# Molecular Genetics of Cornelia de Lange Syndrome

Tzu-Jou Wang

A thesis submitted for the degree of
Doctor of Philosophy
in the Institute of Human Genetics
at the University of Newcastle-upon-Tyne



December 2006

# ABSTRACT

Cornelia de Lange syndrome (CdLS) is a multiple malformation disorder characterized by peculiar facial features, growth and mental retardation, and a variety of other abnormalities affecting multiple organ systems. The molecular pathology underlying the disease phenotypes is not known. In the past, 3q26.3 was considered as a candidate region. However, no patient-specific mutations in the giant gene, *NAALADL2*, truncated by the 3q26.3 breakpoint were found. Thus, in this study, we tried to find the disease gene in loci other than 3q26.3 mainly by FISH mapping in patients with t(5;13), t(14;21) *de novo* balanced translocation. Of all the three regions (5p13, 13q12, 14q32) we studied, the novel gene that crosses the 5p translocation breakpoint was found to be the most likely candidate gene for CdLS. Subsequent finer FISH mapping using fosmid clones as probes confirmed the novel gene was truncated by the translocation breakpoint in intron 1. Standard point mutational screening of the newly found gene detected patient-specific mutations. The gene was named, *NIPBL*, which is the major determinant for this rare syndrome (Tonkin et al., 2004). Analysis of the *NIPBL* gene in further patients with CdLS revealed a wide variety of pathogenic mutations. More than half of the mutations resulted in premature termination. Loss of function causing haploinsufficiency is supported by patients with heterozygous deletion involving the whole *NIPBL* gene. No genotype-phenotype correlation was observed. The exact functions of delangin are not known.

Tissue *in situ* studies of *NIPBL*, using riboprobes in early human embryos, showed expression in developing brain, limbs, muscle, bone, renal tubules and lung bronchioles. In our preliminary data, the expression of *NIPBL* is mainly in the active

proliferating tissue in early mammalian development. However, the tissue expression pattern in later mammalian development is not known.

The predicted protein sequence of the *NIPBL* product, delangin, is 2804 amino acids for the long isoform and 2697 amino acids for the short isoform. In the carboxy-terminal region there is a clustering of HEAT repeats and these motifs (also called adherin) are highly conserved in the carboxy-terminal region of delangin homologues, Nipped-B of *Drosophila*, Scc2 of *Saccharomyces cerevisiae* and Mis4 of *Schizosaccharomyces pombe*, which are required for mitotic sister chromatid cohesion. Delangin is also related to Rad9 of *Coprinus cinereus*, which is required for DNA repair and meiotic chromosome paring. The Scc2 and Mis4 yeast, and Drosophila Nipped-B adherin homologues of human delangin are required for the Cohesin protein complex that mediates sister chromatid cohesin to associate with chromosomes. The major role in yeast and Xenopus Scc2 is to interact with Scc4 to help in loading the cohesin ring onto chromatin in the early S phase during cell replication. We also identified the sequence of the human Scc4 homologue, which we call human *Mau-2*. Tissue in situ hybridization data using a probe to dectect human *Mau-2* showed a similar expression pattern to *NIPBL* in human embryos. However, no *hMau-2* mutations were detected in CdLS patients in our mutation screen.

In Drosophila, Nipped-B was also found to participate in remote activation of the *cut* and *Ultrabithorax* genes. Reducing the *Nipped-B* dosage reduces activation of the wild-type *cut* gene by the remote wing margin enhancer and causes the wing margin defect in Drosophila. Furthermore, Mau-2, the *C. elegans* homologue of yeast Scc4, which interacts with Scc2 to help in the establishment of the cohesion loading onto the chromatin, was found to guide axonal migrations in the CNS. However, the exact

functions of both *NIPBL* and *hMau-2* need to be further elucidated.

# TABLE OF CONTENTS

**Abstract**.....i

**Table of Contents**.....iv

**Table of Figures**.....xii

**Tables**.....xv

**Acknowledgements**..... xvi

**Common Abbreviations Used**.....xvii

**CHAPTER 1: INTRODUCTION**..... 1

1. 1 CLINICAL FEATURES OF CORNELIA DE LANGE SYNDROME..... 1

1.1.1 Growth..... 4

1.1.2 Facial characteristics..... 4

1.1.2.1 Head and Neck..... 4

1.1.2.2 Face..... 5

1.1.3 Ears..... 5

1.1.4 Eyes..... 6

1.1.5 Mouth..... 7

1.1.6 Cardiovascular System..... 7

1.1.7 Chest..... 7

1.1.8 Abdomen..... 8

1.1.9 Genitourinary..... 8

1.1.10 Skeletal..... 9

1.1.11 Skin, Nails, Hair..... 9

1.1.12 Neurological Anormalies..... 10

1.1.13 Development and Behavior Phenotypes..... 10

1.1.14 Others..... 11

1.1.15 Mortality ..... 11

1.2 DIAGNOSIS ..... 14

1.3 INHERITANCE .....16

1.4 STATUS OF GENE IDENTIFICATION IN OCTOBER 2002..... 18

1.4.1 Background on the search for the CdLS genes ..... 18

1.4.1.1 Chromosome anomalies associated with CdLS .....18

1.4.1.2 Linkage studies .....22

1.4.1.3 Candidate gene approach .....23

1.4.1.3.1 Syndromes with phenotypes similar to CdLS .....24

1.4.1.4.2 Other possible explanations for CdLS .....28

1.4.1.4.3 Predicted expression patterns and functions of CdLS genes.29

1.4.2 Strategies of Gene Identification in Newcastle Group .....31

1.5 AIMS OF THE PROJECT .....33

1.6 REFERENCES .....34

**CHAPTER 2: FISH MAPPING OF CdLS PATIENTS WITH *de novo* BALANCED RECIPROCAL TRANSLOCATIONS and GENOMIC ORGANIZATION OF THE GENE TRUNCATED BY THE t(5;13) TRANSLOCATION .....47**

2.1 INTRODUCTION .....47

2.1.1 Standard FISH .....48

2.1.2 Array-CGH .....50

2.2 AIMS OF THIS PROJECT .....52

2.3 MATERIALS AND METHODS .....53

2.3.1 Data interpretation of FISH mapping in metaphase spreads .....53

2.3.1.1 5p region .....53

2.3.1.2 13q region .....54

2.3.1.3 14q region .....55

2.3.2 Preparation of metaphase slides .....56

2.3.3 Preparation of BAC/fosmid DNA .....56

2.3.3.1 Growth of bacteria .....56

2.3.3.2 Preparation of DNA .....57

2.3.4 DNA labeling with fluorochromes-Direct labeling .....58

2.3.5 Purification of the labeled probe .....59

2.3.6 Competition of probe and resuspension in hybridization mix .....59

2.3.7 Commercial pre-labeled chromosome 5, 13, 14 marker probes .....60

2.3.8 *In situ* hybridization to chromosome spreads .....60

2.3.9 Slide washing and detection of probe signal .....61

2.3.10 Microscopic visualization of FISH signals .....62

2.4 RESULTS .....63

2.4.1 Phenotypes and Giema chromosome banding of the patient with *de novo* t(5;13)(p13.1;q12.1) translocation .....63

2.4.2 Molecular definition of the t(5;13) breakpoint by chromosome FISH .....64

2.4.3 Results of clones mapped to the 5p translocation breakpoint Region .....65

2.4.4 Hierarchy of the contig map in 5p13.1 .....76

2.4.5 Results of the clones mapped to the translocation breakpoint around the 13q12.1 area .....77

2.4.6 Results of clones mapped to the translocation breakpoint around the 14q32 region .....80

2.5 DISCUSSION .....84

2.5.1 Power of bioinformatics .....84

2.5.2 Assignment of clones relative to the chromosome breakpoints

using standard chromosome FISH .....86
2.5.3 Clues from the linkage studies .....86
2.5.4 Data interpretations .....89
2.5.4.1 13q12.1 region .....90
2.5.4.1.1 Genes disrupted by the translocation .....90
2.5.4.1.2 Genes residing 1Mb on each side of the 13q12.1 Breakpoint .....92
2.5.4.2 Genes residing 1Mb on each side of the 14q32 breakpoint .....97
2.5.4.3 5p13 region .....100
2.6 CONCLUSION .....105
2.7 REFERENCES .....106

**CHAPTER 3: MUTATION SCREENING OF *NIPBL* GENE LOCATED WITHIN 5p13.1 .....112**
3.1 INTRODUCTION –*NIPBL* AS CANDIDATE FOR CDLS .....112
3.2 APPROACHES TO MUTATION SCREENING .....112
3.2.1 SSCP .....113
3.2.2 Heteroduplex Analysis (HET) .....114
3.2.3 PCR Direct sequencing .....115
3.2.4 Cloning of PCR product into the pGem-T easy system .....115
3.3 MATERIALS AND METHODS .....118
3.3.1 Materials .....118
3.3.2 Amplification of exons from the *NIPBL* gene .....118
3.3.3 Single stranded conformation polymorphism (SSCP) and heteroduplex gel mobility shift analyses .....119
3.3.3.1 Preparation of the gel and electophoresis of the samples .....119
3.3.3.2 Silver staining .....120
3.3.4 Direct sequencing of PCR products .....121
3.3.4.1 Enzymatic purification of PCR products .....121

3.3.4.2 Cycle sequencing .....121

3.3.5 Cloning of PCR product into pGem-T- Easy vector system .....122

3.3.6 DNA preparation of the transformants by using QIAprep kit .....124

3.3.7 Sequences of the cloned product .....125

3.3.8 Assessing the sequencing results .....125

3.4 RESULTS .....126

3.4.1 Study population .....126

3.4.2 Mutation screening approach used in exons of different sizes .....126

3.4.3 Spectrum of *NIPBL* mutation detected .....127

3.4.4 Genotype-Phenotype correlation .....129

3.5 DISCUSSION .....136

3.6 REFERENCES .....148

**CHAPTER 4: EXPRESSION PATTERNS OF THE *NIPBL* GENE.....153**

4.1 INTRODUCTION.....153

4.2 APPROACHES TO EXPRESSION STUDIES OF *NIPBL* .....153

4.2.1 Northern blot analysis .....153

4.2.2 Tissue *in situ* hybridization .....156

4.2.3 Development of different organ systems during early human development .....158

4.2.3.1 The developing nervous system .....158

4.2.3.2 The developing limbs .....162

4.2.3.3 The developing craniofacial system .....165

4.2.3.4 The developing heart .....167

4.2.3.5 The developing lung, kidney and pancreas .....170

4.2.3.6 The developing gastrointestinal system .....171

4.2.3.7 The developing germ cells .....172

4.3 MATERIAL AND METHODS .....174

4.3.1 Tissue *in situ* Hybridization .....174

4.3.1.1Collection of human embryos .....174

4.3.1.2 Tissue fixation and embedding .....174

4.3.1.3 Tissue sectioning .....175

4.3.1.4 Tissue pre-treatment .....175

4.3.1.5 Probes used for *in situ* hybridization .....176

4.3.1.6 Probe labeling .....178

4.3.1.6.1 Temperate linearilisation .....178

4.3.1.6.2 RNA transcription .....178

4.3.1.6.3 Control Riboprobes .....179

4.3.1.7 Probe hybridization .....179

4.3.1.8 Post-hybridisation washes .....179

4.3.1.9 Antibody Detection .....180

4.3.2 Immunohistochemical (IHC) staining of the slides .....180

4.3.2.1 Dewaxing and Antigen retrieval .....181

4.3.2.2 Staining of mouse monoclonal primary PCNA antibody .....182

4.3.2.3 Secondary Antibody staining and antibody detection .....182

4.3.3 Image analysis .....183

4.4 RESULTS .....184

4.4.1 Expression of NIPBL in human embryonic nervous system .....184

4.4.2 Expression of NIPBL in human embryonic limbs .....189

4.4.3 Craniofacial and ophthalmic expressions of NIPBL in human embryos .....193

4.4.3.1 Craniofacial expression of NIPBL .....193

4.4.3.2 Expression of NIPBL in developing eyes5.3.2.3.1 Craniofacial expression

of *NIPBL* .....196

4.4.3.3 Expression of NIPBL in developing ears5.3.2.3.3 Expression of *NIPBL* in developing ears .....196

4.4.4 Expression of NIPBL in developing heart .....198

4.4.5 Expression of NIPBL in developing lung, kidney and pancreas .....201

4.4.5.1 Expression of NIPBL in developing lung .....201

4.4.5.2 Expression of NIPBL in developing kidney .....203

4.4.5.3 Expression of NIPBL in developing pancreas .....205

4.4.6 Expression of NIPBL in developing gastrointestinal tract .....205

4.4.7 Expression of *NIPBL* in developing gonads .....207

4.5 DISCUSSION .....209

4.5.1 Expression of *NIPBL* in nervous system .....209

4.5.2 Expression of *NIPBL* in developing limb .....210

4.5.3 Expression of *NIPBL* in developing craniofacial skeletons, eye and ear .....212

4.5.4 Expression of *NIPBL* in developing heart .....214

4.5.5 Expression of *NIPBL* in developing lung, kidney and pancreas .....216

4.5.6 Expression of *NIPBL* in developing gut .....218

4.5.7 Expression of *NIPBL* in developing gonads .....219

4.6. CONCLUSION .....220

4.7 REFERENCES .....223

**CHAPTER 5: THE EXPRESSION PATTERNS AND MUTATION SCREENING OF HUMAN *MAU-2*.....234**

5.1 INTRODUCTION .....234

5.2. MATERIAL AND METHODS .....242

5.2.1. Northern blot analysis .....242

5.2.1.1 Probe used for Northern blot analysis and Tissue *in situ* hybridization.....242

5.2.2.2 Probe labelling-random priming .....242

5.2.2.3 Prehybridization of the RNA blots .....243

5.2.2.4 Hybridization .....243

5.2.2.5 Post-hybridization wash and film development .....243

5.2.2 Tissue *in situ* hybridization .....244

5.3 RESULTS .....245

5.3.1 Northern blot analysis .....245

5.3.2 Tissue *in situ* hybridization by using riboprobes .....246

5.3.3 Mutation screening .....250

5.4 GENERAL DISCUSSION .....252

5.5 REFERENCES .....254

**CHAPTER6: CONCLUSION.....257**

**APPENDIX A.....260**

**APPENDIX B.....262**

**APPENDIX C.....264**

# TABLE OF FIGURES


Fig. 1.1 Phenotypes of patients with CdLS.....13

Fig. 2.1 Illustration of the three types of chromosome FISH results expected after hybridizing 5p probes to the t(5;13) metaphase chromosome preparation.....54

Fig. 2.2 Illustration of the three types of chromosome FISH results expected after hybridizing 13q probes to the t(5;13) metaphase chromosome preparation.....55

Fig. 2.3 Phenotypes of the patient with t(5;13)(p13.1;q12.1) *de novo* translocation.....63

Fig 2.4 A BAC clone mapping telomeric to the 5p13.1 breakpoint (CTD-2353F22).....68

Fig. 2.5 A BAC clone mapped centromeric to the 5p13.1 breakpoint (CTD-2249G23).....69

Fig. 2.6 A BAC clone mapped across the 5p13.1 breakpoint (CTD-2653M23).....70

Fig 2.7 Results of FISH mapping. The colored 5p13 fosmid clones against DNA from the t(5;13) (p13.1;q12.1) *de novo* translocation case.....73

Fig. 2.8 FISH mapping of clone G248P84262B4.....74

Fig. 2.9 FISH mapping of clone G248P8840C10.....75

Fig. 2.10 Genomic organization, FISH clones and expressed products of the gene located at the 5p13.1 breakpoint.....76

Fig. 2.11 A BAC clone that spans the 13q12.1 breakpoint (RP11-172H24).....79

Fig. 2.12 A BAC clone mapped centromeric to the 14q32 breakpoint ( RP11-796G6).....82

Fig. 2.13 A BAC clone mapped telomeric to the 14q32 breakpoint (RP11-10e13).....83

Fig. 2.14 Genes located within 1Mb of each side of the 13q12.1 breakpoint.....92

Fig. 2.15 Genes residing within 1Mb of each side of the 14q32 50Kb gap near the breakpoint.....97

Fig. 2.16 Genes residing within the 2Mb-region of the 5p13.1 breakpoint.....101

Fig. 3.1 Map of the pGem-T-Easy Vector .....117

Fig. 3.2 Examples of chromograms of patients with *NIPBL* mutations .....135

Fig. 3.3 Missense mutations of the NIPBL. Most of the missense mutations were located within the highly conserved regions of NIPBL .....140

Fig. 3.4 Sequences conservation -2.5kb upstream of the NIPBL promoter and 2kb downstream of the exon 1 .....143

Fig. 4.1 Results of *NIPBL* expression by Northern blot .....156

Fig. 4.2 Elevator movement and the relationship with the cell-cycle of the by neuroepithelial cells .....161

Fig. 4.3 Key stages in limb skeletal development .....164

Fig. 4.4 Expression pattern of *NIPBL* transcript on CS13 (A) and CS20 (B) embryo sections .....186

Fig. 4.5 *NIPBL* expression in developing brain of CS20 embryo.....187

Fig. 4.6 *NIPBL* expression in developing brain of CS23 human embryonic section .....188

Fig. 4.7 *NIPBL* expression in early limb bud .....190

Fig. 4.8 *NIPBL* expression in hand plate of CS18 human embryo .....191

Fig. 4.9 *NIPBL* expression in upper limb of CS23 human embryo .....192

Fig. 4.10 *NIPBL* expression of in developing face .....194

Fig. 4.11 *NIPBL* expression of in developing pharynx and esophagus ..... 195

Fig. 4.12 *NIPBL* expression in developing otic canal ..... 197

Fig. 4.13. *NIPBL* expression in developing heart ..... 200

Fig. 4.14 *NIPBL* expression in developing lung ..... 202

Fig. 4.15 *NIPBL* expression in developing embryonic kidney .....204

Fig. 4.16 *NIPBL* expression in developing stomach and pancreas .....206

Fig. 4.17 *NIPBL* expression in developing gonads .....208

Fig. 5.1 The cohesin cycles.....237

Fig. 5.2 Northern analysis of *KIAA 0896* transcript .....245

Fig. 5.2 *hMau-2* expression on CS23 embryo .....247

Fig. 5.3 *hMau-2* expression on transverse section of CS18 embryo brain.....248

Fig. 5.4 *hMau-2* expression on CS18 embryonic section.....249

# TABLES

Table 1.1 Chromosome anomalies associated with CdLS.....19

Table 2.1 Results of mapping clones to the 5p13.1 breakpoint region by FISH.....66

Table 2.2 Results of mapping clones to the 13q12.1 breakpoint region by FISH....78

Table 2.3 Results of mapping clones to the 14q32 breakpoint region by FISH.....81

Table 2.4 Linkage analysis of genome-scan data in 10 pedigrees .....88

Table 2.5 Possible function of the genes residing within the 2 Mb candidate area...93

Table 2.6 Possible function of the genes within the 2Mb candidate area (14q32)....98

Table 2.7 Genes located within the 2Mb candidate region (5p13.1).....102

Table 3.1 Summery of CdLS patients with NIPBL mutations.....131

Table 3.2 conservations of sequences at sites of NIPBL missense mutations and in-frame deletion between different species.....133

Table 3.3 Chromosomal location of the human cohesin complexes and associated proteins .....146

# ACKNOWLEDGEMENTS

Particular thanks are due of course to my supervision team, Prof. Tom Strachan and Dr Susan Lindsay of the Institute of Human Genetics at Newcastle University. Without their ideas, help and encouragement, this thesis would never have been possible. I would especially like to thank Prof. Tom Strachan for his extensive critical suggestions on most of my thesis.

I would also like to thank, in particular, Dr. Emma Tonkin who taught me all the experiment procedures related to genomics and lots of associated knowledge including how to assess the internet information and experiment designs.

I would also like to thank Steve Lisgo for teaching me the techniques involved in *in situ* hybridization and for his help in interpreting the data and solving the problems related to *in situ* hybridization.

I would also like to thank Alexander Raymond for his thorough proof reading and moral support throughout the formation of the thesis. I also would like to thank John Fairlie, a former PhD student, who gave me lots of support for my dissertation writing.

Most importantly, I would like to thank the Kaohsiung Medical Centre at Chang-Gung Memorial Hospital for providing me with the grant to pay for the tuition fee and the expenses in the UK for at least 3 years.

Finally I have to thank my parents for their warming and financial support during these 4 years. Without their support, I would never have had a chance to go to the UK to pursue my PhD.

# COMMON ABBREVIATION USED:

| | |
|---|---|
| Ab | antibody |
| BAC | bacterial artificial chromosome |
| BCIP | 5-bromo-4-chloro-3-indolyl-phosphate |
| BMP | bone morphogenetic protein |
| bp | base pair |
| BSA | bovine serum albumin |
| CdLS | Cornelia de Lange Syndrome |
| cDNA | copy or complementary DNA |
| CGH | comparative genomic hybridization |
| CNS | central nervous system |
| CS | Carnegie stage |
| DAPI | 4 ,6-diamino-2-phenyl-indole |
| dATP | deoxy adenosine triphosphate |
| dCTP | deoxy cytosine triphosphate |
| ddNTP | dedeoxy nucleotide triphosphate |
| DEPC | diethyl pyrocarbonate |
| dGTP | deoxy guanosine triphosphate |
| DNA | deoxyribonucleic acid |
| dNTP | deoxy nucleotide triphosphate |
| DTT | dithiothreitol |
| dTTP | deoxy thymidine triphosphate |
| dUTP | deoxy uridine triphophate |
| EDTA | ethylene diamine tetra acetic acid |

| | |
|---|---|
| FGF | fibroblast growth factor |
| FISH | fluorescent *in situ* hybridization |
| IPTG | isopropyl-b-D-thiogalactoside |
| ISH | *in situ* hybridization |
| kb | kilo base pairs |
| kD | kilo Daltons |
| m | mili ($10^{-3}$) |
| M | molar |
| mg | milligram |
| mJ | milliJoules |
| ml | milliliter |
| mM | millimolar |
| mRNA | messenger RNA |
| u | micro ($10^{-6}$) |
| $\mu$g | microgram |
| $\mu$l | microliter |
| $\mu$M | micromolar |
| $\eta$ | nano ($10^{-9}$) |
| NBT | nitro blue tetrazolium |
| $\eta$g | nanogram |
| MIN | Mendelian Inheritance in Man |
| ORF | open reading frame |
| p | pico ($10^{-12}$) |
| PAC | Phagemid artificial chromosome |
| PBS | phosphate buffer saline |
| PCNA | proliferating cell nuclear antigen |

| | |
|---|---|
| PCR | polymerase chain reaction |
| PFA | paraformaldehyde |
| pg | pictogram |
| RNA | ribonucleic acid |
| SDS | sodium dodecyl sulphate |
| UTR | untranslated region |
| X-gal | 5-bromo-4-chloro-3-indolyl-b-D-galactoside |
| YAC | yeast artificial chromosome |

# CHAPTER 1 INTRODUCTION

## 1.1 CLINICAL FEATURES OF CORNELIA DE LANGE SYNDROME

The Cornelia de Lange syndrome (CdLS), also termed the Brachmann-de Lange syndrome (BDLS) (OMIM# 122470) is a congenital multiple malformation disorder. In 1916, Brachmann (Brachmann, 1916) reported on a foetus with severe ulnar defects and monodactylous hands resembling a very severe form of what is now known as the de Lange syndrome. In 1933, De Lange (deLange, 1933), a paediatrician in Amsterdam, described 2 infant girls with a mental deficiency and other features with a less severe form of the same syndrome. From that time, the term "Brachmann-de Lange syndrome" was used officially to describe patients with this congenital malformation condition. Over 300 patients with this syndrome have been reported in literature.

The diagnosis of CdLS is mainly based on clinical observations. The key criteria for the diagnosis are (1) the phenotypic facial changes and other minor anomalies occurring with increased frequency in affected individuals (Leavitt et al., 1985; Opitz, 1985; Preus and Rex, 1983); (2) pre- and postnatal growth deficiencies including (a) a feeding dysfunction (Cates et al., 1989; Kawai et al., 2004; Sommer, 1993); (b) a psychomotor delay and (c) a distinctive behaviour profile (Berney et al., 1999; Bryson et al., 1971; Johnson et al., 1976); and (3) associated major malformations most frequently developed in the musculoskeletal system (Filippi, 1989; Lee and Kenny, 1967; Roposch et al., 2004), but also in other organ systems such as the cardiac and gastrointestinal systems (Braddock et al., 1993; Greenwood et al., 1977; Husain et al., 1994; Selicorni et al., 2005).

Although it is said that the diagnosis of CdLS is not a challenge to the experienced clinicians, controversy has continued over the minimal criteria for the diagnosis of CdLS. Uncertainty arises in physically less affected patients. In 1991, at the DW Smith Workshop on Malformation and Morphogenesis, consensus was reached amongst clinical geneticists on the existence of the "mild" CdLS phenotype. Based on the clinical variability in the de Lange syndrome, Van Allen et al. (Van Allen et al., 1993) proposed a classification system in which he subdivided CdLS into three types. These consist of Type I, or "classic" CdLS; Type II, or "mild" CdLS; and Type III, or "phenocopies" CdLS.

The key criteria for Type I, or "classic" CdLS, are (1) growth retardation with a birth weight normally less than 2.5 std (standard deviation) below the mean for gestation, which becomes more severe postnatally (<3.5 std for the chronological age); (2) moderate to profound psychomotor retardation; (3) the presence of minor facial anomalies, skeletal and other components of the disorder giving a positive score using the diagnostic index of Preus and Rex(Preus and Rex, 1983); and (4) the presence of malformations, which are lethal or associated with severe disabilities. These criteria would place the patient into CdLS type I independent of the severity of psychomotor retardation or growth deficiency.

Type II, or "mild" CdLS is characterised by the presence of typical facial changes and other distinct minor anomalies apparent in classic CdLS patients. The most significant distinctive findings of type II CdLS are: (1) less severe psychomotor retardation; (2) milder pre- and postnatal growth deficiencies; and (3) the absence of major malformations or - if present - a malformation that is surgically correctable or does not cause a severe disability. Patients with Type II CdLS have been reported to have three types of clinical presentations. These include (1) patients with the "classic"

CdLS facial features at birth who have had an unexpected positive outcome with mild psychomotor impairment (Gadoth et al., 1982; Hawley et al., 1985; Opitz, 1985); (2) a diagnostically difficult group of patients that develop the typical facial changes of "classic" CdLS during the first 2-4 years of life, and usually have less severe manifestations than those patients with "classic" CdLS (Greenberg and Robinson, 1989); (3) a group consisting of the previously undiagnosed adults, usually with a low to normal range of performance, who have offspring with the "classic" CdLS phenotype (Leavitt et al., 1985; McConnell et al., 2003; Mosher et al., 1985). Behavioural problems may be significant in patients with Type II CdLS. However, as these patients are enduring a mild form of CdLS, better developmental skills permit more mobility and independence (Berney et al., 1999).

A general consensus for the minimal criteria for the diagnosis of "mild" CdLS has not been reached. However, Allen et. al. recommend that at least 2 of the other major diagnostic criteria alongside the characteristic facial features be present (Filippi, 1989; Van Allen et al., 1993).

Type III CdLS, or phenocopies of CdLS, includes patients who have phenotypic manifestations of CdLS, which are causally related to chromosomal aneuploidies or teratogenic exposures, such as foetal alcohol sequences. Allanson et al. speculated that the "classic" and "mild" phenotype might be due to allele specificity or to modifying genes. Least likely, in their opinion, was the possibility that the mild de Lange syndrome might be a phenocopy of the classic phenotype (Allanson et al., 1997).

The clinical phenotypes of CdLS are described below

### 1.1.1 Growth:

Growth retardation and short stature are seen in almost all the patients with CdLS. Growth retardation is prenatally onset and becomes more significant by the age of six months. In a study of 180 CdLS patients, Kline et. al. (Kline et al., 1993a) reported a mean height and weight below the 5$^{th}$ centile throughout life, with the mean head circumference below the 2$^{nd}$ centile. While the height velocity is comparable to that of unaffected children, it begins to slow once they reach adolescence. On the other hand, the weight velocity is below that of unaffected individuals until late adolescence. As is the case with normal spurts, maximum pubertal growth occurs at 15 years in males and 13 years in females. Kline et al. concluded that the patients with CdLS are of a proportionate small stature, with relatively normal or mild microcephaly. Thus, the diagnosis of failure to thrive would not be appropriate in these patients, since their small size is uniform and their weight gain is not unduly low for their height. On the other hand, the growth curves for many of the infants with the severe form of CdLS are far below the –2SD in comparison with normal children.

### 1.1.2 Facial characteristics:

#### 1.1.2.1 Head and Neck:

Almost all the patients with CdLS have microcephaly (a small head) and brachycephaly (a decreased anterior-posterior diameter of the head) (Hall, 1971; Jackson, 1993). Moreover, a short neck was another characteristic of patients with CdLS (Jackson et al., 1993).

1.1.2.2 Face:

Facial features of patients include (1) "pencilled" eyebrows (98%), (2) long eyelashes (99%), (3) a long philtrum (94%), (4) a thin upper lip (94%), (5) a down-turned mouth (94%), (6) a broad and/or depressed bridge and anteverted nostrils (85%), (7) micrognathia and a high arched palate with or without a cleft palate/lips (84%). Widely spaced teeth and late eruption of the teeth were also observed in CdLS patients (Ireland and Burn, 1993; Ireland et al., 1993; Verloes et al., 1996). The facial features combined with facial hypertrichosis are probably the most diagnostic of all the physical signs (Jackson et al., 1993, Bay et al., 1993). In a report by Ireland et. al. (1993), they found that the depressed nasal bridge, which is common in infants with CdLS, may disappear after the age of four. A finding that had not been highlighted in most literature was that underdeveloped orbital ridges and zygomatic arches were almost universal features in patients with CdLS. In that report, they also found that the crescent shaped mouth with down turned corners may disappear in male CdLS patients after puberty. This phenotypical change was not observed in postpubertal females.

**1.1.3 Ears:**

Low-set ears in patients with CdLS are not striking (Bay et al., 1993; Jackson et al., 1993), while disproportionately small auricles have been reported (Leroy et al., 1993). Moreover, sensorineural hearing loss has been reported to be associated with CdLS. Sataloff et al. (Sataloff et al., 1990) studied the otolarngologic manifestation of CdLS in 45 patients and discovered that more than 90% of children with CdLS had a sensorineural hearing loss tested by audiometry. In 2002, Sakai et al. (2002) tested 13

children with CdLS by behavioral observation audiometry (BOA) and auditory brainstem evoked response (ABR). In the ABR examination, a delayed wave I peak latency was found in almost all the cases, and a prolonged peak interval of waves I-V was found in only one patient.

Their results indicate a very high prevalence of hearing loss at the middle and inner ear levels in patients with CdLS. However, the finding in 2 patients without peaks after wave III in one ear suggested that they have brainstem impairment. Two patients who initially had no ABR in either ear, but later exhibited clear ABRs in one ear, were thought to have brainstem immaturity, such as delayed myelination. Further evidence of brainstem immaturity can be found in an earlier histopathological study performed in 1996. Sasaki et. al. ( 1996) conducted a histopathological study on the temporal bone and brainstem of a 9-month-old infant with CdLS. They located a malformation of the bony labyrinth, as no mesenchyme was seen in the perilymphatic space of cochlea. The size of the brain stem, particularly the pons, was below average.

**1.1.4 Eyes:**

Levin et al. (1990) examined 22 children with CdLS to ascertain the relative frequencies of ophthalmologic abnormalities. In addition to the frequent findings of brow hypertrichosis, synophrys (eyebrows growing across the base of the nose), and long arcuate eyelashes, they documented the association of CdLS with ptosis, nystagmus, and high myopia. Optic nerve pallor, poor macular reflex, chin lifts induced by ptosis, hypertropia, and nasolacrimal duct fistula were also described in their report. Moreover, hypertelorism and telecanthus were found amongst CdLS patients. However, contrary to previous reports, they did not find a link between

corneal, pupillary, or retinal disorders and this syndrome. Furthermore, they found no association between blue sclera and the syndrome.

### 1.1.5 Mouth:

Patients with CdLS are also characterized by a thin upper lip, down-turned corners of the mouth and a high arched palate with or without a cleft lip or palate. In contrast to the depressed nose, the crescent shaped mouth with down-turned corners may remain in the mild form of CdLS after puberty.

### 1.1.6 Cardiovascular System:

As a result of reviewing the literature on CdLS, Rao and Sissman (1971) could confirm that 15% of the affected individuals had a congenital heart defect. The cardiovascular diseases associated with CdLS patients are mainly ventricular septal defect and pulmonary stenosis. Other possible defects include tetralogy of Fallot, persistent arterial duct, coarctation, rudimentary left ventricle, anomalous systemic venous drainage, overriding aorta and ventricular fibro-elastosis (Mehta and Ambalavanan, 1997; Tsukahara et al., 1998). These may contribute to the early demise of patients with severe CdLS.

### 1.1.7 Chest:

Small nipples were noted in patients with CdLS, while congenital diaphragmatic hernia was reported by Pankau and Janig, (1993). The latter contributes to the high lethality caused by this syndrome more often seen in severe CdLS patients (Cunniff et al., 1993).

### 1.1.8 Abdomen:

Gastroesophageal reflux (GER) is a frequent and severe medical complication of the Cornelia de Lange syndrome (Luzzani et al., 2003). The most severe complications of reflux include torticollis, opisthotonus, and paroxysmal dystonic posture (known as the Sandifer complex). Recently, Luzzani et al. (2003) evaluated a series of 43 patients and concluded that pathological GER was evident in 65% of CdLS patients. The incidence of the complication was not significantly different in patients with the classical (93.3%) versus the mild phenotype (82.3%). However, a strong correlation was present between the degree of the esophageal damage and the clinical phenotype. Hyperactivity was the most frequent sign associated with the condition (85%). Moreover, clinically significant feeding problems included tracheal aspiration, esophageal dysmotility, gastroesophageal reflux, hiatal hernia, and esophagitis. Adenocarcinoma following complicated GER was also reported (DuVall and Walden, 1996). Improved body weight gain was noted with early recognition of gastroesophageal reflux in these patients when medical treatment or surgical interventions were instituted appropriately. Another associated abdominal anomaly includes cecal volvulus (Husain et al., 1994; Masumoto et al., 2001).

### 1.1.9 Genitourinary:

Cryptochidism and hypoplastic genitals have been reported in more than 50% of males with CdLS, while renal abnormalities (primarily ureteral reflux) have been reported in 12% of males (Jackson et al., 1993). In a series reported by Selicorni et. al. (2005), structural anomalies of the kidney and urinary tract were detected either by sonography or voiding cystoureterography in 25 (41%) of the 61 patients. They

included absent or poor corticomedullary differentiation (13%), pelvic dilation (10%), small kidneys (5%), isolated renal cyst (5%), and renal ectopia (3%). Renal function was reduced in 15% of patients (Selicorni et al., 2005).

### 1.1.10 Skeletal:

Limb defects are noted in around 20% of patients with CdLS (Bozner et al., 1995; Jackson et al., 1993; Pashayan et al., 1975). The limb defects in patients with CdLS consist primarily of limited elbow extensions, dislocation of the radial head and phocomelia. Abnormalities developed on the hands and feet include single transverse palmar crease, proximally placed thumbs, fifth finger clinodactyly, oligodactyly and syndactyly of the 2$^{nd}$ and 3$^{rd}$ toes. Braddock et al. (1993) reviewed roentgenograms in 21 cases of the Brachmann-de Lange syndrome and published descriptions of the radiological manifestations. They found that unusual radiologic manifestations were related mainly to the limb anomalies, and were often asymmetric. These manifestations included digital abnormalities, which ranged from acheiria to oligodactyly, hypoplasia of the thumb and first metacarpal, clinodactyly of the fifth finger, or ectrodactyly. Long bone abnormalities included ulnar aplasia/hypoplasia, and dysplasia of the radial head, or fusion of the elbow. Other radiological manifestations included 13 ribs with a precocious sternal fusion, and micrognathia.

### 1.1.11 Skin, Nails, Hair:

Cutis marmorata, hirsutism and a low posterior hair line are frequent findings in patients with CdLS (Jackson et al., 1993).

### 1.1.12 Neurological Anormalies:

Yamaguchi and Ishitobi, (1999) described congenital dysgenesis of the brain including microbrachycephaly, immature or simple convolution pattern of the cerebral gyri, thickened leptomeninges, persistent subpial granule cells, hypoplasia of the anterior thalamic nuclei, neurohypophysis, lateral geniculate body, cerebral peduncle, ventral pons and cerebellar internal granular layers, and heterotopic cell nests in the cerebellar white matter. They concluded that congenital dysgenesis of the brain, especially that found in the diencephalon and the cortico-ponto-cerebellar system, may constitute morphologic evidence explaining the severe growth retardation and neurological abnormalities in CdLS.

### 1.1.13 Development and Behavior Phenotypes:

In a series conducted by Kline at al., most of the 122 patients with CdLS were found to be in the mild to moderate region of mental retardation, with I.Q.s ranging from 30 to 85. Moreover their average I.Q. level of 53 (Kline et al., 1993b) was higher than previously reported (Barr et al., 1971), while CdLS patients with normal IQ levels have also been reported (Saal et al., 1993). In a survey of speech and language skills in 116 patients with CdLS, Goodban (Goodban, 1993) discovered that 53% of the children who were four years or older combined two or more words into sentences, while 33% uttered no words or only 1-2 words. Meanwhile, 4% of the children in this range were judged to have language skills within normal to low normal limits. They suggested that the children who were not talking at all or who were severely delayed in their language skills had at least one of the following characteristics: (1) a birth weight under 5lbs; (2) a moderate to severe hearing impairment; (3) upper lim defects;

(4) poor abilities to relate socially; (5) severe motor delays. The behavioural phenotype of CdLS includes hyperactivity (40%), self-injury (44%), daily aggression (49%), and sleep disturbance (55%). The phenotypes correlate with the presence of autistic like behaviour and a degree of mental retardation (Berney et al., 1999).

### 1.1.14 Others:

In addition to the above mentioned clinical phenotypes, low-pitched, growling cries in the infancy stage (Jackson et al., 1993), panhypopituitarism (Kousseff et al., 1993; Tonini and Marinoni, 1990) and dysfunctions of gonadotropin and prolactin secretion and of osmoregulatory mechanisms (Schwartz et al., 1990) have been described. In 1993, Froster and Gortner (Froster and Gortner, 1993) reported on a severely affected infant with an upper limb defect who also had thrombocytopenia. Furthermore,Fryns and Vinken, (1994) described thrombocytopenia in 2 out of 50 patients with CdLS over a period of 25 years.

### 1.1.15 Mortality:

The mortality rate of de Lange patients was found to have slightly risen considering the expected number of deaths. However, no differences in the mortality rate were observed between the opposite genders. During their evaluation of the causes of death and pathological findings, Beck and Fenger, (1985) found pneumonia to be the most frequent cause of death. In some patients, congenital malformations seemed also to be contributory causes of death, the most frequent being congenital heart disease and gastro-intestinal malformations. Moreover, there is an increased incidence of aspiration related deaths among the CdLS population, which may be

related to the prevalence of GER in patients with CdLS.

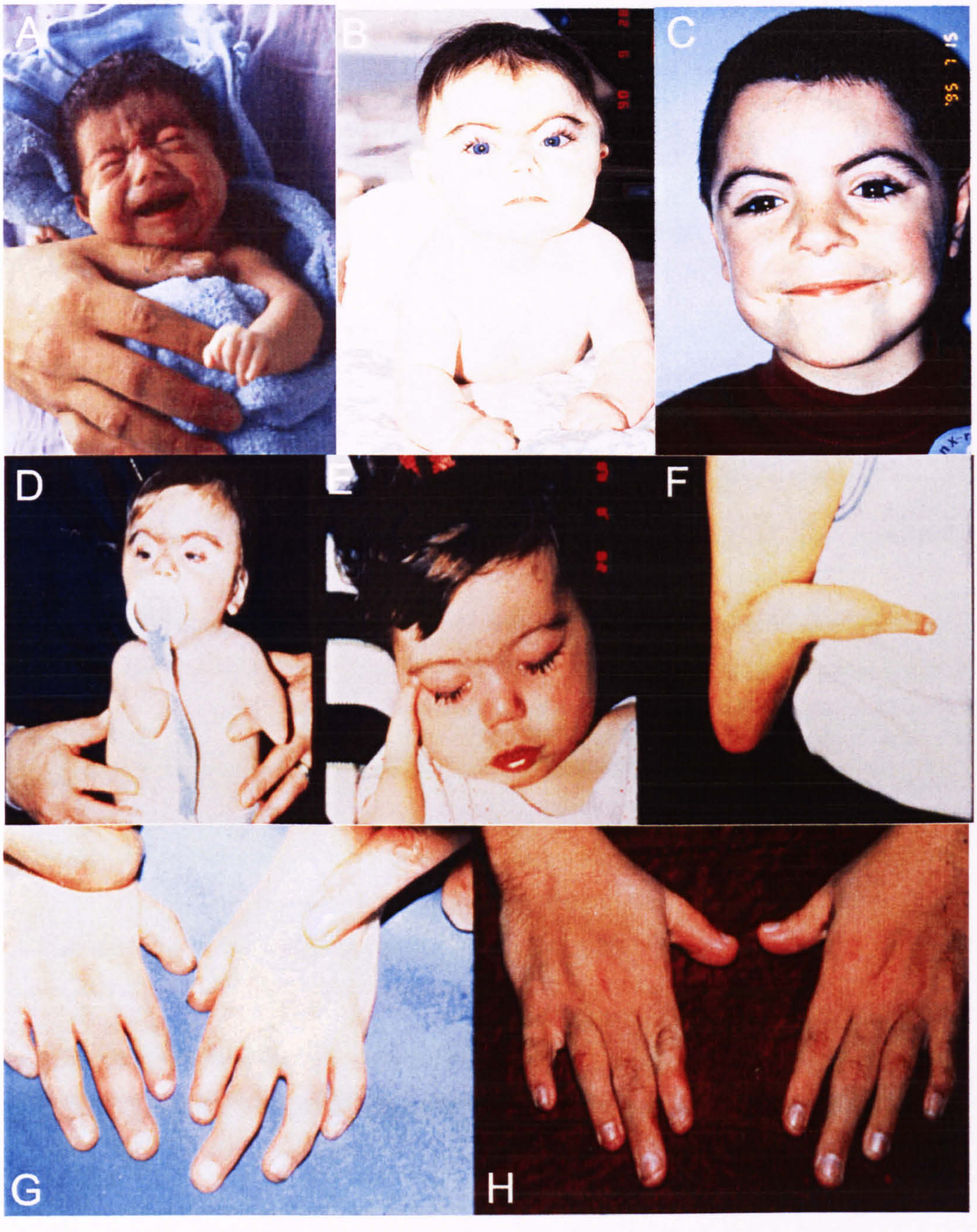



Fig. 1.1 Phenotypes of patients with CdLS. (A) A newborn baby with severe prenatal onset growth deficiency of oriental origin. No severe limb reduction defect was noted (B) Typical facial characteristics of CdLS, note the synphorys, flat nasal root with anteverted nostrils and long philtrum. (C) A mild CdLS patient. (D, E, F) Severe limb reduction defects in patients with severe CdLS. (G, H) Minor upper limb abnormality in patients with CdLS, note the bilateral clinodactyly of the 5$^{th}$ fingers and short phanlanges of the 2$^{nd}$ and 4$^{th}$ fingers.

