable mutations associated with Noonan syndrome. In the absence of a much higher mutation detection rate, this would remain the case even if we were able to sequence larger numbers of patients. The important conclusion I was able to draw from this work is that NFNS can represent a variant form of NF1 and can be caused by different mutations, some of which cause classical NF1 in other individuals. Consequently, these results add NFNS to the phenotypic variation associated with mutations in the *NF1* gene.

## Part 3: *Effect of different mutations in NF1 on splicing efficiency.*

### 7.1 The splicing assay, and diagnostic usefulness.

Previous studies examining the effects on splicing of *NF1* mutations have analysed RNA extracted from individual patients to confirm possible splicing defects. This is, however, a laborious technique requiring a potentially difficult RNA extraction from a cell line or tissue, and complicated by sources of variability such as, splice site leakiness in some tissues and lower levels of expression of mRNA from the mutant alleles. However, in many cases the necessary further samples may be hard to obtain or unavailable for the laboratory performing the sequence analysis. As an alternative technique utilising genomic DNA and circumventing these problems of variability or the need for further sampling, I took advantage of a simple minigene splicing assay. In this assay individual exons are inserted, along with flanking intronic sequences, into a fibronectin (Fn) minigene containing upstream and downstream intronic and exonic sequences sufficient to allow splicing. It had been previously shown how this minigene system can be used to identify sequences required for the correct regulation of alternative splicing in the cystic fibrosis (*CFTR*) gene (Pagani *et al.*, 2000), and also to identify abnormalities of splicing of the *ATM* gene as a disease mechanism in ataxia-telangectasia (Pagani *et al.*, 2002).
As techniques for the identification of sequence variation become faster and cheaper, the distinction between polymorphisms and pathogenic mutations will be an increasing challenge. The assay I have used here in the analysis of NF1 families is a potentially valuable tool for the identification of those mutations that cause splicing defects. No

RNA or further samples are required from the patient, making it feasible to carry out this further step in the molecular genetic diagnostic laboratory as part of the analysis of the DNA sample provided by the referring clinicians. In addition, the minigene system effectively recreates within the cell line the splicing defect of each patient, so facilitating further studies on the relationship between genotype and phenotype in this disease.

## 7.2 An intronic mutation affects exon 3 splicing with the involvement of a context-specific hnRNP H1 binding site hampering its interaction with U1 snRNA.

Here genomic DNA comprising exon 3 and flanking intronic sequences from the potentially abnormal *NF1* gene in this family was amplified and inserted into the minigene. Following transfection and expression of the construct in Hep3B cells, the mRNA produced by the cells was analysed for splicing pattern by RT-PCR. The exon 3 +5 G>C mutation dramatically affected pre-mRNA processing, causing exon 3 to be completely skipped.

To further confirm the role of this mutation I constructed and expressed in the cells a U1-snRNA complementary to the mutation observed in the patient. One of the early events in the process of intron removal from mRNA precursors involves recognition of the 5' splice site by U1 small nuclear ribonucleoprotein (snRNP). This recognition involves complementary base pairing, and the substitution of the guanosine by a cytidine in position +5 of IVS 3 lessens the degree of U1-snRNA base pairing with the 5' splice site. Altering the complementary cytidine to guanosine in the U1 snRNA was found to restore

normal base pairing, and as predicted co-expression of this altered U1 snRNA with the minigene carrying the mutation resulted in rescue of exon 3 splicing. In contrast, expression of WT U1 snRNA did not rescue splicing, showing that this does not simply result from increased levels of U1 snRNA. This experiment proved that the exon 3 +5G>C variation is a disease-causing mutation that induces aberrant skipping of exon 3. The demonstration of a corrected splicing defect in human NF1 cells opens the way for further -omic studies in which corrected cells represent the controls for the *NF1* mutation, so eliminating the problems of genetic and molecular heterogeneity in different human cells and facilitating the identification of critical molecular changes in *NF1*.

Th functional characterisation of the NF1 exon 3 +5G>C mutation was undertaken by Dr E. Burratti, ICGEB, Italy to begin to address some unanswered questions: Does +5G>C substitution simply abolishes U1 binding to pre mRNA or do other 5'SS recognition mechanisms also play a role? Indeed why is complete complementarity not needed for U1 binding in the cases of exons 7 and 37 in the same gene? Therefore further studies on the snRNP U1 and 5'SS interactions were undertaken by pull-down analysis using RNA extracted from cells with mutant and wild-type exon 3. The observation that a U1 snRNA complementary to the mutations could rescue NF1 exon 3 skipping (Baralle *et al* 2003) suggests that it is indeed the U1snRNP-exon 3 5' ss interaction which is involved. This hypothesis was confirmed using pulldown and band shift analyses which demonstrated that U1 snRNP was indeed capable of binding to the NF1 exon 3 (wt) donor site and that this binding was abolished by the +5G>C mutation. Interestingly, the pulldown analysis also showed that the GGGgu donor site sequence of NF1 exon 3 could also bind another

protein, hnRNP H, which has a well-known ability to affect the splicing process (Caputi *et al* 2002, Chou *et al* 1999). We have shown using nuclear extracts depleted of hnRNP H that binding of this protein to the GGGgu sequence can reduce, but not abolish, U1 snRNP binding.

Therefore the sequence of events is as follows; if hnRNP H1/H' binds to the GGGG motif then snRNP U1 does not have enough sequential bases to pair with and therefore the splicesome will not assemble and the exon is skipped. If however, as shown by the -2G>A mutant, the hnRNP H1/ H' does not bind then snRNP U1, even in the mutant +5 G>A can find enough bases to pair with and thus recognise the 5'SS, **Figure 28.**