## 1.2 DIAGNOSIS

The diagnosis of CdLS is mainly based on the recognition of the clinical features. The diagnosis for the severe form of CdLS is seldom in doubt. However, uncertainties may be raised in patients with mild phenotypes, in which the limb defects and growth deficiency are not prominent. The differential diagnosis of CdLS includes patients with the partial 3q duplication syndrome. Wilson et al. (1978) proposed clinical criteria for a distinction between these two syndromes. Intrauterine growth retardation, a prominent philtrum, proximally placed thumbs, and oligodactyly/phocomelia are more frequent in CdLS, while craniosynostosis, a cleft palate, and urinary tract anomalies are more characteristic of the dup (3q) syndrome. In most cases, CdLS patients have apparently normal chromosomes. Thin upper lips, a depressed nasal bridge, and hirsutism may also be frequent findings in patients with the fetal alcoholic syndrome or fetal teratogenic exposure. The differential diagnosis of CdLS with other syndromes also includes patients with the Rubinstein-Taybi syndrome (MIN 180846) (Partington, 1990) and the Ruvalcaba syndrome (MIN 180870).

As for postnatal diagnosis, prenatal diagnosis also depends mainly on the phenotypic features such as the presence of nuchal fold by ultrasonography and also by certain serum markers such as PAPP-A and FßhCG (Bruner and Hsia, 1990). The pregnancy-associated plasma protein A (PAPP-A) is a large molecular-weight, placenta-derived glycoprotein. During pregnancy it is produced in high concentrations by the trophoblast and released into the maternal circulation. The functional significance of PAPP-A is unclear. The major clinical role of PAPP-A is as a marker of chromosomal aneuploidy (trisomies 21, 18 and 13), complicated pregnancies and the Cornelia De Lange syndrome (Westergaard et al., 1983). The PAPP-A levels have

been reported to be decreased in first trimester (Arbuzova et al., 2003) and second trimester pregnancies (Aitken et al., 1999). In Aitken's series, they suggested that, in the second trimester, maternal serum PAPP-A levels are low but not deficient in pregnancies with the classical form of CdLS (Aitken et al., 1999). Although the median InhA level was significantly reduced in the CdLS cases, the magnitude of the reduction was less marked than for PAPP-A. FßhCG levels were also reduced but showed a wider range with several high values. They also pointed out that the cases in their study were biased towards the severe end of the phenotypic presentations. In other words, the rate of limb defects was higher than average.

So it still remains to be established if similar reductions in PAPP-A levels are present in maternal serum when the fetus is only mildly affected. However, the recurrence rate for a couple with a previous affected CdLS pregnancy is approximately 1%. These are generally offered prenatal assessment with serial and detailed ultrasonography. The most common ultrasound abnormality associated with the syndrome is IUGR (Bruner and Hsia, 1990), but this is usually not identifiable until the third trimester. Abnormalities which may be detected at the 16-20 week scan include severe limb defects (14%), diaphragmatic hernias, structural heart abnormalities (18%), fixed flexion deformities of the elbows (84%) and characteristic facial profiles (Bruner and Hsia, 1990; Drolshagen et al., 1992) and increased nuchal thickness during first and second trimesters. Thus, the adjuvant use of PAPP-A maternal serum measurement during the first and second trimester of pregnancy with detailed ultrasonography may provide a useful tool to detect pregnancies at risk of CdLS. It is important for the counselor to consider the possibility of CdLS, especially in women with positive first-trimester screening results and a normal karyotype, in which case the PAPP-A level is very low.

## 1.3 INHERITANCE

Despite the fact that the vast majority of CdLS cases are sporadic, multiple affected cases in the same families have been reported (Beck, 1974; Kumar et al., 1985; Opitz, 1985; Robinson et al., 1985). In 1963, Ptacek et al., (1963) suggested a dominant mode of transmission of the mutation. In a recent review Russell et al., (2001) reported that 60% of the familial cases appear to demonstrate a dominant mode of transmission of the phenotype, while 40% may be consistent with recessive inheritance, or do not point directly to a dominant pattern of inheritance. In addition to these cases, another report by Krantz et al., (2001) described nine previously unreported families. He found that four families indicated dominant inheritance, four suggested recessive inheritance, and one family was inconclusive. However, it is difficult to decide whether the diagnosis of CdLS is correct or not simply based on the clinical descriptions and photographs provided in several of the familial cases.

Although several reported familial cases suggested autosomal recessive inheritance (Naguib et al., 1987; Opitz, 1985), reasons exist against the assumption that CdLS is produced by the homozygous state of recessive mutation. These include the absence of consanguinity in the parents, and the high proportion of sporadic cases. It is possible that "gonadal mosaicism" is the mechanism in those families with recurrent affected siblings (de Die-Smulders et al., 1992; Ptacek et al., 1963). In addition, the preponderance of maternal transmission of the trait observed in the dominant inherited pedigrees may suggest a possible imprinting effect (de Die-Smulders et al., 1994). However, both paternal and maternal transmissions have been demonstrated and thus exclude the possibility of X-linked inheritance and imprinting (McKenney et al., 1996; Russell et al., 2001). Furthermore, Shaffer et al.

found in one study that uniparental disomy is not present in patients with CdLS (Kotzot, 2002; Shaffer et al., 1993). Similarly, mitochondrial inheritance seems unlikely, although there has been a CdLS case with multiple mitochondrial deletions (Melegh et al., 1996).

Thus we can conclude from the literature that if CdLS were to be genetically homogeneous, most of the reported familial cases are consistent with a dominant transmission of the disease trait. However, recessive forms of inheritance still can not be excluded. With the recognition of the milder CdLS phenotype (Allanson et al., 1997), it is possible that a greater number of familial cases will arise and will provide more information about the mode of inheritance. Thus, it is reasonable to speculate that heterogeneity exists in patients affected with de Lange syndrome. Hence, the definite diagnosis could not be established, especially for mild CdLS, until the molecular etiologies associated with the de Lange syndrome were found.

The exact incidence of CdLS is unclear, but the rate amongst live births is thought to be between 1:10,000 and 1:30,000 (Opitz, 1985). The life expectancy of severe CdLS patients is shortened, while this reduction in the life span in less severe patients is not known. In the past, many children died of serious medical problems in the infancy stage because their needs were not anticipated. This is no longer the case, and it is expected that most CdLS patients will live into adulthood.

## 1.4 STATUS OF GENE IDENTIFICATION IN OCTOBER 2002

### 1.4.1 Background on the search for the CdLS genes

There are many ways of identifying disease genes. Most commonly, a candidate chromosome region is pinpointed, and then candidate genes can be identified within that region through the help of bioinformatics and data obtained. Strategies for this position dependent approach include positional cloning by use of chromosome abnormalities and linkage analysis. Alternatively, we also can use positional independent candidate gene approach to identify the disease genes.

#### 1.4.1.1 Chromosome anomalies associated with CdLS

Chromosome abnormalities can provide clues to the localization of candidate genes. Balanced chromosome abnormalities (translocations or inversions) are of particular use. As no DNA materials are missing or added, a carrier is not expected to have any phenotypic effect. If a person has a balanced chromosomal abnormality with abnormal phenotypes, there are three possible explanations: (1) the finding is coincidental; (2) the rearrangement is not really balanced – there is unnoticed gain or loss of DNA material; (3) one of the breakpoints causes the disease. A chromosomal breakpoint can cause a loss-of-function phenotype by disrupting the coding sequence of the gene or separating it from nearby regulatory elements. On the other hand, it could cause a gain of function by bringing together exons from two genes to form a new chimeric gene, or cause inappropriate activation of a gene. Thus, the breakpoint provides a valuable information for the location of disease gene, especially in *de novo* balanced chromosome rearrangement.

A few individuals with CdLS have been reported to have chromosome abnormalities (Kousseff et al., 1994), but the gene responsible for the syndrome remains to be identified due to the lack of common breakpoints. However, chromosome abnormalities that rise *de novo* in sporadic CdLS patients may be specific and crucial. Review of the reported CdLS cases with chromosome derangement are as follows:

**Table 1.1 Chromosome anomalies associated with CdLS**

| chromosome | karyotype | reference |
|---|---|---|
| 1 | 46,XX, del(1)(q44) | Borck et al., (2004) |
| 2 | 46,XY, t(2;12) *de novo* | Craig and Luzzatti, (1965) |
| 3 | 46,XY,?del(3q) | Szemere et al., (1972) |
| 3 | 46,XX,dup(3q)mat | Schwanitz et al., (1977) |
| 3 | inv dup(3)(q29q25) | Wilson et al. (1978) |
| 3 | dup(3)(q25q29) | Wilson et al. (1978) |
| 3 | 46,XX,dup(3)(q23q27) | Sciorra et al. (1979) |
| 3 | 46,XY,r(3)de novo | Lakshminarayana and Nallasivam, ( 1990) |
| 3 | 46,XX,t(3;5)(q21;p13) | Price et al. (2005) |
| 3 | 46,XX,t(3;17)(q26.3;q23.1)de novo | Ireland et al. (1991) |
| 3 | 46,XX,ins(10;3)(q21.2;q25.1q26.2) | Holder et al. (1994) |
| 3 | 46,XY,der(3)(3;12)(p25.3;p13.3)mat | Descipio et al. (2005) |
| 4 | 46,XY,t(4;11)/46,XY mosaism, *de novo* | Payne and Maeda (1965) |
| 4 | 46, XY, der(9)t(4;9)mat | Hersh et al. (1985) |
| 5 | 46,XY,t(5;12)/46, XY *de novo* | Geudeke et al. (1963) |
| 5 | 46,XX, der(5)t(5;9;13)(q23;p24;q22)mat | Eeg-Olofsson and Liedgren (1981) |
| 5 | 46, XX, del(5)(p13) *de novo* | Taylor and Josifek (1981) |
| 5 | 46,XX,t(5;13)(p13.2;q12.1) | Tonkin et al. (2004b) |
| 5 | 46,XX, der(5)t(1;5)(q41;q35) | Telvi et al. (1999) |

| | | |
|---|---|---|
| 5 | 46,XY, del(5)(p13.1p14.2) *de novo* | Hulinsky et al. (2005) |
| 7 | 46,XY, der(7)t(7;10)(q32;q24) pat | Randall-Pinto et al. (2000) |
| 8 | 46,XX,t(8;?) *de novo* | D'Oelsnitz et al. (1971) |
| 8 | t (X;8)(p11.2;q24.3) *de novo* | Egemen et al.(2005) |
| 9 | 46,XY, 9qh+pat | Berg et al. (1967) |
| 9 | 46,XX,r(9) *de novo* | Motl and Opitz (1971) |
| 9 | 46,XY, 9qh+qh+patmat | Babu et al. (1985) |
| 11 | 46,XY, del(11q)/46,XY *de novo* | Payne and Maeda (1965) |
| 13 | 46,XY,t(13q;14q)mat | Beck and Mikkelsen (1981) |
| 14 | 46,XX,t(14;21)(q32;q11.2) *de novo* | Wilson et al. (1983) |
| 17 | 46,XY,?del(17q) *de novo* | Gans and Thurston (1965) |
| 18 | 47,XX,+i(18)(p10) | Borck et al.(2004) |
| X | 45,X/46,XX *de novo* | Klosovskii et al. (1968) |
| X | 45,X/46,XY *de novo* | Calo et al. (1968); Beck and Mikkelen (1981) |
| D group | 45,XY,t(Dq;Dq)mat | Ott et al. (1968) |
| D group | 46,XX,del(Dq)/46,XX *de novo* | Westermann et al. (1977) |
| G Group | 46,XY,Gp+ *de novo* | Hooft et al. (1965) |

The diversity of disease-associated chromosome abnormalities in patients with CdLS may suggest a variety of candidate chromosomal regions that could harbour the CdLS genes, and there are only very few consistently associated chromosome anomalies. It is difficult to determine if the chromosome rearrangements are causative of the CdLS phenotypes or just a coincidence. In cases where the chromosomal abnormalities have a well defined phenotype such as 45, X/46,XX or (XY), the CdLS phenotypes are likely to have no association with the chromosome anomalies. In many cases reported above, the phenotypes were not well described or the phenotypes were only partially overlapped with the de Lange syndrome. However, in the reported cases, three cases diagnosed as de Lange syndrome have rearrangements involving

5p13. It should be noted, however, that two of these cases were reported only recently. Several cases with 3q rearrangement were also reported to have CdLS phenotypes.

The regions involved in chromosome aberrations in CdLS patients with balanced reciprocal translocation are inconsistent and often, coincident. However, the 3q26.3 breakpoint was initially considered to be particularly worthy of further investigation due to the perceived overlap in phenotype for the dup (3q) syndrome whose crticial region was mapped to 3q26.3-q27 (Aqua et al., 1995). Individuals who are not trisomic for this region, but are trisomic for regions immediately proximal or distal to it, do not show a phenotype resembling that of mild CdLS (Aqua et al., 1995; Ireland et al., 1995). While some CdLS-associated features, notably prenatal growth retardation and severe limb abnormalities, are not seen in the dup 3q syndrome, phenotypic divergence between the two syndromes could be explained by differential dosage of active gene copies (given that the overexpression expected in trisomy 3q is unlikely to be consistently paralleled by heterozygous mutations in CdLS). Genes residing in other areas may also contribute to the CdLS phenotypes.

A variety of different studies in distal 3q have been conducted in CdLS patients. Early in 1993, Shaffer et al., (1993) speculated that uniparental disomy (UPD) might be responsible for CdLS as seen in patients with Prader-Willi, Angelman, and Beckwith-Wiedemann syndromes based on the observation of more than one mode of inheritance and occasionally associated with a cytogenetic abnormality. However, none of the 16 probands studied had UPD using polymorphic markers in the distal 3q region and all demonstrated normal biparental inheritance for at least one locus. Mutation screening in SOX2, CHRD genes located within 3q26.3/q27 and the GSC gene located within 14q32 failed to identify any CdLS patient-specific mutations (Smith et al., 1999).

Nevertheless, the majority of the CdLS patients with chromosomal abnormalities showed no consistent regions involved. This may suggest that the multiple loci implicated in CdLS are just coincidental. Another possible explanation for the multiple chromosomal anomalies observed in CdLS cases is that the CdLS disease gene may play an important role in the repair of double strand DNA. An example of this possibility is Seckel syndrome type I. Mutations in the *ATR* gene cause increased chromosome breakage at fragile sites. The malformations seen in the affected individuals with Seckel syndrome may result from inappropriate checkpoint responses and chromosome breakage, thus resulting in cell death at time of replicative stress during development (Casper, 2004).

### 1.4.1.2 Linkage studies

Collecting familial cases for genetic mapping provides an alternative route to disease gene identification. However, for this approach, a large number of familial cases, clear diagnosis and the mode of Mendelian transmission are necessary.

Even though the Cornelia de Lange syndrome was described 70 years ago, the exact molecular mechanisms that lead to CdLS had not been identified for the following three reasons: (1) lack of consistent chromosomal rearrangement; (2) paucity of familial cases; (3) small size and small affected cases in familial cases.

In 2001, Krantz et al., (2001) performed a linkage analysis to the minimal critical region for the dup (3q) syndrome that encompasses the translocation breakpoint in 10 rare familial cases with CdLS and excluded CDL1 region as the candidate region in 4 out of 10 familial CdLS cases. However, due to the paucity of the familial cases, linkage analysis has not been an easy approach.

### 1.4.1.3 Candidate gene approach

Another approach to identify disease genes is through positional independent strategies, especially for diseases without relevant mapping information. In this situation, the candidate must be suggested by the (1) gene products; (2) similar organism models; (3) comparison of phenotypic similarities with other related human diseases through possible expression array experiments.

It was important to consider which developmental processes are involved with the CdLS phenotypes and how the roles of the processes are played in lower species. As CdLS affects many different organ systems and generalised and symmetrical growth deficiency is a hallmark of the disease, it is expected that the causative gene is likely to be expressed early in development and well conserved between divergent species and might be associated with the cell cycle. Genes, which have fundamental roles in development are subject to selective pressures to conserve the sequences of the gene product. Examples of such genes include cell regulators such as cyclins which have homologues conserved between *Drosophila* and humans.

The most obvious phenotypes associated with CdLS are generalised growth deficiency and limb defects of variable severity. The microcephaly associated with CdLS actually is not severe relative to the generalised growth deficiency they have. Microcephaly seen in patients with CdLS is due to generalised underdevelopment of the brain content. Other syndromes caused by defects in fibroblast growth factor receptors usually manifest normal brain development and most of them have a nearly normal mentality. Their impairment of brain development is a result of premature craniosynostosis rather than defect in the brain itself. Thus, genes that cause microcephaly due to primary craniosynostosis are not considered directly linked to

CdLS. Nevertheless, syndromes with resemblance to CdLS may be considered candidates for CdLS, as they involve genes with generalised growth deficiency and limb defects.

#### 1.4.1.4.1 Syndromes with phenotypes similar to CdLS

The genetic basis of syndrome producing phenotypes similar to CdLS might give clues to the genetic connection with CdLS gene(s). There are some syndromes that share phenotypic overlap with CdLS. The most obvious is the 3q duplication syndrome, as described in section 1.2. Other syndromes characterised by growth deficiency with multiple malformations include (1) Coffin-Siris syndrome (MIN135900), which is characterised by growth retardation of prenatal onset, mental retardation, microcephaly, coarse faces, bushy eyebrows and long eyelashes, a flat nasal bridge, long philtrum and hypoplasia of the digits (especially 5$^{th}$ digits); (2) Scott Craniodigital syndrome with mental retardation (MIN 312860), which is characterised by mental retardation, brachycephaly, a small and narrow nose, a 'startled' appearance, thick head hair extending unusually far on the temples and sideburn areas, long eyelashes, thick eyebrows, a short mandible, digital syndactyly and X-linked inheritance; (3) Ruvalcaba syndrome (MIN 180870), which is characterised by mental retardation, short stature, microcephaly, a peculiar face with a hooked nose and a small mouth, a narrow thoracic cage with pectus carinatum, hypoplastic genitalia, hypoplastic 'onion skin' cutaneous lesions, and skeletal deformities including short metatarsals and metacarpals and epiphysitis of the spine; (4) Fryns syndrome (MIN 229850), which is characterised by coarse faces with cloudy corneae, diaphragmatic defects, absence of lung lobulation, distal digital hypoplasia and early death; (5) Rubinstein-Taybi syndrome (MIN180849) (RSTS), which is characterised by growth deficiency of postnatal onset, broad thumbs and toes,

slanted palpabral fissure, microcephaly, hypoplastic maxilla and highly arched eyebrows; (6) Russell-Silver syndrome (MIN180860) (RSS), which is characterised by short stature of prenatal onset, skeletal asymmetry, small triangular face, and bilateral clinodactyly of bilateral 5$^{th}$ fingers; (7) Seckel syndrome I (MIN 210600), which is characterised by growth retardation, microcephaly with mental retardation, and a characteristic 'bird-headed' facial appearance; (8) Roberts syndrome (MIN268300), which is characterised by hypomelia, midfacial defect, and severe growth retardation. The molecular defects associated with Coffin-Siris syndrome, Scott Craniodigital syndrome with mental retardation, Ruvalcaba syndrome and Fryns syndrome are still not recognised.

Due to the presence of neat eyebrows in patients with Rubinstein-Taybi syndrome, the syndrome is frequently mis-diagnosed as a mild form of CdLS. Rubinstein-Taybi syndrome is caused by the heterozygous mutation in the CREB-binding protein (CBP) gene located on chromosome 16p13.3 and EP300 (Roelfsema et al., 2005). CBP serves as a transcriptional coactivator and acts as an integrator of the signals from various pathways. Haploinsufficiency is the ultimate cause of the syndrome (Goodman and Smolik, 2000). Furthermore, the majority of the patients with mutations are located within the domain that would affect the intrinsic histone acetyl transferase activity (HAT) (Roelfsema et al., 2005), and the loss of HAT activity leads to the phenotypes of RSTS. This is an example for multiple congenital malformations with mental retardation caused by generalised dysregulation of gene expressions during early human development. An additional example would be the thalassemia/mental retardation syndrome (MIN301040), in which the mutation is located in the gene encoding X-linked helicase-2 (Gibbons et al., 1995).

In the Russell-silver syndrome (RSS), around 10% of RSS patients were found to have maternal uniparental disomy of chromosome 7, which indicated that an imprinted gene located on chromosome 7 may cause the disease phenotypes of RSS (Bernard et al., 1999; Yoshihashi et al., 2000). Moreover, the human growth factor receptor–bound protein 10/maternally expressed gene-1 (GRB10/MEG1) gene was found to be associated with RSS. Bilateral maternal expression of the imprinted *GRB 10* gene due to maternal uniparental disomy leads to RSS phenotypes in these patients. *In vitro* studies indicate that by interacting with either the insulin-like growth factor I (IGFI) receptor (Morrione et al., 1997) or the growth hormone receptor (Moutoussamy et al., 1998), GRB10 has a suppressive effect on growth.(Yoshihashi et al., 2000). Recently, hypomethylation of the *H19* region in 11p15 was found to cause RSS and associated phenotypes (Gicquel et al., 2005). Hypermethylation of the *H19* gene, on the contrary, causes the gigantism syndrome, Beckwith-Wiederman syndrome (MIN 130650). Due to the phenotypic overlap with the 3q duplication syndrome, the CdLS gene was once suspected to be an imprinted gene(s) located within the 3q region. However, no UPD in the 3q region was found in patients with CdLS (De Marchi et al., 1994). Nevertheless, genes that are regulated through epimutation are still strong candidates for causative CdLS gene(s).

Seckel syndrome (type I) shares features in common with disorders involving impaired DNA-damage responses that lead to chromosome instability, such as Nijmegen breakage syndrome (OMIM 251260) and LIG4 syndrome (OMIM 606593) (Casper et al., 2004). To date, the only reported genetic defect identified in Seckel syndrome is a mutation in ataxia telangiectasia and Rad3-related protein (ATR)(O'Driscoll et al., 2003). ATR is a member of the PI3K family, and is closely related to ATM. Moreover, it functions in the cell-cycle checkpoint and DNA repair

pathways. It is a key member of the intra-S and G2/M checkpoints and responds primarily to replication stress, such as that caused by hydroxyurea, aphidicolin, and hypoxia (Alderton et al., 2004). When activated, ATR stabilizes stalled replication forks, inhibits late origin firing, and blocks progression from G2 into mitosis (Cobb et al., 2005). Although lymphoma has been reported in some Seckel patients, they tend not to have ataxia or immunodeficiency. However, Seckel syndrome patients display features, including microcephaly and dysmorphic faces, commonly found in other syndromes associated with impaired responses to DNA damage(O'Driscoll et al., 2004). Similar to the *CREBBP* gene that causes Rubinstein-Taybi syndrome and epimutation of *H19* imprinted gene that causes Russell-Silver syndrome, it is possible that disease phenotypes caused by defects in essential genes involving basic cellular functions through interactions with multiple genes result in multiple malformations with generalised growth deficiency.

Another example of this is Roberts syndrome which is also characterised by severe growth deficiency with meromelia (partial limb defects) in four limbs and a peculiar face with microcephaly. The causative gene has recently been recognised and named *ESCO2*. The ESCO2 protein product is a member of a conserved protein family that is required for the establishment of sister chromatid cohesion during S phase and has putative acetyltransferase activity. Defects in chromatid cohesion lead to mitotic spindle checkpoint activation and impaired cell growth. The mitotic spindle checkpoint may be activated by the defect, resulting in the mitotic delay and impaired cell proliferation observed in cell lines derived from patients with Roberts syndrome. During embryogenesis, the loss of progenitor cells could preclude a sufficient number of cells required for development of structures affected in Roberts syndrome (Vega et al., 2005).

#### 1.4.1.4.2 Other possible explanations for CdLS

Another fascinating mechanism that might be involved in CdLS is the non-coding MicroRNAs (miRNAs) pathway. MicroRNAs (miRNAs) are 22-nucleotide, short, noncoding RNAs that are thought to regulate gene expression through sequence-specific base pairing with target mRNAs. MicroRNAs are transcribed as long RNA precursors (pri-miRNAs) that contain a stem loop structure of about 80 bases. Pri-miRNAs are processed in the nucleus by the RNase III enzyme Drosha and DGCR8/Pasha, which excises the stem-loop to form the pre-miRNA. Pre-miRNAs are then exported from the nucleus by Exportin-5. In the cytoplasm, another RNase III enzyme, Dicer, cuts the pre-miRNA to generate the mature microRNA as part of a short RNA duplex. The RNA is subsequently unwound by a helicase activity and incorporated into a RNA-induced silencing complex (RISC) (Alvarez-Garcia and Miska, 2005; Pasquinelli, 2006). Most microRNAs in animals are thought to be functioned through the inhibition of effective mRNA translation of target genes through imperfect base pairing with the 3'-untranslated region (3'UTR) of target mRNAs (Yekta et al., 2004). The underlying mechanism is still poorly understood. MicroRNA targets are largely unknown, but estimates range from one to hundreds of target genes for a given microRNA(Enright et al., 2003). Studies in model organisms have shown that miRNAs are involved in the control of developmental timing, cell proliferation, left-right patterning, neuronal cell fate, apoptosis and fat metabolism in invertebrates (Alvarez-Garcia and Miska, 2005; Mattick and Makunin, 2005). Therefore, it is reasonable to expect a similar range of functions in the developing vertebrate. Dicer defective ES cells exhibit severe defects in differentiation *in vitro* as well as in centromeric silencing (Kanellopoulou et al., 2005). miRNAs embedded in *Hox* clusters exhibit expression patterns that are

comparable of *Hox* genes (Chopra and Mishra, 2006; Pearson et al., 2005). This suggests a miRNA-mediated mechanism for the post-transcriptional restriction of *Hox* gene expression during vertebrate development (Tanzer et al., 2005; Yekta et al., 2004). Although no evidence suggests the association between miRNAs machinery and the CdLS, the regulatory role of miRNAs during early vertebrate development has gradually caught scientists' attention. Thus, we can not exclude the possibility that CdLS is caused by the defect in molecules such as dicer or helicase, which is involved in miRNAs biogenesis. The CdLS phenotype might also be caused by yet identified miRNAs that exert their effects through posttranscriptional modulation of a range of genes important to progenitor cells proliferation and differentiation.

Another syndrome with phenotypic similarity to CdLS is fetal alcohol syndrome, which is caused by chronically maternal alcohol consumption. Babies affected with fetal alcohol syndrome are characterised by growth deficiency of prenatal onset, microcephaly, and short palpebral fissure. This syndrome is caused by the environmental factor rather than by the genetic factors. It is possible that teratogenic agents might also contribute to the generalised disregulation of certain important developmental steps associated with CdLS.

1.4.1.4.3 Predicted expression patterns and functions of CdLS genes

. The inheritance pattern in most of the familial cases suggested that CdLS appears to be a dominant disorder and it seems that in most cases the phenotype is expressed in the heterozygote. Dominantly inherited genes are likely to cause a phenotype through two actions. The first is haploinsufficiency, a defect in one of the alleles, which leads to the insufficient expression of the gene product that is required for the minimal function of the gene. Another possibility is a gain-of-function. Here,

mutations in the gene, especially missense mutations, may alter the properties of an important functional domain leading to altered gene function. The resulting product either functions in an inappropriate way or interferes with the functions of the normal gene product - a so called dominant-negative effect. As gradually more genes involved in early human developmental process are identified, the identification of the genes that might be involved with the CdLS becomes difficult. The predicted function of the putative CdLS gene is regulation of cell growth, since prenatal and postnatal growth retardation are the hallmarks of CdLS. Thus, a gene that is associated with cell proliferation is strongly considered as a candidate. Although severe limb reduction defects are reported in around 10-20% of CdLS patients (Jackson et al., 1993), severe limb reduction defects are very rare in CdLS patients of Taiwanese origin, even in patients with very severe phenotypes (personal communication in local meeting, unpublished data). Most of the patients in Taiwan have small hands and feet which are proportionate to their small build. Unless several different genes are involved in the pathogenesis of the CdLS phenotype from different ethnic backgrounds, we believe that the CdLS causing gene must play a role in basic cellular functions such as cell cycle regulators during early embryonic development that affect cell proliferation during organogenesis. Hence, genes that cause limb defects only are not considered likely candidates. The CdLS gene may exert its developmental regulatory effects through interaction with other molecules as seen in *CREBBP* gene in Rubinstein-Taybi syndrome.

Unlike with diseases such as enzymopathies, it was impossible for us to use the gene product to map the gene, as this was not feasible in CdLS due to the lack of protein product information. Moreover, no known mouse models that are similar to

the phenotypes of CdLS had been reported. Thus, it was hard to find the CdLS disease gene through mouse models.

However, due to the lack of some consistent clues such as some particular biological markers or animal models reminiscent of CdLS, gene mapping through a candidate gene approach is presently impractical.

### 1.4.2 Strategies of Gene Identification in Newcastle Group

Among these reported cases, only three patients with clinical features typical of severe CdLS who had a *de novo* balanced reciprocal translocation with blood samples were available to us. These include a female infant with a t(14;21)(q32;q11) (Wilson et al., 1983), a female newborn with a t(3;17)(q26.3;q23.1) (Ireland et al., 1991), and an unreported girl with a t(5;13)(p13.1;q12.1) balanced translocation. All of them are sporadic cases. No familial inheritance was noted.

Previously, a novel gene was found to be interrupted by the t (3; 17) (q26.3; q23.1) translocation breakpoint. However, patient-specific mutations were not found in this gene (Tonkin et al., 2004a). Linkage studies also exclude 3q26.3 as a candidate region (Krantz et al., 2001). Thus, we were prompted to investigate other chromosomal regions. Linkage studies using the rare familial cases are ideal for the purpose. Moreover, collection of the familial CdLS cases was extremely difficult for us. At that time, the Newcastle group only had two families affected by familial CdLS and a decision was taken to donate the blood samples from members of these two families to Drs. Krantz and Jackson who were committed to a global search for CdLS families that were suitable for linakge analysis. Meanwhile, the Newcastle group decided to continue the gene mapping by seeking to define, at the molecular level, the breakpoints in the t(5;13) and t(14;21) *de novo* balanced translocations, that would

then be expected to lead to candidate gene identification. A second approach undertaken by the Newcastle group in collaboration with Nigel Carter and co-workers at the Wellcome Trust Sanger Institute was to use microarray CGH to investigate the possibility that CdLS was due to a chromosomal microdeletion/microduplication that would involve deletion/duplication of multiple contiguous genes but that could not be detected by standard cytogenetic approaches. The initial approach was to use BAC clones selected from the candidate regions for the FISH mapping. The details for our positional mapping are described in chapter 2.

## 1.5 AIMS OF THE PROJECT

The aims of this project were:

1. To map the CdLS disease gene by using FISH mapping on three patients with *de novo* balanced translocation as described above.

2. Patient-specific mutation screening of the most compelling candidate CdLS gene to determine if the mapped gene is the causative gene for CdLS

3. Ongoing expression studies of the CdLS disease gene(s).

## 1.6 REFERENCES


Aitken, D. A., Ireland, M., Berry, E., Crossley, J. A., Macri, J. N., Burn, J., and Connor, J. M. (1999). Second-trimester pregnancy associated plasma protein-A levels are reduced in Cornelia de Lange syndrome pregnancies. Prenat Diagn *19*, 706-710.

Alderton, G. K., Joenje, H., Varon, R., Borglum, A. D., Jeggo, P. A., and O'Driscoll, M. (2004). Seckel syndrome exhibits cellular features demonstrating defects in the ATR-signalling pathway. Hum Mol Genet *13*, 3127-3138.

Allanson, J. E., Hennekam, R. C., and Ireland, M. (1997). De Lange syndrome: subjective and objective comparison of the classical and mild phenotypes. J Med Genet *34*, 645-650.

Alvarez-Garcia, I., and Miska, E. A. (2005). MicroRNA functions in animal development and human disease. Development *132*, 4653-4662.

Aqua, M. S., Rizzu, P., Lindsay, E. A., Shaffer, L. G., Zackai, E. H., Overhauser, J., and Baldini, A. (1995). Duplication 3q syndrome: molecular delineation of the critical region. Am J Med Genet *55*, 33-37.

Arbuzova, S., Nikolenko, M., Krantz, D., Hallahan, T., and Macri, J. (2003). Low first-trimester pregnancy-associated plasma protein-A and Cornelia de Lange syndrome. Prenat Diagn *23*, 864.

Babu, K. A., Verma, R. S., Rodriguez, J., Rosenfeld, W., and Jhaveri, R. C. (1985). A possible clinical implication of homozygous inversions of 9qh regions with Cornelia de Lange syndrome (CLS). Hum Hered *35*, 265-267.

Barr, A. N., Grabow, J. D., Matthews, C. G., Grosse, F. R., Motl, M. L., and Opitz, J. M. (1971). Neurologic and psychometric findings in the Brachmann-De Lange syndrome. Neuropadiatrie *3*, 46-66.

Bay, C., Mauk, J., Radcliffe, J., and Kaplan, P. (1993). Mild Brachmann-de Lange syndrome. Delineation of the clinical phenotype, and characteristic behaviors in a six-year-old boy. Am J Med Genet *47*, 965-968.

Beck, B., and Fenger, K. (1985). Mortality, pathological findings and causes of death

in the de Lange syndrome. Acta Paediatr Scand *74*, 765-769.

Beck, B., and Mikkelsen, M. (1981). Chromosomes in the Cornelia de Lange syndrome. Hum Genet *59*, 271-276.

Berg J, S. G., Ridler M, McCreary B, Faunch J, Franham F, Allen M. (1967). De Lange syndrome: Report of a case with an unusual karyotype. J Med Genet *4*, 184-189.

Bernard, L. E., Penaherrera, M. S., Van Allen, M. I., Wang, M. S., Yong, S. L., Gareis, F., Langlois, S., and Robinson, W. P. (1999). Clinical and molecular findings in two patients with russell-silver syndrome and UPD7: comparison with non-UPD7 cases. Am J Med Genet *87*, 230-236.

Berney, T. P., Ireland, M., and Burn, J. (1999). Behavioural phenotype of Cornelia de Lange syndrome. Arch Dis Child *81*, 333-336.

Borck, G., Redon, R., Sanlaville, D., Rio, M., Prieur, M., Lyonnet, S., Vekemans, M., Carter, N. P., Munnich, A., Colleaux, L., and Cormier-Daire, V. (2004). NIPBL mutations and genetic heterogeneity in Cornelia de Lange syndrome. J Med Genet *41*, e128.

Bozner, P., Blackburn, W., and Cooley, N. R., Jr. (1995). Bilateral ulnar agenesis: case report and review of the literature. Pediatr Pathol Lab Med *15*, 895-913.

Brachmann (1916). Ein fall von symmetrischer monodaktylie durch Ulnadefekt, mit symmetrischer flughautbildung in den ellenbeugen, sowie anderen abnormitaten (zwerghaftogkeit, halsrippen, behaarung). . Jarb Kinder Phys Erzie *84* 225-235.

Braddock, S. R., Lachman, R. S., Stoppenhagen, C. C., Carey, J. C., Ireland, M., Moeschler, J. B., Cunniff, C., and Graham, J. M., Jr. (1993). Radiological features in Brachmann-de Lange syndrome. Am J Med Genet *47*, 1006-1013.

Bruner, J. P., and Hsia, Y. E. (1990). Prenatal findings in Brachmann-de Lange syndrome. Obstet Gynecol *76*, 966-968.

Bryson, Y., Sakati, N., Nyhan, W. L., and Fish, C. H. (1971). Self-mutilative behavior in the Cornelia de Lange syndrome. Am J Ment Defic *76*, 319-324.

Calo, S., Gualandri, W., and Radice, C. (1968). [XY-XO mosaicism in a case of Cornelia De Lange syndrome]. Minerva Pediatr *20*, 2600-2604.

Casper, A. M., Durkin, S. G., Arlt, M. F., and Glover, T. W. (2004). Chromosomal instability at common fragile sites in Seckel syndrome. Am J Hum Genet *75*, 654-660.

Cates, M., Billmire, D. F., Bull, M. J., and Grosfeld, J. L. (1989). Gastroesophageal dysfunction in Cornelia de Lange syndrome. J Pediatr Surg *24*, 248-250.

Chopra, V. S., and Mishra, R. K. (2006). "Mir"acles in hox gene regulation. Bioessays *28*, 445-448.

Cobb, J. A., Schleker, T., Rojas, V., Bjergbaek, L., Tercero, J. A., and Gasser, S. M. (2005). Replisome instability, fork collapse, and gross chromosomal rearrangements arise synergistically from Mec1 kinase and RecQ helicase mutations. Genes Dev *19*, 3055-3069.

Craig, A. P., and Luzzatti, L. (1965). Translocation in De Lange's Syndrome? Lancet *15*, 445-446.

Cunniff, C., Curry, C. J., Carey, J. C., Graham, J. M., Jr., Williams, C. A., Stengel-Rutkowski, S., Luttgen, S., and Meinecke, P. (1993). Congenital diaphragmatic hernia in the Brachmann-de Lange syndrome. Am J Med Genet *47*, 1018-1021.

D'Oelsnitz, M., Ayraud, N., Vaillant, J. M., and de Swarte, M. (1971). [Cornelia de Lange's syndrome. 4 neonatal cases]. Ann Pediatr (Paris) *18*, 7-16.

De Marchi, N., Antonarakis, S. E., and Jackson, L. (1994). No uniparental disomy for chromosome 3 in Brachmann-De Lange syndrome. Am J Med Genet *49*, 133-135.

deLange (1933). Sur un type nouveau de degenerescence (typus Amstelodamensis). Arch Med Enfants *36*, 713-719.

Descipio, C., Kaur, M., Yaeger, D., Innis, J. W., Spinner, N. B., Jackson, L. G., and Krantz, I. D. (2005). Chromosome rearrangements in Cornelia de Lange syndrome (CdLS): Report of a der(3)t(3;12)(p25.3;p13.3) in two half sibs with features of CdLS and review of reported CdLS cases with chromosome rearrangements. Am J Med

Genet A.

Drolshagen, L. F., Durmon, G., Berumen, M., and Burks, D. D. (1992). Prenatal ultrasonographic appearance of "Cornelia de Lange" syndrome. J Clin Ultrasound *20*, 470-474.

DuVall, G. A., and Walden, D. T. (1996). Adenocarcinoma of the esophagus complicating Cornelia de Lange syndrome. J Clin Gastroenterol *22*, 131-133.

Eeg-Olofsson, O., and Liedgren, S. (1981). A re-investigation of an inherited chromosome aberration in a girl with signs of De Lange syndrome. Acta Paediatr Scand *70*, 581-582.

Egemen, A., Ulger, Z., Ozkinay, F., Gulen, F., and Cogulu, O. (2005). A de novo t (X;8)(p11.2;q24.3) demonstrating Cornelia de Lange syndrome phenotype. Genet Couns *16*, 27-30.

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2003). MicroRNA targets in Drosophila. Genome Biol *5*, R1.

Filippi, G. (1989). The de Lange syndrome. Report of 15 cases. Clin Genet *35*, 343-363.

Froster, U. G., and Gortner, L. (1993). Thrombocytopenia in the Brachmann-de Lange syndrome. Am J Med Genet *46*, 730-731.

Fryns, J. P., and Vinken, L. (1994). Thrombocytopenia in the Brachmann-de Lange syndrome. Am J Med Genet *49*, 360.

Gadoth, N., Lerman, M., Garty, B. Z., and Shmuelewitz, O. (1982). Normal intelligence in the Cornelia de Lange syndrome. Johns Hopkins Med J *150*, 70-72.

Gans, B., Thurston, J. G. (1965). De Lange's Amsterdam Dwarfs Syndrome. Report of four cases. Dev Med Child Neurol *38*, 42-45.

Geudeke, M., Bijlsma, J. B., and De Bruijne, J. I. (1963). [Chromosome Study on Typus Degenerativus Amstelodamensis (De Lange's Syndrome).]. Maandschr Kindergeneeskd *31*, 248-258.

Gibbons, R. J., Picketts, D. J., Villard, L., and Higgs, D. R. (1995). Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). Cell *80*, 837-845.

Gicquel, C., Rossignol, S., Cabrol, S., Houang, M., Steunou, V., Barbu, V., Danton, F., Thibaud, N., Le Merrer, M., Burglen, L., *et al.* (2005). Epimutation of the telomeric imprinting center region on chromosome 11p15 in Silver-Russell syndrome. Nat Genet *37*, 1003-1007.

Goodban, M. T. (1993). Survey of speech and language skills with prognostic indicators in 116 patients with Cornelia de Lange syndrome. Am J Med Genet *47*, 1059-1063.

Goodman, R. H., and Smolik, S. (2000). CBP/p300 in cell growth, transformation, and development. Genes Dev *14*, 1553-1577.

Greenberg, F., and Robinson, L. K. (1989). Mild Brachmann-de Lange syndrome: changes of phenotype with age. Am J Med Genet *32*, 90-92.

Greenwood, R. D., Sommer, A., Craenen, J., Waldman, J. D., and Rosenthal, A. (1977). Congenital heart disease in de Lange's syndrome. South Med J *70*, 80-81.

Hawley, P. P., Jackson, L. G., and Kurnit, D. M. (1985). Sixty-four patients with Brachmann-de Lange syndrome: a survey. Am J Med Genet *20*, 453-459.

Hersh, J. H., Dale, K. S., Gerald, P. S., Yen, F. F., Weisskopf, B., and Dinno, N. D. (1985). Dup(4p)del(9p) in a familial mental retardation syndrome. Resemblance to de Lange syndrome detected by high-resolution banding. Am J Dis Child *139*, 81-84.

Holder, S. E., Grimsley, L. M., Palmer, R. W., Butler, L. J., and Baraitser, M. (1994). Partial trisomy 3q causing mild Cornelia de Lange phenotype. J Med Genet *31*, 150-152.

Hooft, C., Lormans, J., and Jongbloet, P. (1965). [Typus Degenerativus Amstelodamensis or the Cornelia De Lange Syndrome.]. Acta Paediatr Belg *19*, 5-37.

Hulinsky, R., Byrne, J. L., Lowichik, A., and Viskochil, D. H. (2005). Fetus with interstitial del(5)(p13.1p14.2) diagnosed postnatally with Cornelia de Lange

syndrome. Am J Med Genet A.

Husain, K., Fitzgerald, P., and Lau, G. (1994). Cecal volvulus in the Cornelia de Lange syndrome. J Pediatr Surg *29*, 1245-1247.

Ireland, M., and Burn, J. (1993). Cornelia de Lange syndrome--photo essay. Clin Dysmorphol *2*, 151-160.

Ireland, M., Donnai, D., and Burn, J. (1993). Brachmann-de Lange syndrome. Delineation of the clinical phenotype. Am J Med Genet *47*, 959-964.

Ireland, M., English, C., Cross, I., Houlsby, W. T., and Burn, J. (1991). A de novo translocation t(3;17)(q26.3;q23.1) in a child with Cornelia de Lange syndrome. J Med Genet *28*, 639-640.

Ireland, M., English, C., Cross, I., Lindsay, S., and Strachan, T. (1995). Partial trisomy 3q and the mild Cornelia de Lange syndrome phenotype. J Med Genet *32*, 837-838.

Jackson, L., Kline, A. D., Barr, M. A., and Koch, S. (1993). de Lange syndrome: a clinical review of 310 individuals. Am J Med Genet *47*, 940-946.

Johnson, H. G., Ekman, P., and Friesen, W. (1976). A behavioral phenotype in the de Lange syndrome. Pediatr Res *10*, 843-850.

Kanellopoulou, C., Muljo, S. A., Kung, A. L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D. M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev *19*, 489-501.

Kawai, M., Kawahara, H., Hirayama, S., Yoshimura, N., and Ida, S. (2004). Effect of baclofen on emesis and 24-hour esophageal pH in neurologically impaired children with gastroesophageal reflux disease. J Pediatr Gastroenterol Nutr *38*, 317-323.

Kline, A. D., Barr, M., and Jackson, L. G. (1993a). Growth manifestations in the Brachmann-de Lange syndrome. Am J Med Genet *47*, 1042-1049.

Kline, A. D., Stanley, C., Belevich, J., Brodsky, K., Barr, M., and Jackson, L. G. (1993b). Developmental data on individuals with the Brachmann-de Lange syndrome.

Am J Med Genet *47*, 1053-1058.

Klosovskii, B. N., Iankova, M. F., Fateeva, E. M., and Damanskaia, L. (1968). [On the problem of the De Lange's syndrome]. Pediatriia *47*, 33-39.

Kousseff, B. G., Newkirk, P., and Root, A. W. (1994). Brachmann-de Lange syndrome. 1994 update. Arch Pediatr Adolesc Med *148*, 749-755.

Kousseff, B. G., Thomson-Meares, J., Newkirk, P., and Root, A. W. (1993). Physical growth in Brachmann-de Lange syndrome. Am J Med Genet *47*, 1050-1052.

Krantz, I. D., Tonkin, E., Smith, M., Devoto, M., Bottani, A., Simpson, C., Hofreiter, M., Abraham, V., Jukofsky, L., Conti, B. P., *et al.* (2001). Exclusion of linkage to the CDL1 gene region on chromosome 3q26.3 in some familial cases of Cornelia de Lange syndrome. Am J Med Genet *101*, 120-129.

Lakshminarayana, P., and Nallasivam, P. (1990). Cornelia de Lange syndrome with ring chromosome 3. J Med Genet *27*, 405-406.

Leavitt, A., Dinno, N., and Davis, C. (1985). Cornelia de Lange syndrome in a mother and daughter. Clin Genet *28*, 157-161.

Lee, F. A., and Kenny, F. M. (1967). Skeletal changes in the Cornelia de Lange syndrome. Am J Roentgenol Radium Ther Nucl Med *100*, 27-39.

Leroy, J. G., Persijn, J., Van de Weghe, V., Van Hecke, R., Oostra, A., De Bie, S., and Craen, M. (1993). On the variability of the Brachmann-de Lange syndrome in seven patients. Am J Med Genet *47*, 983-991.

Levin, A. V., Seidman, D. J., Nelson, L. B., and Jackson, L. G. (1990). Ophthalmologic findings in the Cornelia de Lange syndrome. J Pediatr Ophthalmol Strabismus *27*, 94-102.

Luzzani, S., Macchini, F., Valade, A., Milani, D., and Selicorni, A. (2003). Gastroesophageal reflux and Cornelia de Lange syndrome: typical and atypical symptoms. Am J Med Genet A *119*, 283-287.

Masumoto, K., Izaki, T., and Arima, T. (2001). Cornelia de Lange syndrome associated with cecal volvulus: report of a case. Acta Paediatr *90*, 701-703.

Matsuo, N., Smith, R. J., and Kosaki, K. (2000). Imprinting of human GRB10 and its mutations in two patients with Russell-Silver syndrome. Am J Hum Genet *67*, 476-482.

Mattick, J. S., and Makunin, I. V. (2005). Small regulatory RNAs in mammals. Hum Mol Genet *14 Spec No 1*, R121-132.

McConnell, V., Brown, T., and Morrison, P. J. (2003). An Irish three-generation family of Cornelia de Lange syndrome displaying autosomal dominant inheritance. Clin Dysmorphol *12*, 241-244.

Mehta, A. V., and Ambalavanan, S. K. (1997). Occurrence of congenital heart disease in children with Brachmann-de Lange syndrome. Am J Med Genet *71*, 434-435.

Morrione, A., Valentinis, B., Resnicoff, M., Xu, S., and Baserga, R. (1997). The role of mGrb10alpha in insulin-like growth factor I-mediated growth. J Biol Chem *272*, 26382-26387.

Mosher, G. A., Schulte, R. L., Kaplan, P. A., Buehler, B. A., and Sanger, W. G. (1985). Pregnancy in a woman with the Brachmann-de Lange syndrome. Am J Med Genet *22*, 103-107.

MOtl, M. L., Opitz, J. M. (1971). Studies of malformation syndromes XXVA. Phenotypic and genetic studies of the Brachmann-de Lange syndrome. Hum Hered *21*, 1-16.

Moutoussamy, S., Renaudie, F., Lago, F., Kelly, P. A., and Finidori, J. (1998). Grb10 identified as a potential regulator of growth hormone (GH) signaling by cloning of GH receptor target proteins. J Biol Chem *273*, 15906-15912.

O'Driscoll, M., Gennery, A. R., Seidel, J., Concannon, P., and Jeggo, P. A. (2004). An overview of three new disorders associated with genetic instability: LIG4 syndrome, RS-SCID and ATR-Seckel syndrome. DNA Repair (Amst) *3*, 1227-1235.

O'Driscoll, M., Ruiz-Perez, V. L., Woods, C. G., Jeggo, P. A., and Goodship, J. A.

(2003). A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. Nat Genet *33*, 497-501.

Opitz, J. M. (1985). The Brachmann-de Lange syndrome. Am J Med Genet *22*, 89-102.

Ott, J., Robinson, A., Peakman, D. C. (1968). D/D balanced translocations. Lancet *2*, 1352-1353.

Pankau, R., and Janig, U. (1993). Diaphragmatic defect in Brachmann-de Lange syndrome: a further observation. Am J Med Genet *47*, 1024-1025.

Partington, M. W. (1990). Rubinstein-Taybi syndrome: a follow-up study. Am J Med Genet Suppl *6*, 65-68.

Pashayan, H. M., Fraser, F. C., and Pruzansky, S. (1975). Variable limb malformations in the Brachmann-Cornelia de Lange syndrome. Birth Defects Orig Artic Ser *11*, 147-156.

Pasquinelli, A. E. (2006). Demystifying small RNA pathways. Dev Cell *10*, 419-424.

Payne, H. W., and Maeda, W. K. (1965). The Cornelia de Lange syndrome: clinical and cytogenetic interpretations. Can Med Assoc J *93*, 577-586.

Pearson, J. C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nat Rev Genet *6*, 893-904.

Preus, M., and Rex, A. P. (1983). Definition and diagnosis of the Brachmann-De Lange syndrome. Am J Med Genet *16*, 301-312.

Price, N., Bahra, M., Griffin, D., Hanna, G., and Stock, A. (2005). Cornelia de Lange Syndrome in association with a balanced reciprocal translocation involving chromosomes 3 and 5. Prenat Diagn *25*, 602-603.

Randall-Pinto S, J. G., Lopes V, Teshima I, Wyatt P. (2000). Two cases of partial trisomy 10q and monosomy 7q with features similar to Cornelia de Lange syndrome, including split hand/split foot deformity. Am J Hum Genet *suppl 67*, 126.

Rao, P. S., and Sissman, N. J. (1971). Congenital heart disease in the de Lange syndrome. J Pediatr *79*, 674-677.

Roelfsema, J. H., White, S. J., Ariyurek, Y., Bartholdi, D., Niedrist, D., Papadia, F., Bacino, C. A., den Dunnen, J. T., van Ommen, G. J., Breuning, M. H., *et al.* (2005). Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. Am J Hum Genet *76*, 572-580.

Roposch, A., Bhaskar, A. R., Lee, F., Adedapo, S., Mousny, M., and Alman, B. A. (2004). Orthopaedic manifestations of Brachmann-de Lange syndrome: a report of 34 patients. J Pediatr Orthop B *13*, 118-122.

Saal, H. M., Samango-Sprouse, C. A., Rodnan, L. A., Rosenbaum, K. N., and Custer, D. A. (1993). Brachmann-de Lange syndrome with normal IQ. Am J Med Genet *47*, 995-998.

Sakai, Y., Watanabe, T., and Kaga, K. (2002). Auditory brainstem responses and usefulness of hearing aids in hearing impaired children with Cornelia de Lange syndrome. Int J Pediatr Otorhinolaryngol *66*, 63-69.

Sasaki, T., Kaga, K., Ohira, Y., Ogawa, Y., and Fukushima, Y. (1996). Temporal bone and brain stem histopathological findings in Cornelia de Lange syndrome. Int J Pediatr Otorhinolaryngol *36*, 195-204.

Sataloff, R. T., Spiegel, J. R., Hawkshaw, M., Epstein, J. M., and Jackson, L. (1990). Cornelia de Lange syndrome. Otolaryngologic manifestations. Arch Otolaryngol Head Neck Surg *116*, 1044-1046.

Schwanitz, G., Schmid, R. D., Grosse, G., and Grahn-Liebe, E. (1977). [Familial translocation 3/22 MAT with partial trisomy 3q (author's transl)]. J Genet Hum *25*, 141-150.

Schwartz, I. D., Schwartz, K. J., Kousseff, B. G., Bercu, B. B., and Root, A. W. (1990). Endocrinopathies in Cornelia de Lange syndrome. J Pediatr *117*, 920-923.

Sciorra, L. J., Bahng, K., and Lee, M. L. (1979). Trisomy in the distal end of the long arm of chromosome 3. A condition clinically similar to the Cornelia de Lange syndrome. Am J Dis Child *133*, 727-730.

Selicorni, A., Sforzini, C., Milani, D., Cagnoli, G., Fossali, E., and Bianchetti, M. G. (2005). Anomalies of the kidney and urinary tract are common in de Lange syndrome.

Am J Med Genet A *132*, 395-397.

Shaffer, L. G., Overhauser, J., Jackson, L. G., and Ledbetter, D. H. (1993). Genetic syndromes and uniparental disomy: a study of 16 cases of Brachmann-de Lange syndrome. Am J Med Genet *47*, 383-386.

Smith, M., Herrell, S., Lusher, M., Lako, L., Simpson, C., Wiestner, A., Skoda, R., Ireland, M., and Strachan, T. (1999). Genomic organisation of the human chordin gene and mutation screening of candidate Cornelia de Lange syndrome genes. Hum Genet *105*, 104-111.

Sommer, A. (1993). Occurrence of the Sandifer complex in the Brachmann-de Lange syndrome. Am J Med Genet *47*, 1026-1028.

Szemere, G., Godo, B., Osvath, P., Lehrner, J., and Pataki, O. (1972). Chromosomal anomaly associated with Cornelia de Lange's syndrome. Acta Paediatr Acad Sci Hung *13*, 51-55.

Tanzer, A., Amemiya, C. T., Kim, C. B., and Stadler, P. F. (2005). Evolution of microRNAs located within Hox gene clusters. J Exp Zoolog B Mol Dev Evol *304*, 75-85.

Taylor, M. J., and Josifek, K. (1981). Multiple congenital anomalies, thymic dysplasia, severe congenital heart disease, and oligosyndactyly with a deletion of the short arm of chromosome 5. Am J Med Genet *9*, 5-11.

Telvi, L., Lebbar, A., Del Pino, O., Barbet, J. P., and Chaussain, J. L. (1999). 45,X/46,XY mosaicism: report of 27 cases. Pediatrics *104*, 304-308.

Tonini, G., and Marinoni, S. (1990). Neonatal-onset panhypopituitarism in a girl with Brachmann-De Lange syndrome. Am J Med Genet *36*, 102-103.

Tonkin, E. T., Smith, M., Eichhorn, P., Jones, S., Imamwerdi, B., Lindsay, S., Jackson, M., Wang, T. J., Ireland, M., Burn, J., *et al.* (2004a). A giant novel gene undergoing extensive alternative splicing is severed by a Cornelia de Lange-associated translocation breakpoint at 3q26.3. Hum Genet *115*, 139-148.

Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J., and Strachan, T. (2004b). NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins

and fly Nipped-B, is mutated in Cornelia de Lange syndrome. Nat Genet *36*, 636-641.

Tsukahara, M., Okamoto, N., Ohashi, H., Kuwajima, K., Kondo, I., Sugie, H., Nagai, T., Naritomi, K., Hasegawa, T., Fukushima, Y., *et al.* (1998). Brachmann-de Lange syndrome and congenital heart disease. Am J Med Genet *75*, 441-442.

Van Allen, M. I., Filippi, G., Siegel-Bartelt, J., Yong, S. L., McGillivray, B., Zuker, R. M., Smith, C. R., Magee, J. F., Ritchie, S., Toi, A., and et al. (1993). Clinical variability within Brachmann-de Lange syndrome: a proposed classification system. Am J Med Genet *47*, 947-958.

Vega, H., Waisfisz, Q., Gordillo, M., Sakai, N., Yanagihara, I., Yamada, M., van Gosliga, D., Kayserili, H., Xu, C., Ozono, K., *et al.* (2005). Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nat Genet *37*, 468-470.

Verloes, A., Lesenfants, S., Philippet, B., Iyawa, A., Laloux, F., and Koulischer, L. (1996). Microcephaly, macrotia, unusual mimics and mental retardation syndrome: new syndrome or variant of De Lange type 2 syndrome. Genet Couns *7*, 277-282.

Westergaard, J. G., Chemnitz, J., Teisner, B., Poulsen, H. K., Ipsen, L., Beck, B., and Grudzinskas, J. G. (1983). Pregnancy-associated plasma protein A: a possible marker in the classification and prenatal diagnosis of Cornelia de Lange syndrome. Prenat Diagn *3*, 225-232.

Westermann, P., Rossi, M. V., Merlo, G., and Talarico, R. (1977). [The Cornelia De Lange syndrome. Study of 2 siblings, children of a diabetic mother]. Minerva Pediatr *29*, 1155-1160.

Wilson, G. N., Hieber, V. C., and Schmickel, R. D. (1978). The association of chromosome 3 duplication and the Cornelia de Lange syndrome. J Pediatr *93*, 783-788.

Wilson, W. G., Kennaugh, J. M., Kugler, J. P., and Wyandt, H. E. (1983). Reciprocal translocation 14q;21q in a patient with the Brachmann-de Lange syndrome. J Med Genet *20*, 469-471.

Yamaguchi, K., and Ishitobi, F. (1999). Brain dysgenesis in Cornelia de Lange syndrome. Clin Neuropathol *18*, 99-105.

Yekta, S., Shih, I. H., and Bartel, D. P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science *304*, 594-596.

Yoshihashi, H., Maeyama, K., Kosaki, R., Ogata, T., Tsukahara, M., Goto, Y., Hata, J.,

# CHAPTER 2: FISH MAPPING OF CdLS PATIENTS WITH *de novo* BALANCED RECIPROCAL TRANSLOCATIONS and GENOMIC ORGANIZATION OF THE GENE TRUNCATED BY THE t(5;13) TRANSLOCATION

## 2.1 INTRODUCTION:

Chromosome translocations can be extremely valuable for assessing the contributions of individual loci to the phenotype of certain syndromes, especially when occurring *de novo*. Compared with cytogenetically visible deletions or duplications, or unbalanced translocations which could leave tens or hundreds of genes to search through, the location of the *de novo* breakpoint within a balanced translocation should provide a quick entry into the gene which will most likely be severed by the translocation or immediately adjacent to it, since there is no apparent loss or gain of DNA. From the wide range of CdLS-associated chromosome abnormalities (Table 1.1), our Newcastle group were able to collect samples from three CdLS individuals with de novo balanced reciprocal translocations that had been reported before the onset of our study or had been referred as an unpublished case: a t(3; 17) (q26.3;q23.1) case (Ireland et al, 1990); a t(14; 21)(q32;q11.2) case (Wilson et al., 1983) and a t(5; 13)(p13.1; q12.1) From the wide range of CdLS-associated chromosome abnormalities (Table 1.1), our Newcastle group were able to collect samples from three CdLS individuals with *de novo* balanced reciprocal translocations that had been reported before the onset of our study or had been referred as an unpublished case: a t(3; 17) (q26.3;q23.1) case (Ireland et al, 1990); a t(14; 21)(q32;q11.2) case (Wilson et al., 1983) and a t(5; 13)(p13.1; q12.1) case (M.

Bamshad, pers. comm). After initial prioritization of the 3q26.3 breakpoint (based on the perceived similarities in phenotype between CdLS and individuals that were trisomic for 3q26.3, as discussed in the previous chapter) the case for a CdLS gene at 3q26 was weakened by the observation that in at least some multicase families linkage was excluded from the 3q26.2-q27 region ((Krantz et al., 2001). Accordingly, we also investigated the other breakpoints in the available translocation cases.

### 2.1.1 Standard FISH

Fluorescent *in situ* Hybridization (FISH) is a relatively new and important technology utilizing fluorescently labeled DNA probes for identifying chromosomes and part of chromosomes, deciphering chromosome rearrangements and locating genes on chromosomes. By using a fluorophore, which absorbs energy of a specific wavelength and re-emits energy at a different (but equally specific) wavelength, as well as the techniques of *in situ* hybridization, this technique allows the use of non-radioactive DNA probes to be hybridized with the chromosomes, cells or tissue in place (*in situ*) on a microscope slide.

FISH uses nucleic acid probes such as YACs, BACs, PACs, fosmids, cosmids, as well as cDNA sequences and Alu-PCR products. If the denatured probe and the target DNA sequences are complimentary to each other, fluorescent *in situ* hybridization will give a clear signal. With the completion of the human genome project and advances in bio-informatics technology, the information of the probes is easy to find and then information on the target DNA segments can be elucidated.

Up to now, two kinds of FISH technologies have been commonly used – the

indirect and the direct method. The indirect FISH method involves probes pre-labeled with a hapten, most commonly digoxigenin. After hybridization, fluorochrome-labeled antibodies to the hapten are used for probe detection. The indirect FISH method requires additional blocking reagents and amplification steps. Furthermore, it has been reported to have a higher degree of background fluorescence problems.

The direct FISH method involves directly labeled probes that have been pre-labeled with a specific fluorophore, allowing the fluorescent signal to be bound to the target in a single hybridization step. This method is more advanced, involves fewer steps, is easier to read and has a faster time-to-result. Therefore, direct label FISH represents a superior technology and offers many advantages. However, when analyzing very small probes, the indirect method is still preferred since the intensity of the fluorescence by using the direct method is not as strong as the intensity provided by the indirect method (Wiegant et al., 1991).

FISH can be used to detect genetic abnormalities, providing valuable information about prenatal disorders, cancer, as well as other genetic diseases and chromosome derangements. Unlike other molecular DNA-based tests, which require cell lysis to free nucleic acids for analysis, FISH allows the analysis of DNA *in situ*, that is, in its native, chromosomal form within the cell nucleus. This attribute permits the analysis of chromosomes and genes of individual cells.

Conventional chromosomal FISH uses metaphase spreads (metaphase FISH), with positive hybridization signals often showing double spots corresponding to probes hybridized to both sister chromatids. The maximum resolution for metaphase

FISH is around 2Mb. The use of interphase FISH can provide higher resolution analyses, since the chromosomes are extended longer. However, the problems with chromatin folding will make the signals from two different probes appear side-by-side and interfere with the data interpretation. Moreover, with the completion of the human genome draft, the relative positions of most BACs, PACs and fosmid clones are determined and can be viewed easily on most of the genome browsers. Therefore, the procedure to decide the relative positions of the two different probes becomes unnecessary. Thus, we employed the metaphase FISH as the main strategy for FISH mapping in our patients with balanced translocations.

In the past, once a BAC clone had been shown to map near the breakpoint, a cosmid library would be screened for higher resolution mapping. However, there are some disadvantages associated with cosmid clones that use bacteria as hosts for propagation. The presence of repeated elements, palindromes, Z-DNA forming sequences and methylated bases in recombinant constructs can cause deletion or rearrangement of cloned DNA in bacterial hosts. This instability has also been observed for a subset of the human chromosome 19-specific cosmid library (Yokobata et al., 1991). Therefore Kim et. al. developed a new cosmid cloning system (referred to as the Fosmid system) that employs the use of a low copy number plasmid based on the F factor replication origin (Kim et al., 1992). The use of host strains with various deficiencies in the recombination system could improve the stability of mammalian genomic libraries. Moreover, with more and more fosmid libraries being constructed and reported, we decided to use fosmids for finer mapping once the BAC clone that crosses the breakpoints was found.

### 2.1.2 Array-CGH

Another approach for gene mapping is to use comparative genome hybridization (CGH). CGH involves simultaneous painting of chromosomes in two different colors by using total DNA from two different sources as probes. CGH to metaphase chromosomes is a relatively rapid method for screening the entire genome. However, the resolution is limited to gains or losses of 2-10Mb, rendering it unsuitable in the analysis of small segmental deletions and duplications (Bentz et al., 1998).

A more sensitive technique designed for this purpose is microarrays based CGH (array CGH), in which individual BAC/PAC clones are arranged in arrays, rather than whole genomes as metaphases. They are subsequently hybridized with genomic DNA to detect dosage changes with the resolution down to 0.2-0.4 Mb. The increased resolution of array-CGH allows detection of deletions and duplications of a single BAC/PAC clone (Cheung et al., 2005). However, in 2002, the resolution for BAC array-CGH was still low. High resolution of BAC (~3000) array-CGH only became available later in 2003 (Carter et al., 2002; Watson et al., 2004).

## 2.2 AIMS OF THIS PROJECT

The linkage data performed by Krantz et al. suggested the exclusion of 3q26.3 as a candidate region for the CdLS gene (Krantz et al., 2001). Therefore, in this project, we focused on the translocation breakpoints on t(5;13) and t(14;21). FISH was first employed to map BAC clones relative to the chromosome 5p, 13q, 14q translocation breakpoints to narrow down the breakpoint regions because the breakpoint on chromosome 21 was expected to be in a region that had extremely few genes. The targets of the FISH analyses were metaphase chromosome spreads from blood lymphocytes collected from patients with translocations t(5;13) and t(14;21) respectively. Once the BAC clones that cross the respective translocation breakpoints were identified, the region was further narrowed down by FISH mapping using fosmid clones as probes to help with positional cloning.

## 2.3 MATERIALS AND METHODS

### 2.3.1 Data interpretation of FISH mapping in metaphase spreads:

#### 2.3.1.1 5p region

Clones that were mapped centromeric to the 5p translocation breakpoint gave two signals - one on normal chromosome 5 and one on translocated (derivative) chromosome 5. Clones that were mapped telomeric to the translocation breakpoint gave one signal on normal chromosome 5 and one on translocated chromosome 13. Clones, which spanned the translocation breakpoint, would give three signals – one on normal chromosome 5, one on translocated chromosome 5 and one on translocated chromosome 13. This is shown in the diagram below.

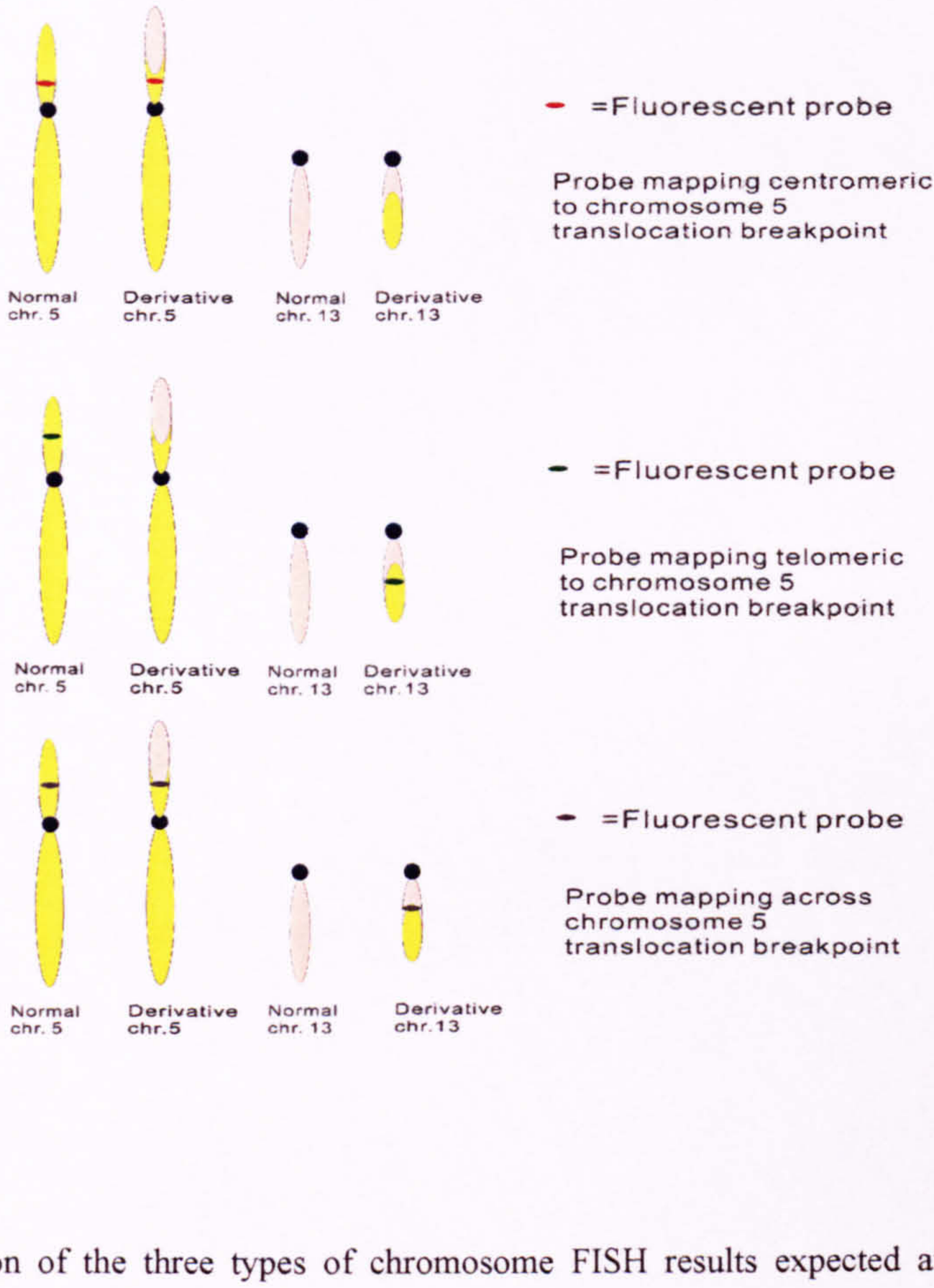



Fig. 2.1 Illustration of the three types of chromosome FISH results expected after hybridizing 5p probes to the t(5;13) metaphase chromosome preparation.

### 2.3.1.2 13q region

Clones that were centromeric to the 13q translocation breakpoint gave two signals - one on the normal chromosome 13 and one on the translocated (derivative) chromosome 13. Clones that were telomeric to the translocation breakpoint would give one signal on normal chromosome 13 and one on translocated chromosome 5. Clones, which spanned the translocation breakpoint, would give three signals – one on normal chromosome 13, one on translocated chromosome 5 and one on translocated chromosome 13. This is shown in the diagram below.

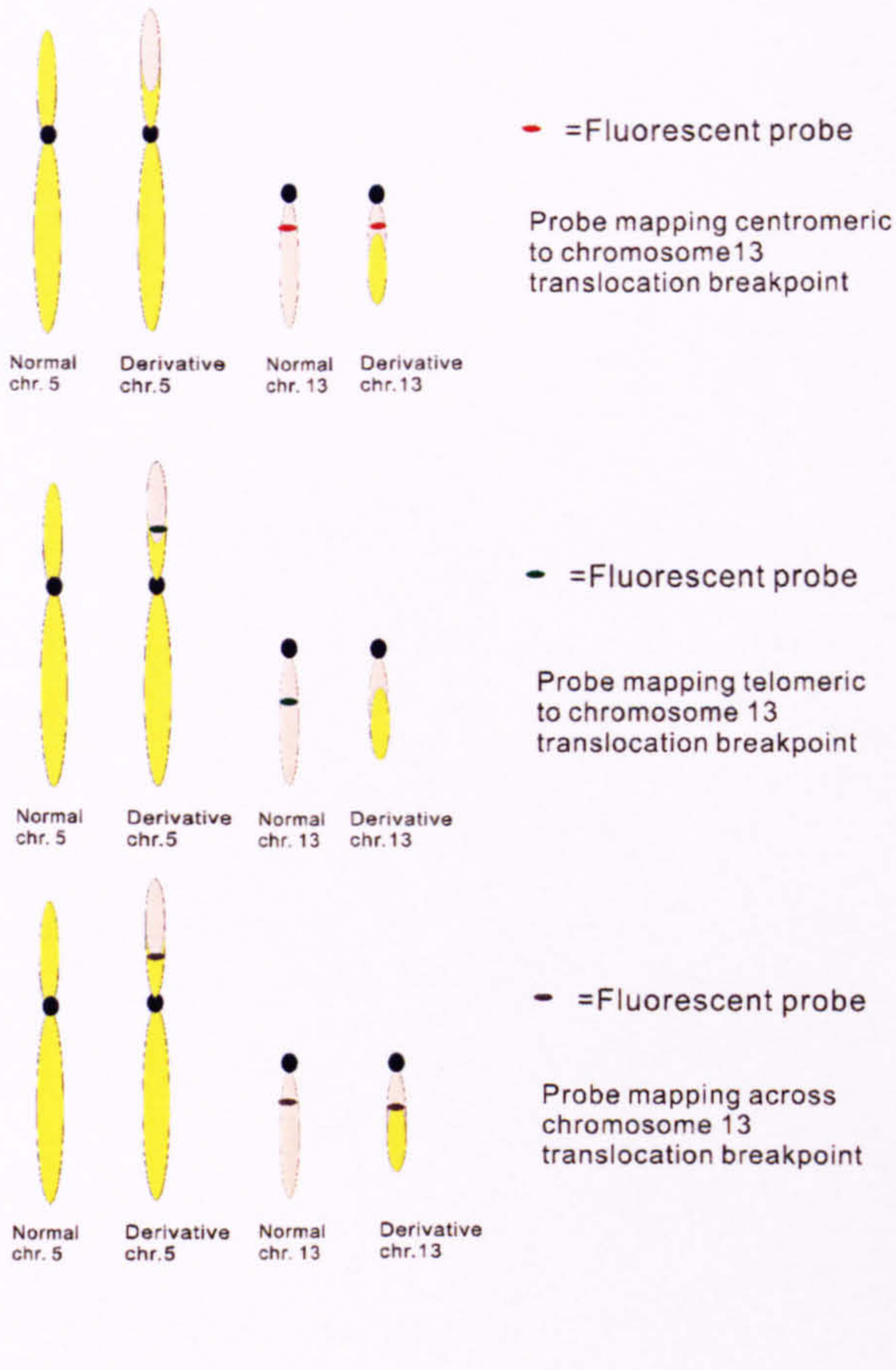



Fig. 2.2 Illustration of the three types of chromosome FISH results expected after hybridizing 13q probes to the t(5;13) metaphase chromosome preparation

### 2.3.1.3 14q region

Clones that were mapped centromeric to the 14q translocation breakpoint gave two signals - one on normal chromosome 14 and one on translocated (derivative) chromosome 14. Clones that were mapped telomeric to the translocation breakpoint would give one signal on normal chromosome 14 and one on translocated chromosome 21. Clones that spanned the translocation breakpoint, would give three signals – one on normal chromosome 14, one on translocated chromosome 14 and one on translocated chromosome 21.

### 2.3.2 Preparation of metaphase slides

Slides containing metaphase spreads were prepared by staff from the NHS Cytogenetics Service within the Institute of Human Genetics the day before the FISH experiment and the remaining slides were stored at 4℃.

Cells stored in a fixative solution (methanol: acetic acid = 3:1) at -20℃ formed a pellet during storage. The fixative was discarded and the cell pellet was resuspended in 6ml of a fresh fixative solution. The cells were centrifuged at 1500rpm for 5 min and the supernatant was then discarded. This procedure was repeated 3 times. A small amount of fresh fixative was added to resuspend the cells. Approximately 25 $\mu$l of cell suspension was dropped onto a clean glass microscopic slide from a height of 3 cm. The slides containing the metaphase spreads were then air-dried. The sum of cells on each slide amounted to approximately $2x10^5$.

### 2.3.3 Preparation of BAC/fosmid DNA

The QIAGEN Plasmid Midi Kit was employed in the preparation of the BAC and fosmid DNA.

#### 2.3.3.1 Growth of bacteria

Bacteria containing BAC DNA were grown on solid LB media containing 12.5 $\mu$ g/ml chloramphenicol at 37℃ overnight. A single colony was picked and inoculated into a starter culture of a 5ml LB medium containing 12.5 $\mu$ g/ml chloramphenicol and grown at 37℃ for 8 hours by vigorous shaking (~250rpm). In addition, 0.5ml of

pre-culture was inoculated into a 100ml LB medium containing 12.5 $\mu$ g/ml chloramphenicol and grown at 37°C for 14 hours by vigorous shaking. The cells were divided into two 50ml tubes and harvested by centrifugation at 4500 x g for 20 min.

### 2.3.3.2 Preparation of DNA

Each bacterial pellet was resuspended in 20 ml of Buffer P1 (50mM Tris.HCl pH8, 10mMEDTA, 100 $\mu$g/ml RNase A). Next, 20 ml of Buffer P2 (200mM NaOH, 1%SDS) was added to each tube, followed by gentle inverting of the tube 4–6 times. The mixture was incubated at room temperature for 5 minutes, before 20 ml of chilled Buffer P3 (3M potassium acetate pH5.5) was added. Each tube was gently inverted 4–6 times, and the mixture was incubated on ice for 15 minutes. After the mixture was centrifuged at =20,000 x g for 30 min at 4°C, the supernatant containing plasmid DNA was removed promptly and filtered through pre-wetted folded filter paper.

A QIAGEN-tip 100 was equilibrated through the application of 4 ml of Buffer QBT (750mM NaCl, 50mM MOPS pH7, 15% isopropanol, 0.15% Triton X-100) and allowed the column to empty by gravity flow. The filtered supernatant was applied to the QIAGEN-tip and allowed it to enter the resin by gravity flow. The QIAGEN-tip was washed with 2 x 10 ml of Buffer QC (1M NaCl, 50mM MOPS pH7, 15% isopropanol), and the DNA was eluted with 5 x 1 ml of Buffer QF (1.25M NaCl, 50mM Tris.HCL pH8.5, 15% isopropanol). To increase the yields, Buffer QF was pre-warmed to 65℃ and applied to the column in 5 aliquots of 1 ml instead of 1 aliquot of 5 ml to prevent cooling of the elution buffer.

The DNA was precipitated by adding 3.5 ml of room-temperature isopropanol to

the eluted DNA. The precipitated DNA was centrifuged immediately at =15,000 x g for 30 min at 4°C, followed by careful decanting of the supernatant. The DNA pellet was washed with 2 ml of room-temperature 70% ethanol and centrifuged at =15,000 x g for 10 min. The supernatant was decanted carefully without disturbing the pellet. The pellet was air-dried for 5–10 min, and redissolved in a suitable volume of buffer (e.g., TE, pH 8.0, or 10 mM Tris·Cl, pH 8.5). The concentration and purification of the dissolved DNA was determined by using the ND-1000 Spectrophotometer (NanoDrop Technologies).

### 2.3.4 DNA labeling with fluorochromes-Direct labeling

The DNA used for FISH probes was labeled with either SpectrumGreen or SpectrumRed by Nick translation (Nick translation kit, Vysis). 1 $\mu$l of DNA was labeled in a 50 $\mu$l reaction containing 5 $\mu$l of the 10x Nick translation buffer (500mM Tris-HCl, pH7.2, 100mM $MgSO_4$, 1mM DTT ), 5 $\mu$l 0.1mM dTTP, 10 $\mu$l dNTP mix, 10 $\mu$l of Nick translation enzyme (DNA polymerase I, DNase I in 50% glycerol, 50mM Tris-HCl pH7.2, 10mM $MgSO_4$, 0.1mM DTT, 0.5 mg/mL nuclease-free BSA), 2.5 $\mu$l of 0.2mM SpectrumGreen or SpectrumRed and sd$H_2O$ to a final volume of 50 $\mu$l. The reaction was incubated at 15℃ for approximately 6 hours for BACs, and approximately 3 hours for fosmids. A 3 $\mu$l of aliquot was then size fractionated on a 1.5% agarose gel. For FISH assays, the goal is to produce labeled probe in a range of 50-600bp, with an average size of 300 bp. If the fragments were still larger than this, the reaction was incubated further at 15℃ until the desired fragments of the correct size had been produced. The reaction was stopped by allowing it to heat in a 70℃ water bath for 10 minutes.

### 2.3.5 Purification of the labeled probe

SpectrumGreen or SpectrumRed labeled DNA was purified in a NICK$^{TM}$ column containing Sephadex G-50 gel to separate the probe from unincorporated nucleotides. The column was washed twice with 3ml of sd$H_2O$. The probe sample was added to the column and 400 $\mu$l of sd$H_2O$ was supplied to allow the probe to enter the gel. Subsequently, the purified probe was eluted in a second fraction of 400 $\mu$l sd$H_2O$ and the fraction was collected in a clean eppindorf.

### 2.3.6 Competition of probe and resuspension in hybridization mix

High copy repetitive sequences in the DNA probe were competed with human Cot-I DNA and salmon sperm DNA. 5 $\mu$l of sheared salmon sperm DNA (10mg/ml) and 10 $\mu$l of CotI DNA (GIBCO 1mg/ml) were added to each eluted 400 $\mu$l purified labeled probe. 1/10$^{th}$ volume of 3M sodium acetate (42 $\mu$l, pH 4.8-7) and 2 volumes (915 $\mu$l) of 100% ethanol were then added to the DNA, permitting it to precipate at -20℃ overnight.

The DNA was recovered by centrifugation at 13000rpm for 20 minutes in a microcentrifuge. The supernatant was subsequently removed and the DNA pellet was washed with 500 $\mu$l of 70% ethanol. The sample was centrifuged at 13000rpm for 10 minutes and the pellet was air-dried for 5 minutes. 16 $\mu$l of the hybridization mix (50% deionised formamide, 2xSSC, 1% Tween-20, 10% dextran sulfate, pH7-7.5) was thereafter added to the pellet, which was allowed to rehydrate at room temperature for 1 hour. The probe was either used immediately for FISH or stored at -20℃. If probes were stored, they were thawed at room temperature for 1 hour before

usage.

### 2.3.7 Commercial pre-labeled chromosome 5, 13, 14 marker probes

The majority of FISH experiments were carried out on metaphase spreads prepared from the blood lymphocytes of patients with t(5;13) and t(14;21) translocations. This allowed us to determine if the probes are mapped centromeric, telomeric or across the translocation breakpoint. Hence, it is very important to accurately determine the exact chromosome on the metaphase spreads. With the help of DAPI (1.5 $\mu$ g/ml 4',6-diamidino-2-phenylindole) counterstaining, it is easy to identify normal chromosomes 13, 14 and 21 since they are acrocentric. It is also easy to identify derivative chromosome 13 and derivative chromosome 21 since they are very small in size (smaller than normal chromosome 21). However, chromosome 5, derivative chromosome 5, and derivative chromosome 14 are not so easily identifiable. To overcome this problem, direct labeled marker 5q, 13q telomeric probes and a chromosome 14, 21 $\alpha$-satellite direct labeled probes were added to the hybridization reactions.

### 2.3.8 *In situ* hybridization to chromosome spreads

Metaphase and prometaphase slides were provided by staff from the MHS Cytogenetics Service of the Institute of Human Genetics.

Slides were incubated in a pre-warmed 2xSSC solution, which contained 100 $\mu$ g/ml of ribonuclease A, for 1 hour at 37℃. After being briefly dipped in a 2xSSC solution, the chromosomes were denatured by incubating slides in a 70 ℃

pre-warmed 2xSSC solution containing 70% of formamide for 4 minutes in a waterbath in a fume hood. The chromosomes were dehydrated by washing the slides for 5 minutes each in a series of ice-cold 70%, 90%, and 100% ethanol solutions. The slides were dried at room temperature and warmed to 37℃ on a hot plate. 20x40mm glass coverslips were cleaned with 100% ethanol and also warmed to 37℃.

The competed probe that was previously dissolved in a hybridization mix was denatured at 75℃ for 5 minutes, snap cooled on ice for 2 minutes and then pre-annealed at 37℃ for 30 minutes. Then, if applicable, 2 $\mu$l of the direct labeled marker probe was added to each probe. After brief centrifugation to collect the probe, 15 $\mu$l of the preannealed probe was applied to the centre of the denatured metaphase spreads. A glass coverslip was carefully lowered to the slide. The slides were placed in a plastic slide holder and put in a Tupperware box containing water-soaked towels. They were then hybridized in an oven at 37℃ overnight.

### 2.3.9 Slide washing and detection of probe signal

After overnight hybridization, the slides were washed three times for 5 minutes in 50% formamide, 2xSSC at 45℃ and then twice for 5 minutes in a 2xSSC solution at 45℃. The coverslips were removed from the slides by gentle shaking. The slides were then dehydrated in a series of room temperature 70%, 85%, and 100% ethanol solutions. Next, they were air-dried and mounted by applying one drop (40 $\mu$l) of Vectashield antifadent containing DAPI (1.5 $\mu$g/ml 4',6-diamidino-2-phenylindole). The chromosomes were stained blue. A clean glass coverslip was lowered to each slide, upon which the slides were stored in the dark at 4℃ until visualization.

### 2.3.10 Microscopic visualization of FISH signals

The slides were analysed using a Zeiss Axioskop epifluorescence microscope equipped with a cooled charge coupled device (CCD) camera. The filter sets within this microscope allow detection of specific fluorescent signals through color visualization. These signals include FITC (green signal), Rhodamine (red signal) and DAPI (blue signal). While FITC has an excitation minimum of 494nm and an emission maximum of 523nm, Rhodamine has an excitation maximum of 570nm and an emission maximum of 590nm. Meanwhile DAPI has an excitation maximum of 345nm and an emission maximum of 425nm. Images were captured at 1000x magnification and analysed using AxioVision 4.1 image analysis software (Zeiss).