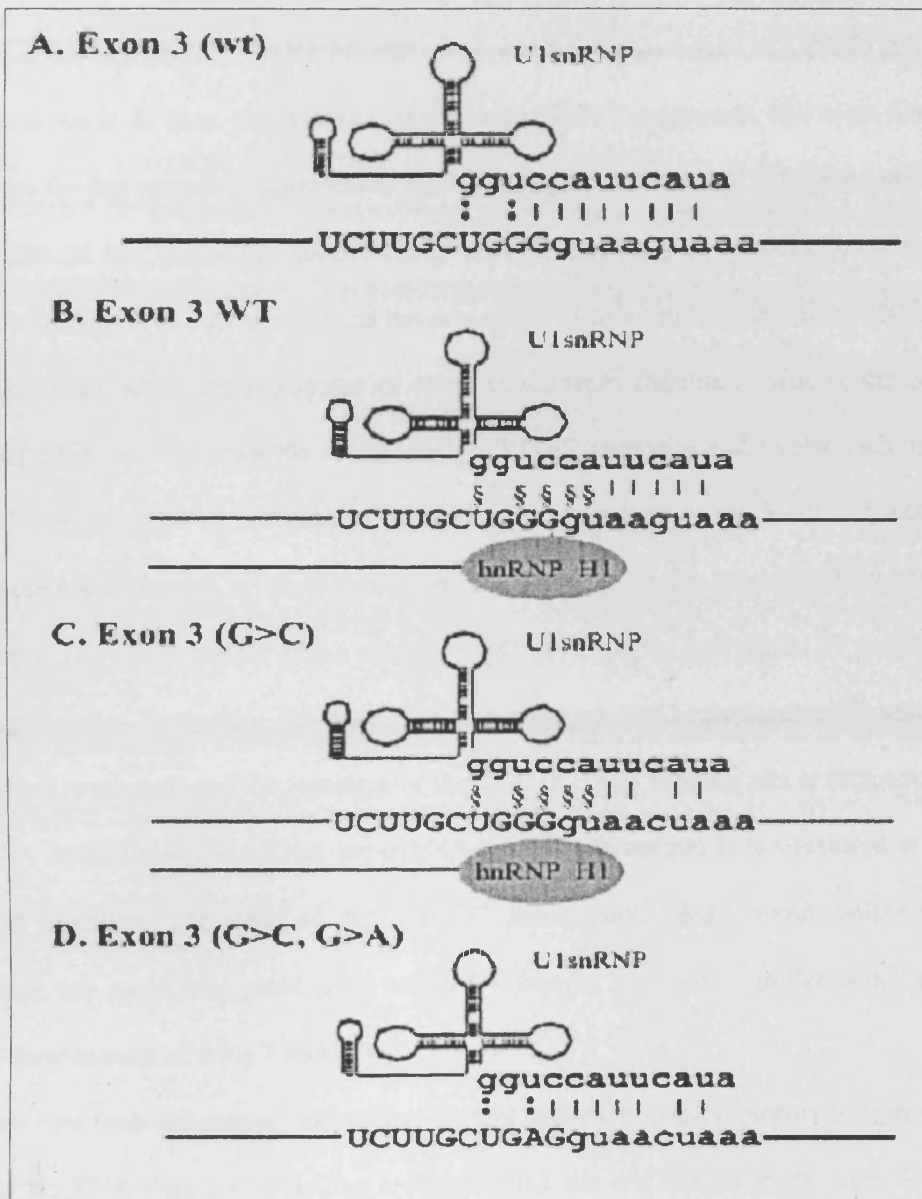



**Figure 28.** A model that explains these observations through the restricted availability of U1 snRNA base paring sequences with the 5'ss due to hnRNPH steric hindrance is shown. Removal of the GGGG motif (D) results in an extension of the pairing and in this context the +5 G>C mutation becomes non pathogenic.

The binding requirements for hnRNPH1/H' involve either the presence of GGGX motifs and/or consecutive G runs. On analysis of the exon 3 RNA sequences, the most likely binding site for this protein is represented by the GGGgu run occurring in correpondance with the 5'SS. In fact, the loss of hnRNP H1/H' binding following the introduction of the -2G>A substitution is in agreement with the disruption of this binding site: from GGGgu to GAGgu. Also, when the sequences of other exons were examined closely, the only significant difference between the mutated exon 3 G>C sequence and exons such as 7, 10b or 37 is represented by the absence of G runs in the vicinity of the 5' SS, suggesting that in these exons there is no binding site for hnRNPH1/H'. In this model, only the very stable occupancy of U1 snRNP at the wild type 5'SS prevents the hnRNP H1/H' molecule inhibiting complex formation. However, following the +5 G>C substitution U1 snRNP occupancy is weakened and the presence of the hnRNP H1/H' binding site is removed by the -2G>A substitution U1 snRNA-pre-mRNA complex formation is not restored to the wild-type condition (because of the +5G>C substitution) but correct splice site recognition can again take place (as it evidently does in other very similar donor sites such as those in exon of exon 7 and 37).

Therefore data from this project shows that distinct sequence specific protein complexes can bind the RNA region surrounding a donor splice site and that in some cases these proteins can affect donor site recognition by the splicing machinery. This may be one of the reasons why, at the molecular level, nucleotide frequencies are selected on additional factors rather than just U1 snRNA complementarity and provides an indication that early protein binding events can influence the outcome of later RNA processing steps. In addition, considering the evr-increasing need for distinguishing

innocuous polymorphisms from potential disease-causing mutations this finding highlights the need for careful 'local context' analyses when predicting the possible effects of mutations introduced in donor sites.

### 7.3 A nonsense mutation that causes skipping of exon 37 in NF1 and definition of an exonic splice enhancer.

*NF1* exon 37 skipping caused a by premature termination codon (PTC) had been previously reported due to the nucleotide substitution C6792A that causes the change Y2264X. However, mutations in the same area, such as 6790insTT and 6789delTTAC, both of which create PTC's did not result in the skipping of this exon- **Figure 25**. It seems then that the splicing defect is not due to the presence of PTC but probably to the disruption of a cis-acting element necessary for correct splicing such as an exonic splicing enhancer (ESE).

ESEs are present in constitutively and alternatively spliced exons, are distinct from the splice site sequences and are required for efficient splicing of many exons. ESEs may act as binding sites for serine/arginine–rich proteins (SR proteins), a family of highly conserved and structurally similar splicing factors.

When *NF1* exons 34-38 were inserted into a minigene system and mutagenesis of this region performed. The results show that, as hypothesised, even a transitionally silent mutation in nucleotide 6792 (C6792T) causes skipping. This observation indicates that the nucleotide change itself rather than the creation of a stop codon produces aberrant splicing. I also see that a cis-acting element affected by the nucleotide change at position

6792 is also necessary for the correct splicing of exon 36.

The molecular basis of the function of the exon 37 ESE appears very complex. It is possible that secondary structure constraints in addition to potential interactions with splicing factors elsewhere in the pre-mRNA are involved in correct processing of exons 36 and 37.

## 7.4 Silent mutations that did not affect splicing and a new form of recessive cancer syndrome.

Two silent mutations generated by sequence analysis were studied using the minigene assay. The first in exon 28 (5172 G>A) in a large NF1 family with various features of the disorder and no other sequence variation found. However this sequence change was not found to cause skipping of exon 28. It may be that this family have a large deletion that cannot be detected by ACSA, and that this sequence variation is truly a polymorphism. However, this case was useful as a negative control for the assay.

The second was in exon 6 (846 G>A) in a family that did fulfil NF1 diagnostic criteria but also had a more extended phenotype with PNETs in two sisters. The sequence change did not cause skipping of the exon. However, further studies showed a nonsense mutation in exon 6 of the *PMS2* gene, changing the diagnosis completely from NF1 to a new cancer syndrome. *PMS2* is a post replication DNA repair protein and this mutation finding significantly widens the spectrum of disorders associated with mismatch repair gene defects, and suggests a new form of recessive cancer syndrome. Care should be taken when assessing families with consanguinity, features of NF1 and rare tumours and

consider this syndrome in the differential diagnosis.

## References

Abeliovich, D., Gelman-Kohan, Z., Silverstein, S., Lerer, I., Chemke, J., Merin, S., and Zlotogora, J. (1995): Familial cafe au lait spots: a variant of neurofibromatosis type 1. *J Med Genet* **32**, 985-986.

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Luhrmann, R. (1999): A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *Embo J* **18**, 5789-5802.

Aebi, M., Hornig, H., and Weissmann, C. (1987): 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved 5' GU. *Cell* **50**, 237-246.

Ahlbom, B. E., Dahl, N., Zetterqvist, P., and Anneren, G. (1995): Noonan syndrome with cafe-au-lait spots and multiple lentigines syndrome are not linked to the neurofibromatosis type 1 locus. *Clin Genet* **48**, 85-89.

Ainsworth, P. J., Rodenhiser, D. I., and Costa, M. T. (1993): Identification and characterization of sporadic and inherited mutations in exon 31 of the neurofibromatosis (NF1) gene. *Hum Genet* **91**, 151-156.

Allanson, J. E., Hall, J. G., and Van Allen, M. I. (1985): Noonan phenotype associated with neurofibromatosis. *Am J Med Genet* **21**, 457-462.

Ars, E., Serra, E., Garcia, J., Kruyer, H., Gaona, A., Lazaro, C., and 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.

Atit, R. P., Crowe, M. J., Greenhalgh, D. G., Wenstrup, R. J., and Ratner, N. (1999): The Nf1 tumor suppressor regulates mouse skin wound healing, fibroblast proliferation, and collagen deposited by fibroblasts. *J Invest Dermatol* **112**, 835-842.

Bahuau, M., Houdayer, C., Assouline, B., Blanchet-Bardon, C., Le Merrer, M., Lyonnet, S., Giraud, S., Recan, D., Lakhdar, H., Vidaud, M., and 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.

Bai, Y., Lee, D., Yu, T., and Chasin, L. A. (1999): Control of 3' splice site choice in vivo by ASF/SF2 and hnRNP A1. *Nucleic Acids Res* **27**, 1126-1134.

Berget, S. M., Moore, C., and Sharp, P. A. (1977): Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* **74**, 3171-3175.

Berglund, J. A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997): The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**, 781-787.

Bernards, A. (1995): Neurofibromatosis type 1 and Ras-mediated signaling: filling in the GAPs. *Biochim Biophys Acta* **1242**, 43-59.

Bernards, A., Snijders, A. J., Hannigan, G. E., Murthy, A. E., and Gusella, J. F. (1993): Mouse neurofibromatosis type 1 cDNA sequence reveals high degree of conservation of both coding and non-coding mRNA segments. *Hum Mol Genet* **2**, 645-650.

Black, D. L. (1991): Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? *Genes Dev* **5**, 389-402.

Blencowe, B. J., Issner, R., Nickerson, J. A., and Sharp, P. A. (1998): A coactivator of pre-mRNA splicing. *Genes Dev* **12**, 996-1009.

Boddrich, A., Griesser, J., Horn, D., Kaufmann, D., Krone, W., and Nurnberg, P. (1995): Reduced neurofibromin content but normal GAP activity in a patient with neurofibromatosis type 1 caused by a five base pair duplication in exon 12b of the NF1 gene. *Biochem Biophys Res Commun* **214**, 895-904.

Boguski, M. S., and McCormick, F. (1993): Proteins regulating Ras and its relatives. *Nature* **366**, 643-654.

Brambilla, R., Gnesutta, N., Minichiello, L., White, G., Roylance, A. J., Herron, C. E., Ramsey, M., Wolfer, D. P., Cestari, V., Rossi-Arnaud, C., *et al.* (1997): A role for the Ras signalling pathway in synaptic transmission and long-term memory. *Nature* **390**, 281-286.

Brannan, C. I., Perkins, A. S., Vogel, K. S., Ratner, N., Nordlund, M. L., Reid, S. W., Buchberg, A. M., Jenkins, N. A., Parada, L. F., and Copeland, N. G. (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.

Brosi, R., Hauri, H. P., and Kramer, A. (1993): Separation of splicing factor SF3 into two components and purification of SF3a activity. *J Biol Chem* **268**, 17640-17646.

Buchberg, A. M., Cleveland, L. S., Jenkins, N. A., and Copeland, N. G. (1990): Sequence homology shared by neurofibromatosis type-1 gene and IRA-1 and IRA-2 negative regulators of the RAS cyclic AMP pathway. *Nature* **347**, 291-294.

Buratti, E., and Baralle, F. E. (2001): Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J Biol Chem* **276**, 36337-36343.

Buratti, E., Baralle, M., De Conti, L., Baralle, D., Romano, M., Ayala Y.M., Baralle, F.E. (2004): hnRNP binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF1 and TSHβ genes. *NAR* **32** (14): 4224-4236.

Burge, C. B., Padgett, R. A., and Sharp, P. A. (1998): Evolutionary fates and origins of U12-type introns. *Mol Cell* **2**, 773-785.

Burge, C. B., Sharp P.A. (1999). Splicing of precursor to mRNAs by the spliceosome. In The RNA World , edn 2, R. Gesteland, Atkins JF, ed. (New York, Cold Spring Harbor Laboratory Press).

Cappione, A. J., French, B. L., and Skuse, G. R. (1997): A potential role for NF1 mRNA editing in the pathogenesis of NF1 tumors. *Am J Hum Genet* **60**, 305-312.

Caputi, M., Casari, G., Guenzi, S., Tagliabue, R., Sidoli, A., Melo, C. A., and Baralle, F. E. (1994): A novel bipartite splicing enhancer modulates the differential processing of the human fibronectin EDA exon. *Nucleic Acids Res* **22**, 1018-1022.

Caputi, M., Zahler, A.M. (2002): SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. *EMBO J.* **21**: 845-855

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 (Semin. Med. Genet.)* **89**: 7-13.

Carothers, A. M., Urlaub, G., Grunberger, D., and Chasin, L. A. (1993): Splicing mutants and their second-site suppressors at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol Cell Biol*

13, 5085-5098.

Cartegni, L., Chew, S. L., and Krainer, A. R. (2002): Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**, 285-298.

Cavanee, W.K., Dryja, T.P, Phillips, R.A., Benedict, W.F., Godbout, R., Galie, B.L., Murphree, A.L., Strong, L.C., White, R.L. (1983): Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305**: 779-784.

Cawthon, R. M., Weiss, R., Xu, G. F., Viskochil, D., Culver, M., Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P., and et al. (1990): A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. *Cell* **62**, 193-201.

Champion-Arnaud, P., and Reed, R. (1994): The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes Dev* **8**, 1974-1983.

Chan, G.C., Nicholls, J.M., Lee, A.C., Chan, L.C., Lau, Y.L. (1996): Malignant peripheral neuroectodermal tumor in an infant with neurofibromatosis type 1. *Med Pediatr Oncol* **26**:215-219.

Chou, M.Y., Rooke, N., Turck, C.W., Black, D.L. (1999): hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol Cell Biol* **19**: 69-77.

Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977): An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1-8.

Cichowski, K., Shih, T. S., Schmitt, E., Santiago, S., Reilly, K., McLaughlin, M. E., Bronson, R. T., and 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., Halley, D. J., and 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.

Coleman, T. P., and Roesser, J. R. (1998): RNA secondary structure: an important cis-element in rat calcitonin/CGRP pre-messenger RNA splicing. *Biochemistry* **37**, 15941-15950.

Colley, A., Donnai, D., and Evans, D. G. (1996): Neurofibromatosis/Noonan phenotype: a variable feature of type 1 neurofibromatosis. *Clin Genet* **49**, 59-64.

Collins, C. A., and Guthrie, C. (1999): Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev* **13**, 1970-1982.

Colman, S. D., Williams, C. A., and Wallace, M. R. (1995): Benign neurofibromas in type 1 neurofibromatosis (NF1) show somatic deletions of the NF1 gene. *Nat Genet* **11**, 90-92.

Coolidge, C. J., Seely, R. J., and Patton, J. G. (1997): Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res* **25**, 888-896.

Cooper, T. A., and Mattox, W. (1997): The regulation of splice-site selection, and its role in human disease. *Am J Hum Genet* **61**, 259-266.

Costa, R. M., Federov, N. B., Kogan, J. H., Murphy, G. G., Stern, J., Ohno, M., Kucherlapati, R., Jacks, T., and Silva, A. J. (2002): Mechanism for the learning deficits in a mouse model of neurofibromatosis type 1. *Nature* **415**, 526-530.