## 2.4 RESULTS

### 2.4.1 Phenotypes and Giema chromosome banding of the patient with *de novo* t(5;13)(p13.1;q12.1) translocation.

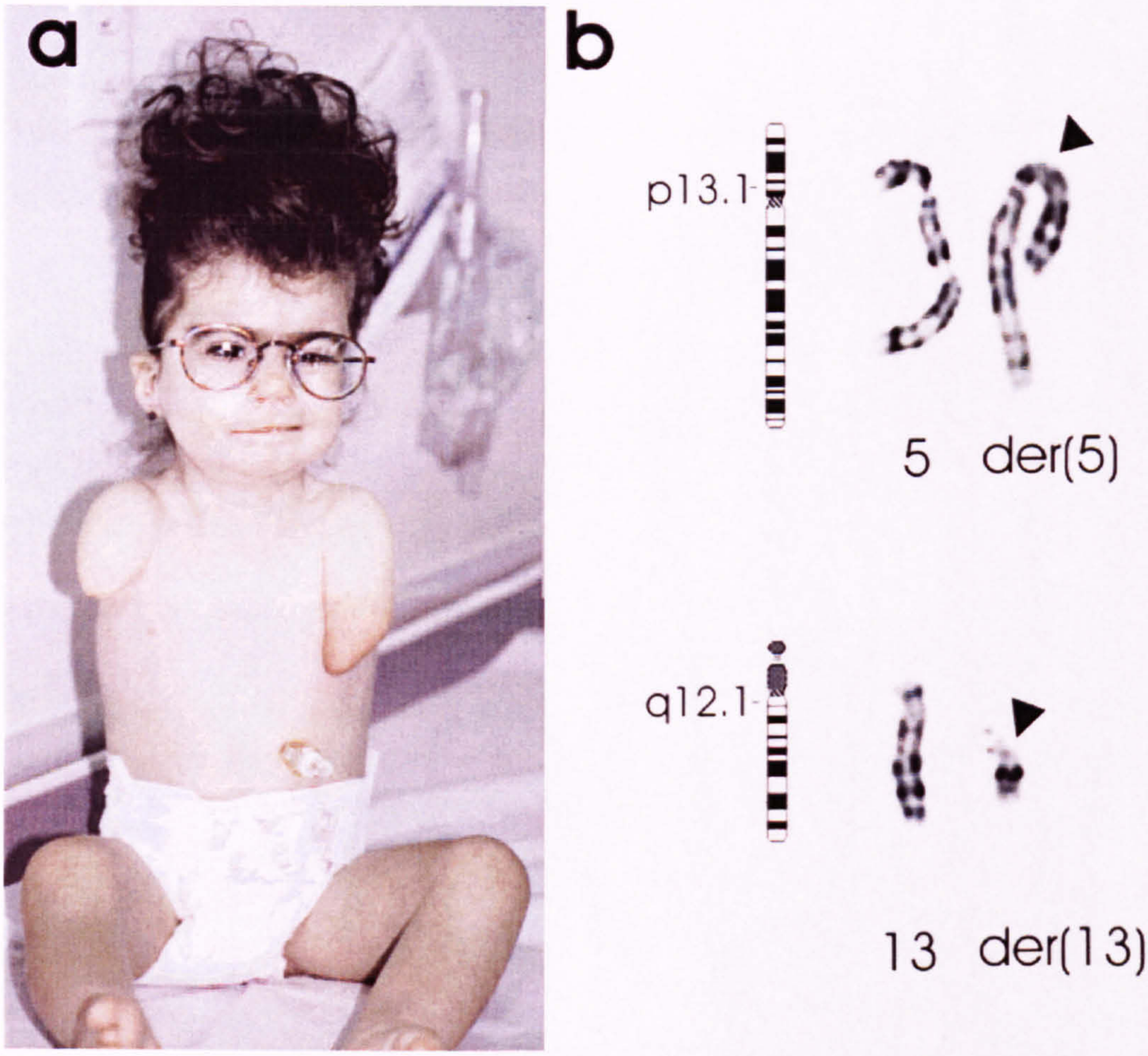



Fig 2.3 (a) Phenotypes of the patient with t(5;13)(p13.1;q12.1) *de novo* translocation. The index case showed typical characteristic facial features and severe limb defects with a generalized growth deficiency (provided with courtesy of Dr. M. Bamshad). (b) Giemsa chromosome banding showed a balanced *de novo* t(5;13)(p13.1;q12.1) translocation (the karyotype was done by Mike Bamshad's team). The chromosome defect was not seen in either parent.

### 2.4.2 Molecular definition of the t(5;13) breakpoint by chromosome FISH

FISH was carried out using a single probe per slide on metaphase spreads from the t(5;13) translocation fibroblasts. Chromosome 5q and 13q telomere marker probes were also used to allow accurate identification of the chromosomes of interest.

For the initial mapping, the clones used as probes were BACs sourced from the RP-11, CTD libraries and selected using the UCSC Genome Browser, Assemblies: July 2003 and May 2004. The selected clones were directly labeled either with florochromes SpectrumRed or SpectrumGreen by Nick™ translation. BACs have advantages over PACs in that they can have slightly larger inserts and are more stable. Therefore, they are significantly less likely to contain chiremic inserts, i.e. inserts from more than one region of the genome.

The aim of the FISH experiments was to determine whether clones were mapped centromeric, telomeric or spanned across the breakpoints on the translocated chromosomes 5, 13, and 14. A clone mapped centromeric or telomeric to the breakpoint will give two signals - one on the normal chromosome and one on the derivative chromosome. Thus, if a probe is mapped centromeric to the translocation breakpoint on chromosome 5, it will result in two signals - one on normal chromosome 5 and one on derivative chromosome 5. If a probe is mapped telomeric to the breakpoint on chromosome 5, it will also result in two signals - one on normal chromosome 5 and the other on derivative chromosome 13. On the other hand, if a probe is spanned across the breakpoint, it will give three signals - one on normal chromosome 5, one on derivative chromosome 5 and the other on derivative chromosome 13. The intensity of the signals on the derivative chromosomes of the

clone that crosses the breakpoint depends on the amounts of the corresponding DNA that are split by the translocation.

### 2.4.3 Results of clones mapped to the 5p translocation breakpoint region

Previous work done by E. Tonkins and M. Smith identified that on chromosome 5p, BAC clones RP11-29K6 and CTD-2039P12 are mapped telomeric to the translocation breakpoint. However, repeated FISH experiments showed that BAC clone RP11- 29K6 mapped centromeric to the translocation breakpoint on 5p13.1. Based on the new results, I selected BAC clones that are located in the interval between RP11-29K6 and CTD-2039P12 as probes to identify the clone that crosses the translocation breakpoint. For the selection of clones, we used the human BAC based physical map generated through the Human Genome Project. Information on BAC clones was obtained from both the ensembl and the NCBI genome browser web sites (http://www.ensembl.org/Homo_sapiens/index.html; http://www.ncbi.nlm.nih.gov/genome/guide/human/;). We restricted our clone selection to those that were readily available in the RPCI-11 and Caltech D1/D2 (informa.bio.caltech.edu/Bac_info.html) libraries.

The results of mapping BAC clones to the 5p13.1 breakpoint region by FISH on t(5;13) metaphase chromosome spreads are shown in Table 2.1 See Figures 2.4, 2.5 and 2.6 for examples of clones mapped telomeric, centromeric and across the chromosome 5p breakpoint, respectively.

Table 2.1 Results of mapping clones to the 5p13.1 breakpoint region by FISH

| | End(bp) | Clone name | Position relative to the breakpoint |
|---|---|---|---|
| 31634946 | 31773870 | CTD2039P12 | Telomeric |
| 33048353 | 33198517 | CTD-2017D15 | Telomeric |
| 34134813 | 34296260 | RP11-110H4 | Telomeric |
| 35164437 | 35322042 | CTD2258D22 | Telomeric |
| 35322159 | 35472737 | CTD-2562I21 | Telomeric |
| 36659329 | 36809502 | CTD-2353F22 | Telomeric |
| 36832297 | 37028067 | CTD-2653M23 | Across |
| 37002744 | 37106095 | CTD-2249G23 | Centromeric |
| 37183020 | 37333544 | CTD-2124B8 | Centromeric |
| 37755520 | 37985498 | CTD-2194L12 | Centromeric |
| 38535555 | 38685966 | RP11-29K6 | Centromeric |

Once the BAC clone that crosses the translocation breakpoint was determined, we queried the genes located within the area covered by the clone CTD-2653M23 through the databases from both the ensembl and the NCBI databases (http://www.ensembl.org/Homo_sapiens/index.html; http://www.ncbi.nlm.nih.gov/genome/guide/human/;).

Within this area, possible candidate genes in the regions were identified, including an uncharacterized gene that had been named as *LOC345513*, which is similar to Keratin, type I cytoskeletal 18 (Cytokeratin 18) (K18) (CK 18) and *IDN3*, a

sequence that corresponded to part of a previously uncharacterized gene . The exact function of these genes was unknown, but *IDN3* was of interest because a predicted protein sequence that could be derived from it exhibits 37% sequence identity to the *Nipped-B* gene product of *Drosophila*, that facilitates enhancer-promoter communication of the remote wing margin enhancer in the *cut* and *Ultrabithorax* homeobox genes, both of which are important in limb development (Rollins et al., 1999). The Drosophila Nipped-B protein is also homologous to a family of chromosomal adherins with broad roles in sister chromatid cohesion, chromosome condensation, and DNA repair. Transcript variants resulting from alternative splicing, and from use of alternative polyadenylation sites, have been reported for this gene; two of these encode different isoforms (Neuwald and Hirano, 2000; Rollins et al., 1999). As growth deficiency exists among essentially all patients with CdLS, and limb defects were observed in around 10-25% of patients with CdLS, we recognized the human homologue of *Drosophila Nipped-B* as a potential candidate gene. The functions of the *Drosophila Nipped-B* gene product suggest that they are cell-cycle related and can also be developmentally regulated, and so the case for involvement of *IDN3* in CdLS was considerable.

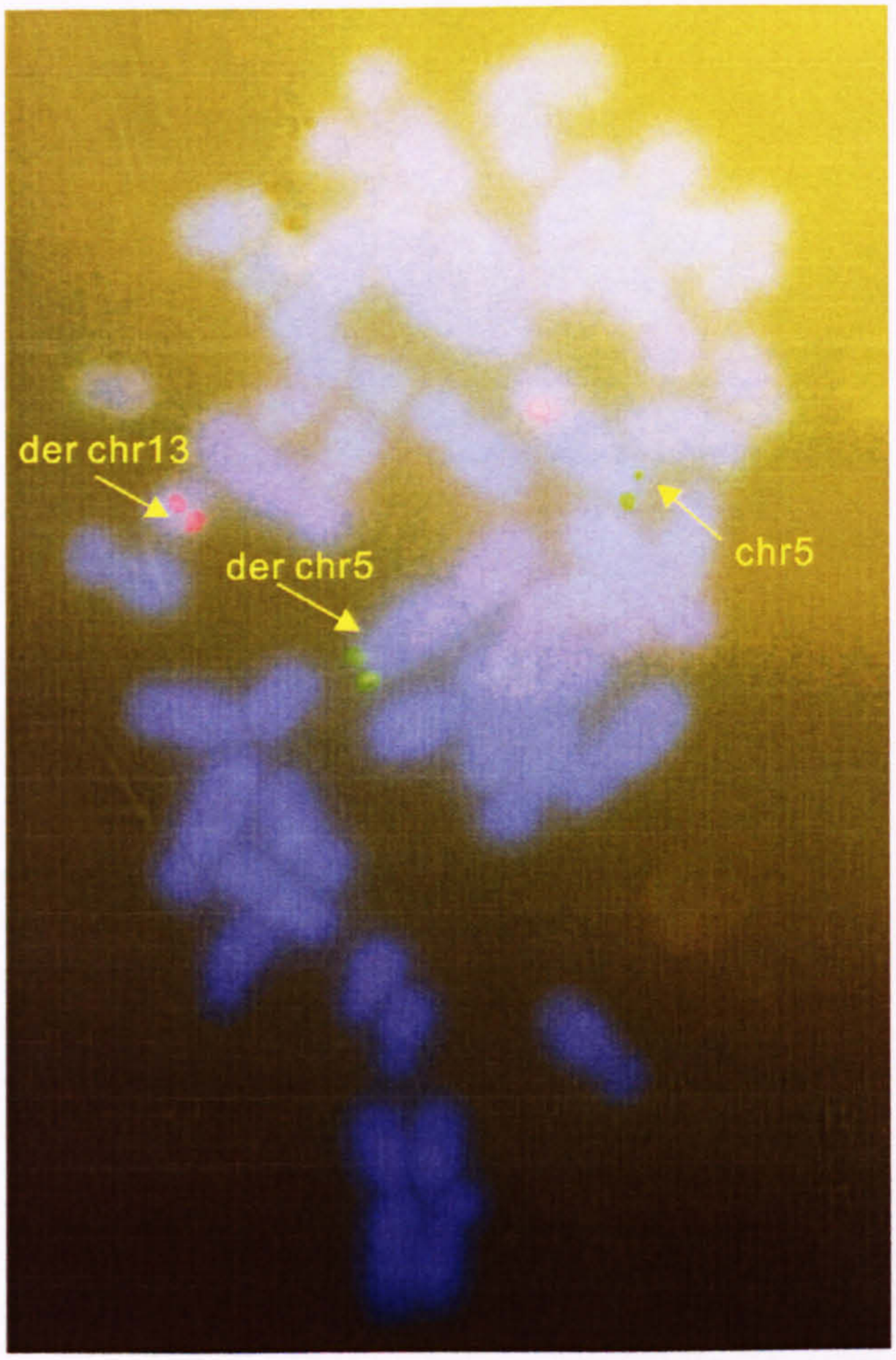



Fig 2.4 A BAC clone mapping telomeric to the 5p13.1 breakpoint (CTD-2353F22). FISH mapping of the chromosome breakpoint in the t(5;13) metaphase spreads. The green signals are from the chromosome 5 telomere marker. The BAC clone CTD-2353F22 that was mapped telomeric to the 5p breakpoint shows red signals on normal chromosome 5 and derivative chromosome 13.

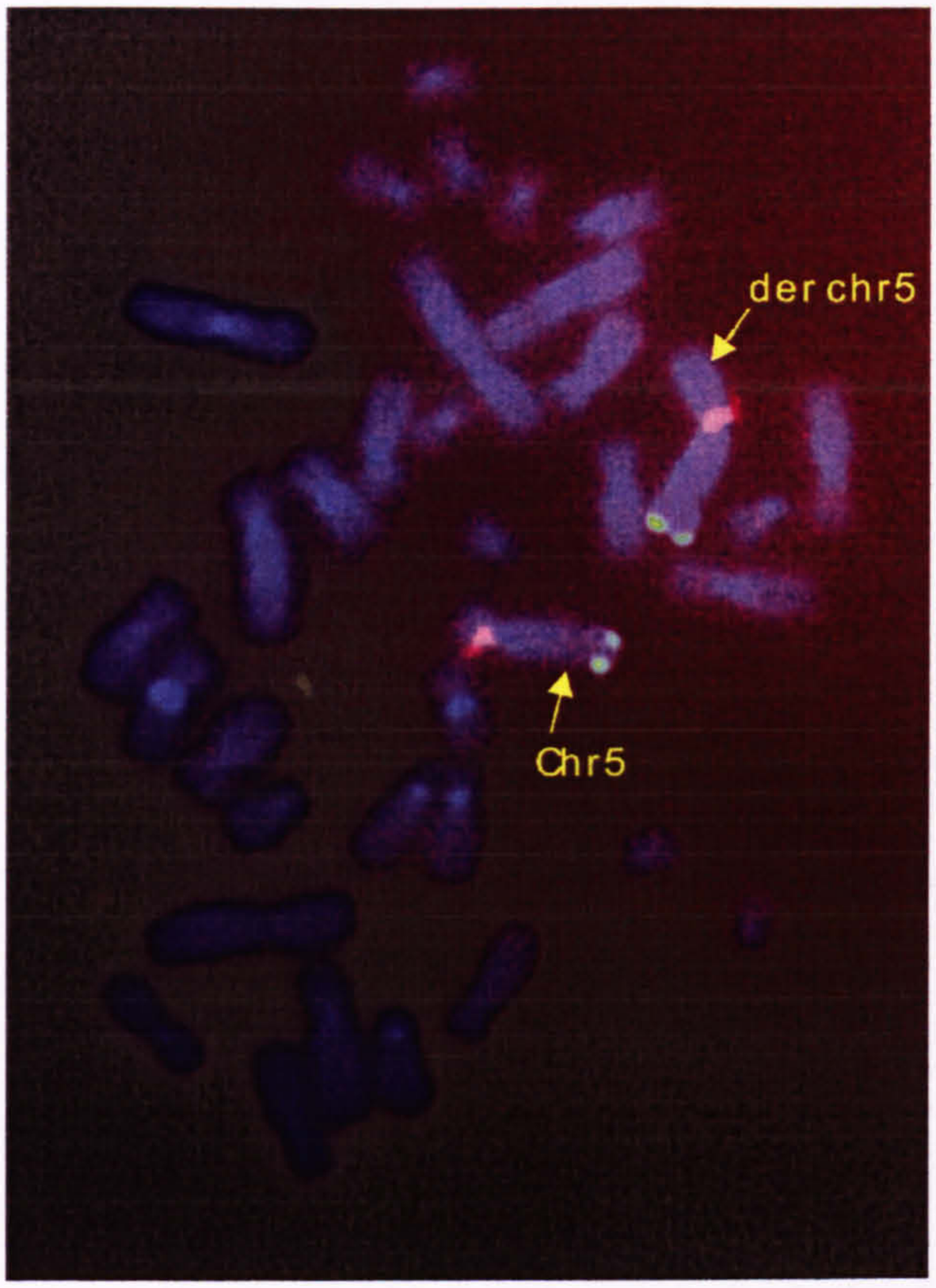



Fig 2.5 A BAC clone mapping centromeric to the 5p13.1 breakpoint (CTD-2249G23). FISH mapping of chromosome breakpoint in t(5;13) metaphase spreads. The green signals are from the chromosome 5 telomere marker. The BAC clone CTD-2249G23 that was mapped centromeric to the 5p breakpoint shows red signals on normal chromosome 5 and derivative chromosome 5.

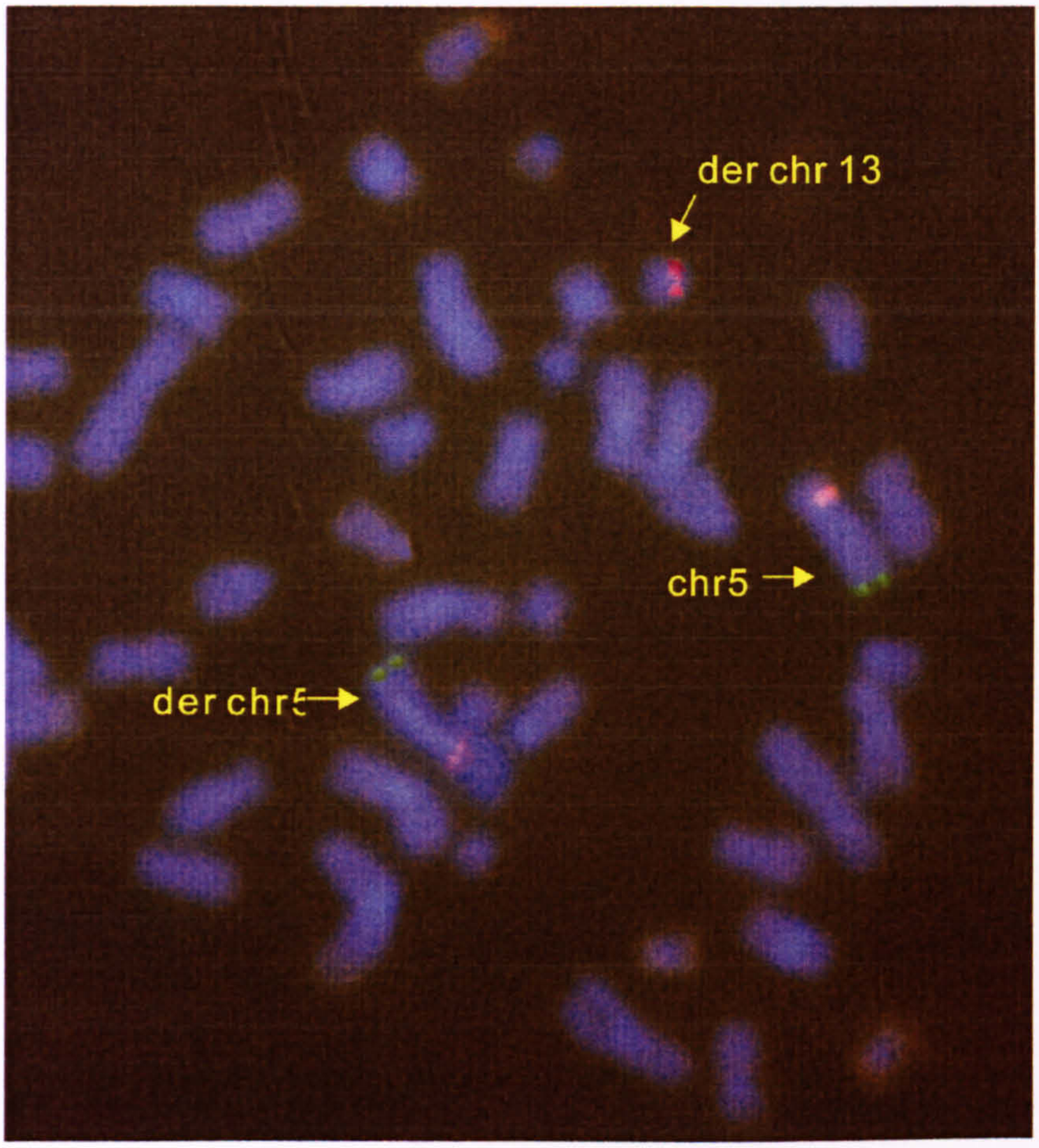



Fig 2.6 A BAC clone spanning the 5p13.1 breakpoint (CTD-2653M23). FISH mapping of chromosome breakpoint in t(5;13) metaphase spreads. The green signals are from the chromosome 5 telomere marker. The BAC clone CTD-2653M23 mapped across the 5p breakpoint showed red signals on normal chromosome 5, derivative chromosome 5 and derivative chromosome 13.The red signal on derivative chromosome 5 is fainter than the one on derivative chromosome 13.

Due to the high probability of 5p13 as a candidate region harboring the CdLS disease gene (discussed in section 2.6), we decided to perform higher resolution mapping by using fosmids as probes. The results of FISH using the BAC clone as a probe highly suggested that the *IDN3* sequence was disrupted by the 5p13 breakpoint in the t(5;13) translocation patient. However, further investigation suggested that the *IDN3* sequence was a gene fragment and that it was part of a larger gene that extended at least 100kb 5' upstream. The 3'end of a novel transcript, *BX 640644(DKFZp686J10142)*, comprising 13 exons, overlapped with the 5' end (up to exon 5) of the original *IDN3*. Further data searches identified a cDNA clone, *AJ627564*, whose 3'end partially overlapped (by 743bp) with the gene fragment *IDN3* isoform A (*AB019494*, 39 exons). In-house sequencing revealed sequence-identity over 743bp between *AJ627564* and gene fragment *AB019494* suggesting that they both were part of a larger transcript, which we called *NIPBL* (*Nipped-B like*). The protein product of *NIPBL* was called delangin. Further analysis extended the previous *IDN3* (39 exons) up to 47 exons as follows. BLAT analysis by using the UCSC genome browser (2003) (http://genome.ucsc.edu/) suggested that the transcript of the cDNA clone *AJ627564* consisted of 10 exons and overlapped with the original *IDN3* gene fragment at exon 9 of *AJ627564* (IMAGE 5784375 as seen in Fig. 2.10). Thus, the total number of exons in *NIPBL* increased up to 47 exons.

Subsequent mutation screening identified a patient-specific mutation in exon 4 of *NIPBL* and together with the previous FISH result suggested that the breakpoint of our patient with t(5:13) was within intron 1 of *NIPBL*. Thus, the full exon-intron organization of the *NIPBL* gene was determined.

To verify that the *IDN3* sequence was a gene fragment, we decided to select

fosmids corresponding to this region (mainly located within intron 1 of the larger gene (now named *NIPBL*) for finer mapping. In this project, I initially selected 6 fosmid clones spanning intron 1 of this gene for FISH mapping. The results are shown in Figure 2.7, 2.8 and 2.9.

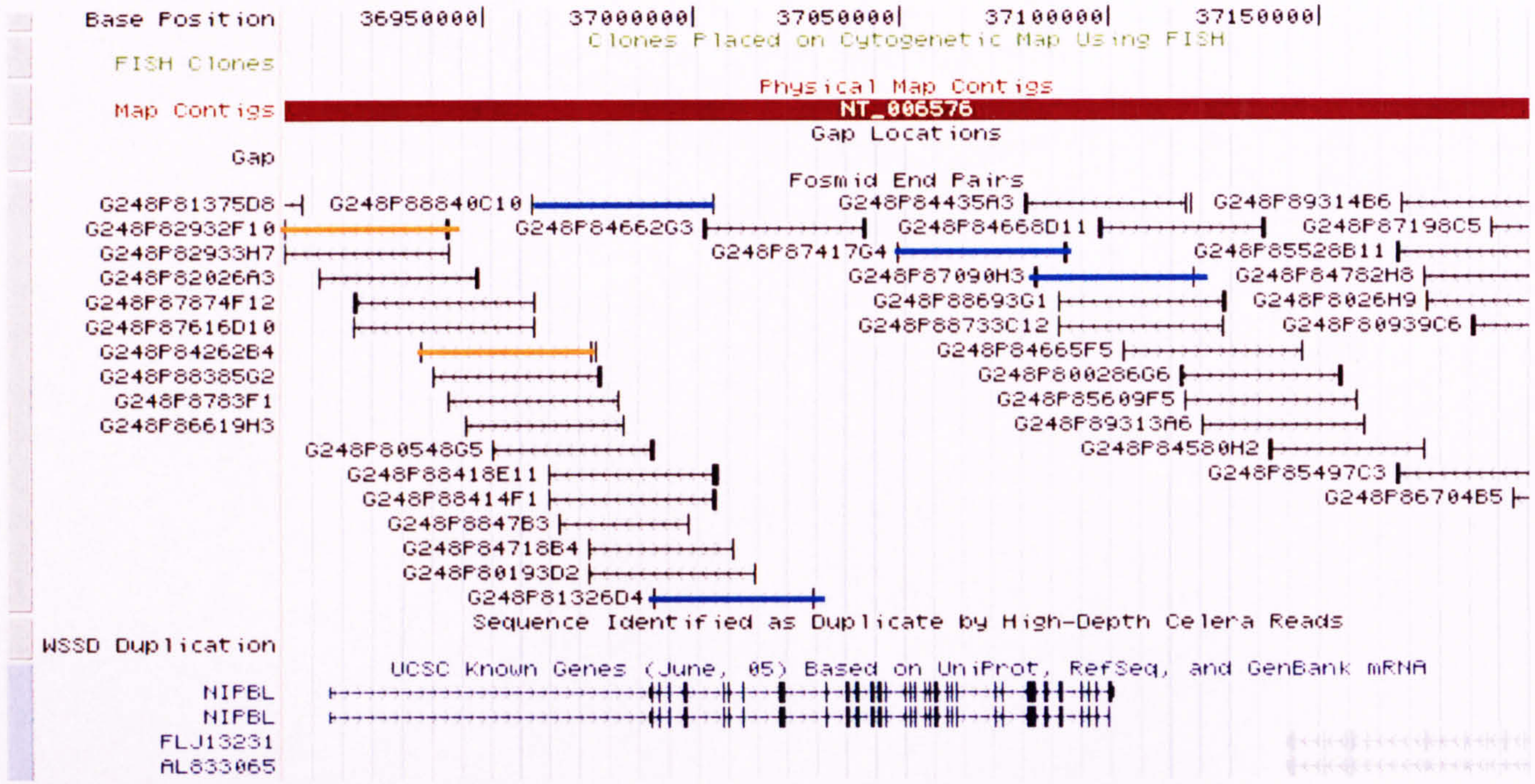



Fig 2.7 Results of FISH mapping. The colored 5p13 fosmid clones against DNA from the t(5;13) (p13.1;q12.1) *de novo* translocation case. Names for fosmid clones appear to the immediate left of the indicated bars. The orange and the light blue bars indicate clones that were mapped telomeric and centromeric to the 5p13 breakpoint, respectively. The most likely location of the breakpoint is close to the region of overlap for insert G248P84262B4 and G248P8840C10, which is located within the large intron I of the gene *NIPBL* (the map contig was adapted from UCSC genome browser, 2004)

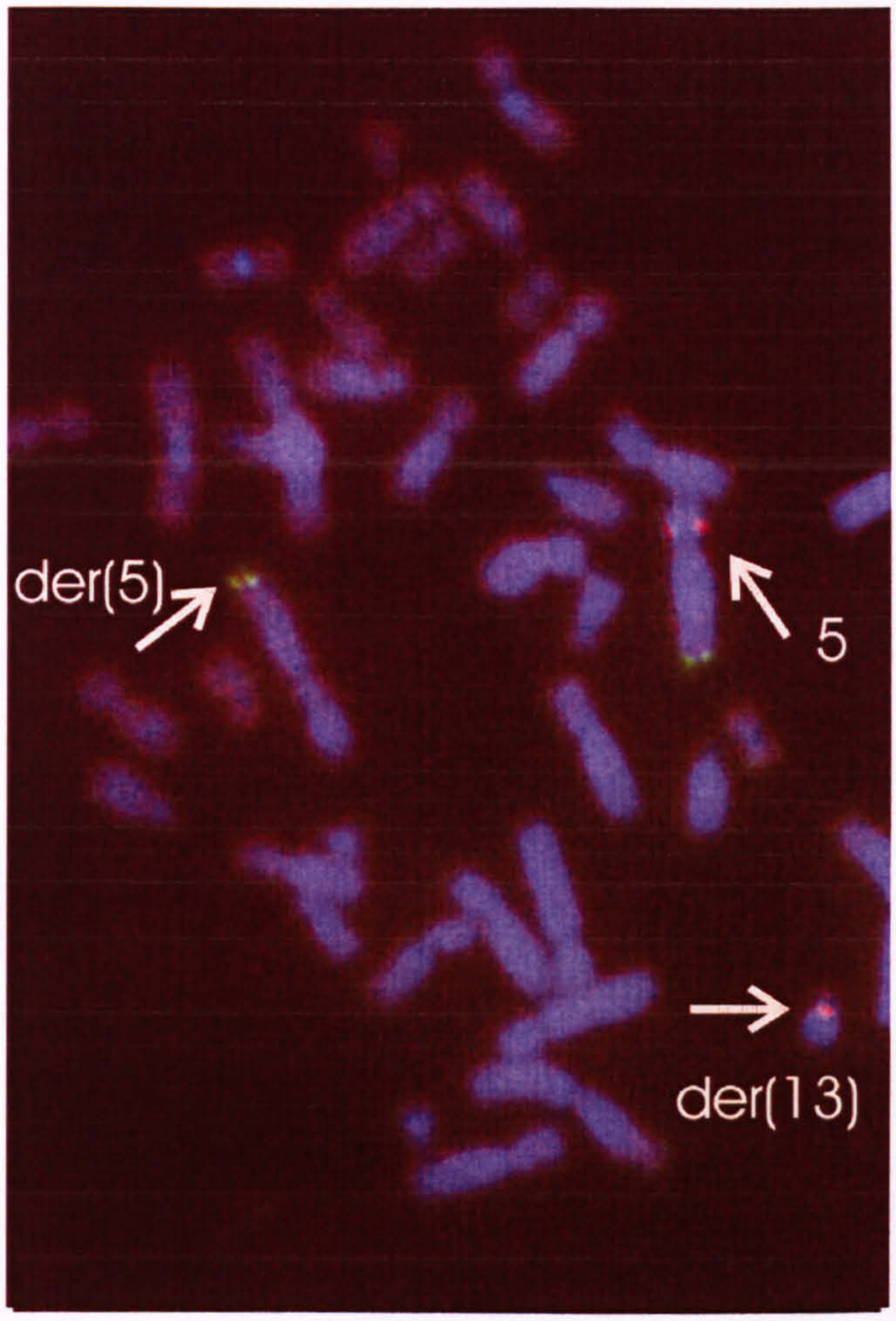



Fig. 2.8 FISH mapping of clone G248P84262B4. FISH mapping of chromosome breakpoint in t(5;13) metaphase spreads. The green signals are from the chromosome 5 telomere marker. Fosmid clone G248P84262B4, which was mapped telomeric to the 5p breakpoint showed red signals on normal chromosome 5 and derivative chromosome 13.

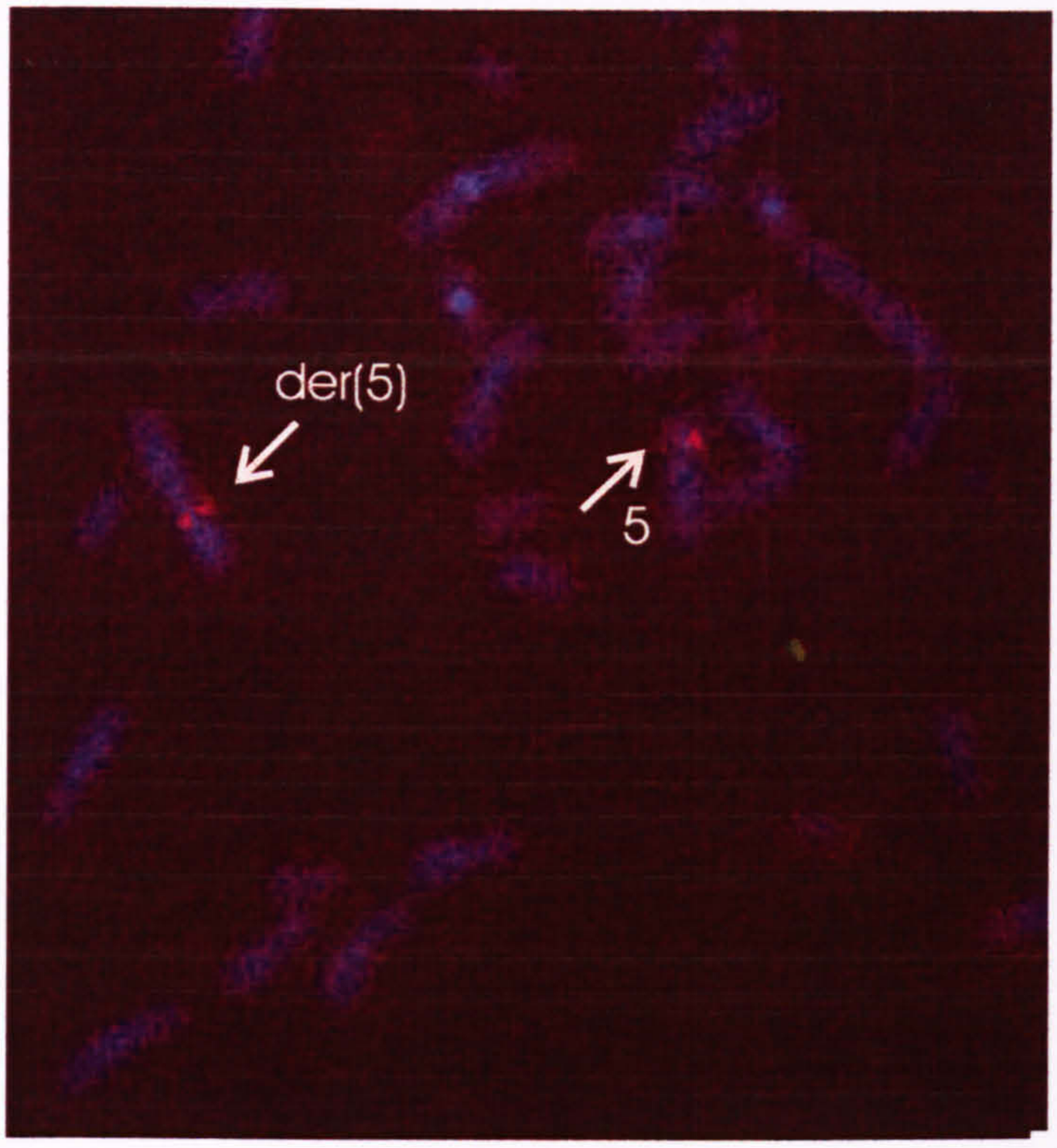



Fig 2.9 FISH mapping of clone G248P8840C10. FISH mapping of chromosome breakpoint in t(5;13) metaphase spreads. The green signals are from the chromosome 5 telomere marker. Fosmid clone G248P8840C10, which was mapped centromeric to the 5p breakpoint showed red signals on normal chromosome 5 and derivative chromosome 5.

### 2.4.4 Hierarchy of the contig map in 5p13.1

Previous FISH mapping enabled us to pinpoint the possible candidates in this 5p13 area to a number of limited genes. We then subsequently analyzed the expression patterns and phenotypes of mutants of other model organisms and sequence conservations to determine the priority gene for mutation screens.

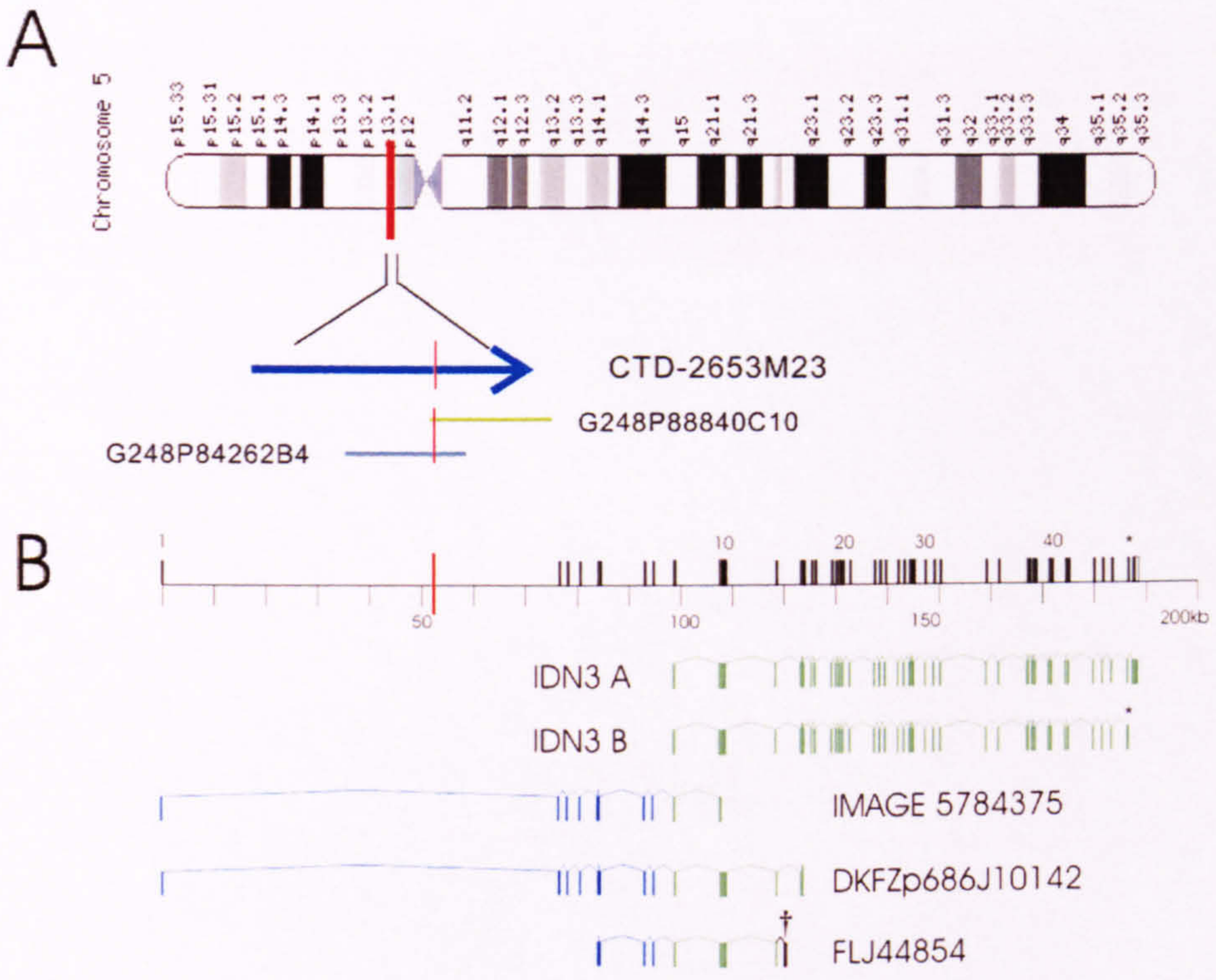



Fig. 2.10 Genomic organization, FISH clones and expressed products of the gene located at the 5p13.1 breakpoint. (A) Ideogram of human chromosome 5 and FISH clones. Red vertical bars indicate the site of disruption at 5p13.1 by the t(5;13) translocation. G248P88840C10 was mapped centromeric to the breakpoint and G248P84262B4 was mapped telomeric to the breakpoint. The breakpoint was located within intron 1 of *NIPBL*. (B) Genomic organization of *NIPBL* and the cognate cDNAs. The linked vertical bars illustrate the exon selection in representative cDNAs used to determine the full exon complement including the original *IDN3 A* and *B* clones. The start coding site of the *NIPBL* is located in exon 2 (adapted from E.T Tonkin, with permission).

*NIPBL* has two isoforms, of which the longer isoform A is 9717bp long and comprises of 47 exons. The protein product of *NIPBL* isoform A is 2804 a.a. The shorter isoform, *NIPBL* isoform B, is 8648bp long and comprises of 46 exons. The protein product of isoform B is 2697 a.a. The differences between isoform A and isoform B is their 3' end. *NIPBL* isoform B uses a different polyadenylation site from that located within intron 46 of *NIBPL* isoform A. The functional differences between these two isoforms are not known.

### 2.4.5 Results of the clones mapped to the translocation breakpoint around the 13q12.1 area

Region 13q12.1, which was trancated by the *de vono* t(5;13) translocation was also studies. Preliminary work done by E.T. Tonkin and M.Smith on haplotyping genetic markers around the breakpoint region showed that marker D13S175 mapped centromeric and D13S120 mapped telomeric to the breakpoint in the 13q12.1 region. The interval between these two markers is around 9.5Mb. Based on these results, we selected the BAC clones as the probes for FISH mapping through the information provided by MapViewer from NCBI (http://www.ncbi.nlm.nih.gov/mapview/).

The results of mapping clones to the 13q12.1 breakpoint region by FISH on t(5;13) metaphase chromosome spreads are shown in Table 2.2. See Figures 2.8 for example of clones mapped across the chromosome 13q breakpoint.