Costa, R. M., Yang, T., Huynh, D. P., Pulst, S. M., Viskochil, D. H., Silva, A. J., and Brannan, C. I. (2001): Learning deficits, but normal development and tumor predisposition, in mice lacking exon 23a of Nf1. *Nat Genet* **27**, 399-405.

Croft, L., Schandorff, S., Clark, F., Burrage, K., Arctander, P., and Mattick, J. S. (2000): ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. *Nat Genet* **24**, 340-341.

Danglot, G., Regnier, V., Fauvet, D., Vassal, G., Kujas, M., and Bernheim, A. (1995): Neurofibromatosis 1 (NF1) mRNAs expressed in the central nervous system are differentially spliced in the 5' part of the gene. *Hum Mol Genet* **4**, 915-920.

Daston, M. M., Scrable, H., Nordlund, M., Sturbaum, A. K., Nissen, L. M., and Ratner, N. (1992): The protein product of the neurofibromatosis type 1 gene is expressed at highest abundance in neurons, Schwann cells, and oligodendrocytes. *Neuron* **8**, 415-428.

DeClue, J. E., Cohen, B. D., and Lowy, D. R. (1991): Identification and characterization of the neurofibromatosis type 1 protein product. *Proc Natl Acad Sci U S A* **88**, 9914-9918.

De Rosa, M., Fasano, C., Panariello, L., Scarano, M.I., Belli, G., Iannelli, A., Izzo, P. (2000): Evidence for a recessive inheritance of Turcot's syndrome caused by compound heterozygous mutations within the PMS2 gene. *Oncogene* **19**(13): 1719-1723.

De Vos, M., Hayward, B.E., Picton, S., Sheridan, E., Bonthron, D.T. (2004): Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. *Am J Hum Genet* **74** (5): 954-964.

Digilio, M. C., Conti, E., Sarkozy, A., Mingarelli, R., Dottorini, T., Marino, B., Pizzuti, A., and Dallapiccola, B. (2002): Grouping of multiple-lentigines/LEOPARD and Noonan syndromes on the PTPN11 gene. *Am J Hum Genet* **71**, 389-394.

DiMario, F. J., Jr., Ramsby, G., Greenstein, R., Langshur, S., and Dunham, B. (1993): Neurofibromatosis type 1: magnetic resonance imaging findings. *J Child Neurol* **8**, 32-39.

Dominski, Z., and Kole, R. (1991): Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol* **11**, 6075-6083.

Dominski, Z., and Kole, R. (1994): Identification and characterization by antisense oligonucleotides of exon and intron sequences required for splicing. *Mol Cell Biol* **14**, 7445-7454.

Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., and Burd, C. G. (1993): hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* **62**, 289-321.

Easton, D. F., Ponder, M. A., Huson, S. M., and Ponder, B. A. (1993): An analysis of variation in expression of neurofibromatosis (NF) type 1 (NF1): evidence for modifying genes. *Am J Hum Genet* **53**, 305-313.

Eldridge, A. G., Li, Y., Sharp, P. A., and Blencowe, B. J. (1999): The SRm160/300 splicing coactivator is required for exon-enhancer function. *Proc Natl Acad Sci U S A* **96**, 6125-6130.

Eperon, I. C., Ireland, D. C., Smith, R. A., Mayeda, A., and Krainer, A. R. (1993): Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. *Embo J* **12**, 3607-3617.

Fahsold, R., Hoffmeyer, S., Mischung, C., Gille, C., Ehlers, C., Kucukceylan, N., Abdel-Nour, M., Gewies, A., Peters, H., Kaufmann, D., *et al.* (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.

Fairbrother, W. G., and Chasin, L. A. (2000): Human genomic sequences that inhibit splicing. *Mol Cell Biol* **20**, 6816-6825.

Fischer, U., Sumpter, V., Sekine, M., Satoh, T., and Luhrmann, R. (1993): Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *Embo J* **12**, 573-583.

Fu, X. D. (1993): Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature* **365**, 82-85.

Gersappe, A., and Pintel, D. J. (1999): CA- and purine-rich elements form a novel bipartite exon enhancer which governs inclusion of the minute virus of mice NS2-specific exon in both singly and doubly spliced mRNAs. *Mol Cell Biol* **19**, 364-375.

Giordano, M. J., Mahadeo, D. K., He, Y. Y., Geist, R. T., Hsu, C., and Gutmann, D. H. (1996): Increased expression of the neurofibromatosis 1 (NF1) gene product, neurofibromin, in astrocytes in response to cerebral ischemia. *J Neurosci Res* **43**, 246-253.

Goguel, V., and Rosbash, M. (1993): Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell* **72**, 893-901.

Gozani, O., Feld, R., and Reed, R. (1996): Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev* **10**, 233-243.

Graveley, B. R. (2000): Sorting out the complexity of SR protein functions. *Rna* **6**, 1197-1211.

Graveley, B. R. (2001): Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* **17**, 100-107.

Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K., and Hutton, M. (1999): 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J Biol Chem* **274**, 15134-15143.

Guha, A., Lau, N., Huvar, I., Gutmann, D., Provias, J., Pawson, T., and Boss, G. (1996): Ras-GTP levels are elevated in human NF1 peripheral nerve tumors. *Oncogene* **12**, 507-513.

Gui, J. F., Lane, W. S., and Fu, X. D. (1994): A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature* **369**, 678-682.

Gunderson, K. L., and Kopito, R. R. (1994): Effects of pyrophosphate and nucleotide analogs suggest a role for ATP hydrolysis in cystic fibrosis transmembrane regulator channel gating. *J Biol Chem* **269**, 19349-19353.

Guo, H. F., Tong, J., Hannan, F., Luo, L., and Zhong, Y. (2000): A neurofibromatosis-1-regulated pathway is required for learning in Drosophila. *Nature* **403**, 895-898.

Gutmann, D. H., Aylsworth, A., Carey, J. C., Korf, B., Marks, J., Pyeritz, R. E., Rubenstein, A., and Viskochil, D. (1997): The diagnostic evaluation and multidisciplinary management of neurofibromatosis 1 and neurofibromatosis 2. *Jama* **278**, 51-57.

Gutmann, D. H., and Collins, F. S. (1993): The neurofibromatosis type 1 gene and its protein product,

neurofibromin. *Neuron* **10**, 335-343.

Gutmann, D. H., Wood, D. L., and Collins, F. S. (1991): Identification of the neurofibromatosis type 1 gene product. *Proc Natl Acad Sci U S A* **88**, 9658-9662.

Hajra, A., Martin-Gallardo, A., Tarle, S. A., Freedman, M., Wilson-Gunn, S., Bernards, A., and Collins, F. S. (1994): DNA sequences in the promoter region of the NF1 gene are highly conserved between human and mouse. *Genomics* **21**, 649-652.

Hanahan, D., and Weinberg, R. A. (2000): The hallmarks of cancer. *Cell* **100**, 57-70.

Hawkins, J. D. (1988): A survey on intron and exon lengths. *Nucleic Acids Res* **16**, 9893-9908.

Heim, R. A., Kam-Morgan, L. N., Binnie, C. G., Corns, D. D., Cayouette, M. C., Farber, R. A., Aylsworth, A. S., Silverman, L. M., and Luce, M. C. (1995): Distribution of 13 truncating mutations in the neurofibromatosis 1 gene. *Hum Mol Genet* **4**, 975-981.

Hoffman, B. E., and Grabowski, P. J. (1992): U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes Dev* **6**, 2554-2568.

Hoffmeyer, S., Assum, G., Griesser, J., Kaufmann, D., Nurnberg, P., and Krone, W. (1995): On unequal allelic expression of the neurofibromin gene in neurofibromatosis type 1. *Hum Mol Genet* **4**, 1267-1272.

Hofmann, Y., and Wirth, B. (2002): hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1. *Hum Mol Genet* **11**, 2037-2049.

Horii, A., Han, H. J., Sasaki, S., Shimada, M., and Nakamura, Y. (1994): Cloning, characterization and chromosomal assignment of the human genes homologous to yeast PMS1, a member of mismatch repair genes. *Biochem Biophys Res Commun* **204**, 1257-1264.

Huh, G. S., and Hynes, R. O. (1993): Elements regulating an alternatively spliced exon of the rat fibronectin gene. *Mol Cell Biol* **13**, 5301-5314.

Huson, S. M. (1994). The Neurofibromatosis: A pathogenetic and clinical overview. In, L. Chapman and Hall, ed.

Huson, S. M., Compston, D. A., Harper, P. S. (1989): A genetic study of von Recklinghausen neurofibromatosis in east Wales. II. Guidelines for genetic counselling. . *J Med Genet* **26,** 712-721.

Huson, S. M., Compston, D. A., Clark, P., and Harper, P. S. (1989): A genetic study of von Recklinghausen neurofibromatosis in south east Wales. I. Prevalence, fitness, mutation rate, and effect of parental transmission on severity. *J Med Genet* **26**, 704-711.

Huson, S. M., Harper, P. S., and Compston, D. A. (1988): Von Recklinghausen neurofibromatosis. A clinical and population study in south-east Wales. *Brain* **111**, 1355-1381.

Ingram, D. A., Yang, F. C., Travers, J. B., Wenning, M. J., Hiatt, K., New, S., Hood, A., Shannon, K., Williams, D. A., and Clapp, D. W. (2000): Genetic and biochemical evidence that haploinsufficiency of the Nf1 tumor suppressor gene modulates melanocyte and mast cell fates in vivo. *J Exp Med* **191**, 181-188.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, I. W. (1994): A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**, 657-668.

Izawa, I., Tamaki, N., and Saya, H. (1996): Phosphorylation of neurofibromatosis type 1 gene product (neurofibromin) by cAMP-dependent protein kinase. *FEBS Lett* **382**, 53-59.

Jacks, T., Shih, T. S., Schmitt, E. M., Bronson, R. T., Bernards, A., and Weinberg, R. A. (1994): Tumour predisposition in mice heterozygous for a targeted mutation in Nf1. *Nat Genet* **7**, 353-361.

Jeffreys, A. J., and Flavell, R. A. (1977): The rabbit beta-globin gene contains a large large insert in the coding sequence. *Cell* **12**, 1097-1108.

Kambach, C., Walke, S., and Nagai, K. (1999): Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr Opin Struct Biol* **9**, 222-230.

Kawano, T., Fujita, M., and Sakamoto, H. (2000): Unique and redundant functions of SR proteins, a conserved family of splicing factors, in Caenorhabditis elegans development. *Mech Dev* **95**, 67-76.

Kayes, L. M., Burke, W., Riccardi, V. M., Bennett, R., Ehrlich, P., Rubenstein, A., and Stephens, K. (1994): Deletions spanning the neurofibromatosis 1 gene: identification and phenotype of five patients. *Am J Hum Genet* **54**, 424-436.

Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A., and Manley, J. L. (1994): Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368**, 119-124.

Konarska, M. M., and Sharp, P. A. (1987): Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell* **49**, 763-774.

Krainer, A. R., Conway, G. C., and Kozak, D. (1990): The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* **62**, 35-42.

Krawczak, M., Reiss, J., and 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.

Krecic, A. M., and Swanson, M. S. (1999): hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* **11**, 363-371.

Lamond, A. I. (1993): The spliceosome. *Bioessays* **15**, 595-603.

Lang, K. M., and Spritz, R. A. (1983): RNA splice site selection: evidence for a 5' leads to 3' scanning model. *Science* **220**, 1351-1355.

Langford, C. J., Klinz, F. J., Donath, C., and Gallwitz, D. (1984): Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. *Cell* **36**, 645-653.

Legius, E., Marchuk, D. A., Hall, B. K., Andersen, L. B., Wallace, M. R., Collins, F. S., and Glover, T. W. (1992): NF1-related locus on chromosome 15. *Genomics* **13**, 1316-1318.

Li, G.M., Modrich, P. (1995): Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells bya heterodimer of human MutL homologs. *PNAS* **92**(6): 1950-1954.

Li, Y., O'Connell, P., Breidenbach, H. H., Cawthon, R., Stevens, J., Xu, G., Neil, S., Robertson, M., White, R., and Viskochil, D. (1995): Genomic organization of the neurofibromatosis 1 gene (NF1). *Genomics* **25**, 9-18.

Listernick, R., Charrow, J., Greenwald, M., and Mets, M. (1994): Natural history of optic pathway tumors

in children with neurofibromatosis type 1: a longitudinal study. *J Pediatr* **125**, 63-66.

Longman, D., Johnstone, I. L., and Caceres, J. F. (2000): Functional characterization of SR and SR-related genes in Caenorhabditis elegans. *Embo J* **19**, 1625-1637.

Lopez, A. J. (1998): Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu Rev Genet* **32**, 279-305.

Madhani, H. D., and Guthrie, C. (1994): Dynamic RNA-RNA interactions in the spliceosome. *Annu Rev Genet* **28**, 1-26.

Malhotra, R., and Ratner, N. (1994): Localization of neurofibromin to keratinocytes and melanocytes in developing rat and human skin. *J Invest Dermatol* **102**, 812-818.

Marchuk, D. A., Saulino, A. M., Tavakkol, R., Swaroop, M., Wallace, M. R., Andersen, L. B., Mitchell, A. L., Gutmann, D. H., Boguski, M., and Collins, F. S. (1991): cDNA cloning of the type 1 neurofibromatosis gene: complete sequence of the NF1 gene product. *Genomics* **11**, 931-940.

Marshall, M. (1995): Interactions between Ras and Raf: key regulatory proteins in cellular transformation. *Mol Reprod Dev* **42**, 493-499.

Marti, T.M., Kunz, C., Fleck, O. (2002): DNA mismatch repair and mutation avoidance pathways. *J Cell Physiol* **191**(1): 28-41.

Martin, G. A., Viskochil, D., Bollag, G., McCabe, P. C., Crosier, W. J., Haubruck, H., Conroy, L., Clark, R., O'Connell, P., Cawthon, R. M., and et al. (1990): The GAP-related domain of the neurofibromatosis type 1 gene product interacts with ras p21. *Cell* **63**, 843-849.

Martinez, J. M., Breidenbach, H. H., and Cawthon, R. (1996): Long RT-PCR of the entire 8.5-kb NF1 open reading frame and mutation detection on agarose gels. *Genome Res* **6**, 58-66.

Mattocks, C., Tarpey, P., Bobrow, M., and Whittaker, J. (2000): Comparative sequence analysis (CSA): a new sequence-based method for the identification and characterization of mutations in DNA. *Hum Mutat* **16**, 437-443.

Mattocks, C., Baralle, D., Tarpey, P., ffrench-Constant, C., Bobrow, M., Whittaker, J. (2004): Automated comparative sequence analysis identifies mutations in 89% of NF1 patients and confirms a mutation cluster in exons 11-17 distinct from the GAP related domain. *J Med Genet* **41**:e48.