Table 2.2 Results of mapping clones to the 13q12.1 breakpoint region by FISH

| Start (bp) | End (bp) | Clone name | Position relative to the breakpoint |
|---|---|---|---|
| 19023711 | 19043428 | RP11-76K19 | Centromeric |
| 19573840 | 19576369 | RP11-264J4 | Centromeric |
| 20038208 | 20164503 | RP11-172H24 | Across |
| 20647784 | 20650871 | RP11-101P17 | Telomeric |
| 21099552 | 21102154 | RP11-271B5 | Telomeric |
| 24216405 | 24219032 | RP11-534K14 | Telomeric |
| 28478469 | 28671391 | RP11-60N6 | Telomeric |
| 34899905 | 35090902 | RP11-98D3 | Telomeric |

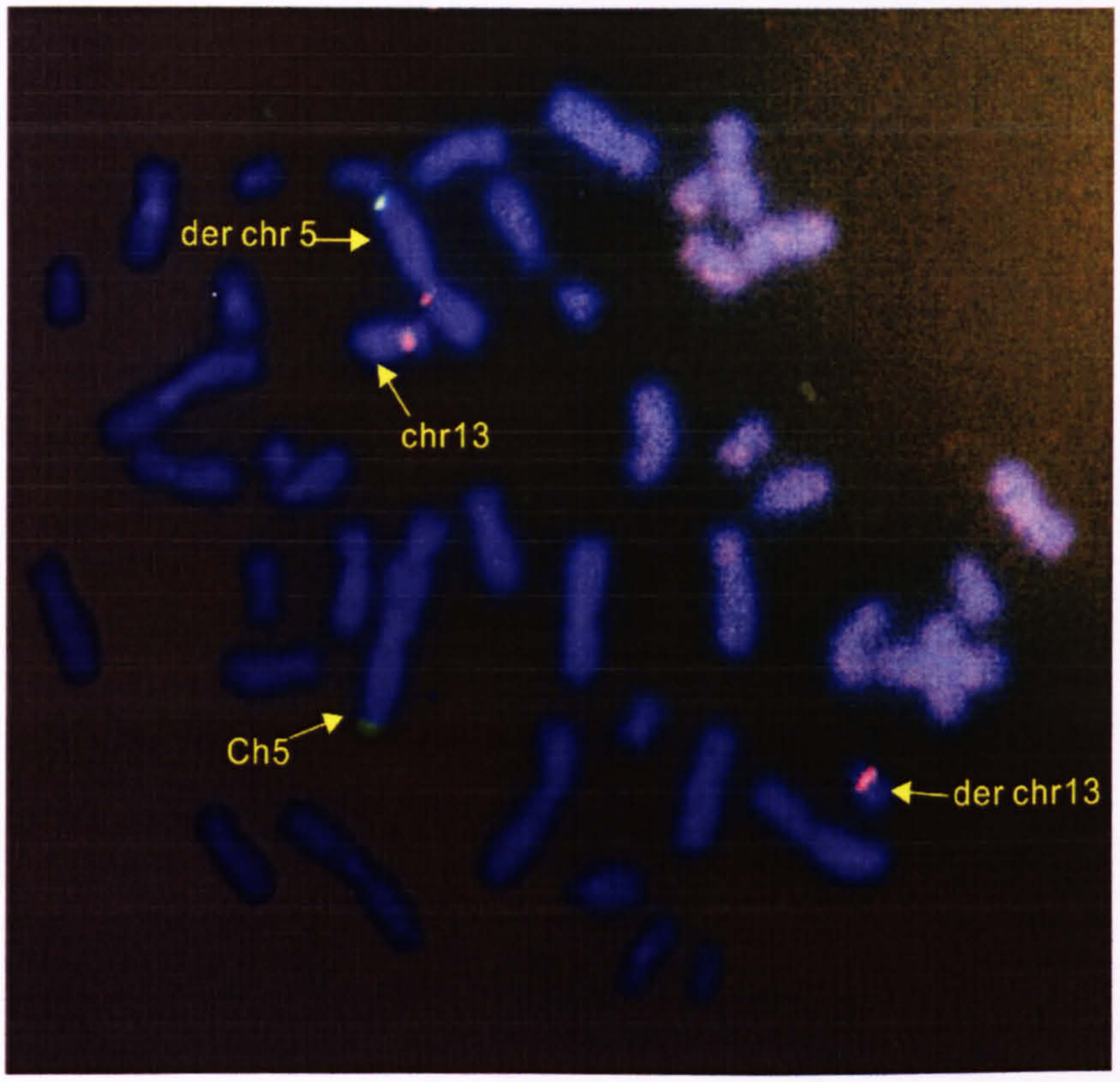



Fig 2.11 A BAC clone that spans the 13q12.1 breakpoint (RP11-172H24). FISH mapping of chromosome breakpoint in t(5;13) metaphase spreads. The green signals are from the chromosome 5 telomere marker. BAC clone RP11-172H24 that was mapped crossing the 13q breakpoint gave red signals on normal chromosome 13, derivative chromosome 5 and derivative chromosome 13. The red signal on derivative chromosome 5 is fainter than the one on derivative chromosome 13.

### 2.4.6 Results of clones mapped to the translocation breakpoint around the 14q32 region

Preliminary work done by E.T. Tonkin and M. Smith found that the marker D14S865 mapped centromeric to the 14q breakpoint. By query on the MapViewer (http://www.ncbi.nlm.nih.gov/mapview) the marker hit the BAC clone RP11-860p12 on the 14q32 region. FISH mapping assigned the clone centromeric to the translocation breakpoint. Moreover, another marker, D14S1227, mapped telomeric to the breakpoint. An inquiry on the MapViewer arrived at the BAC clones RP11-894P9 and RP11-600F2. FISH mapping showed that BAC clone RP11-600F24 is mapped telomeric to the breakpoint on 14q. Based on these results, we selected BAC clones located between these 2 BAC clones suggested by MapViewer as probes to narrow down the breakpoint. Through step-by-step FISH mapping, we showed that RP11-796G6 and RP11-1017G21 mapped most closely to the breakpoint, centromeric and telomeric, respectively.

According to the results from the genome browser, CTD-2017C7 should be crossing the translocation breakpoint in 14q. However, ongoing FISH mapping by using CTD-2017C7 as a probe showed that clone CTD-2017C7 is mapped outside of chromosome 14. It is more likely to be located in chromosome 19. Nevertheless, research on the human genome resource failed to find another BAC clone that encompasses the segment covered by CTD-2017C7. The only clone that we could find is RP11-10e13, which encompasses about 3/4 of the region covered by clone CTD-2017C7 and was mapped telomeric to the breakpoint. Nevertheless, there still existed a gap of 50kb between clones RP11-796G6 and RP11-10e13. To overcome

this problem, we considered to perform cosmid library screening to choose cosmid clones for mapping. At that time, we found a patient-specific mutation in the gene located in the 5p13 translocation breakpoint. Moreover, in this same patient with a t(14;21) balanced translocation, a mutation was found in the *NIPBL* gene (Gillis et al., 2004) and also in the patient with a t(3; 17) *de novo* translocation that we had studied for a long time (by TJ Wang and ET Tonkins). Thus, this region was not considered a potential candidate for the gene responsible for CdLS.

The results of mapping clones to the 14q32 breakpoint region by FISH on t(14;21) metaphase chromosome spreads are shown in Table 2.3. See Figures 2.11 and 2.12 for examples of clones mapped telomeric and centromeric to the chromosome 14q breakpoint, respectively.

**Table 2.3 Results of mapping clones to the 14q32 breakpoint region by FISH**

| Start(bp) | End(bp) | Clone name | Position relative to the breakpoint |
|---|---|---|---|
| 100980568 | 101190341 | RP11-1029J19 | Centromeric |
| 101069362 | 101274469 | RP11-796G6 | Centromeric |
| | | 50kb gap | |
| 101330K | 101510K | RP11-10E13 | Telomeric |
| 101225028 | 101416143 | CTD-2017C7 | Telomeric |
| 101407260 | 101604130 | RP11-1017G21 | Telomeric |
| 101652975 | 101774270 | CTD-2273N11 | Telomeric |
| 102583047 | 102777508 | RP11-736N17 | Telomeric |

Red bars indicate the breakpoint on the corresponding chromosomes

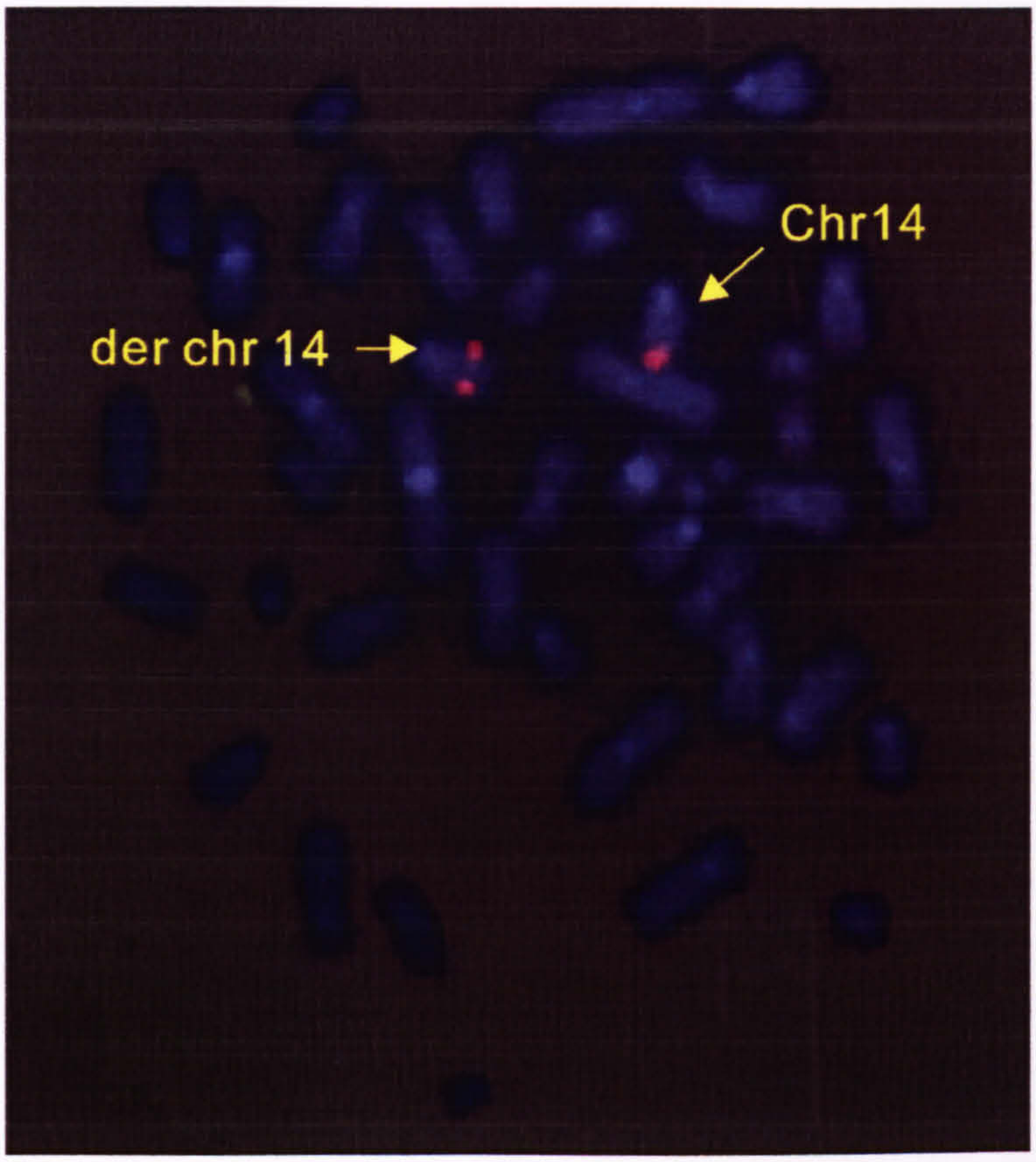



Fig 2.12 A BAC clone mapped centromeric to the 14q32 breakpoint (RP11-796G6).FISH mapping of chromosome breakpoint in the t(14;21) metaphase spreads. BAC clone RP11-796G6 that was mapped centromeric to the 14q breakpoint gave red signals on normal chromosome 14 and derivative chromosome 14, respectively.

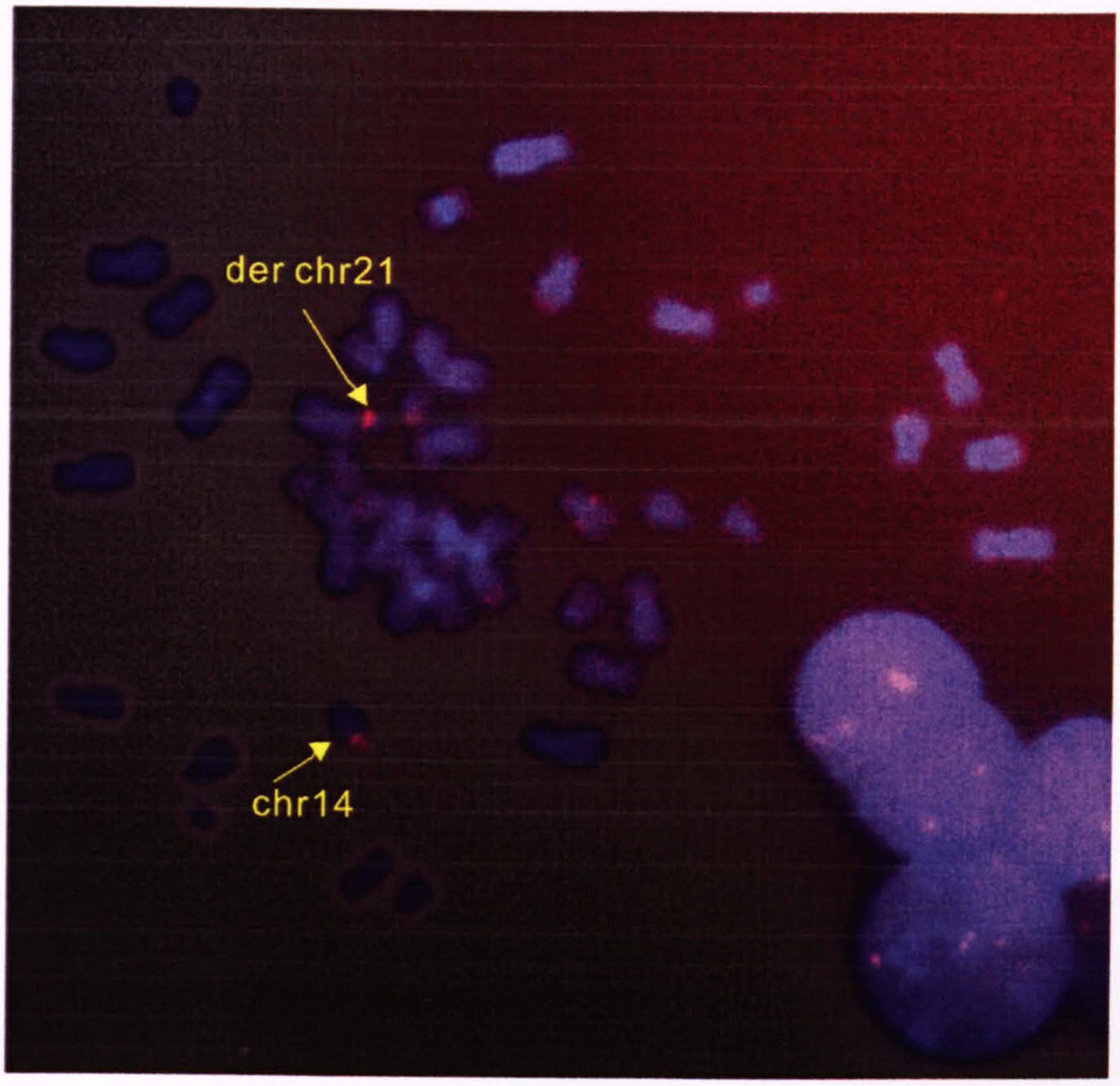



Fig 2.13 A BAC clone mapped telomeric to the 14q32 breakpoint (RP11-10e13). FISH mapping of chromosome breakpoint in t(14;21) metaphase spreads. BAC clone RP11-10e13 that was mapped telomeric to the 14q breakpoint gave red signals on normal chromosome 14 and derivative chromosome 21, respectively.

## 2.5 DISCUSSION

In the beginning, the CdLS phenotypes involving multiple organ systems enabled us to consider it as a microduplication or microdeletion syndrome affecting several genes. However, a whole genome high-density BAC microarray comparative genomic hybridization screen found no evidence for a consistent pattern of microdeletion or micropuplication in a panel of 11 classical CdLS patients (H. Fiegler and N. Carter, personal communication; subsequently published in Tonkin et al., 2004b). Thus, we considered microdeletion or microduplication an unlikely cause of CdLS. After the 3q breakpoint had been exhaustively analysed, other lab members also defined the 17q breakpoint, but no obvious candidate genes mapped in the vicinity of this breakpoint (Tonkin et al, 2004b). These conclusions prompted investigation of the other two *de novo* balanced translocation cases. The 21q breakpoint was expected to map to a gene desert and most attention was based on progressive FISH mapping to identify clones that spanned the 5p, 13q and 14q breakpoints.

### 2.5.1 Power of bioinformatics

With the completion of the Human Genome draft, the choice for the BAC clone used for high-resolution FISH mapping has become more straightforward. Scientists can now save lots of time by spending hours in front of the computer and querying the necessary information through the Internet. (Tilghman, 1996). Once a candidate chromosome region has been pinpointed, candidate genes can be identified from within that region. As the more novel human genes have been identified, the task of identifying candidates has become easier. Once the contig across the candidate region has been established, the next step is to catalog the genes within. Genome browsers

such as Ensembl (http://www.ensembl.org) and the UCSC genome browser (http://genome.cse.ucsc.edu) are frequently used to display and analyze all the definite and possible genes in the candidate region. From the lists of genes that have been mapped to the candidate region, it is possible to look for a gene that has the appropriate expression and/or appropriate function. Furthermore, one can investigate the possibility of significant sequence homology to some other human or nonhuman genes that have similar expressions or functions, and to identify other animal models related to the gene of interest and see if there are mutants with related phenotypes. After gaining a greater understanding of the structural and functional homologies extending across different species, even distantly related species like *Drosophila* or yeast, the top candidates can be shortlisted from multiple candidate genes.

Nevertheless, the data needs to be interpreted with great caution. In general, the genomic DNA sequence is very accurate, but the exons of a gene can be dispersed, complicating the identification of transcripts. Although the cDNA sequence is a good source of transcript information, cDNAs are often incomplete and seldom extend to the 5'end. Expressed sequence tags (ESTs) are single-pass sequencing reads generated from either the 5' or 3' end of a cDNA clone. Due to the paucity of full length cDNA sequences and the abundance of ESTs, several groups have used these sequences to assemble electronic transcripts such as Unigene, STACK and HGI data sets. However, the relative abundance of various types of data varies greatly. The amount of genomic DNA identified as a coding region is limited as is the number of full length cDNA insert sequences. The most abundant source of gene information is the ESTs data. The advantage of using ESTs is that a large number of ESTs derive from different tissue and different developmental and disease states. The disadvantage of using ESTs is that many of them have low-quality sequencing reads. Efforts made by most of the groups

working on the electronic assembly of the ESTs into transcripts are trying to prevent the incorporation of very low quality ESTs and non-gene sequences (Bouck et al., 1999; Miller et al., 1999). However, not all the clustering databases contain the same information and consensus sequences. So in our project, we consulted many database providers and found that the original *IDN3* was a gene fragment, as the first 9 exons are missing in the transcript described both in Ensembl and the NCBI genome browsers. However, with the help of the ESTs databases, we identified at least two clones (*AJ627564 and BX640644*) that were partially overlapping with the original *IDN3* at their 3'end. Also with the in-house sequences of the clones, this allowed us to extend the gene from the original *IDN3 A* to 47 exons (46 exons for *IDN3 B*).

### 2.5.2 Assignment of clones relative to the chromosome breakpoints using standard chromosome FISH

Standard color FISH on metaphase chromosomes from the t(5;13) metaphase spreads was successful in mapping clones relative to the chromosome 5 and chromosome 13 breakpoints. Unfortunately, no BAC clone that spanned the breakpoint in t(14;21) metaphase spreads was identified, but the data suggest that the gap around the translocation breakpoint in the 14q region was narrowed down to 50kb.

### 2.5.3 Clues from the linkage studies

In December 2002 the Krantz group informally communicated the results of their genome wide linkage studies (which had included families supplied by the Newcastle group). Their data indicated that five regions of the genome could not be excluded and

so were candidate CdLS gene regions: 3q13; 8p21.3-p23.1; 10p11.23-p12.31; 14q21.2; and 17p13.2-13.3 (see Table 2.4). However, in July 2003, a member of the this group showed a summary slide of their linkage data at the world CdLS meeting in Sydney, Australia where it was revealed, unexpectedly, that 5p13 was now considered to be one of five likely locations for a CdLS gene. At the same time, FISH mapping of the metaphase spreads in a patient with the t(5;13) translocation showed that a novel gene, at that time named *IDN3,* was likely to be disrupted by the translocation in that patient. Also in 2003 , a CdLS case with deletion involving 5p13.1 was reported at the American Human Genetics Society annual meeting (Hulinsky et al., 2005). The surprising revelation from the linkage data that 5p13 could now be considered a possibility, but especially the report of a fetus resembling CdLS with interstitial del(5)(p13.1p14.2) chromosome derangement prompted targeting of the 5p13.1 area as a top priority. The gene fragment, *IDN3*, which mapped just proximal to the 5p13 breakpoint in the t(5;13) translocation was of particular interest since it is involved in long distance regulation of the *cut* phenotype in *Drosophila* and is evolutionally well conserved.

**Table 2.4 Linkage analysis of genome-scan data in 10 pedigrees (data provided by Dr. Ian Krantz, pers. Comm.)**

| Marker | cM | Lods | alpha | HLods | NPL | p-val | Information |
|---|---|---|---|---|---|---|---|
| **d3s1271** | 117 | -0.96 | 0.71 | 0.93 | 1.33 | 0.09 | 0.90 |
| | 118 | 0.59 | 0.75 | 0.97 | 1.31 | 0.09 | 0.87 |
| | 119 | 1.01 | 0.85 | 1.08 | 1.33 | 0.09 | 0.79 |
| | 120 | 1.22 | 0.96 | 1.22 | 1.35 | 0.09 | 0.75 |
| | 121 | 1.35 | 1.00 | 1.35 | 1.37 | 0.09 | 0.72 |
| | 122 | 1.45 | 1.00 | 1.45 | 1.40 | 0.09 | 0.70 |
| | 123 | 1.51 | 1.00 | 1.51 | 1.42 | 0.08 | 0.70 |
| | **124** | **1.55** | **1.00** | **1.55** | **1.45** | **0.07** | **0.70** |
| | 125 | 1.54 | 1.00 | 1.54 | 1.47 | 0.07 | 0.73 |
| | 126 | 1.49 | 1.00 | 1.49 | 1.50 | 0.07 | 0.76 |
| | 127 | 1.35 | 0.97 | 1.34 | 1.53 | 0.07 | 0.82 |
| **d3s1278** | 128 | 0.88 | 0.78 | 1.15 | 1.57 | 0.06 | 0.90 |

| Marker | cM | Lods | alpha | HLods | NPL | p-val | Information |
|---|---|---|---|---|---|---|---|
| **d8s550** | 16 | -2.22 | 0.72 | 1.23 | 1.68 | 0.05 | 0.90 |
| | 17 | 1.30 | 0.85 | 1.39 | 1.69 | 0.05 | 0.78 |
| | 18 | 1.57 | 0.99 | 1.57 | 1.69 | 0.05 | 0.70 |
| | 19 | 1.72 | 1.00 | 1.71 | 1.69 | 0.05 | 0.64 |
| | 20 | 1.80 | 1.00 | 1.80 | 1.70 | 0.05 | 0.60 |
| | 21 | 1.85 | 1.00 | 1.84 | 1.71 | 0.04 | 0.58 |
| | **22** | **1.86** | **1.00** | **1.86** | **1.72** | **0.04** | **0.56** |
| | 23 | 1.85 | 1.00 | 1.85 | 1.73 | 0.04 | 0.56 |
| | 24 | 1.79 | 1.00 | 1.79 | 1.74 | 0.04 | 0.58 |
| | 25 | 1.68 | 1.00 | 1.68 | 1.76 | 0.04 | 0.60 |
| | 26 | 1.45 | 0.93 | 1.46 | 1.77 | 0.04 | 0.65 |
| **d8s549** | 27 | -1.99 | 0.75 | 1.23 | 1.79 | 0.04 | 0.74 |

| Marker | cM | Lods | alpha | HLods | NPL | p-val | Information |
|---|---|---|---|---|---|---|---|
| **d10s548** | 46 | -0.65 | 0.72 | 0.99 | 1.74 | 0.04 | 0.85 |
| | 47 | 1.55 | 1.00 | 1.54 | 1.75 | 0.04 | 0.82 |
| | 48 | 1.86 | 1.00 | 1.86 | 1.79 | 0.04 | 0.79 |
| | 49 | 1.99 | 1.00 | 1.99 | 1.83 | 0.04 | 0.78 |
| | 50 | 2.06 | 1.00 | 2.06 | 1.87 | 0.03 | 0.78 |
| | 51 | 2.09 | 1.00 | 2.09 | 1.92 | 0.03 | 0.81 |
| **d10s197** | **52** | **2.10** | **1.00** | **2.10** | **1.98** | **0.03** | **0.86** |
| | 53 | 2.06 | 1.00 | 2.06 | 1.94 | 0.03 | 0.82 |
| | 54 | 1.98 | 1.00 | 1.98 | 1.86 | 0.03 | 0.77 |
| | 55 | 1.91 | 1.00 | 1.91 | 1.78 | 0.04 | 0.73 |
| | 56 | 1.82 | 1.00 | 1.82 | 1.70 | 0.05 | 0.72 |
| | 57 | 1.73 | 1.00 | 1.72 | 1.63 | 0.05 | 0.71 |
| | 58 | 1.61 | 1.00 | 1.61 | 1.56 | 0.07 | 0.72 |
| | 59 | 1.47 | 1.00 | 1.47 | 1.49 | 0.07 | 0.75 |
| | 60 | 1.28 | 0.95 | 1.28 | 1.43 | 0.07 | 0.79 |
| | 61 | 0.96 | 0.80 | 1.08 | 1.37 | 0.09 | 0.85 |
| **d10s208** | 62 | -2.40 | 0.68 | 0.93 | 1.32 | 0.09 | 0.96 |

| Marker | cM | Lods | alpha | HLods | NPL | p-val | Information |
|---|---|---|---|---|---|---|---|
| **d14s288** | 40 | -3.87 | 0.33 | 0.13 | 0.61 | 0.27 | 0.88 |
| | 41 | -0.29 | 0.51 | 0.30 | 0.76 | 0.23 | 0.84 |
| | 42 | 0.45 | 0.75 | 0.57 | 0.91 | 0.19 | 0.79 |
| | 43 | 0.87 | 0.93 | 0.87 | 1.06 | 0.15 | 0.76 |
| | 44 | 1.17 | 1.00 | 1.17 | 1.21 | 0.12 | 0.75 |
| | 45 | 1.40 | 1.00 | 1.40 | 1.37 | 0.09 | 0.74 |
| | 46 | 1.59 | 1.00 | 1.59 | 1.52 | 0.07 | 0.75 |
| | 47 | 1.76 | 1.00 | 1.76 | 1.68 | 0.05 | 0.77 |
| **d14s276** | 48 | 1.90 | 1.00 | 1.90 | 1.84 | 0.04 | 0.81 |
| | **49** | **1.91** | **1.00** | **1.91** | **1.80** | **0.04** | **0.75** |
| | 50 | 1.91 | 1.00 | 1.91 | 1.76 | 0.04 | 0.71 |
| | 51 | 1.90 | 1.00 | 1.90 | 1.72 | 0.04 | 0.68 |
| | 52 | 1.89 | 1.00 | 1.89 | 1.68 | 0.05 | 0.66 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 53 | 1.87 | 1.00 | 1.87 | 1.64 | 0.05 | 0.65 |
| | 54 | 1.84 | 1.00 | 1.84 | 1.61 | 0.05 | 0.64 |
| | 55 | 1.81 | 1.00 | 1.81 | 1.57 | 0.06 | 0.64 |
| | 56 | 1.77 | 1.00 | 1.77 | 1.54 | 0.07 | 0.65 |
| | 57 | 1.71 | 1.00 | 1.71 | 1.50 | 0.07 | 0.66 |
| | 58 | 1.63 | 1.00 | 1.63 | 1.47 | 0.07 | 0.68 |
| | 59 | 1.54 | 1.00 | 1.53 | 1.44 | 0.07 | 0.71 |
| | 60 | 1.39 | 1.00 | 1.39 | 1.40 | 0.08 | 0.74 |
| | 61 | 1.16 | 0.88 | 1.19 | 1.37 | 0.09 | 0.79 |
| **d14s63** | 62 | 0.53 | 0.73 | 1.03 | 1.34 | 0.09 | 0.86 |

| **Marker** | **cM** | **Lods** | **alpha** | **HLods** | **NPL** | **p-val** | **Information** |
|---|---|---|---|---|---|---|---|
| **d17s831** | 6 | -1.63 | 0.79 | 1.67 | 2.73 | 0.00 | 0.97 |
| | 7 | 1.90 | 0.96 | 1.90 | 2.67 | 0.00 | 0.90 |
| | 8 | 2.08 | 1.00 | 2.08 | 2.60 | 0.00 | 0.86 |
| | 9 | 2.14 | 1.00 | 2.14 | 2.55 | 0.01 | 0.84 |
| | **10** | **2.12** | **1.00** | **2.12** | **2.49** | **0.01** | **0.82** |
| | 11 | 2.07 | 1.00 | 2.07 | 2.45 | 0.01 | 0.82 |
| | 12 | 1.96 | 1.00 | 1.96 | 2.40 | 0.01 | 0.82 |
| | 13 | 1.79 | 1.00 | 1.79 | 2.36 | 0.01 | 0.83 |
| | 14 | 1.48 | 0.97 | 1.48 | 2.33 | 0.01 | 0.86 |
| **d17s938** | 15 | -1.65 | 0.74 | 1.15 | 2.30 | 0.01 | 0.92 |

**The size of the candidate regions listed above are as follows:**

| Region | Distance spanned | Approximate number of genes contained |
|---|---|---|
| **3q13** | **~10 Mb** | **over 30 genes,** |
| **17p13.2-13.3** | **~6Mb** | **over 53 genes** |
| **8p21.3-p23.1** | **~10Mb** | **over 40 genes** |
| **10p11.23-p12.31** | **~13Mb** | **over 50 genes** |
| **14q21.2** | **~20Mb** | **over 100 genes.** |

### 2.5.4 Data interpretation

As discussed in chapter 1, chromosomal rearrangements can facilitate positional cloning of human disease genes. However, in some instances, the translocation breakpoint occurs outside of the putative gene. In these cases, the malfunction of the gene can be caused by a 'position effect', a deleterious change in the level of gene expression brought about by a change in the position of the gene relative to its normal chromosomal environment, but not associated with an intragenic mutation or deletion. Thus, the transcription unit and minimal promoter of the rearranged gene are expected to remain intact (Kleinjan and van Heyningen, 1998). It is well known that the local megabase-scale chromatin organization can affect gene expression, particularly for genes that are located within or near the heterochromatin (Cook and Karpen, 1994).

Chromosomal rearrangements frequently lead to alteration of the gene's environment and this may be reflected in a change of expression, referred to as a position effect. A number of factors can affect the level of gene expression: (1) the chromosomal rearrangement that separates the promoter/transcription unit from a distant regulatory element, (2) a rearrangement may bring together the gene with an enhancer element from another gene, (3) the translocation can place the gene and its regulatory elements next to a gene located near the translocation. Competition between the disease gene and the second gene for the regulatory elements may result in reduction of the level of gene expression, (4) the rearrangement could give rise to classical position effect variegation *(PEV)* - that is, the rearrangement bring together of an euchromatic gene to a heterochromatic region (Cook and Karpen, 1994; Kleinjan and van Heyningen, 1998).The position effect can stretch as far as 900Kb, as observed in *POU3F4,* a gene involved in X-linked deafness (de Kok et al., 1996). Thus, it is rational for us to examine genes residing within 1Mb on each side of the translocation of the breakpoint.

### 2.5.4.1 13q12.1 region:

#### 2.5.4.1.1 Genes disrupted by the translocation

As shown in the results, RP11-172H24 was mapped across the 13q12.1 breakpoint. Human DNA sequence from clone RP11-172H24 on chromosome 13 Contains the 3' end of the gene for polycystic kidney disease, autosomal recessive *TG737,* now termed *IFT88* (intraflagellar transport 88 homolog, *TTC10* in Fig.2.14) and the 5' end of the *IL17D* gene for interleukin 17D. Mutations in mouse *Tg737* cause random left-right axis specification, polycystic kidney disease, liver and

pancreatic defects, hydrocephalus and skeletal patterning abnormalities such as polydactyly (Taulman et al., 2001). However, the abnormal phenotypes observed in *Tg737* are not observed in CdLS patients and together with the recessive mode of inheritance in *Tg737,* rendering *Tg737* an unlikely candidate gene for de Lange phenotypes. Another gene, *IL-17D*, a member of the interleukin 17 family, is associated with rheumatoid arthritis, lupus, asthma and allograft rejection (Starnes et al., 2002). As the defects of the immune system in CdLS patients are not apparent, the phenotypes associated with the *IL-17D* mutants render this gene an unlikely candidate for de Lange phenotypes.

2.5.4.1.2 Genes residing 1Mb on each side of the 13q12.1 breakpoint

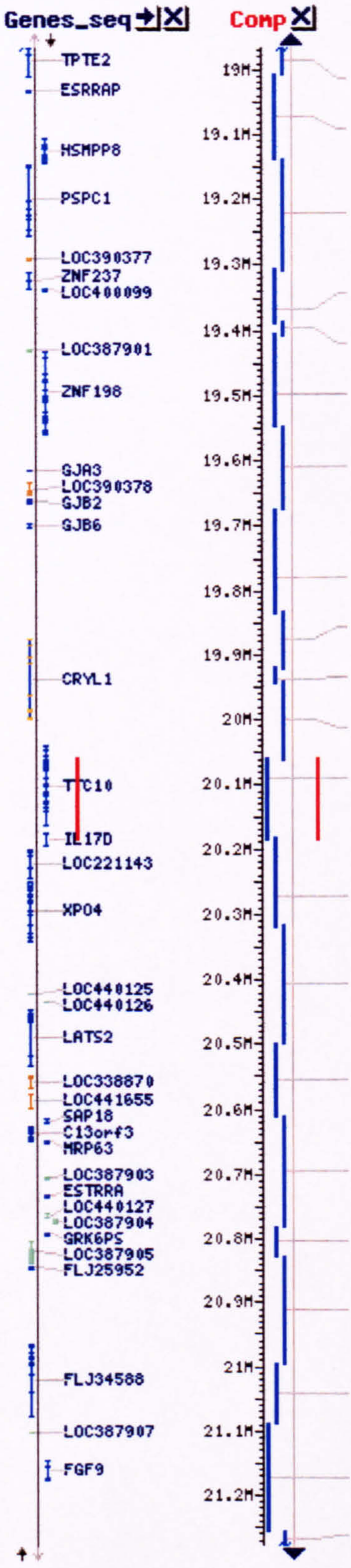



Fig. 2.14 Genes located within 1Mb of each side of the 13q12.1 breakpoint (adapted from NCBI genome browser, Jun 04, 2004).
The red bar indicates site of breakpoint.

**Table 2.5 Possible function of the genes residing within a 2 Mb candidate area spanning a translocation breakpoint at 13q12.1**

| Gene | Function/phenotypes | References |
|---|---|---|
| *FGF9* | Male-to-female sex reversal in mouse model | Colvin et al. (Colvin et al., 2001) |
| *GRK6PS* | G protein-coupled receptor kinase 6 pseudogene | |
| *ESTRRA* | estrogen-related receptor alpha pseudogene | |
| *MRP63* | Helps in protein synthesis within the mitochondria, function not known | |
| *SAP18* | mediated repression of transcription involves the modification of histone polypeptides | Zhang et al. (Zhang et al., 1997) |
| *LATS2* | encodes a putative serine/threonine kinase in *Drosophila*, essential role in the integrity of processes that govern centrosome duplication, maintenance of mitotic fidelity and genomic stability | Yabuta et al. (Yabuta et al., 2000); McPhenrson et al. (McPherson et al., 2004) |

| | | |
|---|---|---|
| *XPO4* | exportin 4, function not known | |
| *IL-17D* | As previously described | |
| *Tg737* | As previously described | |
| *CRYL1* | crystallin, lambda 1, function not known | |
| *GJB6* | CONNEXIN 30, codes for the protein subunits of gap junction channels that mediate direct diffusion of ions and metabolites between the cytoplasm of adjacent cells | Dahl et al. (Dahl et al., 1996) |
| *GJB2* | associated with autosomal recessive neurosensory deafness | Mese et al. (Mese et al., 2004) |
| *GJA3* | Maintains normal lens transparency by providing a cell-cell signaling pathway or structural component for the proper organization of lens membrane and cytoplasmic proteins. | Gong et al. (Gong et al., 1997) |
| *ZNF198* | stem cell | Abruzzo et al. (Abruzzo et |

| | | |
|---|---|---|
| | leukemia/lymphoma (SCLL) syndrome | al., 1992) |
| *ZMYM5* | zinc finger, MYM-type 5, function not known | |
| *PSPC1* | paraspeckle component 1, function not known | |
| *HSMPP8* | M-phase phosphoprotein, mpp8, function not known | |
| *ESRRAP* | estrogen-related receptor alpha pseudogene | |
| *TPIP* | as a phosphatidylinositol 3-phosphatase with no activity against the soluble inositol phosphate | Walker et al. (Walker et al., 2001) |

Two CdLS patients have been described with sex reversal. One was an XX male with the presence of an *SRY* gene, which suggested X-Y interchange, indicating that the sex reversal and the CdLS phenotypes were co-incidental (Chen, 1994). The other case was an XY female without mutations in the HMG box of the SRY gene (N. Affara, per. comm.). Thus we cannot exclude the *FGF 9* gene as a candidate, which might have genetic interaction with the CdLS disease gene.

Other adjacent genes, *SAP18*, *LATS2, TPIP*, which are associated with the basic chromosome functions, could not be excluded as candidates and may be considered

affected by the "position effect" brought by the *de novo* translocation in our patient with a t(5;13) translocation. As described previously, the major characteristic of CdLS phenotypes is general growth deficiency. And the growth deficiency is profound in CdLS with a t(5;13) translocation. Thus, the genes described here might be affected through the "position effect" in this patient. However, this requires further experiments in order to be proven, which can only be undertaken if further patient samples are available.

## 2.5.4.2 Genes residing 1Mb on each side of the 14q32 breakpoint

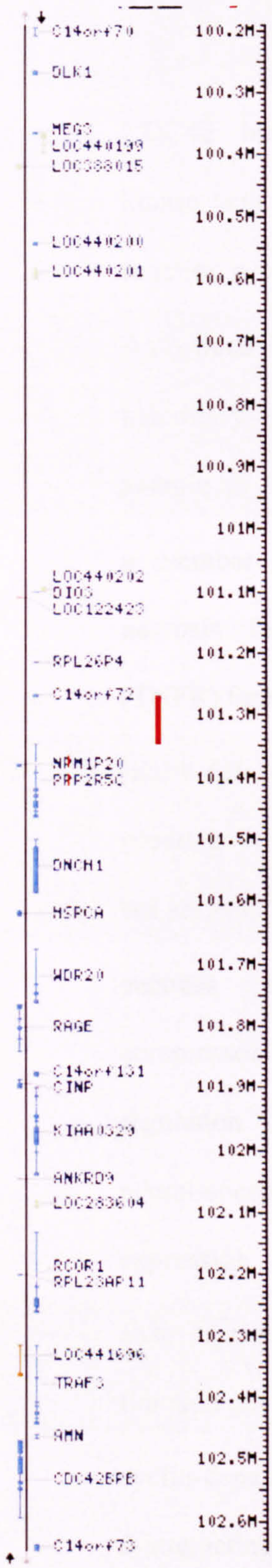



Fig. 2.15 Genes residing within 1Mb of each side of the 14q32 50Kb gap near the breakpoint (adapted from NCBI genome browser, Jun 04, 2004). The red bar indicates site of breakpoint.

**Table 2.6 Possible function of the genes within the 2Mb candidate area (14q32)**

| Gene | Function/phenotype | References |
|---|---|---|
| *CDC42BPB* | CDC42 binding protein kinase beta (DMPK-like), function not known | |
| *AMN* | Autosomal recessive hereditary megaloblastic anemia | Aminoff et al. (Aminoff et al., 1999) |
| *TRAF3* | a member of the tumor necrosis factor receptor (TNFR) family | Hu et al. (Hu et al., 1994) |
| *RPL23AP11* | ribosomal protein L23a pseudogene 11 , function not known | |
| *RCOR1* | encodes a functional corepressor required for regulation of neural-specific gene expression | Lunyak et al. (Lunyak et al., 2002) |
| *ANKRD9* | ankyrin repeat domain 9, function not known | |
| *CINP* | cyclin-dependent kinase 2-interacting protein, function not known | |

| | | |
|---|---|---|
| *RAGE* | Renal tumor antigen | Gaugler et al. (Gaugler et al., 1996) |
| *WDR20* | WD repeat domain 20, function not known | |
| *HSPCA* | reduced activity of Hsp90 induces a heritably altered chromatin state in *Drosophila* | Sollars et al. (Sollars et al., 2003) |
| *DNCH1* | transmitted as autosomal dominant traits and give rise to age-related progressive loss of muscle tone and locomotor ability in heterozygous mice without a major reduction in life span in mouse | Hafezparast et al. (Hafezparast et al., 2003) |
| *PPP2R5C* | Protein phosphatase 2A (PP2A), a heterotrimeric serine/threonine phosphatase, has been implicated in a variety of regulatory processes including cell growth and division, muscle contraction, and gene | McCright et al. (McCright and Virshup, 1995) |

| | | |
|---|---|---|
| | transcription | |
| *NPM1P20* | nucleophosmin 1 (nucleolar phosphoprotein B23, numatrin) pseudogene 20 | |
| *RPL26P4* | ribosomal protein L26 pseudogene 4 | |
| *DIO3* | Thyroxine Deiodinase, Type III, congenital hypothyroidism | Huang et al. (Huang et al., 2000) |
| *MEG3* | Maternally expressed gene 3 | Miyoshi et al. (Miyoshi et al., 2000) |
| *DLK1* | inhibition of adipogenesis | Smas et al.(Smas and Sul, 1993) |

*PPP2R5C* can be considered as a potential candidate as it is involved in the cell growth and division. However, (Gillis et al., 2004) found a nonsense mutation in the CdLS patient with t(14;21) balanced translocation. Thus, the 14q32 qnd 21q11 regions did not have any strong case to be included as regions containing candidate CdLS genes.

2.5.4.3 5p13 region:

For all the candidate regions, we found that the 5p breakpoint had top priority for harboring the CdLS disease gene after report of a CdLS fetus with interstitial

deletion involving 5p13.1 (Hulinsky et al., 2005). The genes residing within 1Mb of the 5p13.1 translocation breakpoint are summarized below:

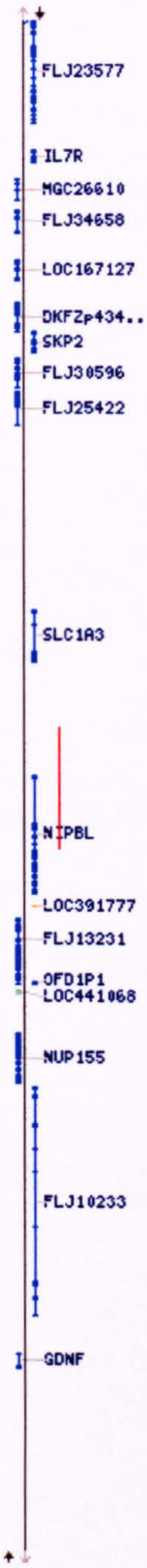



Fig 2.16 Genes residing within the 2Mb-region of the 5p13.1 breakpoint (adapted from NCBI genome browser, 2004)

**Table 2.7 Genes located within the 2Mb candidate region (5p13.1)**

| Gene | Function/phenotype | References |
|---|---|---|
| *GDNF* | promoted the survival and differentiation of dopaminergic neurons in rat embryonic midbrain cell cultures, rescued avian motor neurons from programmed cell death, reduced ectopic discharges within sensory neurons after nerve injury in mouse model, promoted the migration of neural crest stem cells, is associated with human Hirschsprung disease | Lin et al. (Lin et al., 1993); Oppenheim et al. (Oppenheim et al., 1995); Boucher et al.(Boucher et al., 2000); Eketjall et al. (Eketjall and Ibanez, 2002). |
| *NUP155* | involved in bidirectional trafficking of molecules, especially mRNAs and proteins, through the nuclear pores between the nucleus and the cytoplasm in eukaryotic cells | Gorlich et al. (Gorlich and Mattaj, 1996) |
| *OFD1P1* | OFD1 pseudogene 1 | |
| *NIPBL* | Described in the text | |
| *SLC1A3* | reuptake of secreted amino acid neurotransmitter, possibly maintaining extracellular amino acid concentrations at nontoxic and nonepileptogenic levels; related to the syndrome of microcephaly and mental retardation observed in association with interstitial deletion of | Keppen et al. (Keppen et al., 1992); (Shashidharan et al., 1994) |

| | | |
|---|---|---|
| | distal band 5p13 | |
| *SKP2* | CDK2/cyclin A associated protein 45, binds to cyclin F and potentially to other regulatory proteins that may be involved in ubiquitin proteolysis through its F box | Bai et al. (Bai et al., 1996) |
| *CAPSL* | calcyphosine-like , function not known | |
| *IL7R* | encoded a secreted form of the receptor capable of binding IL7 in solution | Goodwin et al.(Goodwin et al., 1990) |

Of all the genes listed above, *IL7R* and *GDNF* were not considered potential candidates since immunodeficiency and the Hirschsprung disease (MIN 142623) are improbable major characteristics for CdLS phenotypes. However, we could not exclude *NUP155*, *SLC1A3*, *SKP2* as potential candidates, that could be inappropriately expressed in the t(5;13) translocation case through a "position effect". As described in chapter 1, most of the CdLS patients have phenotypes of growth retardation, microcephaly and mental retardation. As most of the genes listed above are expected to have either associated cell-cycle functions or related to neuronal function, it was reasonable to consider them as potential candidates. However, it was impossible for us to test such assumptions because we had limited material to study, because the cell line established from the t(5;13) translocation case was a fibroblast cell line.

As the *NIPBL* gene was disrupted by the translocation, we considered it the most compelling CdLS candidate gene within the 5p13.1 2Mb candidate region. This

reasoning was based on: (1) The *NIPBL* gene encodes the homolog of the *Drosophila melanogaster Nipped-B* gene product that is known to be important in regulating various developmentally significant genes, including genes such as *cut* and *Ultrabithorax* that are involved in limb development, and fungal *Scc2*-type sister chromatid cohesion proteins, The CdLS phenotypes involve multiple organs and systems and are assumed to be a result of defect occurring during early human development. The severe growth retardation and limb defects observed in classic patients with CdLS suggest that the disease gene might have functions related to cell proliferations and wide-range developmental regulation. (2) The protein sequences of the *NIPBL* gene product, delangin, are well conserved among different species. The possible association of the CdLS phenotype with the Drosophila mutants prompted us to perform mutation screening in the *NIPBL* gene.

## 2.6 CONCLUSION

Through sequential rounds of chromosome FISH mapping, genes spanning the various breakpoints of the three translocation cases were identified. Of all of these, the gene severed by the 5p13 translocation breakpoint, appeared the strongest CdLS gene candidate.

We considered *NIPBL* the top priority for ongoing mutation due to (1) evidence of linkage to the 5p13.1 region (Krantz et al. world CdLS meeting in Australia, per. comm.); (2) report of the CdLS case with interstitial deletion involving 5p13.1 (ASHG annual meeting, 2003); (3) *NIPBL* being found to be disrupted in our patient with de novo balanced t(5;13) translocation; (4) potential phenotype overlap in *Drosophila* model with CdLS (Rollins et al., 1999); (5) the sequences of the *NIPBL* gene product, delangin, being well conserved among different species (see chapter 3).

Due to the large number of coding exons and the large size of some of the exons (notably exons 9 and 10), mutation screening posed a considerable challenge both in terms of time and expense. As mutation screening in large genes is time consuming and the cost is high, a detailed and thoughtful investigation of the information that databases provide is very important nowadays. We therefore had to select the ideal mutation screening method for our giant *NIPBL* gene, which initiated our mutation screening sessions discussed in further detail in Chapter 3.

## 2.7 REFERENCES:


Abruzzo, L. V., Jaffe, E. S., Cotelingam, J. D., Whang-Peng, J., Del Duca, V., Jr., and Medeiros, L. J. (1992). T-cell lymphoblastic lymphoma with eosinophilia associated with subsequent myeloid malignancy. Am J Surg Pathol *16*, 236-245.

Aminoff, M., Carter, J. E., Chadwick, R. B., Johnson, C., Grasbeck, R., Abdelaal, M. A., Broch, H., Jenner, L. B., Verroust, P. J., Moestrup, S. K., *et al.* (1999). Mutations in CUBN, encoding the intrinsic factor-vitamin B12 receptor, cubilin, cause hereditary megaloblastic anaemia 1. Nat Genet *21*, 309-313.

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J. W., and Elledge, S. J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell *86*, 263-274.

Bentz, M., Plesch, A., Stilgenbauer, S., Dohner, H., and Lichter, P. (1998). Minimal sizes of deletions detected by comparative genomic hybridization. Genes Chromosomes Cancer *21*, 172-175.

Boucher, T. J., Okuse, K., Bennett, D. L., Munson, J. B., Wood, J. N., and McMahon, S. B. (2000). Potent analgesic effects of GDNF in neuropathic pain states. Science *290*, 124-127.

Bouck, J., Yu, W., Gibbs, R., and Worley, K. (1999). Comparison of gene indexing databases. Trends Genet *15*, 159-162.

Carter, N. P., Fiegler, H., and Piper, J. (2002). Comparative analysis of comparative genomic hybridization microarray technologies: report of a workshop sponsored by the Wellcome Trust. Cytometry *49*, 43-48.

Chen, H. (1994). An approach to work-up of dysmorphic patients: clinical, cytogenetic, and molecular aspects. Keio J Med *43*, 98-107.

Cheung, S. W., Shaw, C. A., Yu, W., Li, J., Ou, Z., Patel, A., Yatsenko, S. A., Cooper, M. L., Furman, P., Stankiewicz, P., *et al.* (2005). Development and validation of a CGH microarray for clinical cytogenetic diagnosis. Genet Med *7*, 422-432.

Colvin, J. S., Green, R. P., Schmahl, J., Capel, B., and Ornitz, D. M. (2001). Male-to-female sex reversal in mice lacking fibroblast growth factor 9. Cell *104*, 875-889.

Cook, K. R., and Karpen, G. H. (1994). A rosy future for heterochromatin. Proc Natl Acad Sci U S A *91*, 5219-5221.

Dahl, E., Manthey, D., Chen, Y., Schwarz, H. J., Chang, Y. S., Lalley, P. A., Nicholson, B. J., and Willecke, K. (1996). Molecular cloning and functional expression of mouse connexin-30,a gap junction gene highly expressed in adult brain and skin. J Biol Chem *271*, 17903-17910.

de Kok, Y. J., Vossenaar, E. R., Cremers, C. W., Dahl, N., Laporte, J., Hu, L. J., Lacombe, D., Fischel-Ghodsian, N., Friedman, R. A., Parnes, L. S., *et al.* (1996). Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. Hum Mol Genet *5*, 1229-1235.

Eketjall, S., and Ibanez, C. F. (2002). Functional characterization of mutations in the GDNF gene of patients with Hirschsprung disease. Hum Mol Genet *11*, 325-329.

Gaugler, B., Brouwenstijn, N., Vantomme, V., Szikora, J. P., Van der Spek, C. W., Patard, J. J., Boon, T., Schrier, P., and Van den Eynde, B. J. (1996). A new gene coding for an antigen recognized by autologous cytolytic T lymphocytes on a human renal carcinoma. Immunogenetics *44*, 323-330.

Gillis, L. A., McCallum, J., Kaur, M., DeScipio, C., Yaeger, D., Mariani, A., Kline, A. D., Li, H. H., Devoto, M., Jackson, L. G., and Krantz, I. D. (2004). NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. Am J Hum Genet *75*, 610-623.

Gong, X., Li, E., Klier, G., Huang, Q., Wu, Y., Lei, H., Kumar, N. M., Horwitz, J., and Gilula, N. B. (1997). Disruption of alpha3 connexin gene leads to proteolysis and cataractogenesis in mice. Cell *91*, 833-843.

Goodwin, R. G., Friend, D., Ziegler, S. F., Jerzy, R., Falk, B. A., Gimpel, S., Cosman, D., Dower, S. K., March, C. J., Namen, A. E., and et al. (1990). Cloning of the human and murine interleukin-7 receptors: demonstration of a soluble form and homology to a new receptor superfamily. Cell *60*, 941-951.

Gorlich, D., and Mattaj, I. W. (1996). Nucleocytoplasmic transport. Science *271*, 1513-1518.

Hafezparast, M., Klocke, R., Ruhrberg, C., Marquardt, A., Ahmad-Annuar, A., Bowen, S., Lalli, G., Witherden, A. S., Hummerich, H., Nicholson, S., *et al.* (2003). Mutations in dynein link motor neuron degeneration to defects in retrograde transport. Science *300*, 808-812.

Hu, H. M., O'Rourke, K., Boguski, M. S., and Dixit, V. M. (1994). A novel RING finger protein interacts with the cytoplasmic domain of CD40. J Biol Chem *269*, 30069-30072.

Huang, S. A., Tu, H. M., Harney, J. W., Venihaki, M., Butte, A. J., Kozakewich, H. P., Fishman, S. J., and Larsen, P. R. (2000). Severe hypothyroidism caused by type 3 iodothyronine deiodinase in infantile hemangiomas. N Engl J Med *343*, 185-189.

Hulinsky, R., Byrne, J. L., Lowichik, A., and Viskochil, D. H. (2005). Fetus with interstitial del(5)(p13.1p14.2) diagnosed postnatally with Cornelia de Lange syndrome. Am J Med Genet A.

Keppen, L. D., Gollin, S. M., Edwards, D., Sawyer, J., Wilson, W., and Overhauser, J. (1992). Clinical phenotype and molecular analysis of a three-generation family with an interstitial deletion of the short arm of chromosome 5. Am J Med Genet *44*, 356-360.

Kim, U. J., Shizuya, H., de Jong, P. J., Birren, B., and Simon, M. I. (1992). Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucleic Acids Res *20*, 1083-1085.

Kleinjan, D. J., and van Heyningen, V. (1998). Position effect in human genetic disease. Hum Mol Genet *7*, 1611-1618.

Krantz, I. D., Tonkin, E., Smith, M., Devoto, M., Bottani, A., Simpson, C., Hofreiter, M., Abraham, V., Jukofsky, L., Conti, B. P., *et al.* (2001). Exclusion of linkage to the CDL1 gene region on chromosome 3q26.3 in some familial cases of Cornelia de Lange syndrome. Am J Med Genet *101*, 120-129.

Lin, L. F., Doherty, D. H., Lile, J. D., Bektesh, S., and Collins, F. (1993). GDNF: a

glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. Science *260*, 1130-1132.

Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S. H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., Mandel, G., and Rosenfeld, M. G. (2002). Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. Science *298*, 1747-1752.

McCright, B., and Virshup, D. M. (1995). Identification of a new family of protein phosphatase 2A regulatory subunits. J Biol Chem *270*, 26123-26128.

McPherson, J. P., Tamblyn, L., Elia, A., Migon, E., Shehabeldin, A., Matysiak-Zablocki, E., Lemmers, B., Salmena, L., Hakem, A., Fish, J., *et al.* (2004). Lats2/Kpm is required for embryonic development, proliferation control and genomic integrity. Embo J *23*, 3677-3688.

Mese, G., Londin, E., Mui, R., Brink, P. R., and White, T. W. (2004). Altered gating properties of functional Cx26 mutants associated with recessive non-syndromic hearing loss. Hum Genet *115*, 191-199.

Miller, R. T., Christoffels, A. G., Gopalakrishnan, C., Burke, J., Ptitsyn, A. A., Broveak, T. R., and Hide, W. A. (1999). A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. Genome Res *9*, 1143-1155.

Miyoshi, N., Wagatsuma, H., Wakana, S., Shiroishi, T., Nomura, M., Aisaka, K., Kohda, T., Surani, M. A., Kaneko-Ishino, T., and Ishino, F. (2000). Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q. Genes Cells *5*, 211-220.

Neuwald, A. F., and Hirano, T. (2000). HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. Genome Res *10*, 1445-1452.

Oppenheim, R. W., Houenou, L. J., Johnson, J. E., Lin, L. F., Li, L., Lo, A. C., Newsome, A. L., Prevette, D. M., and Wang, S. (1995). Developing motor neurons rescued from programmed and axotomy-induced cell death by GDNF. Nature *373*, 344-346.

Rollins, R. A., Morcillo, P., and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics *152*, 577-593.

Shashidharan, P., Huntley, G. W., Meyer, T., Morrison, J. H., and Plaitakis, A. (1994). Neuron-specific human glutamate transporter: molecular cloning, characterization and expression in human brain. Brain Res *662*, 245-250.

Smas, C. M., and Sul, H. S. (1993). Pref-1, a protein containing EGF-like repeats, inhibits adipocyte differentiation. Cell *73*, 725-734.

Sollars, V., Lu, X., Xiao, L., Wang, X., Garfinkel, M. D., and Ruden, D. M. (2003). Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. Nat Genet *33*, 70-74.

Starnes, T., Broxmeyer, H. E., Robertson, M. J., and Hromas, R. (2002). Cutting edge: IL-17D, a novel member of the IL-17 family, stimulates cytokine production and inhibits hemopoiesis. J Immunol *169*, 642-646.

Taulman, P. D., Haycraft, C. J., Balkovetz, D. F., and Yoder, B. K. (2001). Polaris, a protein involved in left-right axis patterning, localizes to basal bodies and cilia. Mol Biol Cell *12*, 589-599.

Tilghman, S. M. (1996). Lessons learned, promises kept: a biologist's eye view of the Genome Project. Genome Res *6*, 773-780.

Walker, S. M., Downes, C. P., and Leslie, N. R. (2001). TPIP: a novel phosphoinositide 3-phosphatase. Biochem J *360*, 277-283.

Watson, S. K., deLeeuw, R. J., Ishkanian, A. S., Malloff, C. A., and Lam, W. L. (2004). Methods for high throughput validation of amplified fragment pools of BAC DNA for constructing high resolution CGH arrays. BMC Genomics *5*, 6.

Wiegant, J., Ried, T., Nederlof, P. M., van der Ploeg, M., Tanke, H. J., and Raap, A. K. (1991). In situ hybridization with fluoresceinated DNA. Nucleic Acids Res *19*, 3237-3241.

Wilson, W. G., Kennaugh, J. M., Kugler, J. P., and Wyandt, H. E. (1983). Reciprocal

translocation 14q;21q in a patient with the Brachmann-de Lange syndrome. J Med Genet *20*, 469-471.

Yabuta, N., Fujii, T., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Nishiguchi, H., Endo, Y., Toji, S., Tanaka, H., Nishimune, Y., and Nojima, H. (2000). Structure, expression, and chromosome mapping of LATS2, a mammalian homologue of the Drosophila tumor suppressor gene lats/warts. Genomics *63*, 263-270.

Yokobata, K., Trenchak, B., and de Jong, P. J. (1991). Rescue of unstable cosmids by in vitro packaging. Nucleic Acids Res *19*, 403-404.

Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (1997). Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. Cell *89*, 357-364.

# CHAPTER 3: MUTATION SCREENING OF THE *NIPBL* GENE LOCATED WITHIN 5p13.1

## 3.1 INTRODUCTION –*NIPBL* AS CANDIDATE FOR CDLS

As described in Chapter 2, we considered *NIPBL* as the top priority for mutation screening. *NIPBL* has two isoforms; the longer isoform A consists of 47 exons that specify a 2804 a.a. polypeptide; and the shorter isoform B consists of 46 exons that encode a 2697 a.a. polypeptide. The predicted A and B polypeptide isoforms have a common sequence from a.a. 1 to a.a. 2683 but differ at the extreme C-terminus ends. The transcription start site is in Exon 2. Because of the large number of coding exons and the large size of some of the exons (notably exons 9 and 10), mutation screening posed a considerable challenge both in terms of time and expense.

## 3.2 APPROACHES TO MUTATION SCREENING

The ideal mutation screen method should consist of the following features: (1) ability to scan kilobase lengths (2) a 100% detection rate (3) no false positives or negatives (4) usage of a single step (5) no complex equipment or facilities (6) no harmful reagents (7) no electrophoresis (8) high throughput (9) low elapse and little bench time (10) mutation identification (11) low cost (Cotton, 1997). Hence, we considered DHLPC a very promising method for mutation screening. However, due to the cost and facilities available then, our initial screening method consisted of single strand conformation polymorphism and gel-based heteroduplex analyses for fragments of less than 400bp. Also, direct sequencing was employed for apparent shifts and larger fragments (Petersen et al., 1994). Primers were designed to span the

exon-intron boundary to detect the splice site mutations. Every time a sample demonstrated the presence of a deletion or insertion by the reading of the chromogram from the sequencing results, the PCR product was cloned into vector pGem-T easy vector systems (Promega) and a number of transformants were sequenced to confirm the change. All mutations were re-amplified and re-sequenced to confirm that the change observed was not the result of base mis-incorporation by the Taq polymerase. Parental DNA, where available, was analyzed to confirm that the observed mutation had occurred *de novo* in sporadic cases or in familial cases if the mutation was inherited. For each of the mutations, 200 normal chromosomes were also screened.

### 3.2.1 Single strand conformation polymorphism (SSCP)

Orita et al. (Orita et al., 1989) developed the PCR-SSCP method as a rapid and sensitive method to detect mutations or polymorphisms in DNA. PCR is used to amplify the region of interest. The resultant PCR product is heat denatured and separated as single-stranded molecules by electrophoresis in a non-denaturing polyacrylamide gel. The mutations or polymorphisms can be detected electrophoretically as mobility shifts in non-denaturing gel resulting from a change in the conformation of single stranded DNA. The change in the mobility shift depends on the secondary structure that may be altered by a single nucleotide substitution. The mutations are detected as the appearance of new bands on autoradiograms (radioactive detection), by silver staining of bands or the use of fluorescent PCR primers. These are subsequently detected by an automated DNA sequencer (non-radioactive detection).

Mutation detection for PCR-SSCP is around >80% in a single run for fragments

shorter than 300bp (Hayashi and Yandell, 1993). However, the sensitility for fragments larger than 400bp is decreased. There are two major structural factors responsible for the reduced mutation detection rate in long amplicons. The first is increased contribution from other secondary structure modules and domains in longer fragments, which mask the structural change induced by the mutation. The second is higher frequency of single-nucleotide polymorphisms (SNPs) including common polymorphisms in longer fragments. Since the sensitility for mutation detection is not 100%, the nonappearance of a new band does not prove the absence of a mutation in the analyzed fragments.

### 3.2.2 Heteroduplex Analysis (HET).

Heteroduplexes are generated by heat denaturation and reannealing of a mixture of wild type and mutant DNA molecules. In nondenaturing polyacrylamide gels, homoduplexes and heteroduplexes exhibit distinct electrophoretic mobilities. They are detected by ethidium bromide or silver staining. The optimal fragment length for heteroduplex analyses is between 200 and 600 bps. The detection of mutations mainly depends on the position and type of mismatch. The sensitivity of HET analyses is around 80%. However, through the combination of SSCP, HET, running on the same gel, and adjustments of the temperature, gel composition, ionic strength and additives, the mutation detection rate can be as high as 97.5% (Liechti-Gallati et al., 1999).

### 3.2.3 PCR Direct sequencing

PCR direct sequencing is an efficient method providing more sequence data in limited lab time. Indeed, there are some distinct advantages to this approach. However,

some drawbacks of direct sequencing have to be recognized. They include (1) amplification of the wrong fragments (2) unclear removal of residual PCR primers and unincorporated nucleotides (3) inadequate design of primers for automated sequencing (4) excessively high concentration of DNA. However, direct sequencing is quicker and more accurate and has much less PCR mutations by cloning artifacts compared with the cloned based sequencing methods. Moreover, the cost for the sequencing kits has recently decreased considerably. Thus, direct sequencing has become a mainstream for large scale mutation screens (Rao, 1994).

### 3.2.4 Cloning of PCR product into the pGem-T easy system

We employed the pGEM®-T Easy Vector System as the RCP product cloning system, which is a convenient system for cloning PCR products. The vectors are prepared by adding a 3′ terminal thymidine to both ends. These single 3′-T overhangs at the insertion site greatly improve the efficiency of ligation of a PCR product into the plasmids by preventing recircularization of the vector and providing a compatible overhang for PCR products generated by certain thermostable polymerases (supplier's manual).

The high copy number pGEM®-T Easy Vectors contain T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the a-peptide coding region of the enzyme ß-galactosidase. Insertional inactivation of the a-peptide allows recombinant clones to be directly identified by color screening on indicator plates (supplier's manual). Thus, the recombinants will appear as white colonies and the non-recombinants will appeared as blue colonies. This enabled us to select clones for further sequencing. The pGEM-T Easy vector also contains the origin of replication of

the filamentous phage f1 for the preparation of single-stranded DNA. We presumed that the changes observed in *NIPBL* gene in CdLS patients would be heterozygous as most of the familial cases suggested an autosomal mode of inheritance. Particularly when a deletion or insertion is observed, confirmation of the clones with single-stranded DNA product is very important for interpreting the data.

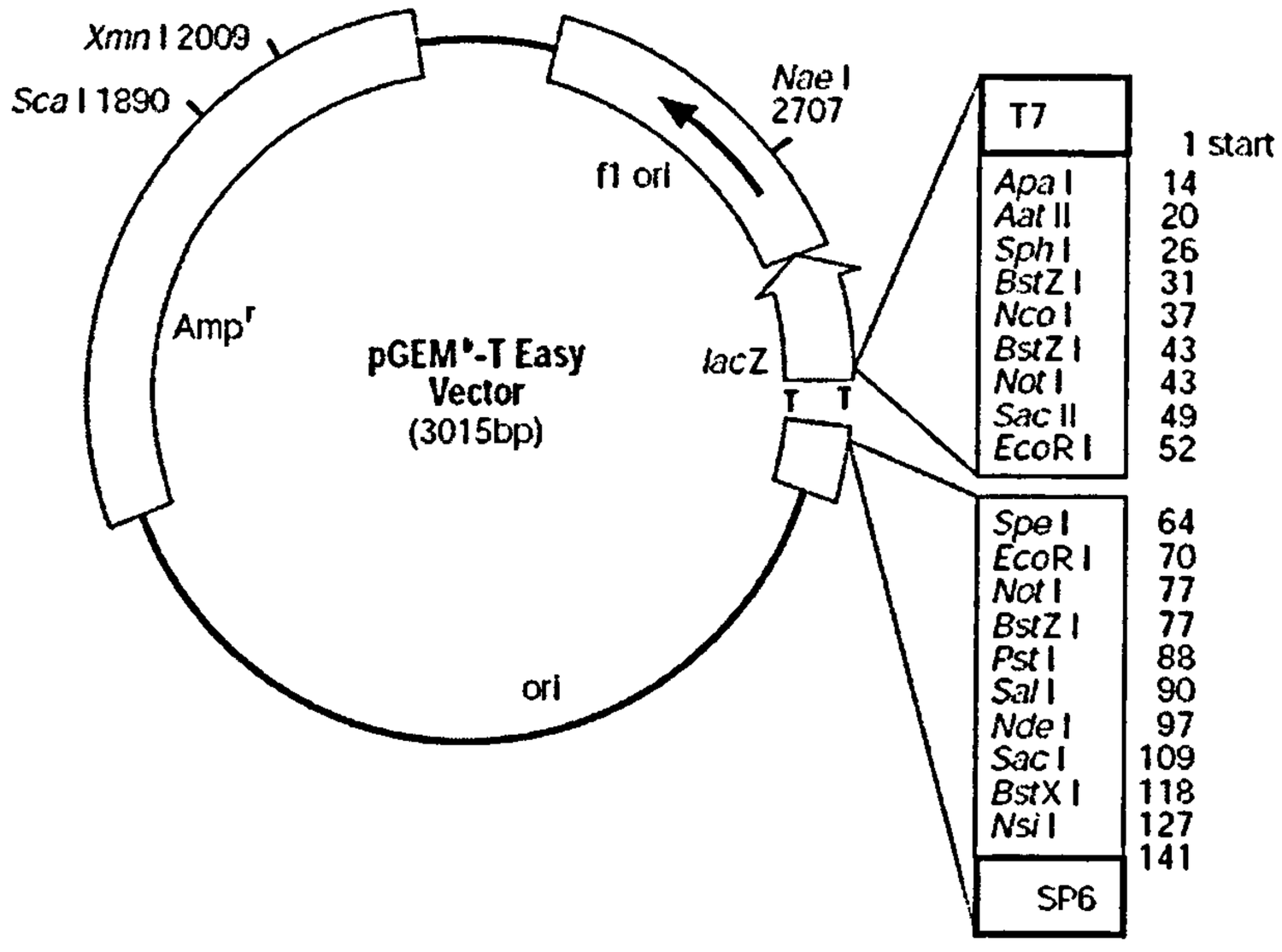



**Fig 3.1** Map of the pGem-T-Easy Vector. The pGEM®-T Easy Vector contain multiple restriction sites within the multiple cloning region. These restriction sites allow for the release of insert by digestion with a single restriction enzyme. The pGEM®-T Easy Vector multiple cloning region is flanked by recognition sites for the restriction enzymes EcoR BstZ I and Not I, thus providing three single-enzyme digestions for release of the insert.

## 3.3 MATERIALS AND METHODS

### 3.3.1 Materials

The sample of CdLS patients selected for the screening consisted of 68 patients that were divided into three groups. These included 63 sporadic and 2 familial cases. The study population also included 2 previously reported, unrelated cases with *de novo* balanced translocation cases involving t(3;17) and t(14;21) respectively. Most of the cases are of Caucasian ethnic origin except for two that were of Arabic origin. All the DNA samples were extracted from blood lymphocytes from the referring centres.

### 3.3.2Amplification of exons from the *NIPBL* gene

PCR primers were designed from the intronic regions flanking the exon-intron boundaries by using the primer designing program *Primer 3* (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3). The amplification reactions were carried out on DNA attained from 45 patients diagnosed with CdLS. We first analyzed exons from the 3' end, since its sequences were more conserved among different species. PCR amplifications for each exon were carried out in a 20 $\mu$l reaction for short exons that allow a single reaction to amplify them. For long exons, sets of primers were designed and amplified in several reactions. For every amplification, a 3 $\mu$l aliquot of each PCR product was analyzed on 1 % agarose gel to ensure a single band before being used for SSCP, heteroduplex, or direct sequencing analyses. However, amplification was hindered by the high quanidine/cytosine content of some exons. To overcome this problem, DMSO (dimethyl sulphoxide) was added to the reaction in the concentration of 7.5%. This helps to maintain the single stranded state

during the elongation reaction and thereby allows the use of more stringent amplification conditions to produce specific products. The sequences of PCR primers for the *NIPBL* gene mutation detection were listed in Appendix A.

### 3.3.3 Single stranded conformation polymorphism (SSCP) and heteroduplex gel mobility shift analyses

#### 3.3.3.1 Preparation of the gel and electophoresis of the samples

SSCP and heteroduplex analyses were carried out simultaneously on a single non-denaturizing polyacryamide gel using the PROTEAN II xi system and PowerPac 3000 power supply. Polyacryamide gels are formed by co-polymerization of polyacryamide and bis-acryamide (N.N'-methylene-bis-acryamide), initiated by TEMED (tetramethylethylenediamine) and APS (ammonium persulphate). Gels used in our experiments contained 1x MDE (Mutation detection enhancement) gel solution and 5% glycerol.

To make two 2x MDE gel for 40 samples per gel tank, 80 ml of gel mixture was prepared by mixing 40ml 2x MDE gel solution, 8ml of 50% glycerol, 4.8ml 10xTBE buffer (0.89M Tris base, 0.89M orthoboric acid, 0.02M EDTA pH 8), 27.2ml $dH_2O$, 480 $\mu$l 10% (w/v) APS and 4.8 $\mu$l TEMED. A 200x200x1.5mm gel was then poured quickly and a comb was inserted to produce the wells. The gel was allowed to set for at least 1 hour before being assembled into clamps for running. Prior to loading the samples, the combs were removed from the gel and rinsed with 0.6xTBE buffer by using a small plastic pastette.

A 5 $\mu$ l aliquot of each PCR product was mixed with 5 $\mu$ l denaturing buffer (95% formamide, 20mMEDTA, 20mM NaOH, 0.025% bromophenol blue, 0.025% xylene cyanol) in a 96 wee microtiter plate. The DNA was denatured by heating at 95℃ for 5 minutes and snap cooled on ice. Each sample was loaded into a separate well of the gel using a long gel-loading pipette tip. The gel plates were sealed onto the gel cooling core and placed in the lower buffer chamber containing approximately 1.2 litres 0.6xTBE buffer. The upper buffer reservoir was filled with 0.6xTBE buffer, the lid attached to the apparatus and then connected to the power supply. The gels were run at 250W, 140V for 1 hour, then at 250W, 300V at 4℃ for more than 16 hours. The duration for the gel running depends on the size of the amplified PCR products and until denaturing loading dye runs to the bottom of the gel plates.

3.3.3.2 Silver staining

After electrophoresis, DNA was visualized on the gel by silver staining. The gel apparatus was dissembled and the gel was carefully transferred to a plastic tray. The gel was fixed by incubating in solution I (10% ethanol, 0.5% acetic acid) twice for 3 minutes each, upon which the solution was discarded. Next, the gel was immersed in solution II (0.1% aqueous silver nitrate) for 15 minutes and rinsed quickly in two changes of distilled water. Solution II can be collected and stored in the dark at 4 ℃, and can also be re-used up to 8 times. After the meticulous rinsing, the gel was then incubated in solution III (aqueous solution of 1.5% sodium hydroxide, 0.15% formaldehyde, freshly prepared before use) for 20 minutes. Solution III was subsequently discarded. Finally, the gel was incubated with solution IV (0.75% aqueous solution of sodium carbonate), upon which solution IV was also discarded. During all the steps, the gel was shaken gently and kept submerged. For long term

storage, each gel was sealed in BIORAD GelAir cellophane support membrane and dried.

### 3.3.4 Direct sequencing of PCR products

#### 3.3.4.1 Enzymatic purification of PCR products

The amplified PCR products for cycle sequencing were purified by enzymatic clean-up method to remove residual oligonucleotides and dNTP's (Kim JB, 2001). For each of 3 $\mu$l aliquot of PCR product, a 2 $\mu$l aliquot of the enzyme mixture containing 0.1$\mu$l of Exonuclease I (20u/$\mu$l, NEB), 0.1$\mu$l of Shrimp Alkaline Phosphatase (1u/$\mu$l, Promega ) and 1.8$\mu$l of buffer (50mM Tris-Cl Buffer, pH 8.0) was added. The mixture was incubated at 37℃ for 20 minutes and then heat inactivated at 80℃ for a further 20 minutes.

#### 3.3.4.2 Cycle sequencing

The system we used for sequences analyses was the MegaBACE 1000 DNA Analysis System. It is a high-throughput, fluorescence-based DNA system utilizing capillary electrophoresis with up to 96 capillaries operating in a parallel manner. The system performs sample injection, gel matrix replacement, DNA separation, detection, and data analysis, offering unparalleled flexibility. The sequencing kit used in our experiments was DYEnamic ET Terminator Kits for MegaBACE.

For direct PCR sequencing, once treated with Exonuclease I and Shrimp Alkaline Phosphoptase, 3-5 $\mu$l of the purified PCR product was mixed with 8 $\mu$l of

DYEnamic ET Terminators, 5 pmol of primer previously used for PCR amplification and distilled water to a total of 20 $\mu$l sequencing reactions in a 96-well plate and mixed thoroughly by gentle vortexing. Then the plate was sealed and placed into the thermal cycler for 35 sequencing cycles. Terminator clean up was performed by propanolol precipitation. The samples were injected into MegaBACE at 2 kV for 80 sec, and separation was carried out at 9 kV for 130 min by the in-house sequencing services in the Institute of Human Genetics.

### 3.3.5 Cloning of PCR product into pGem-T- Easy vector system:

The pGem-T-Easy vector and the control DNA were provided by the supplier and the cloning procedures were under the manufacturer's instructions. Briefly, the ligation reaction mixtures were set up in a 0.5ml tube as described below:

| | Standard reaction | Positive control | Background control |
|---|---|---|---|
| T4 DNA ligase (3u/ul) | 1 $\mu$l | 1 $\mu$l | 1 $\mu$l |
| 10x ligation buffer (300mM Tris-HCl pH7.8, 100mM $MgCl_2$, 100mMDTT, 10mM ATP, 10% polyethylene glycol) | 1 $\mu$l | 1 $\mu$l | 1 $\mu$l |
| pGEM-T Easy Vector (50ng) | 1 $\mu$l | 1 $\mu$l | 1 $\mu$l |
| PCR product | X $\mu$l | - | - |
| Control insert DNA | - | 2 $\mu$l | - |
| Deionized water | 7-X $\mu$l | 5 $\mu$l | 7 $\mu$l |

The amount of insert used was following the formula described below:

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{insert:vector molar ratio} = \text{ng of insert}$$

Most of our PCR product size is around 500bp, so the amount of PCR we used was:

$$\frac{50\text{ng vector} \times 0.5\text{kb insert}}{3.0\text{kb vector}} \times \frac{3}{1} = 25\text{ng insert}$$

The mixtures were mixed by pipetting and were then incubated at 4℃ for overnight.

Meanwhile, the LB plates containing ampicillin (100 $\mu$ g/ml), IPTG(0.5mM) and X-Gal (80 $\mu$ g/ul) were prepared for plating the transformants. Briefly, 15g of agar was added to 1 L of LB medium (10gm trypton, 5gm yeast extract, 5gmNaCl) and autoclaved. The LB medium was then allowed to cool to 50℃ before adding ampicillin, IPTG and X-Gal. 30-35 ml of medium was pour into 85mm petridishes and allowed to set. The plates can be stored at 4℃ to up to one month.

Two LB/ampicillin/IPTG/X-Gal plates were prepared for each ligation reaction. The plates were equilibrated to room temperature before plating. The tubes containing the ligation reaction were centrifuged and 2ul of the ligation reaction was added to a 1.5ml eppendorf on ice. Frozen JM109 High Efficiency Competent Cells were removed from - 80°C storage and placed in an ice bath for about 5 minutes until just thawed. The cells were mixed by gently flicking the tube. 50$\mu$l of cells were carefully transferred into each tube containing the ligation reaction mixtures and mixed gently. The mixtures were then placed on ice for 20 minutes. Next, the cells were

heat-shocked for 45-50 seconds in a water bath at exactly 42°C (do not shake), upon which the tubes were immediately returned to ice for 2 minutes. 950$\mu$l of room temperature SOC medium (2.0g Bacto®-tryptone, 0.5g Bacto®-yeast extract, 1ml 1M NaCl, 0.25ml 1M KCl 1ml, 2M Mg2+ stock, and filtersterilized 1ml 2M glucose/100ml) was added to the tubes containing cells transformed with ligation reactions and 900$\mu$l to the tube containing cells transformed with uncut plasmid. The cells were then incubated for 1.5 hours at 37°C by shaking (~200rpm). 100$\mu$l of each transformation culture was plated onto duplicate LB/ampicillin/IPTG/X-Gal plates. A 1:10 dilution with SOC medium was also plated for the transformation control. To obtain a higher number of colonies, the cells were pelleted by centrifugation at 1,000 × g for 10 minutes and resuspended in 200$\mu$l of SOC medium. 100$\mu$l of the resuspended cells were plated on each of the two plates. The plates were incubated overnight (16–24 hours) at 37°C.

White colonies were picked up and inoculated into 5ml of LB medium containing ampicillin and incubated overnight at 37℃.

### 3.3.6 DNA preparation of the transformants by using QIAprep kit

The overnight cultures were centrifuged at 1000rpm for 15 minutes. The pelelted bacterial cells was resuspended in 250 $\mu$l Buffer P1 (50mM Tris.HCl pH8, 10mMEDTA, 100 $\mu$g/ml RNase A) and transferred to a microcentrifuge tube. 250 $\mu$l Buffer P2 (200mM NaOH, 1%SDS) was then added and mixed thoroughly by inverting the tube 4–6 times. 350 $\mu$l Buffer N3 (3M potassium acetate pH5.5) was added and mixed immediately and thoroughly by inverting the tube 4–6 times. The suspension was subsequently centrifuge for 10 min at 13,000 rpm (~17,900 x g) in a table-top microcentrifuge. The supernatants were applied to the QIAprep spin column

by decanting or pipetting and then centrifuged for 30-60 seconds. The flow-through was discarded. The QIA prep spin column was washed by adding 0.75ml buffer PE and centrifuged for 60 seconds. The flow-through was discarded and the spin column was again centrifuged for an additional 60 seconds to remove residual wash buffer. The QIAprep column was placed in a clean 1.5ml microcentrifuge tube. 50 $\mu$l Buffer EB (10 mM Tris•Cl, pH 8.5) was added to the centre of each QIAprep column to elute the DNA. The column was allowed to stand for 60 seconds and was then centrifuged for 60 seconds. The miniprepped DNA was then electrophresed on the 1.5% agarose gel to determine the presence of insert.

### 3.3.7 Sequences of the cloned product

The pGEM-T Easy vector contains M13 site. M13 forward primer was used as a sequencing primer for sequencing reaction. The procedures for the sequencing were similar to that of direct sequencing. However, the primer used was M13 forward primer and the amount of DNA was around 350 $\mu$g (according to the instructions from the supplier).

### 3.3.8 Assessing the sequencing results

The software we used for the sequences analysis is sequencer, version 4.1 (Genecodes).

## 3.4 RESULTS

### 3.4.1 Study population

A total of 68 CdLS patient DNA samples were screened. They are divided into three groups. The first group consisted of 45 patient DNA samples for initial screening, while the second group included the previously reported two *de novo* translocated cases of CdLS involving t(3;17) and t(14;21) respectively to exclude the regions disrupted by the 3q, 17q, 14q and 21q breakpoints as the candidate regions for CdLS. The third group comprised of 20 patient DNA samples from Denmark and the UK, which were screened for the confirmation of the diagnosis. Of the third group, 17 samples were from Denmark and 3 samples were from the UK and Ireland. One familial case (mother and daughter both affected) from Ireland and a sporadic case of Arabic origin were included in the third group. A small number of patients' DNA failed to amplify with any primers, so the actual number of patients tested was less than sixty eight due to the poor quality of the DNA samples.

### 3.4.2 Mutation screening approach used in exons of different sizes

Fragments flanking the exon-intron boundaries of Exons 2-8 were screened by using SSCP and heterduplex analysis due to the smaller fragment size. If a shifted band was found, subsequent direct sequencing was performed. Due to the large fragment size (most of them are more then 400bp), we employed direct PCR sequencing for exons 15-47. The PCR products were sequenced bidirectionally for a better confirmation.

### 3.4.3 Spectrum of *NIPBL* mutations detected

The complete list of mutations identified is shown in Table 3.1. For the first group, a total of 35 exons were screened. The coding sequence sampled in these 35 exons corresponds to around 40% of the total. 2/3 of exon-intron junctions were screened for a splice site mutation. *NIPBL* mutations were identified in 11 patients with CdLS (1 familial and 9 sporadic). A total of 10 different mutations were identified and comprised 2 frameshift, 1 nonsense, 3 missense, 1 non-frame shifting deletion and 3 splice site mutations. All identified mutations were unique.

In the second group, a total of 39 exons were screened. A *NIPBL* mutation was found in the patient with a *de novo* t(3;17) (Ireland et al., 1991), where the 4bps (TTTG) deletion in exon 23 that led to premature truncation of *NIPBL* was identified. In the other CdLS patient with *de novo* t(14;21) translocation (Wilson et al., 1983), the nonsense mutation S1459X in exon 20 of *NIPBL* was identified by Gillis et. al. (Gillis et al., 2004) and confirmed in the present study. Neither of the mutations was present in either of the respective parents, which suggested that the *de novo* balanced translocation may represent events that were not directly related to the pathogenesis, but possibly could be downstream consequences of the *NIPBL* mutations. The sample of the index case with the de novo t(5;13) translocation we used for FISH mapping in this study was found to have disruption in intron 1 (see Chapter 2)

In the third group, a total of 41 exons were screened. The coding sequence sampled in these 41 exons corresponds to around 84% of the total. 41 of the exon-intron juctions were screened for splice-site mutation. *NIPBL* mutations were identified in 4 patients with CdLS (1 familial and 2 sporadic). A total of three different

mutations were identified and comprised a 3bp non-shift deletion, 1 nonsense mutation and 1 missense mutation.

In 14 of the 63 individuals with sporadic CdLS, 13 different mutations were identified. Of these, 2 were small deletions, which resulted in a deletion of 1 amino acid. They included I1206del (exon 14) and N2217del, which are evolutionally well conserved. One small deletion which led to a 4 bp deletion in exon 23 was found in the patient with a t(3;17) balanced translocation and resulted in a prematurely truncated protein product. Two were small insertions which resulted in a prematurely truncated protein product. Three of the mutations had a single-base-pair change that led to immediate stop codons. One mutation contained a major translocation that disrupted the gene in intron 1.

Two patients with sporadic CdLS had different mutations that affect the splice sites and were predicted to lead to alterations in splicing. The splice-site alterations were not found in their parental samples and in the 200 normal chromosomes.

Of the 13 mutations identified in patients with sporadic CdLS, three were single base change that resulted in substitution of a single amino acid. The missense mutations identified included C1311R (exon 17), L1348R (exon 17) and Y2430C (exon 43). These amino acids were highly conserved throughout evolution and such changes were not found in the parental samples and the normal chromosomes. Thus, we considered the changes as pathogenic.