Mayeda, A., Munroe, S. H., Caceres, J. F., and Krainer, A. R. (1994): Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *Embo J* **13**, 5483-5495.

Mayes, A. E., Verdone, L., Legrain, P., and Beggs, J. D. (1999): Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *Embo J* **18**, 4321-4331.

McPherson, J. D., Marra, M., Hillier, L., Waterston, R. H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E. R., Wilson, R. K., *et al.* (2001): A physical map of the human genome. *Nature* **409**, 934-941.

Mendez, M. M. H., Opitz, J. M. (1985): Noonan syndrome: A review. *Am J Med Genet* **21(3)**: 493-506.

Messiaen, L., Callens, T., De Paepe, A., Craen, M., and 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. M., Callens, T., Mortier, G., Beysen, D., Vandenbroucke, I., Van Roy, N., Speleman, F., and

Paepe, A. D. (2000): Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Hum Mutat* **15**, 541-555.

Metheny, L. J., and Skuse, G. R. (1996): NF1 mRNA isoform expression in PC12 cells: modulation by extrinsic factors. *Exp Cell Res* **228**, 44-49.

Misteli, T., Caceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F., and Spector, D. L. (1998): Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J Cell Biol* **143**, 297-307.

Mullen, M. P., Smith, C. W., Patton, J. G., and Nadal-Ginard, B. (1991): Alpha-tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes Dev* **5**, 642-655.

Mulvihill, J. J. (1988): Neurofibromatosis: history, nomenclature, and natural history. *Neurofibromatosis* **1**, 124-131.

Muro, A. F., Caputi, M., Pariyarath, R., Pagani, F., Buratti, E., and Baralle, F. E. (1999): Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Mol Cell Biol* **19**, 2657-2671.

Nakai, K., and Sakamoto, H. (1994): Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* **141**, 171-177.

Nelson, K. K., and Green, M. R. (1989): Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes Dev* **3**, 1562-1571.

Nicolaides, N. C., Carter, K. C., Shell, B. K., Papadopoulos, N., Vogelstein, B., and Kinzler, K. W. (1995): Genomic organization of the human PMS2 gene family. *Genomics* **30**, 195-206.

(NIH), N. I. o. H. (1987). Consensus development conference statement: Neurofibromatosis (Bethesda, Md; USA), pp. 13-15.

Niksic, M., Romano, M., Buratti, E., Pagani, F., and Baralle, F. E. (1999): Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9. *Hum Mol Genet* **8**, 2339-2349.

Nilsen, T. W. (1994): RNA-RNA interactions in the spliceosome: unraveling the ties that bind. *Cell* **78**, 1-4.

Nordlund, M., Gu, X., Shipley, M. T., and Ratner, N. (1993): Neurofibromin is enriched in the endoplasmic reticulum of CNS neurons. *J Neurosci* **13**, 1588-1600.

Nordlund, M. L., Rizvi, T. A., Brannan, C. I., and Ratner, N. (1995): Neurofibromin expression and astrogliosis in neurofibromatosis (type 1) brains. *J Neuropathol Exp Neurol* **54**, 588-600.

North, K., Joy, P., Yuille, D., Cocks, N., and Hutchins, P. (1995): Cognitive function and academic performance in children with neurofibromatosis type 1. *Dev Med Child Neurol* **37**, 427-436.

North, K., Joy, P., Yuille, D., Cocks, N., Mobbs, E., Hutchins, P., McHugh, K., and de Silva, M. (1994): Specific learning disability in children with neurofibromatosis type 1: significance of MRI abnormalities. *Neurology* **44**, 878-883.

Norton, P. A. (1994): Polypyrimidine tract sequences direct selection of alternative branch sites and influence protein binding. *Nucleic Acids Res* **22**, 3854-3860.

Obringer, A. C., Meadows, A. T., and Zackai, E. H. (1989): The diagnosis of neurofibromatosis-1 in the

child under the age of 6 years. *Am J Dis Child* **143**, 717-719.

Pagani, F., Buratti, E., Stuani, C., Bendix, R., Dork, T., and Baralle, F. E. (2002): A new type of mutation causes a splicing defect in ATM. *Nat Genet* **30**, 426-429.

Pagani, F., Buratti, E., Stuani, C., Romano, M., Zuccato, E., Niksic, M., Giglio, L., Faraguna, D., and Baralle, F. E. (2000): Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *J Biol Chem* **275**, 21041-21047.

Park, V. M., and Pivnick, E. K. (1998): Neurofibromatosis type 1 (NF1): a protein truncation assay yielding identification of mutations in 73% of patients. *J Med Genet* **35**, 813-820.

Peltomaki, P. (2001): Deficient DNA mismatch repair: a common etiologic factor for cancer. *Hum Mol Genet* **10**(7): 735-740.

Puig, O., Gottschalk, A., Fabrizio, P., and Seraphin, B. (1999): Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. *Genes Dev* **13**, 569-580.

Purandare, S. M., Huntsman Breidenbach, H., Li, Y., Zhu, X. L., Sawada, S., Neil, S. M., Brothman, A., White, R., Cawthon, R., and Viskochil, D. (1995): Identification of neurofibromatosis 1 (NF1) homologous loci by direct sequencing, fluorescence in situ hybridization, and PCR amplification of somatic cell hybrids. *Genomics* **30**, 476-485.

Quattrin, T., McPherson, E., and Putnam, T. (1987): Vertical transmission of the neurofibromatosis/Noonan syndrome. *Am J Med Genet* **26**, 645-649.

Raffel, C., McComb, J.G., Bodner, S., Gilles, F.E. (1989): Benign brain stem lesions in pediatric patients with neurofinromatosis: case reports. *Neurosurgery* **25**(6): 959-964.

Raker, V. A., Plessel, G., and Luhrmann, R. (1996): The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *Embo J* **15**, 2256-2269.

Reed, R. (1989): The organization of 3' splice-site sequences in mammalian introns. *Genes Dev* **3**, 2113-2123.

Reed, R. (1996): Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr Opin Genet Dev* **6**, 215-220.

Reed, R., and Maniatis, T. (1985): Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell* **41**, 95-105.

Reed, R., and Maniatis, T. (1988): The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev* **2**, 1268-1276.

Riccardi, V. M. (1992): Type 1 neurofibromatosis and the pediatric patient. *Curr Probl Pediatr* **22**, 66-106; discussion 107.

Ricciardone, M.D., Ozcelik, T., Cevher, B., Ozdag, H., Tuncer, M., Gurgey, Uzunalimoglu, O., Cetinkaya, H., Tanyeli, A., Erken, E., Ozturk, M. (1999): Human MLH1 deficiency predisposes to haematological malignancy and neurofibromatosis type 1. *Can Res* **59**(2): 290-293.

Rizvi, T. A., Akunuru, S., de Courten-Myers, G., Switzer, R. C., 3rd, Nordlund, M. L., and Ratner, N. (1999): Region-specific astrogliosis in brains of mice heterozygous for mutations in the neurofibromatosis type 1 (Nf1) tumor suppressor. *Brain Res* **816**, 111-123.

Robberson, B. L., Cote, G. J., and Berget, S. M. (1990): Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* **10**, 84-94.

Romano, M.,Marcucci, R., Baralle, F.E. (2001): Splicing of constitutive upstream introns is essential for the recognition of intra-exonic suboptimal splice sites in the thrombopoietin gene. *NAR* **29**(4): 886-894.

Roscigno, R. F., Weiner, M., and Garcia-Blanco, M. A. (1993): A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. *J Biol Chem* **268**, 11222-11229.

Rosman, N. P., and Pearce, J. (1967): The brain in multiple neurofibromatosis (von Recklinghausen's disease): a suggested neuropathological basis for the associated mental defect. *Brain* **90**, 829-838.

Roth, M. B., Zahler, A. M., and Stolk, J. A. (1991): A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J Cell Biol* **115**, 587-596.

Ruskin, B., Greene, J. M., and Green, M. R. (1985): Cryptic branch point activation allows accurate in vitro splicing of human beta-globin intron mutants. *Cell* **41**, 833-844.

Schaal, T. D., and Maniatis, T. (1999): Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol Cell Biol* **19**, 1705-1719.

Scheurlen, W. G., and 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.

Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, J. E., and Zipursky, S. L. (2000): Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671-684.

Serra, E., Puig, S., Otero, D., Gaona, A., Kruyer, H., Ars, E., Estivill, X., and Lazaro, C. (1997): Confirmation of a double-hit model for the NF1 gene in benign neurofibromas. *Am J Hum Genet* **61**, 512-519.

Shannon, K. M., O'Connell, P., Martin, G. A., Paderanga, D., Olson, K., Dinndorf, P., and McCormick, F. (1994): Loss of the normal NF1 allele from the bone marrow of children with type 1 neurofibromatosis and malignant myeloid disorders. *N Engl J Med* **330**, 597-601.

Shapiro, M. B., and Senapathy, P. (1987): RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* **15**, 7155-7174.

Sharland, M., Burch, M., McKenna, W. M., and Paton, M. A. (1992): A clinical study of Noonan syndrome. *Arch Dis Child* **67**, 178-183.

Shen, M. H., Harper, P. S., and Upadhyaya, M. (1993): Neurofibromatosis type 1 (NF1): the search for mutations by PCR-heteroduplex analysis on Hydrolink gels. *Hum Mol Genet* **2**, 1861-1864.

Sherman, L., Daston, M.M., Ratner, N. (1998): Neurofibromin: distribution, cell biology and role in neurofibromatosis type 1. *Book- Neurofibromatosis type 1 from genotype to phenotype, Bios scientific publishers.* Ed: M. Upadhyaya and D.N. Cooper.

Shiga, N., Takeshima, Y., Sakamoto, H., Inoue, K., Yokota, Y., Yokoyama, M., and Matsuo, M. (1997): Disruption of the splicing enhancer sequence within exon 27 of the dystrophin gene by a nonsense mutation induces partial skipping of the exon and is responsible for Becker muscular dystrophy. *J Clin Invest* **100**, 2204-2210.

Side, L.E., Shannon, K.M. (1998): The NF1 gene as a tumour suppressor. *Book- Neurofibromatosis type 1 from genotype to phenotype, Bios scientific publishers.* Ed: M. Upadhyaya and D.N. Cooper.

Silva, A. J., Frankland, P. W., Marowitz, Z., Friedman, E., Laszlo, G. S., Cioffi, D., Jacks, T., Bourtchuladze, R., and Lazlo, G. (1997): A mouse model for the learning and memory deficits associated with neurofibromatosis type I. *Nat Genet* **15**, 281-284.

Silva, A. J., Kogan, J. H., Frankland, P. W., and Kida, S. (1998): CREB and memory. *Annu Rev Neurosci* **21**, 127-148.

Singh, R., Valcarcel, J., and Green, M. R. (1995): Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**, 1173-1176.

Skuse, G. R., Cappione, A.J. (1997): RNA processing and clinical variability in neurofibromatosis type 1 (NF1). *Hum Mol Genet* **6**(10): 1707-1712.

Skuse, G. R., Cappione, A. J., Sowden, M., Metheny, L. J., and Smith, H. C. (1996): The neurofibromatosis type I messenger RNA undergoes base-modification RNA editing. *Nucleic Acids Res* **24**, 478-485.

Smith, C. W., Porro, E. B., Patton, J. G., and Nadal-Ginard, B. (1989): Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. *Nature* **342**, 243-247.

Sontheimer, E. J., Sun, S., and Piccirilli, J. A. (1997): Metal ion catalysis during splicing of premessenger RNA. *Nature* **388**, 801-805.

Stamm, S., Zhu, J., Nakai, K., Stoilov, P., Stoss, O., and Zhang, M. Q. (2000): An alternative-exon database and its statistical analysis. *DNA Cell Biol* **19**, 739-756.

Stark, H., Dube, P., Luhrmann, R., and Kastner, B. (2001): Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539-542.

Steitz, T. A., and Steitz, J. A. (1993): A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* **90**, 6498-6502.

Sterner, D. A., Carlo, T., and Berget, S. M. (1996): Architectural limits on split genes. *Proc Natl Acad Sci U S A* **93**, 15081-15085.

Suzuki, H., Takahashi, K., Kubota, Y., and 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.

Talerico, M., and Berget, S. M. (1994): Intron definition in splicing of small Drosophila introns. *Mol Cell Biol* **14**, 3434-3445.

Tanaka, K., Lin, B. K., Wood, D. R., and Tamanoi, F. (1991): IRA2, an upstream negative regulator of RAS in yeast, is a RAS GTPase-activating protein. *Proc Natl Acad Sci U S A* **88**, 468-472.

Tartaglia, M., Mehler, E. L., Goldberg, R., Zampino, G., Brunner, H. G., Kremer, H., van der Burgt, I., Crosby, A. H., Ion, A., Jeffery, S., *et al.* (2001): Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat Genet* **29**, 465-468.

Tassabehji, M., Strachan, T., Sharland, M., Colley, A., Donnai, D., Harris, R., and Thakker, N. (1993): Tandem duplication within a neurofibromatosis type 1 (NF1) gene exon in a family with features of Watson syndrome and Noonan syndrome. *Am J Hum Genet* **53**, 90-95.

Teigelkamp, S., Achsel, T., Mundt, C., Gothel, S. F., Cronshagen, U., Lane, W. S., Marahiel, M., and Luhrmann, R. (1998): The 20kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *Rna* **4**, 127-141.

Thiel, G., Marczinek, K., Neumann, R., Witkowski, R., Marchuk, D. A., and Nurnberg, P. (1995): Somatic mutations in the neurofibromatosis 1 gene in gliomas and primitive neuroectodermal tumours. *Anticancer Res* **15**, 2495-2499.

Tian, H., and Kole, R. (1995): Selection of novel exon recognition elements from a pool of random sequences. *Mol Cell Biol* **15**, 6291-6298.

Tian, M., and Maniatis, T. (1993): A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell* **74**, 105-114.

Tian, M., and Maniatis, T. (1994): A splicing enhancer exhibits both constitutive and regulated activities. *Genes Dev* **8**, 1703-1712.

Tong, J., Hannan, F., Zhu, Y., Bernards, A., and Zhong, Y. (2002): Neurofibromin regulates G protein-stimulated adenylyl cyclase activity. *Nat Neurosci* **5**, 95-96.

Trimbath, J.D., Petersen, G.M., Erdman, S.H., Ferre, M., Luce, M.C., Giardello, F.M. (2001): Café-au-lait spots and early onset colorectal neoplasia: a variant of HNPCC? *Fam Can* **1**(2): 101-105.

Turner, D. H., and Sugimoto, N. (1988): RNA structure prediction. *Annu Rev Biophys Biophys Chem* **17**, 167-192.

Umen, J. G., and Guthrie, C. (1995): The second catalytic step of pre-mRNA splicing. *Rna* **1**, 869-885.