Mutations were identified in 2 of 3 familial cases of CdLS. In the first family (from Poland), a splice site mutation in exon 38 (donor site) was observed both in the

affected son and daughter. No mutation was observed in their parents. Thus, germline mosaism is the most likely mechanism in this family. The second CdLS familial case was an Irish family involving three generation with an affected daughter, an affected mother and probably affected maternal grandmother (McConnell et al., 2003). Q2280K (exon 40) missense mutation was observed in both the affected daughter and mother. Hence, autosomal dominant inheritance is the most likely mode of transmission in this family.

No particular mutational hot spots were observed. All the mutations detected in our patients with CdLS are heterozygous. As the inheritance of CdLS from many familial cases suggested autosomal dominant, and the changes in *NIPBL* in CdLS patients were heterozygous, this confirms that autosomal dominant inheritance is the major mode of transmission in de Lange syndrome.

### 3.4.4 Genotype-Phenotype correlation

Because of the clinical heterogeneity observed in patients with CdLS and the lack of patient's information from certain referring samples, genotype-phenotype correlation in our study cohort is not easy. However, we tried to analyze the possible correlations in patients with available phenotypic information. In our series, patients found to have mutations in *NIPBL* that led to truncation of the gene product had more severe phenotypes. Patients with splice site mutation also had severe phenotypes. Patients with in-frame deletion of one amino acid tend to have milder phenotypes. As for patients with missense mutations, both severe and mild phenotypes can be seen.

Moreover, in one of our familial cases that was referred from Poland

(Krajewska-Walasek et al., 1995), phenotypes discrepancy was found between the two affected sibs. However, they both shared the same splice site mutation in intron 38. The variation in the phenotypic severity was also observed in different sporadic CdLS patients who share the same mutation (Gillis et al., 2004). Thus, no clear phenotype-genotype correlation was observed in patients affected with CdLS.

**Table 3.1 Summery of CdLS patients with *NIPBL* mutations**

| Panel identifier | Nature of mutation | Location of mutation | Altered amino acid/premature stop codon | Phenotype |
|---|---|---|---|---|
| **Group I and II patients** | | | | |
| P2 | 1bp insertion | Exon 43<br>7306_7307insG | S2462X | Classical |
| P3 | Missense | Exon 43<br>7289A>G | Y2430C | Classical |
| P11 | Splice acceptor AG dinucleotide | Intron 30<br>5575-2A>G | undertermined | Classical |
| P13 | 3bp deletion | Exon 14<br>3616_3618delATA | I1206del | Mild |
| P16 | Missense | Exon 17<br>4043T→ G | L1348R | Classical |
| P18 | Missense | Exon 17<br>3931 T>G | C1311R | Mild |
| P27 | Nonsense | Exon 46<br>7903G>T | E2635X | Classical |
| P28 | Splice donor GT dinucleotide | Intron 38<br>6589+5G>C | undertermined | Classical<br>The mildly affected sister also has the same change |
| P29 | 1bp insertion | Exon 4<br>355_356 insA | Y116X | Classical |
| P37 | Splice acceptor AG dinucleotide | Intron 44<br>7686-2A>G | undertermined | Classical |
| P46 | t(5;13) | Intron 1 | N/A | Classical |
| P47 | t(3;17)<br>4bp deletion | Exon 23<br>5152_5155 del TTTG | F562X | Classical |

| | | | | |
|---|---|---|---|---|
| P48 | t(14;21) | Exon20 | S1459X(Gillis et al., 2004) | Classical |
| **Group III patients** | | | | |
| D6 | Nonsense | Exon20 | Y1445X | N/A |
| D18 | Missense | Exon40 | Q2280K | Mild Affected mother has same mutation |
| P20 | 3bp deletion | Exon 39 | N2217del | N/A |

**Table 3.2 conservations of sequences at sites of *NIPBL* missense mutations and in-frame deletion between different species**

Q2280K

| Species | Amino acid sequences |
|---|---|
| Human NIPBL | VSSGMSSSIMQLYLKQVLEAFFHTQSSV |
| Mouse | VSSGMSSSIMQLYLKQVLEAFFHTQSSV |
| Chick | ISSGMSSSIMQLYLKQVLEAFFHTQSNV |
| Xenopus | ISSGMSSSIMQLYLKQVLESFFSTQSSV |
| Zebrafish | ISSGMSSSIMQLYLKQVLESFFHAQSSV |
| Drosophila | VSSGMASRIIQLYLEEILECFLNRDDTV |
| C.elegants | SGSGLGSSVIQKYWKAVLESYVDADIQL |

N2217del

| Species | Amino acid sequences |
|---|---|
| Human NIPBL | QEVKNLYNNILSDKNSSVNLKIQVLKNL |
| Mouse | QEVKNLYNSILSDKNSSVNLKIQVLKNL |
| chick | QEVKTLYNSILSDKNCSVNLKIQVLKNL |
| Xenopus | VEVKNLYNNILSDKNSSVNLKIQVLKNL |
| Zebrafish | PEVKTLYNGLLSDKRSSITLKIQVLKNL |
| Drosophila | SELKNLYCEILSSIANDAGFKIICMRNI |

Y2430C

| Species | Amino acid sequences |
|---|---|
| Human NIPBL | DTA---KTDVTMLLYIADNLACFPYQTQEEPLF |
| Mouse | DTA---KTEVTMLLYIADNLACFPYQTQEEPLF |
| Chick | DTA---KTEVNMLLYIADNLACFPYQTQEEPLF |
| Xenopus | DAA---KTEVNMLLYIADNLACFPYQTQEEPLF |
| Zebrafish | DSS---KMEVNMLLFIADNLAYFPYQSQEEPLF |
| Drosophila | DQ----KTSLQQMLYIADNLAYFPYVVQDEPLY |
| C.elegans | EEF---SHDKPQLMEYIFIADNLAMFPYQMIDE |
| O.Sativa | NRS---YPTISFLMYCAEVLASLPFTSPDEPLY |

I1206 del

| Species | Amino acid sequences |
|---|---|
| Human NIPBL | EMMDSSTFKRFTASIENILDN |

| | |
|---|---|
| Muse | EMMDSSTFKRFTASIENILDN |
| Chick | EMMDSSTFKRFIASIENILEN |
| Xenopus | EMMDSSTFKRFSSAVENILEN |
| Zebrafish | ELMDSSTFKRFLTSIDNILEN |
| Drosophila | EVTRSQTYQQFIRNMDHIIEI |
| C.elegans | ELKMPKNKKRRSGGDHHHKGD |

C1311R

| Species | Amino acid sequences |
|---|---|
| Human NIPBL | MERVTKSADACLTTINIMTSPNMPKAV |
| Mouse | MERVTKSADACLTTINIMTSPNMPKAV |
| Chick | MERVTKSADACLTAINIMTSPNMPKAV |
| Xenopus | MERVTKSADACLTAINIMTSPSMPKAV |
| Zbrafish | MERVTKSADACLTALNIMTSAHMPKAV |
| Drosophila | IEKLLNAMDACLLICNIYSTVSDLQFL |
| C.elegans | EERVKRASDAAVVALNIMSSHRMHKQV |

L1348R

| Species | Amino acid sequences |
|---|---|
| Human NIPBL | LQNTLYPQYDPVYR---LDPH |
| Mouse | LQNTLYPQYDPVYR---LDPH |
| Chick | LQNTLYPQYDPVYR---VDPL |
| Xenopus | LQNTLYPQYDPVYR---LDPH |
| Zebrafish | LQNTLYPQYDPVYRVDPHGGG |
| Drosophila | LRETIFPLHDPVYT---AKSI |
| C. elegans | LIHLIYPASDSIYKSVNSKKK |

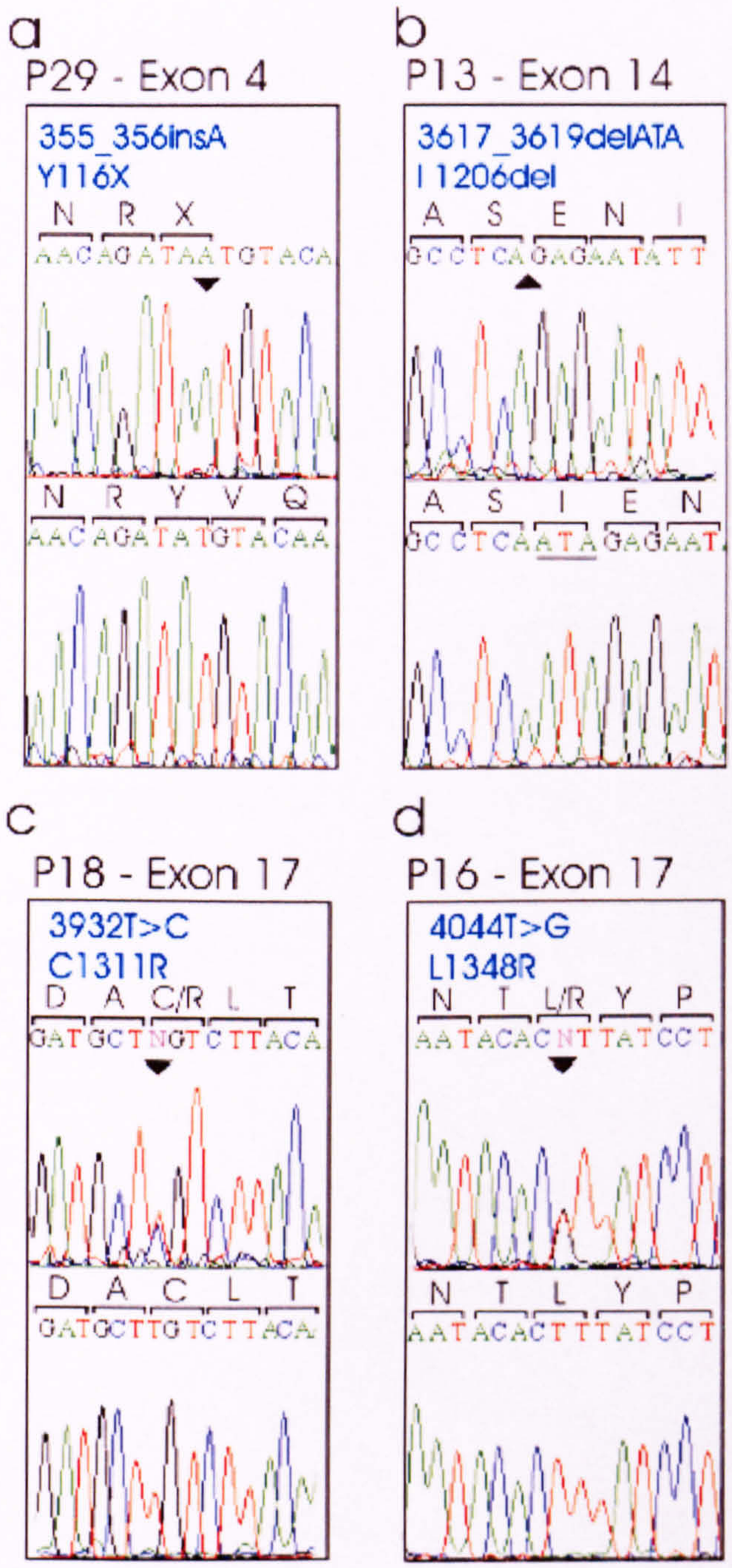



Fig. 3.2 Examples of chromograms of patients with *NIPBL* mutations

## 3.5 DISCUSSION:

Up to now, over 50 causative mutations have been reported that include missense (~22%), frameshift (~43%), nonsense (~20%), and splice-site mutations (~15%) (Borck et al., 2004; Gillis et al., 2004; Krantz et al., 2004; Tonkin et al., 2004). The majority of identified mutations have been private; however, several mutations have been seen in two or three unrelated probands (Gillis et al., 2004). It has been suggested that milder forms of CdLS are more likely to be caused by missense mutations and more severe forms by truncating mutations (Bhuiyan et al., 2006a; Gillis et al., 2004). Mutations in *NIPBL* either lead to haploinsufficiency (frameshift, nonsense, and possibly splice-site mutations) or to altered proteins (missense mutations) whose function and viability is unknown at this time. Haploinsufficiency of the NIPBL protein results in CdLS as demonstrated by two reports of CdLS in infants with deletions of the entire gene (Hulinsky et al., 2005; Taylor and Josifek, 1981).

Missense mutations that impair protein function may result in disease. The disease severity caused by such deleterious mutations is dependent on when and where the protein's function is required by the organism. The disease phenotype is likely determined by three parameters: (1) the degree to which the function of the protein is impaired by the missense mutation; (2) variants of other genes that modulate the effect of the major locus, also referred to as genetic background; and (3) the environment. Several studies suggest that mutations in evolutionary conserved sites tend to impair protein function and lead to disease. The disease severity is also somewhat correlated with the physicochemical difference between the original amino acid and the missense variant (Miller and Kumar, 2001; Sunyaev et al., 2001a;

Sunyaev et al., 2001b; Williams et al., 2003). Residues that are conserved completely in the protein family are expected to be important for function, and even a conservative substitution at one of these residues may affect protein function. At some positions, any amino acid change can be tolerated in the protein if these positions are not involved in protein function or structure. Because these are expected to be neutral substitutions, one might expect amino acids in these positions of the protein aligment to be diverse.

An analysis of the frequency matrix for disease-associated missense mutations reveals some differences from amino acid substitution rates seen in wild-type proteins. There are eight mutations that occur significantly more frequent in disease-associated than in wild-type proteins. Seven of them involve two amino acid types, that is, cysteine (C→R, C→Y and C→G) and arginine (R→G, C→R, H→R, W→R and R→L). Other substitutions commonly found in wild-type proteins such as L→M, L→I, A→S and F→Y are relatively rare causes of disease. Presumably, this is because similarities in physicochemical properties between these amino acids and thus ensure conservation of protein function despite their substitution (Goodstadt and Ponting, 2001). Cysteine and arginine are both more highly substituted and more highly substituting in disease-associated variations. For cysteine, this is likely to arise from its unusual chemical properties, in particular gain or loss of disulphide-bridges or free thiols. The dominance of arginine in disease-associated missense variants may be due, in part, to its participation in salt bridges and its prevalence in solvent-accessible peripheries rather than the hydrophobic interiors of proteins and is important for molecular stability and function (Goodstadt and Ponting, 2001). Protein phosphorylation is a key mechanism for intracellular signal transduction in both prokaryotic and eukaryotic cells. Vertebrate proteins are prevalently phosphorylated

on side chains that contain a hydroxyl group, such as serine, threonine and tyrosine residues. Phosphorylation of serine, threonine or tyrosine results in the formation of a phosphoester linkage. Phosphorylation of histidine residues occurs on nitrogen atoms, producing a phosphoramidate bond. Phosphohistidines have a large standard free energy of hydrolysis making them the most unstable of any known phosphoamino acid (Matthews, 1995). Thus, missense mutation that substitutes these phosphrylation associated amino acids is expected to result in disease phenotypes.

Missense mutations in *NIPBL* that cause substitution of the "wild-type" amino acid to a variant amino acid are also observed in CdLS patients from several research groups (Ben-Asher and Lancet, 2004; Bhuiyan et al., 2005; Bhuiyan et al., 2006b; Borck et al., 2004; Borck et al., 2006; Gillis et al., 2004; Krantz et al., 2004; Tonkin et al., 2004)(see Fig. 3.2 as a summary). Most of the missense mutations are located within the highly conserved region of *NIPBL*.

Of the 32 missense mutations (29 different locations) shown below (Fig. 3.3), 28 mutations are located within the highly conserved region, especially clustered within the HEAT repeat region. Mutations involving the wild-type amino acid into arginine occurred in 4 of the 29 different mutations. Arginine substituted by other amino acids was found in 6 of the 29 different mutations, especially clustered in the HEAT repeat region. Amino acid substitution involving arginine might influence the stability of the protein molecules and result in disease phenotype. Mutation involving substitution of cysteine was found in 3 of the 29 types of mutations. Change in the cysteine amino acid might lead to gain or loss of the disulfide bridges or free thiols. Mutation involving serine, threonine and histidine was found in 6 of the 29 different mutations. Most of such mutations were located within the highly conserved region, especially

near the phophrylated end of NIPBL protein. Mutation involving threonine and histidine might alter the state of phosphrylation and affect the signaling pathway. Aspartic acid (D) change to glycine (G) was observed in 2 of the 29 different mutations. A study to evaluate the exchangiabilty of amino acids in protein showed that the exchangeability from aspartic acid to glycine is low (Yampolsky and Stoltzfus, 2005). Thus, it is reasonable for a substitution from aspartic acid to glycine to alter the protein function.

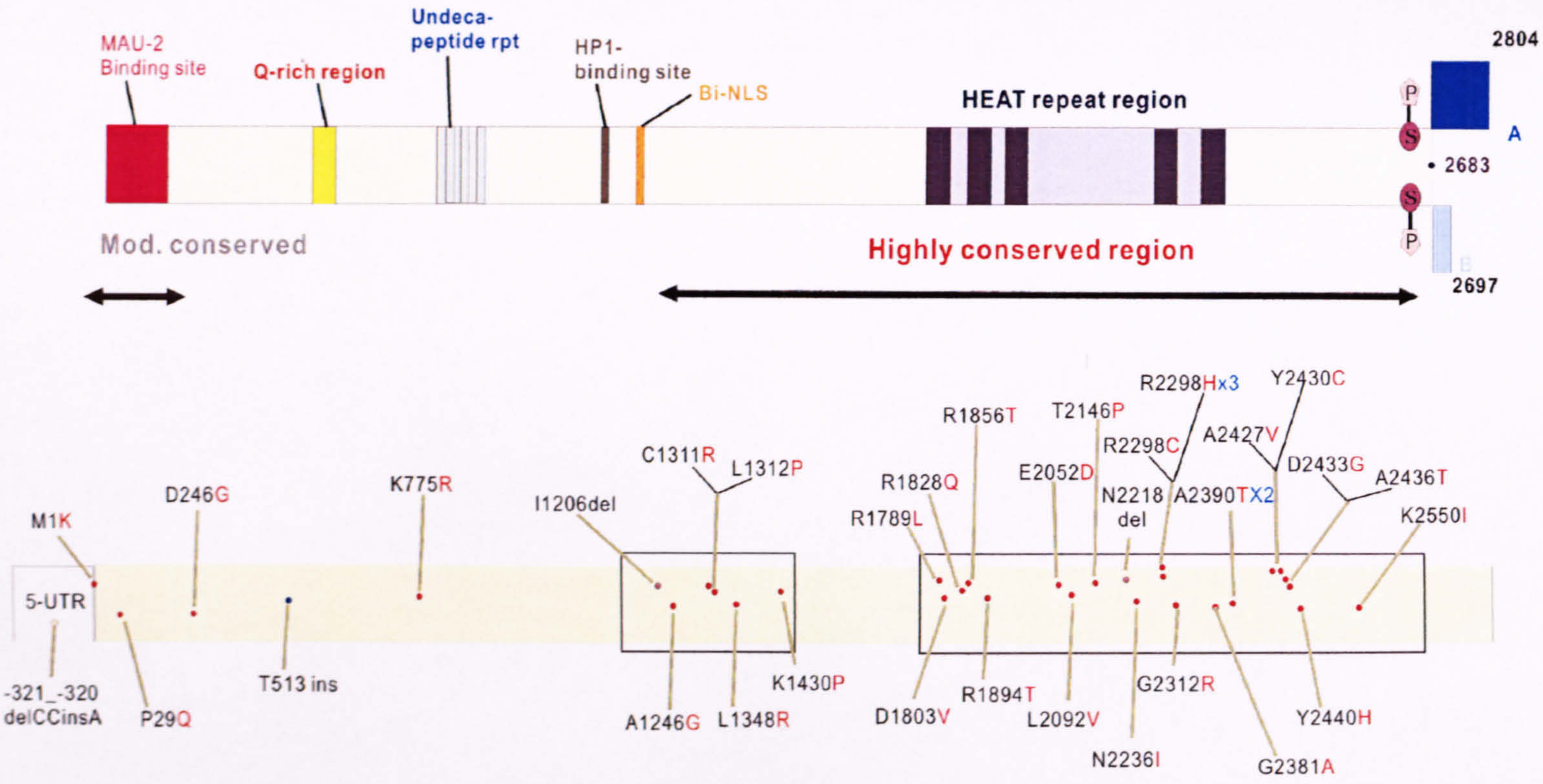



Fig. 3.3. missense mutations of the *NIPBL*. Most of the missense mutations were located within the highly conserved regions of *NIPBL*.

In our series, the overall mutation detection rate in patients with CdLS is relatively low compared with the series reported by (Gillis et al., 2004). In our series, the mutation detection rate was around 40% in group I (when account is taken of the number of exons screened), 100% in group II, but only 18% in group III. The mutation detection rate in the first 2 groups is similar to other series reported (Ben-Asher and Lancet, 2004; Bhuiyan et al., 2005; Borck et al., 2004; Gillis et al., 2004) with the exception of the Japanese group. They reported a mutation detection rate of 26% (Miyake et al., 2005). However, in the third group, the mutation detection rate is relatively low compared to all the other series reported. It should be noted that most of the samples were referred from Denmark, and most of the patients' information was not clear to us in the third group. In this group, 4 patients were referred from Britain and Ireland. Mutations were found in three (1 sporadic, two familial) out of four patients. Only one sample in the remaining 17 samples referred from Denmark was found to have a nonsense mutation. In contrast to the US group's report (Gillis et al., 2004), most of the patients in our cohort studies were sporadic. Only two families had more then one affected member in their families. The low mutation rate in the cohort studies may be due to the following reasons: (1) The uncertainty of the diagnosis in certain cases; (2) large deletion or duplication involving several exons; (3) variation in sequences beyond exon/intron boundaries such as regulatory elements or intronic sequences; (4) difficulty in amplifying certain *NIPBL* exons such as exon 33 and 37; (5) the existence of mosaicism in tissues other than blood cells; (6) other genes involved in the pathogenesis of CdLS. There are several inherited genetic syndromes that have some phenotypic overlap with CdLS and thus might make the diagnosis confusing, particularly with mild CdLS. One of the examples is Rubinstein-Tabi syndrome (MIN 180849). The neat arched eyebrow, mental retardation and growth deficiency seen in Rubinstein-Tabi syndrome (Francois,

1981; Partington, 1990) sometimes will be mistaken for the mild de Lange syndrome. Thus, further clinical reevaluation in the patients without *NIPBL* mutations is important. Deletions/insertions of the entire exon(s) were also considered as a possibility for the low mutation detection rate in our study. However, to find out if the low mutation detection rate is partly due to rearrangements resulting in deletion or duplication of one or more exons, Lott et al. (reported in annual meeting of European Society of Human Genetics, 2006) tested the *NIPBL* gene for copy number changes with the MLPA (multiplex ligation dependent probe amplification) technique. No exon deletions in the 18 CdLS patients without point mutations were detected. Thus, deletions/insertions involving the whole exon(s) seem unlikely. Recently, the mutation 5'UTR region was found to be associated with the CdLS phenotype. The mutation occurred in exon 1 of the *NIPBL* gene, 321 bps upstream of the translation start site (-321_-320 delCCinsA) (Borck et al., 2006). The mutation was located within the highly conservative region of exon 1 (as illustrated in Fig. 3.3). It is quite possible this area might contain certain important *cis* regulatory elements. Moreover, in our patient with t(5;13) balanced translocation, the breakpoint was in intron 1 of the *NIPBL* gene. Due to the location of the transcription start site of *NIPBL* in Exon 2, disruption of intron 1 in the t(5;13) CdLS patient with severe phenotypes suggests that the well conserved untranslated regions might contain certain important *cis* regulatory elements. Disruption of this *cis*-element might reduce the level of gene transcription. Sequences analysis of -2kb 5' of promoter and 2.0kb 3'of exon 1 by using VISTA genome browser (http://pipeline.lbl.gov/cgi-bin/gateway2?bg=phr1&selector=vista) (Brudno et al., 2003; Couronne et al., 2003) suggested that the first 2kb sequence of intron 1 as well as exon 1 are highly conserved (see Fig. 3.4).

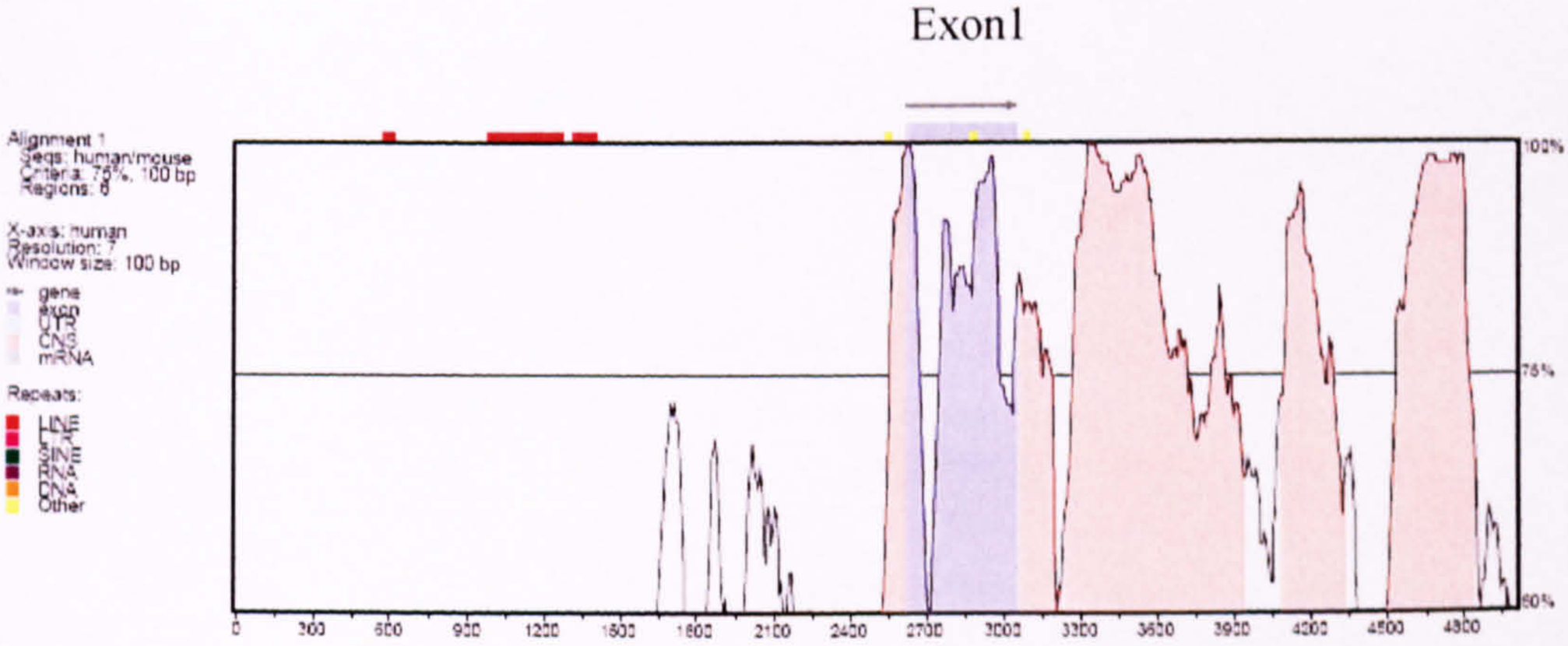



Fig 3.4 sequences conservation -2.5kb upstream of the *NIPBL* promoter and 2kb downstream of the exon 1

For many genes, the region immediately upstream of the minimal promoter contains sufficient transcription factor binding sites to direct correct expression of the gene - these are termed "regulatory promoters." However, many genes also require multiple *cis*-acting distant genomic elements for spatiotemporally correct expression (Howard and Davidson, 2004). These are often defined as enhancers, although some will be repressors or insulators; they can be located upstream, within introns, or downstream of the "transcription unit," which comprises the transcribed exons and introns from the promoter to the polyadenylation site.

Additional complexities include the possible presence of multiple alternative promoters and exons. The genomic regions harboring regulatory elements can stretch as much as 1Mb in either direction from the transcription unit (Pfeifer et al., 1999). Also, in one of our familial cases, both the affected brother and sister harbor the same splice site mutation in intron 38. However, the phenotypes in the brother were milder than his affected sister. There have been reports on phenotypic variations in the affected individuals in the same family (Feingold and Lin, 1993; Gardner, 2003). In

today's studies, the use of molecular genetic analysis to interpret and predict the phenotype in genetic disease initially appeared promising. However, for many human diseases, including malformation syndromes, no clear phenotype-genotype relationship has been demonstrated. For some phenotypes it is becoming clear that mutations in a single gene are not sufficient to cause a phenotype without sequence alterations in another gene. Further variations in phenotype or penetrance can be caused by loss or alteration of contiguous genes, transporter proteins and activator proteins, and other classes of molecules. Modifier genes have become increasingly recognized as an important source of phenotypic variation that may explain the relation of phenotype to genotype. Thus, it is reasonable to speculate the existence of another gene or modifier gene for the phenotypes variation seen in some of our family members with the same mutation in *NIPBL*. Therefore, we still need to try to find other possible genes that cause or modulate the CdLS phenotypes. Thus, it is also important to check for the possible mutation occurred within. Recently, germline mosaism has also been reported in familial cases with D2433G missense mutation. Mutation was found in the father's sperm but not in the blood lymphocytes, and the phenotype tend to be milder (Niu et al., 2006). It is possible that the mutation-negative mild CdLS patients might have somatic mutations other than blood lymphocytes.

Recently, mutations in SMC1L1 (also known as SMC1), which encodes a different subunit of the cohesin complex, have been reported to be responsible for X-linked variant of CdLS (Musio et al., 2006). Krantz et al. screened a cohort of 95 *NIPBL* mutation-negative patients with CdLS and identified SMC1L1 mutations in 8 of them, which are phenotypically milder (abstract from annual meeting in American Society of Human Genetics, 2006). Furthermore, the 20-50% *NIPBL* mutation detection rate in patients with CdLS suggests the existence of heterogeneity of CdLS.

By using karyotyping, FISH and Array-CGH techniques, *de novo* chromosome imbalances include tetrasomy of the entire chromosome 18p and terminal deletion of 1q spanning 6.5Mb was found in two patients out of 14 CdLS patients. Mutation in *NIPBL* was only detected in 5 of the 14 patients (Borck et al., 2004) Thus, it is quite reasonable that genetic heterogeneity existing in CdLS and mutations in other genes involved in the sister chromatic cohesion complex may cause the CdLS phenotypes. The cohesin complexes consist of SMC1, SMC3, Scc1, and Scc3. The human homologues of these are hSMC1L1, hSMC3, hRAD21 and hSA1/hSA2 respectively. Scc2 (hNIPBL)/Scc4(hMau-2) complexes help in the establishment and loading of cohesin ring (Gruber et al., 2003). Recently, another study in Hela cells reported that a subfraction of SNF2h/ISWI, the ATPase subunit of many remodeling complexes, is co-purified with cohesin. hScc1/hRAD21 subunit of the cohesin complex directly interacts with the ATPase subunit SNF2h. SNF2h can help in the loading of hScc1 on chromatin (Hakimi et al., 2002). Table 3.3 summarizes the chromosomal locations of the subunits of cohesin complexes and associated proteins.

**Table 3.3 chromosomal location of the human cohesin complexes and associated proteins.**

| Cohein | Gene location in *Homo sapiens* | Possible roles |
|---|---|---|
| hSMC1(SMC1L1) | Xp11.22 | One of the disease causing gene for CdLS (Musio et al., 2006) |
| hSMC3 | 10q25 | Structural maintenance protein of cohesin |
| hScc1(RAD21) | 8q24 | Part of the cohesin ring complexes |
| SA1(STAG1) | 3q22.3 | assembles with the cohesin proteins (Prieto et al., 2002) |
| SA2(STAG2) | Xq25 | assembles with the cohesin proteins (Prieto et al., 2002) |
| hScc2/NIPBL | 5p13.2 | Major CdLS causing gene(Krantz et al., 2004; Tonkin et al., 2004) |
| hScc4/hMau-2 | 19p13.11 | Help in loading of cohesin onto chromatin (Gruber et al., 2003) |
| hPds5/APRIN | 13q12.3 | inducing cell cycle arrest in the G0/G1 phase (Geck et al., 2000) |

Of the above gene locations described, hScc1/RAD21 (8q24) is located near *de novo* t (X;8)(p11.2;q24.3) translocation breakpoint affected with CdLS (Egemen et al., 2005). The location of the APRIN (13q12.3) is near the CdLS patient with *de novo* t(5;13)(p13.2; q12.1). The translocation in 13q12.1 might interrupt the normal functions of the APRIN. As disorders affecting the sister chromatic cohesion complex likely represent an underappreciated class of clinical disorders, it is necessary to screen for NIPBL and SMC1L1 mutation negative individuals with CdLS for the possibility of patient-specific mutations in other proteins associated with cohesin complexes. According to all the chromosome anomalies reported to be associated with CdLS, hScc1/RAD21 and APRIN seem to be the most possible disease causing genes. Nevertheless, the majority of the CdLS patients with chromosomal abnormalities showed no consistent regions involved. This may suggest that the multiple loci implicated in CdLS are just coincidental. Another possible explanation for the multiple chromosomal anomalies observed in CdLS cases is that the CdLS disease gene may play an important role in the repair of double strand DNA. Mutations in the CdLS gene(s) cause increased chromosome breakage at fragile sites and result in the multiple chromosome abnormalities associated in some CdLS patients.

## 3.6 REFERENCES


Ben-Asher, E., and Lancet, D. (2004). NIPBL gene responsible for Cornelia de Lange syndrome, a severe developmental disorder. Isr Med Assoc J *6*, 571-572.

Bhuiyan, Z., Klein, M., Hammond, P., Mannens, M. M., Van Haeringen, A., Van Berckelaer-Onnes, I., and Hennekam, R. C. (2005). Genotype-Phenotype correlations of 39 patients with cornelia de Lange syndrome: the Dutch experience. J Med Genet.

Bhuiyan, Z. A., Klein, M., Hammond, P., van Haeringen, A., Mannens, M. M., Van Berckelaer-Onnes, I., and Hennekam, R. C. (2006a). Genotype-phenotype correlations of 39 patients with Cornelia De Lange syndrome: the Dutch experience. J Med Genet *43*, 568-575.

Bhuiyan, Z. A., Zilfalil, B. A., and Hennekam, R. C. (2006b). A Malay boy with the Cornelia de Lange syndrome: clinical and molecular findings. Singapore Med J *47*, 724-727.

Borck, G., Redon, R., Sanlaville, D., Rio, M., Prieur, M., Lyonnet, S., Vekemans, M., Carter, N. P., Munnich, A., Colleaux, L., and Cormier-Daire, V. (2004). NIPBL mutations and genetic heterogeneity in Cornelia de Lange syndrome. J Med Genet *41*, e128.

Borck, G., Zarhrate, M., Cluzeau, C., Bal, E., Bonnefont, J. P., Munnich, A., Cormier-Daire, V., and Colleaux, L. (2006). Father-to-daughter transmission of Cornelia de Lange syndrome caused by a mutation in the 5' untranslated region of the NIPBL Gene. Hum Mutat *27*, 731-735.

Brudno, M., Do, C. B., Cooper, G. M., Kim, M. F., Davydov, E., Green, E. D., Sidow, A., and Batzoglou, S. (2003). LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res *13*, 721-731.

Cotton, R. G. (1997). Slowly but surely towards better scanning for mutations. Trends Genet *13*, 43-46.

Couronne, O., Poliakov, A., Bray, N., Ishkhanov, T., Ryaboy, D., Rubin, E., Pachter, L., and Dubchak, I. (2003). Strategies and tools for whole-genome alignments. Genome Res *13*, 73-80.

Egemen, A., Ulger, Z., Ozkinay, F., Gulen, F., and Cogulu, O. (2005). A de novo t (X;8)(p11.2;q24.3) demonstrating Cornelia de Lange syndrome phenotype. Genet Couns *16*, 27-30.

Feingold, M., and Lin, A. E. (1993). Familial Brachmann-de Lange syndrome: further evidence for autosomal dominant inheritance and review of the literature. Am J Med Genet *47*, 1064-1067.

Francois, J. (1981). [Dysmorphic idiopathic dwarfisms and their ocular manifestations]. J Fr Ophtalmol *4*, 511-524.

Gardner, R. J. (2003). Another explanation for familial Cornelia de Lange syndrome. Am J Med Genet A *118*, 198.

Geck, P., Maffini, M. V., Szelei, J., Sonnenschein, C., and Soto, A. M. (2000). Androgen-induced proliferative quiescence in prostate cancer cells: the role of AS3 as its mediator. Proc Natl Acad Sci U S A *97*, 10185-10190.

Gillis, L. A., McCallum, J., Kaur, M., DeScipio, C., Yaeger, D., Mariani, A., Kline, A. D., Li, H. H., Devoto, M., Jackson, L. G., and Krantz, I. D. (2004). NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. Am J Hum Genet *75*, 610-623.

Goodstadt, L., and Ponting, C. P. (2001). Sequence variation and disease in the wake of the draft human genome. Hum Mol Genet *10*, 2209-2214.

Gruber, S., Haering, C. H., and Nasmyth, K. (2003). Chromosomal cohesin forms a ring. Cell *112*, 765-777.

Hakimi, M. A., Bochar, D. A., Schmiesing, J. A., Dong, Y., Barak, O. G., Speicher, D. W., Yokomori, K., and Shiekhattar, R. (2002). A chromatin remodelling complex that loads cohesin onto human chromosomes. Nature *418*, 994-998.

Hayashi, K., and Yandell, D. W. (1993). How sensitive is PCR-SSCP? Hum Mutat *2*, 338-346.

Howard, M. L., and Davidson, E. H. (2004). cis-Regulatory control circuits in development. Dev Biol *271*, 109-118.

Hulinsky, R., Byrne, J. L., Lowichik, A., and Viskochil, D. H. (2005). Fetus with interstitial del(5)(p13.1p14.2) diagnosed postnatally with Cornelia de Lange syndrome. Am J Med Genet A.

Krajewska-Walasek, M., Chrzanowska, K., Tylki-Szymanska, A., and Bialecka, M. (1995). A further report of Brachmann-de Lange syndrome in two sibs with normal parents. Clin Genet *47*, 324-327.

Krantz, I. D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L. A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C. A., *et al.* (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nat Genet *36*, 631-635.

Liechti-Gallati, S., Schneider, V., Neeser, D., and Kraemer, R. (1999). Two buffer PAGE system-based SSCP/HD analysis: a general protocol for rapid and sensitive mutation screening in cystic fibrosis and any other human genetic disease. Eur J Hum Genet *7*, 590-598.

Matthews, H. R. (1995). Protein kinases and phosphatases that act on histidine, lysine, or arginine residues in eukaryotic proteins: a possible regulator of the mitogen-activated protein kinase cascade. Pharmacol Ther *67*, 323-350.

Miller, M. P., and Kumar, S. (2001). Understanding human disease mutations through the use of interspecific genetic variation. Hum Mol Genet *10*, 2319-2328.

Miyake, N., Visser, R., Kinoshita, A., Yoshiura, K., Niikawa, N., Kondoh, T., Matsumoto, N., Harada, N., Okamoto, N., Sonoda, T., *et al.* (2005). Four novel NIPBL mutations in Japanese patients with Cornelia de Lange syndrome. Am J Med Genet A *135*, 103-105.

Musio, A., Selicorni, A., Focarelli, M. L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P., and Larizza, L. (2006). X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat Genet *38*, 528-530.

Niu, D. M., Huang, J. Y., Li, H. Y., Liu, K. M., Wang, S. T., Chen, Y. J., Udaka, T., Izumi, K., and Kosaki, K. (2006). Paternal gonadal mosaicism of NIPBL mutation in a father of siblings with Cornelia de Lange syndrome. Prenat Diagn.

Orita, M., Suzuki, Y., Sekiya, T., and Hayashi, K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics *5*, 874-879.

Partington, M. W. (1990). Rubinstein-Taybi syndrome: a follow-up study. Am J Med Genet Suppl *6*, 65-68.

Petersen, I., Ohgaki, H., Ludeke, B., and Kleihues, P. (1994). Direct DNA sequencing following SSCP analysis. Anal Biochem *218*, 478-479.

Pfeifer, D., Kist, R., Dewar, K., Devon, K., Lander, E. S., Birren, B., Korniszewski, L., Back, E., and Scherer, G. (1999). Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. Am J Hum Genet *65*, 111-124.

Prieto, I., Pezzi, N., Buesa, J. M., Kremer, L., Barthelemy, I., Carreiro, C., Roncal, F., Martinez, A., Gomez, L., Fernandez, R., *et al.* (2002). STAG2 and Rad21 mammalian mitotic cohesins are implicated in meiosis. EMBO Rep *3*, 543-550.

Rao, V. B. (1994). Direct sequencing of polymerase chain reaction-amplified DNA. Anal Biochem *216*, 1-14.

Rollins, R. A., Morcillo, P., and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics *152*, 577-593.

Sunyaev, S., Lathe, W., 3rd, and Bork, P. (2001a). Integration of genome data and protein structures: prediction of protein folds, protein interactions and "molecular phenotypes" of single nucleotide polymorphisms. Curr Opin Struct Biol *11*, 125-130.

Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., 3rd, Kondrashov, A. S., and Bork, P. (2001b). Prediction of deleterious human alleles. Hum Mol Genet *10*, 591-597.

Taylor, M. J., and Josifek, K. (1981). Multiple congenital anomalies, thymic dysplasia, severe congenital heart disease, and oligosyndactyly with a deletion of the short arm of chromosome 5. Am J Med Genet *9*, 5-11.

Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J., and Strachan, T. (2004). NIPBL,

encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. Nat Genet *36*, 636-641.

Williams, R. S., Chasman, D. I., Hau, D. D., Hui, B., Lau, A. Y., and Glover, J. N. (2003). Detection of protein folding defects caused by BRCA1-BRCT truncation and missense mutations. J Biol Chem *278*, 53007-53016.

Yampolsky, L. Y., and Stoltzfus, A. (2005). The exchangeability of amino acids in proteins. Genetics *170*, 1459-1472.

# CHAPTER 4 EXPRESSION PATTERNS OF THE *NIPBL* GENE

## 4.1 INTRODUCTION:

Previously, we have identified a novel gene, *NIPBL*, as the major determinant for CdLS syndrome (Tonkin et al., 2004), however, the functions of this novel gene are not known. To determine if the expression patterns of this novel gene correlate well with CdLS phenotypes, riboprobes were designed to study the expression patterns of *NIPBL*. Northern blot analyses on both the adult and fetal tissues are examined first. When differentiatial tissue expression patterns are observed, ongoing tissue *in situ* hybridization by using riboprobes will be performed. Human embryos sections obtained from different Carnegie stages are chosen for the tissue in situ hybridisation studies. Carnegie stage (CS) 13-14 embryonic sections were chosen to study the expression patterns of *NIPBL* in early limb bud and CNS development to evaluate if *NIPBL* expressions are at sites of active proliferation. CS17-18 embryonic sections were chosen for the ongoing serial *NIPBL* expression studies in different organ systems. CS21-23 embryonic sections were used to study the more detailed CNS, craniofacial and limb expression. No human foetal tissues are available for us to study the expression patterns at later developmental stages due to ethical issues.