Upadhyaya, M., Osborn, M. J., Maynard, J., Kim, M. R., Tamanoi, F., and Cooper, D. N. (1997): Mutational and functional analysis of the neurofibromatosis type 1 (NF1) gene. *Hum Genet* **99**, 88-92.

Upadhyaya, M., Shaw, D. J., and 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., Shen, M., Cherryson, A., Farnham, J., Maynard, J., Huson, S. M., and Harper, P. S. (1992): Analysis of mutations at the neurofibromatosis 1 (NF1) locus. *Hum Mol Genet* **1**, 735-740.

Upadhyaya, M. a. C. D. N. (1998). Neurofibromatosis Type 1, from genotype to phenotype, BIOS Scientific Publishers Limited).

Valero, M. C., Velasco, E., Moreno, F., and Hernandez-Chico, C. (1994): Characterization of four mutations in the neurofibromatosis type 1 gene by denaturing gradient gel electrophoresis (DGGE). *Hum Mol Genet* **3**, 639-641.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., *et al.* (2001): The sequence of the human genome. *Science* **291**, 1304-1351.

Viskochil, D., White, R., and Cawthon, R. (1993): The neurofibromatosis type 1 gene. *Annu Rev Neurosci* **16**, 183-205.

Vogel, K. S., Klesse, L. J., Velasco-Miguel, S., Meyers, K., Rushing, E. J., and Parada, L. F. (1999): Mouse tumor model for neurofibromatosis type 1. *Science* **286**, 2176-2179.

Vuillaumier-Barrot, S., Barnier, A., Cuer, M., Durand, G., Grandchamp, B., and Seta, N. (1999):

Characterization of the 415G>A (E139K) PMM2 mutation in carbohydrate-deficient glycoprotein syndrome type Ia disrupting a splicing enhancer resulting in exon 5 skipping. *Hum Mutat* **14**, 543-544.

Wagner, E. J., and Garcia-Blanco, M. A. (2001): Polypyrimidine tract binding protein antagonizes exon definition. *Mol Cell Biol* **21**, 3281-3288.

Wallace, M. R., Marchuk, D. A., Andersen, L. B., Letcher, R., Odeh, H. M., Saulino, A. M., Fountain, J. W., Brereton, A., Nicholson, J., Mitchell, A. L., and et al. (1990): Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science* **249**, 181-186.

Watakabe, A., Tanaka, K., and Shimura, Y. (1993): The role of exon sequences in splice site selection. *Genes Dev* **7**, 407-418.

Weighardt, F., Biamonti, G., and Riva, S. (1996): The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. *Bioessays* **18**, 747-756.

Will, C. L. (1997). snRNP structure and function. In Eukaryotic mRNA processing, K. A.R., ed. (O.I.P), pp. 130-173.

Will, C. L., and Luhrmann, R. (1997): Protein functions in pre-mRNA splicing. *Curr Opin Cell Biol* **9**, 320-328.

Wu, B. L., Austin, M. A., Schneider, G. H., Boles, R. G., and 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, J. Y., and Maniatis, T. (1993): Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061-1070.

Wu, R., Legius, E., Robberecht, W., Dumoulin, M., Cassiman, J. J., and Fryns, J. P. (1996): Neurofibromatosis type I gene mutation in a patient with features of LEOPARD syndrome. *Hum Mutat* **8**, 51-56.

Wu, R., Lopez-Correa, C., Rutkowski, J. L., Baumbach, L. L., Glover, T. W., and Legius, E. (1999a): Germline mutations in NF1 patients with malignancies. *Genes Chromosomes Cancer* **26**, 376-380.

Wu, S., Romfo, C. M., Nilsen, T. W., and Green, M. R. (1999b): Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**, 832-835.

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

Xu, H., and Gutmann, D. H. (1997): Mutations in the GAP-related domain impair the ability of neurofibromin to associate with microtubules. *Brain Res* **759**, 149-152.

Xu, R., Teng, J., and Cooper, T. A. (1993): The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Mol Cell Biol* **13**, 3660-3674.

Xu, W., Mulligan, L., Ponder, M.A., Liu, L., Smith, B.A., Mathew, C.G., Ponder, B.A. (1992): Loss of alleles in pheochromocytomas from patients with type 1 neurofibromatosis. *Genes Chrom. Cancer* **4**: 337-341.

Yean, S. L., Wuenschell, G., Termini, J., and Lin, R. J. (2000): Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**, 881-884.

Zahler, A. M., Lane, W. S., Stolk, J. A., and Roth, M. B. (1992): SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev* **6**, 837-847.

Zamore, P. D., and Green, M. R. (1991): Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *Embo J* **10**, 207-214.

Zhang, D., and Rosbash, M. (1999): Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev* **13**, 581-592.

Zhu, Y., Ghosh, P., Charnay, P., Burns, D. K., and Parada, L. F. (2002): Neurofibromas in NF1: Schwann cell origin and role of tumor environment. *Science* **296**, 920-922.

Zhu, Y., Romero, M. I., Ghosh, P., Ye, Z., Charnay, P., Rushing, E. J., Marth, J. D., and Parada, L. F. (2001): Ablation of NF1 function in neurons induces abnormal development of cerebral cortex and reactive gliosis in the brain. *Genes Dev* **15**, 859-876.

Zuker, M., and Stiegler, P. (1981): Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* **9**, 133-148.

Zuo, P., and Maniatis, T. (1996): The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev* **10**, 1356-1368.

Zuo, P., and Manley, J. L. (1994): The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites. *Proc Natl Acad Sci U S A* **91**, 3363-3367.

American Journal of Medical Genetics 119A:1–8 (2003)

# Different Mutations in the *NF1* Gene Are Associated With Neurofibromatosis–Noonan Syndrome (NFNS)

Diana Baralle,[1*] Chris Mattocks,[1] Kamini Kalidas,[2] Frances Elmslie,[3] Joanne Whittaker,[1] Melissa Lees,[2] Nicola Ragge,[4] Michael A. Patton,[2] Robin M. Winter,[5] and Charles ffrench-Constant[1]

DOI 10.1002/ajmg.a.20023

© 2003 Wiley-Liss, Inc.

LETTER TO JMG

# Identification of a mutation that perturbs *NF1* gene splicing using genomic DNA samples and a minigene assay

M Baralle, D Baralle, L De Conti, C Mattocks, J Whittaker, A Knezevich, C ffrench-Constant, F E Baralle

*J Med Genet* 2003;**40**:220–222

ONLINE MUTATION REPORT

# Automated comparative sequence analysis identifies mutations in 89% of NF1 patients and confirms a mutation cluster in exons 11–17 distinct from the GAP related domain

C Mattocks, D Baralle, P Tarpey, C ffrench-Constant, M Bobrow, J Whittaker

*J Med Genet* 2004;41:e48 (http://www.jmedgenet.com/cgi/content/full/41/4/e48). doi: 10.1136/jmg.2003.011890

Published online August 6, 2004

*4224–4236 Nucleic Acids Research, 2004, Vol. 32, No. 14*
*doi:10.1093/nar/gkh752*

# hnRNP H binding at the 5′ splice site correlates with the pathological effect of two intronic mutations in the *NF-1* and *TSH*β genes

Emanuele Buratti[1], Marco Baralle[1], Laura De Conti[1], Diana Baralle[2], Maurizio Romano[1,3], Youhna M. Ayala[1] and Francisco E. Baralle[1,*]

*Nucleic Acids Research, Vol. 32 No. 14 © Oxford University Press 2004; all rights reserved*