## 4.2 APPROACHES TO EXPRESSION STUDIES OF *NIPBL*

### 4.2.1 Northern blot analysis

Northern blotting is the RNA counterpart of the Southern blot technique that was

developed for DNA analysis by E. M. Southern in 1975 (Southern, 1975). RNA samples are first separated by size via electrophoresis in an agarose gel under denaturing conditions. The RNA is then transferred to a membrane, crosslinked and hybridized with a labeled probe. RNA sequences of interest are detected on the blot by hybridization to a specific labeled probe. Probes for Northern blot detection generally contain full or partial cDNA sequences and may be labeled by enzymatic incorporation of radiolabeled (usually $^{32}P$) nucleotides. After probe hybridization and washing to remove nonspecific label, the hybridization signal is generally detected by exposing blots to X-ray film. The resulting band identified by the probe indicates the size of the mRNA, and the intensity of the band corresponds to the relative abundance. Autoradiograph band intensities may be quantitated by densitometry, by direct measurement of hybridized radiolabeled probe via storage phosphor imaging or by scintillation counting of excised bands.

As a result of the relative ease of Northern blotting, it has become the most widely used method for characterization of mRNA size, relative abundance and detection of alternative spliced transcript. Of the techniques discussed here, Northern blotting is the only one which allows mRNA size determination. Hence, it has been a key method for detecting mutations that result in abnormal mRNA size. In general, this technique should be considered semi-quantitative, and is most suitable for determining relative concentrations of mRNA species that occur in moderate to high abundance (Reue, 1998).

Despite these advantages, there are limitations associated with Northern analysis. First, if RNA samples are even slightly degraded, the quality of the data and the ability to quantitate expression are severely compromised. Thus, RNase-free reagents

and techniques are essential. Second, a standard Northern procedure is, in general, less sensitive than nuclease protection assays and RT-PCR, although improvements in sensitivity can be achieved by using high specific activity antisense RNA probes, optimized hybridization buffers and positively charged nylon membranes. A third limitation of Northern blotting has been the difficulty associated with multiple probe analysis. To detect more than one message, it is usually necessary to strip the initial probe before hybridizing with a second probe. This process can be time consuming and problematic, since harsh treatment is required to strip conventional probes from blots.

Previously, E.T. Tonkin performed Northern blot analysis and confirmed the predicted 9.8-kb transcript size by using probe designed across exons 10, 11 and 12. *NIPBL* expression was strongly expressed in adult heart, muscle, thymus, kidney and placenta. *NIPBL* was also strongly expressed in fetal kidney, moderately expressed in fetal liver and mildly expressed in fetal brain. The expression was nearly undetectable in adult and fetal lung, adult brain, colon, and small intestine. The result is shown in Fig. 4.1.

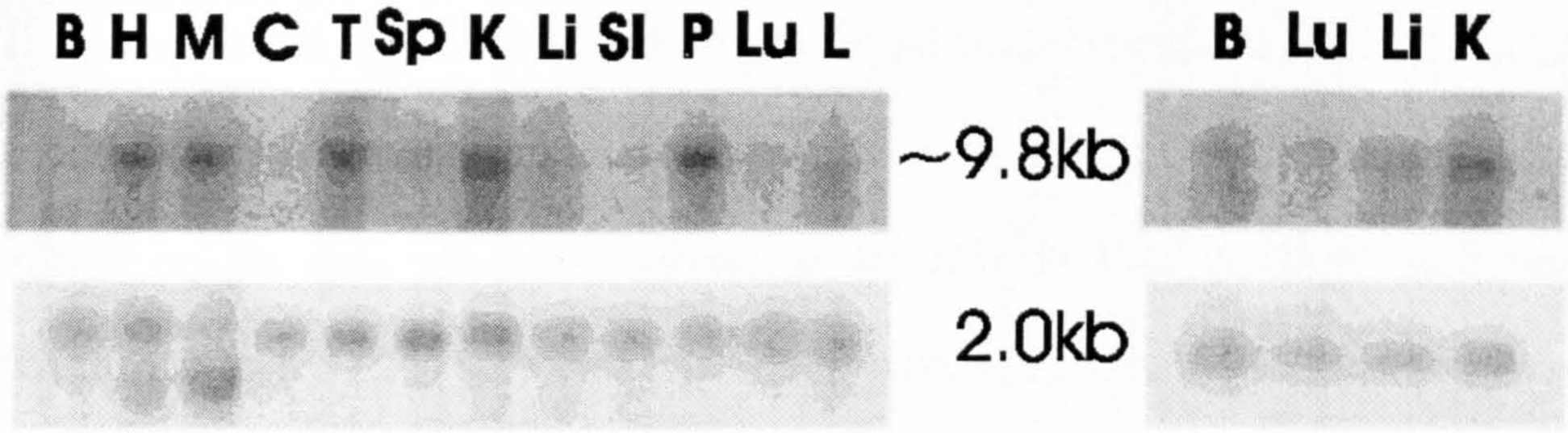



Fig. 4.1. Results of NIPBL expression by Northern blot. Figure on the left panels is the expression data from the commercially available membrane containing fetal RNA, while the figure on the right panels is the expression data of adult tissues. The probe used was corresponding to NIPBL exons 10-12. The NIPBL transcript was around 9.8kb in both fetus and adult tissues. The 2.0kb band was the actin transcript used as the control. Expression was strong in adult heart, muscle, thymus, kidney and placenta and fetal kidney. Expression was nearly undetectable in adult and fetal lung, adult brain, colon and small intestine. B: brain; H: heart; M: muscle; C: colon; T: thymus; Sp: spleen; K: kidney; Li: liver; SI: small intestine; P: placenta; Lu: lung; L: lymphocyte.
(with permission from E.T. Tonkin)

Differential tissue expression was observed in Northern blot. Thus, ongoing Tissue *in situ* hybridization can be performed.

### 4.2.2 Tissue *in situ* hybridization

The introduction in the late 1960s of *in situ* hybridization (ISH) techniques (Buongiorno-Nardelli and Amaldi, 1970) opened a new era in histology and cell biology. The immunocytochemical methods can demonstrate only the presence of synthesized protein molecules. They can not recognize specific DNA or RNA sequences in a tissue and in a cell to define the precise location of a potential or an effective synthesis of a given molecule.

Although Southern blot and Northern blot can identify different classes of DNA and RNA respectively by hybridization to membrane bound nucleic acids, ISH can fill the gap between the detection of a specific sequence and its precise location within the tissue or the cell. A major advantage of ISH is that it allows different hybridizations as it is often performed on thin or ultra thin sections of a piece of tissue (e.g., a single surgical biopsy), which, at times, is not sufficient to allow Northern or Southern blot analysis. It is also possible to make libraries of paraffin-embedded or frozen tissues. No significant loss of the hybridization signal was found in frozen sections kept at -70℃ with desiccant for more than 6 years (Wilcox, 1993).

The sensitivity and efficiency of ISH depend on several variables, for which optimal conditions must be determined: (1) the probe construction and hybridization conditions; (2) the type and efficiency of probe labeling; (3) the tissue preparation (fixation, embedding), which must permit the retention of the target of hybridization and/or the hybridized products; and (d) the method used for signal detection.

ISH is a technique that allows the detection of specific nucleic acid sequences within the intact tissue and cellular architecture. It has the advantage of identifying the gene products from *de novo* synthesis rather than protein uptake or receptor-bound protein, which may result in false-positive immunostaining results. Conversely, because proteins may be rapidly secreted or destroyed, this may lead to false-negative immunostaining results.

Complementary RNA probes or riboprobes are the probes of choice for ISH (Cox et al., 1984), because they are sensitive and specific and there is no competition between hybridization and probe re-annealing, as with double-stranded probes

(Melton et al., 1984). They can be efficiently transcribed from cDNA clones containing T7, SP6, or T3 RNA polymerase promoters (Jorgensen et al., 1991). However this depends on the availability of the cDNA of interest, which would have to be cloned into a suitable vector. Clones must also be linearized by enzyme restriction at opposite ends of the cDNA to allow transcription of sense and antisense probes from the appropriate promoter. The optimal size of the RNA probe is around 500bp. For a probe size larger than 500bps, hydrolysis of the probe is often required to reduce it to a size that is capable of penetrating into tissue and hybridizing with the target (Angerer et al., 1992). Notably, some plasmid-generated probes have also been observed to hybridize nonspecifically due to the presence of sequences similar to human 28S rRNA in the plasmid multiple cloning site (Baumgart et al., 1997). To circumvent these problems, we can generate the probes by using simple PCR technology (Higuchi et al., 1988; Weier and Rosette, 1988). In 1991, Young et al. first produced a DNA fragment with both SP6 and T7 RNA polymerase promoters, enabling production of sense and antisense transcripts from a single template (Young et al., 1991). This technique obviated the need for cloning, screening, and large-scale plasmid culture and the difficulties associated with plasmid restriction digestion. Probe hydrolysis to facilitate ISH is also not required because the PCR-generated template can be designed to yield probe of optimal length. The distinct advantages of this technique are that the probe produced has a user-defined length and sequence without exogeneous vector sequences.

### 4.2.3 Development of different organ systems during early human development

#### 4.2.3.1 The developing nervous system

The central nervous system (CNS) forms from the neural tube. The nervous system develops from the neural plate, a slipper-shaped area of embryonic ectoderm. The notochord causes the overlying ectoderm to differentiate into the neural plate, before the neural folds, neural tube and neural crest are formed from the neural plate. The neural tube differentiates into the brain and spinal cord. The neural crest gives rise to cells that form most of the peripheral nervous system (PNS) and autonomic nervous system (ANS), which consist of cranial, spinal and autonomic ganglia.

Formation of the neural tube (neurulation) begins around Carnegie stage 11 (22-23 days) in the region of the fourth to sixth somites. The cranial two-thirds of the neural plate and tube and the first 4 pairs of somites represent the future brain. The caudal one-third of the neural plate and tube represents the future spinal cord.

The neural tube cranial to the fourth pair of somites develops into brain. Three primary brain vesicles are formed after fusion of the neural folds in the cranial region and closure of the rostral neuropore. These three primary brain vesicles develop into the future forebrain (prosencephalon), midbrain (mesencephalon) and hindbrain (rhomencephalon). During the fifth week the forebrain partly divides into two secondary brain vesicles, the telencephalon and diencephalons; the midbrain does not divide; the hindbrain partly divides into the metencephalon and myelencephalon (Moore, 2003).

The neural tube caudal to the fourth pair of somites develops into the spinal cord. The lateral wall of the neural tube thickens, gradually reducing the size of the neural canal until only a small central canal of the spinal cord is present at 9-10 weeks (Moore, 2003). Initially the wall of the neural tube constitute the ventricular zone

(ependymal layer), the site for active neuronal cell proliferation. Soon a marginal zone composed of the outer parts of the neuroepithelial cells becomes recognizable. This zone gradually becomes the white matter. As these primitive cells continue to divide, subtle decisions begin to be made as to their fate. Some will become neurons while others are fated to become glial cells. The mechanisms controlling these processes still remain unknown.

In early 1960, Fujita et al. noticed a very characteristic mitotic pattern in neural tube of human embryos, which he called it the "matrix cells" cytogenesis theory. In his proposed mechanism, there are three stages of cytogenesis of the multipotent neuroepithelial stem cells (matrix cells). In stage I, the matrix cells perform "elevator-like movements" during proliferation. The nuclei of the cells undergo a cyclic change of position in the neuroepithelial cell layer. That is, the nuclei of intermitotic cells lie deeply, and progressively approach ventricular lumen during G2 and prophase. Nuclear division occurs exclusively at the luminal surface, upon which the daughter nuclei move back again to the deeper positions, as illustrated in Fig.5.5. (Fujita, 2003).

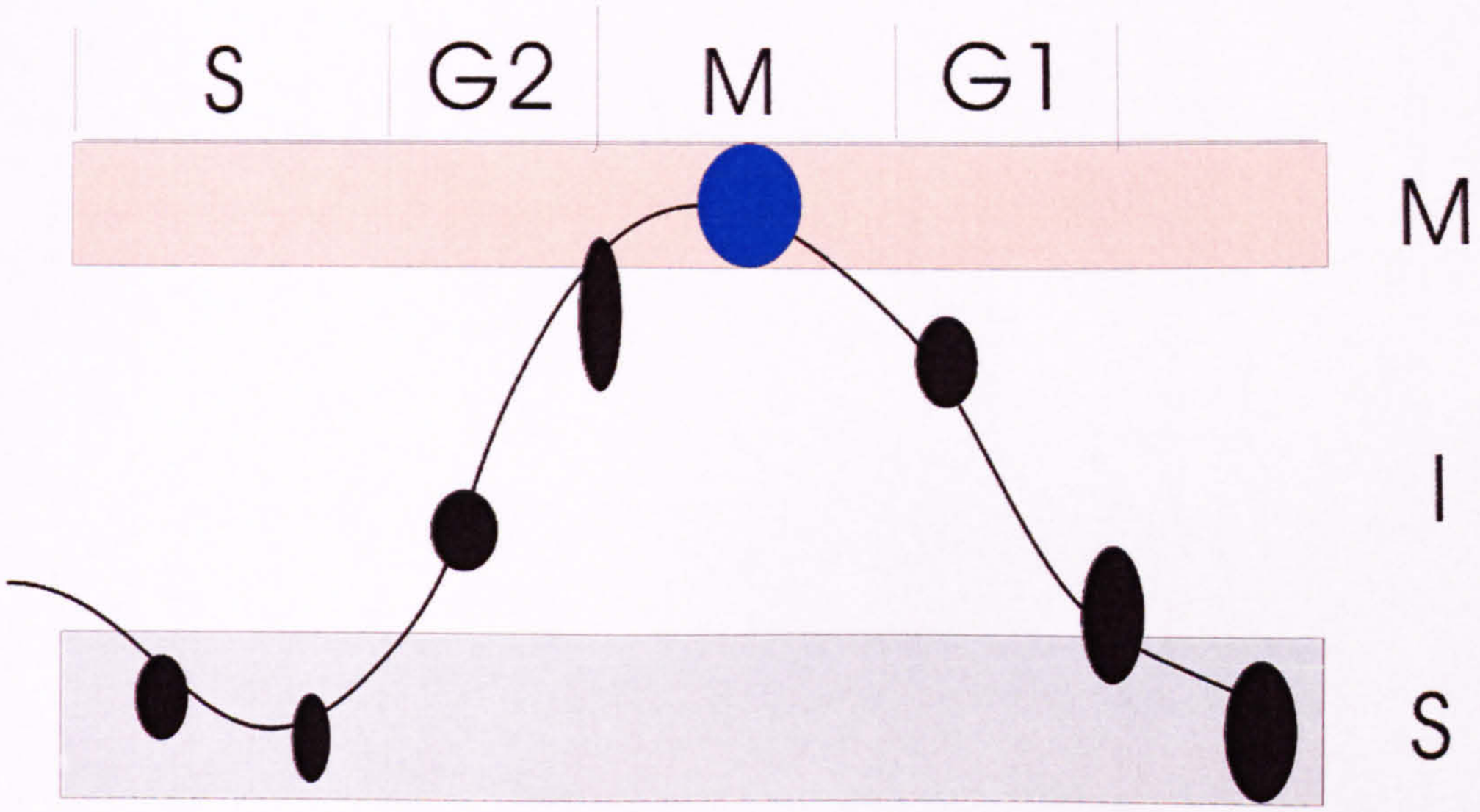



Fig. 4.2 Elevator movement and the relationship with the cell-cycle of the by neuroepithelial cells. M: mitotic zone; I: intermediate zone; S:zone of DNA synthesis. (Adapted from S.Fujta, cell structure and function 28:211)

During stage II cytogenesis, neuroblasts are derived from matrix cells. The neuroblasts subsequently migrate peripherally to the mantle and marginal layers, and are characterized by an absolute lack of $^{3}$H-thymidine incorporation, which represent complete repression of DNA synthesis. The production of neuroblasts in various regions of the brain was found to follow a well-organized temporal and spatial pattern (Lyons et al., 2003).

During stage III of cytogenesis (stage of glial differentiation), $^{3}$H-thymidine autoradiography revealed that the matrix cells switch off to neuroglial differentiation, and change into ependymoglioblasts, which are committed glial progenitors. At this

time the ependymoglioblasts are morphologically indistinguishable from the matrix cells. The ependymoglioblasts are then rapidly differentiated into the ependymal cells and glioblasts. The glioblasts then differentiated into astrocytes, oligodendrocytes and finally the microglia (Fujita, 2003).

There are marked regional differences in the progression of the stage from I to II, and from II to III in various parts of the CNS. In some regions, neuronogenesis (stageII) is prolonged and the transition between stage II and III takes place after birth (Johansson et al., 1999). The postnatal neuronogenesis is reduced as the animals grow older (Zitnik and Martin, 2002). However, modification of neuronal fate can occur only at early stages of matrix cell differentiation. When stage II neurogenesis starts, neuronal fate is irreversibly fixed. The basic mechanism this specification relies on the so called "major differentiation" hypothesis. That is, in the beginning of ontogenesis, the cell is at its maximum totipotent state, and the length of the S-phase is expected to be short. As the cell progresses by steps of differentiation, irreversibly inactivated replicons increased in number and the S-phase become longer. The type of cell differentiation is determined by a specific combination of irreversibly inactivated replicons. With repeat mitoses and entering into G1 phase, irreversibly inactivated replicons increase in number. Cells undergoing neuron-generating division have a longer cell cycle than those undergoing proliferating division and this cell-cycle lengthening mostly concerns G1 (Calegari et al., 2005). However, the molecular basis underlying this specification is still not well known.

### 4.2.3.2 The developing limbs

Basically, the limb consists of four segments: (1) the root (zonoskeleton); (2) a

proximal segment (stylopodium) consisting of a single bone (humerus, femur); (3) a middle segment (zeugopodium) consisting of two bones (radius and ulna, tibia and fibula); and (4) a distal, more complex part (autopodium) consisting of several bones (hand, foot) (Gurrieri et al., 2002). The limb arises from a group of mesodermal cells covered by surface ectoderm. The mesodermal cells contain progenitors of condrocytes and connective tissues that form tendons and muscle sheaths. The surrounding muscles and vessels are developed from cells that migrate into the early limb bud (Kardon et al., 2002). The establishment and outgrowth of the early limb bud depends on the number of cells that constitute it. Once established, the limb bud rapidly increases in size. Expansion of progenitors of proximal, middle and distal compartments is completed at different times, in a proximal to distal sequence (Dudley et al., 2002; Tickle, 2003). Little is known about how it is regulated. It may be a reflection of the timing of condensation (Janners and Searls, 1970).

The key stages in limb skeletal development are initiated as a protrusion from the lateral mesodermal plate of the embryo. Substantial outgrowth and patterning occurs over the course of the next few days, culminating in the establishment of cartilage templates that will form future bones. During this time, chondrocyte progenitors aggregate and form prechonfrogenic condensations. Joint formations subsequently begin and the chondrocytes differentiate and secret cartilage extracellular matrix molecules (as illustrated in Fig. 4.3) (Mariani and Martin, 2003).

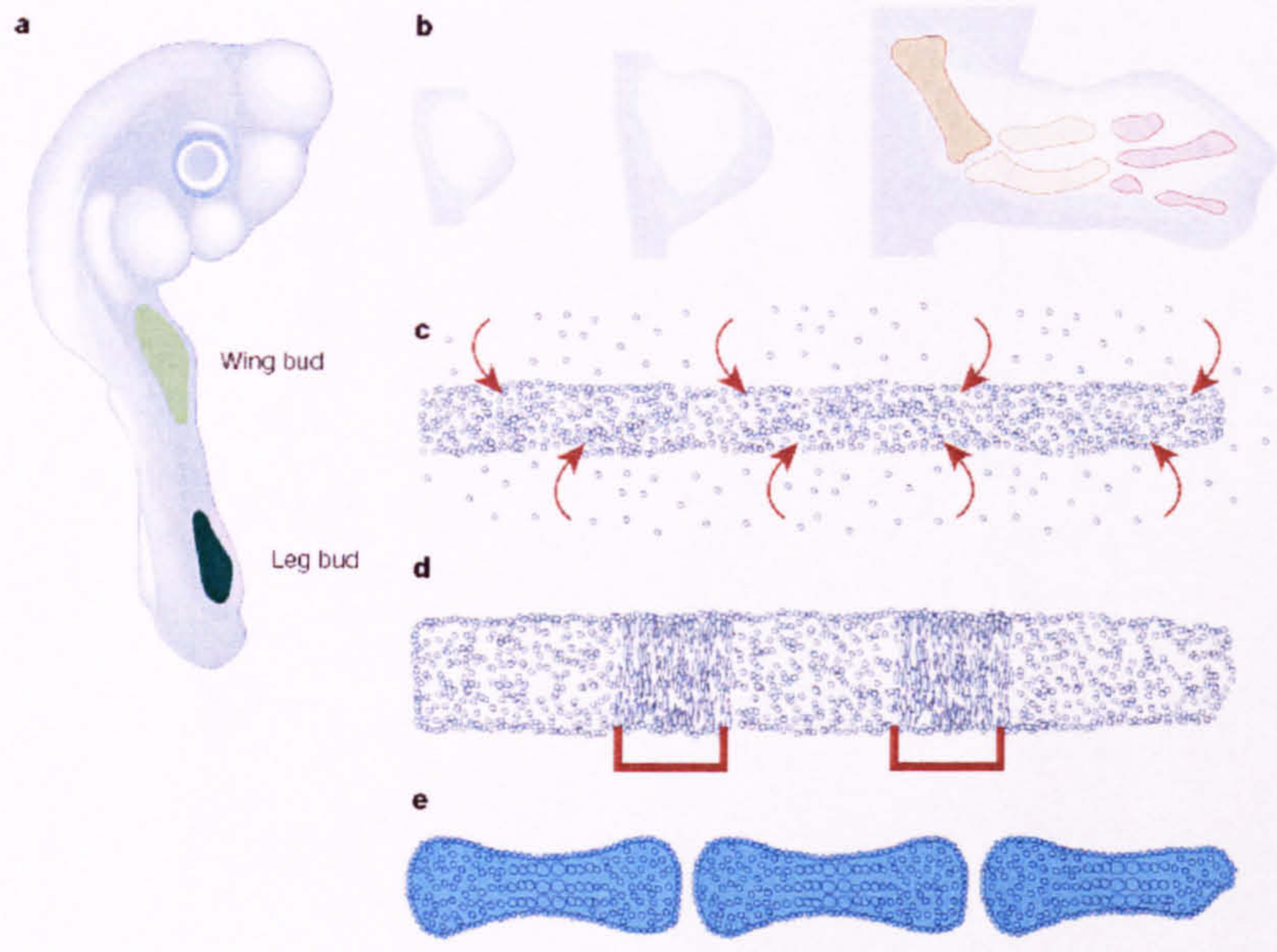



Fig. 4.3 Key stages in limb skeletal development. a, limb formation initiates in the chicken embryo at ~3 days of development. b, substantial outgrowth and patterning of the limb bud that prefigure the bones. c, chondrocyte progenitors aggregate and form prechondrogenic condensations. d, joint formation begins and e, the condrocytes differentiate and begin secreting cartilage extracellular matrix molecules (Figure from F.V. Mariani & G.R. Martin. Nature, 423: p321).

The formation of condensations of chondrocyte progenitors that outline the future limb skeleton patterning is possibly the most critical events in limb development. Prechondrogenic condensations are detectable morphologically as regions of increased cell density. Cell-cell interactions are important in initiating condensation, perhaps by establishing an "aggregation centre" that recruit cells from surrounding tissue (DeLise et al., 2000a; DeLise et al., 2000b). It seems that condensations form at specific stages of limb development, in a proximal to distal sequence. If timing for condensations is indeed fixed, alterations in cell number in the limb bud should affect skeletal patterning. If fewer condrocyte progenitors are present at the time of condensation, one might expect smaller elements to form. It has been

shown that when the limb bud cell number is reduced, the skeletal elements that form are smaller than normal and misshapen (Wolpert et al., 1979). Moreover, when the limb bud cell number drops below a certain threshold, condensations do not occur (Wolpert et al., 1979). The limb reduction defects in patients with CdLS range from amelia (absence of limb), meromelia (absence of part of the limb), and these are often asymmetric. Digital abnormalities seen in CdLS patients include acheiria, oligodactyly, hypoplasia of the thumb and first metacarpal, clinodactyly of the fifth finger, and ectrodactyly. Long bone abnormalities include ulnar a/hypoplasia, dysplasia of the radial head, or fusion of the elbow. Fusion at the elbow and oligodactyly are often observed whenever a single long bone of the middle segment is present (Braddock et al., 1993).

Joint formation is another process crucial for patterning. It occurs at several stages, either between the boundary of two adjacent condensations or within single condensation. Proper positioning of joints within condensations influences the number and the size of skeletal elements (Cohn et al., 2002; Francis-West et al., 1999). Fusion of the elbow joint as described in CdLS phenotypes may result from inappropriate joint formation during limb patterning.

#### 4.2.3.3 The developing craniofacial system

Although the neural crest received a significant amount of attention, it is not the only craniofacial tissue with patterning information. New studies have demonstrated the contribution of epithelia as a source of patterning information for the face. The epithelia can be of ectoderm covering the facial prominences (Hu et al., 2003), or are neural ectoderm (Cordero et al., 2004), or are of endodermal origin that line the

pharynx (Ruhin et al., 2003). In a growing number of cases, epithelial tissues are actually the instigators of morphological change (Tapadia et al., 2005). The craniofacial complexes initially have a more simple geometry, consisting of a series of swelling or prominences that undergo growth, fusion and expansion. These consist of the midline frontonasal prominence, and three paired structures, the lateral nasal, maxillary and mandibular prominences derived from the first pharyngeal (brachial) arch. The frontonasal prominence is derived from a midline primordium that forms on top of the forebrain. The frontonasal prominence contributes to the forehead, midline of the nose, philtrum of the upper lip and primary palate. Until recently, it was thought that the ventral region of the first pharyngeal arch gave rise to the mandibular prominence and therefore the lower jaw. Moreover it was thought that the dorsal region of the first arch gave rise to the maxillary prominences, which contribute to the sides of the middle and lower face, the lateral borders of the lips, and the secondary palate. Two new studies demonstrated that at least part of this fate map is incorrect. Both groups showed that the ventral region of the first arch actually gives rise to both maxillary and mandibular skeletal elements, rather than only to the mandibular elements as previously thought (Cerny et al., 2004; Lee et al., 2004). The cells comprising the face have undergone a massive relocation through both active neural crest cell migration and the passive displacement of tissue that is associated with neurulation and head flexure (Couly et al., 1996).

One of the first crucial steps in craniofacial development occurs when head ectoderm is subdivided into non-neural and neural regions to establish the head epithelium. While the non-neural regions will lie outside of the cranial neural crest, the neural regions will lie inside it. A subset of epithelial cells located at this neural/non-neural boundary adopts a mesenchymal character and initiates the

epithelial cells' migration between these two epithelium layers. This epithelial-mesenchymal transition marks the birth date of the neural crests and has been shown to depend upon cells shifting from G1 to S phase. This shift is dependent upon, at least for trunk neural crest cells, bone morphogenetic protein (Bmp) (Burstyn-Cohen et al., 2004). When Bmp signaling is inhibited by the overexpression of noggin, a Bmp antagonist, the G1/S transition is blocked and neural crest cells are no longer generated from the margins of the neural folds (Burstyn-Cohen et al., 2004). Bmp exerts its effect, in part, though the regulation of the *Wnt 1* transcrition (Garcia-Castro et al., 2002).

Debate still continues over verifying the correct tissue that controls facial patterning. Recent experiments have demonstrated that the pharyngeal endoderm has a profound influence on the morphogenesis of the middle and lower face (Crump et al., 2004a; Crump et al., 2004b; Le Douarin et al., 2004). However, not all cartilages are equally affected. Mandibular cartilages derived from Hox-negative neural crest cells are less affected than are Hox-positive second arch cells. Removal of the endoderm completely blocked the formation of pharyngeal arch skeletons (Ruhin et al., 2003). Even so, the pharyngeal endoderm is not required for normal development of the middle and upper part of the face. As an alternative, the anterior (or forebrain) neuroectoderm and the facial ectoderm appear to have taken over this crucial role (Aoki et al., 2002).

#### 4.2.3.4 The developing heart

The heart arises from cells in the anterior lateral plate mesoderm of the early embryo, where they are arranged in bilateral fields on either side of the prechordal

plate and rostral notochord. These fields include the precursors of both myocardial and endocardial cells, although there is apparently no common pool of bipotential precursors for these two heart cell lineages. The cells of the cardiogenic mesoderm are brought to these positions by gastrulation movements that occur in close association with cells of the rostral endoderm. In mammalians, the bilateral fields of the cardiogenic mesoderm merge at their anterior margins to form the so-called "cardiac crescent". More recent studies have identified a second type of heart field that is located more medially in the splanchnic mesoderm, directly adjacent to the cardiac crescent. The cells from this bilateral field, termed anterior heart-forming field, are fated to generate anterior heart structures of the outflow tracts (Kelly and Buckingham, 2002) The earliest steps of assembly of the heart tube are initiated by the convergence and fusion of the bilateral heart primordia along the midline. The cells of the anterior heart-forming field, after having migrated anteriorly, are added to the anterior end of the linear tube during and after this period. The resulting beating tubular heart is composed of an external myocardial and an internal endocardial layer and also possesses a polarity along the anteroposterior axis, in which the prospective tissues of the aortic sac, outflow tract (conotruncus), right ventricle, left ventricle, and atria are present in an anterior to posterior order along the tube (Kelly and Buckingham, 2002).

The tubular heart undergoes a process known as rightward looping. The morphogenetic steps required to achieve looping are guided by molecular asymmetries that are established in and around the heart by the embryonic left/right axial pathway. Furthermore, in higher vertebrates, septal division of the chambers and formation of the valves, which involves endothelial cells, are essential steps leading to the formation of an integrated 4-chambered heart with separate venous (or inflow) and arterial (or outflow) poles. During the growth process of the cardiac epithelium,

another distinct cell lineage, the migrating cardiac neural crest cells, populate the heart through the outflow channel and contribute to the formation of the great vessels and outflow septum (Christoffels et al., 2000).

The spatial and temporal orchestration of these processes implies a complex program of genetic control. This program is exerted in large part through precisely controlled processes of cell-cell signaling and regulators of gene expression. The major signaling mechanisms controlling early heart development, specification of myocardial cells and heart patterning will be discussed below.

During the stage of early inductive processes in heart specification and determination, cells of the anterior endoderm and/or their precursors have a major role in inducing cardiogenesis. Ectodermal influences on cardiac induction have also been described and are thought to serve mainly in counteracting negative influences from the neural plate. Nevertheless, in vertebrates, the heart-inducing activity is predominantly of endodermal origin (Zaffran and Frasch, 2002).

Genes encoding factors of the NK homeodomain, GATA, T-box, and other families were found to exert the functions of inductive signals during specification, patterning, and differentiation of the heart. Moreover, developmental signaling pathways are required to act in combination with tissue-specific transcriptional cofactors to elicit inductive responses.

Signaling pathways involved during the induction of cardiogenic mesoderm include (1) BMP/Dpp signaling, (2) Wnt/Wingless signaling, (3) Fibroblast growth factor signaling, and (4) Notch signaling pathways (Zaffran and Frasch, 2002).

Complete disruption of any of the major pathways in early heart development results in embryonic lethality. More subtle disruptions by haploinsufficiencies and hypomorph or dominant-negative mutations can lead to malfunctioning of the heart at later stages during the lifespan of the organism (Chien, 2000).

#### 4.2.3.5 The developing lung, kidney and pancreas

Branching morphogenesis, defined as growth and branching of epithelial tubules during embryogenesis, is elemental to the formation of several mammalian tissues including the kidney, lung, salivary gland, mammary gland, and pancreas. During branching morphogenesis, a tubular network is generated via mesenchymal-epithelial interactions that are coordinated by a complex network of several gene products. Signals derived from the mesenchyme instruct the epithelium to grow and branch. Although several classes of growth factors are involved, members of FGF family act in many tissues to positively regulate branching morphogenesis, while members of BMP family might act to limit branching in kidney and lung branching morphogenesis (Angle et al., 2003).

Lung development begins with the appearance of the laryngotracheal groove, which is a small diverticulum that arises from the floor of the primitive pharynx during the fourth week of gestation in human embryos. The groove sepreates dorsoventrally from the primitive esophagus to form the tracheal rudiment that gives rise to two primary bronchial buds. This initial branching event is followed by 23 generations of branches via invasion into the surrounding mesenchyme and bifid branching. Human diseases associated with the pulmonary branching defects include

tracheoesphageal fistula, pulmonary hypoplasia, bronchopulmonary dysplasia, and even asthma (Bartram and Speer, 2004; Chinoy, 2003).

Formation of the kidney starts during the fifth week of gestation in human embryos (CS14) when the ureteric duct is induced to undergo lateral outgrowth from the Wolffian duct and to invade the adjacent metanephric mesenchyme. Reciprocal induction between the ureteric bud and metanephric mesenchyme leads to multiple repetitive branching events and elongation of the ureteric bud to form the collecting system. The mesenchyme is induced to condense and epithelialise around the branch tips and eventually matures into the part of nephron responsible for regulation of iron and organic molecule transport (proximal and distal tubules), as well as the glomerular (filtering) components (Piscione and Rosenblum, 1999; Qiao et al., 1999a).

The pancreas is derived from endodermal cells of the upper duodenal region of the foregut. In the human embryo, the pancreatic primordial first appears during the fifth week of gestation in the form of epithelia buds from dorsal and ventral evaginations of the primitive foregut. The buds undergo lateral and iterative bifid branching to form a complex branched structure.

#### 4.2.3.6 The developing gastrointestinal system

The gut is composed of two tissue types in a tubular arrangement. The outer layer(s) of the tube is primarily smooth muscle derived from lateral plate splanchnic mesoderm. The inner luminal lining is an epithelium derived from ectoderm in the most anterior region of the gut (the mouth) and the most posterior region (the anus).

The majority of the gut epithelium is endodermally derived. The endoderm of early gut tube stages is remarkably uniform in its morphology along the anterior-posterior length of the primitive gut tube. There are no morphologic differences between the portions of tube formed by elongation of the anterior intestinal portal (AIP) or by the caudal intestinal portal (CIP) poles of the embryo, and no distinctions in regions that will eventually form the AP portions of the gut: foregut, midgut, hindgut and their adult phenotypes of esophagus/stomach, intestine, and colon. As the mesoderm grows and differentiates into smooth muscle, the gut tube alters its gross morphology, resulting in clear demarcations among the foregut, midgut, and hindgut. When considering gut formation, it is necessary to pay attention to the interaction between the endoderm and mesoderm. Each of these layers probably contains positional information prior to the establishment of their topographical juxtaposition. Numerous cofactors, enhancers, and modifiers are involved in gut development. Basic events that control gastrointestinal tract development remain still poorly understood (Roberts, 2000).

4.2.3.7 The developing germ cells

The initial stages of gonadal development occur during the fifth week when a thickened area of mesothelium develops on the medial side of the mesonephros. Proliferation of this epithelium and the underlying mesenchyme produce a bulge on the medial side of the mesonephros - the gonadal ridge. During the sixth week of gestation, the primordial germ cells enter the underlying mesenchyme and are incorporated in the gonadal cords (Moore, 2003). Primordial germ cells are migratory cells. They arise very early in embryogenesis and have a similar pattern of migration in *Drosophila*, *Xenopus*, *chick* and *mouse*. Germ cells proliferate mitotically from the

time they begin to migrate to the time they colonize the genital ridges (Gomperts et al., 1994).

The aim of this project is to use human embryos collected from different Carnegie stages to study the expression patterns of *NIPBL* in different major organ systems. The details will be shown in later sections.

## 4.3 MATERIAL AND METHODS

### 4.3.1 Tissue *in situ* Hybridization

#### 4.3.1.1 Collection of human embryos

Collection of human embryos was carried out in accordance with National Ethical Guidelines (Polkinghorne, 1989) and Local Ethical Committee approval by Prof. S. Robson and Dr. P. Bullen (Department of Fetal Medicine, University of Newcastle upon Tyne). Most of the slides were prepared by staffs of Human developmental Genetics in the Institute of Human Genetics, University of Newcastle upon Tyne. Patients who had decided to undergo termination of pregnancy (TOP) were given an information sheet covering the use of embryonic material for research and written consent was gained from those agreeing. The procedure requires the fulfillment of two requirements in order to take place. It is necessary to make sure that (1) the termination of the pregnancy was not motivated by research needs (2) the obstetrician involved in the termination process did not use the embryonic material for research (Polkinghorne, 1989). Embryos from Carnegie stage 10 to 22 (21 to 53 days of development) were collected and staged based on classification of O'Rahilly and Muller (O'Rahilly R, 1999). Embryos were investigated for normal anatomy under stereomicroscope.

#### 4.3.1.2 Tissue fixation and embedding

After dissection from the amniotic sacs, the embryos were placed in RNase free Duran bottles containing 4% PFA in PBS for 16 hours with gentle agitation. The

embryos were then dehydrated by passing through a series of 70%, 80%, 90% and 100% ethanol solutions for 2 hours each at room temperature with gentle agitation. The embryos were treated with chloroform for 16 hours to remove the ethanol and then molten paraffin three times (2 hours each) at 60 ℃ with the last one being in vacuum. Embryos were orientated in paraffin at certain position and the paraffin blocks were allowed to solidify. The embedded embryos were kept at 4℃ until required for sectioning.

4.3.1.3 Tissue sectioning

Paraffin blocks containing fixed embryos were oriented into the required plane to enable transverse or sagittal sectioning and mounted on a microtome stage, which was securely, fitted into the microtome (leica RM 2135, Leica Instruments, Germany). The microtome, the wooden trays and cutting accessories were cleaned with 0.1% DEPC in 96% ethanol to ensure that they are RNase free. Sections were cut at 5um and collected as continuously orientated ribbons on RNase free trays. Four sequential sections were mounted on a single slide and placed them on a meniscus of sterile water covering the coated slide. The sections were warmed up to 51℃ to stretch the tissues. The water was removed from the slide by using a pipette tip attached to a suction tube. The slides were kept in a warm plate for a few hours before drying them overnight at 37℃ in a drying cabinet. Dried sections were stored at 4℃ in RNase free boxes.

4.3.1.4 Tissue pre-treatment

Tissue pre-treatment allows the accessibility of target RNA to the labelled probe

and reduces non specific binding during hybridization (Moorman et al., 1993). All the pre-treatment steps were completed in RNase free conditions by cleaning the surfaces thoroughly with 0.1 % DEPC in 96% ethanol. In addition, all the glassware used in the procedures was prebaked at 180 ℃ for at least 4 hours prior to use. All the solutions were made by using DEPC treated water or PBS. The slides were placed in a metal rack and de-waxed three times with xylene treatment for 5 minutes each. They were then set in a mixture of xylene/ethanol (1:1), before being twice laid in 100% ethanol for 3 minutes each. Finally they were placed in a series of 90%, 70%, 50% ethanol (2 minutes each) and washed twice in PBS solution for 2 minutes each. The sections swell after this treatment, which help both the accessibility of proteins for digestion in the next step and the unfolding of target RNA. Protease treatment was performed using Proteinkinase K (20 $\mu$ g/ml) dissolved in PBS solution (20 $\mu$ l/ml) for 8 minutes. The treatment time varied according to the type of tissue, the age of the embryo (no treatment for CS 13, 8 minutes for embryo staged more than CS17). The tissues were washed twice in PBS solution for 2 minutes each, before being immersed in 4% PFA/PBS solution for 20 minutes to retain their morphology. To reduce the non-specific background staining, the slides were then immersed in a solution containing acetic anhydride (0.25% Acetic anhydride, 0.1M Triethanolamine pH 8.0) for 10 minutes and washed in PBS twice for 2 minutes each. Afterwards, the sections were dehydrated in a series of 50%, 70%, 90%, 100% and fresh 100% of ethanol. The slides were dried under filtered air stream and hybridized within 3 hours.

4.3.1.5 Probes used for *in situ* hybridization

Antisense and sense RNA probes were synthesised from cDNA clones, contained in plasmid vectors, which have two of the RNA polymerase promoter sites. The

cDNA clones used for this are described below:

(1) The human NIPBL cDNA clone was generated by PCR amplifying a 424bp fragment extending exon 10, 11 and 12 of the human *NIPBL*, which should hybridize to transcript encoding both the long and short isoforms. The PCR fragment was amplified from Homo sapiens IMAGE:1952030 cDNA clone. No overlapping sequences were found in the amplified fragment in the whole human genome. The primers used for the amplifying are: 5'GAAAGGAGAGCCGAAAGACA'3 and 5'CGCTCTTCATATTCCCAAGC'3. The amplified 424-bp cDNA fragment was cloned into the pGEM-T Easy vector (Promega) containing Sp6 and T7 RNA polymerase promoter sites. For isoform A specific probe, cDNA was generated by using PCR primers designed to amplify a 348bp cDNA probe in the 3'end isoform A specific region. The sequences of the primers are: 5'CTTGAAAATGTGTTTGGCACA 3' and 5' GAAAAGGGATCCCCAGGTAA 3'. For isoform B specific probe, cDNA was generated by primers designed to amplify a 405bp cDNA probe in the 3'end isoform B specific region. The sequences of the primers are: 5'CAAAAGGATGTATAAACGGGTAA 3' and 5' CCAAAATCACAGAAGAAGAGGTGAA 3'.

(2) Mouse MLC 2V cDNA probe: This probe is a kind gift from Steve Lisgro (Department of Developmental Genetics). The cDNA was cloned into TA vector containing the Sp6 and T7 RNA polymerase sites. The purpose for using this probe is for positive control for *in situ* hybridization. Mouse MLC 2V is stably expressed in the embryonic mouse's heart.

### 4.3.1.6 Probe labeling

#### 4.3.1.6.1 Temperate linearilisation

Single-stranded RNA was synthesised from the cDNA template within a plasmid vector containing Sp6, T7 or T3 RNA polymerase promoter sites (Mierendorf and Pfeffer 1987). The plasmid was linearised by a restriction enzyme which separates the rest of the vector from the insert. Complete digestion was checked by agarose gel electrophoresis. The linearised template was cleaned by phenol/chloroform extraction, precipitated in absolute ethanol and 0.3 M sodium acetate and resuspended in TE buffer to a concentration of 0.3-1$\mu$g/$\mu$l (Ausubel, 2002).

#### 4.3.1.6.2 RNA transcription

Single stranded labelled RNA probes were synthesised using non-radioactive DIG labelling (Roche Applied Science) from 0.5$\mu$g of DNA template in a reaction mixture containing 2$\mu$l of 10X transcription buffer, 2 $\mu$l NTP labelling mixture (10 mM ATP, CTP, GTP,6.5 mM UTP and 3.5 mM DIG-11-UTP, pH 7.510 x conc.), 1 $\mu$l RNase inhibitor, 2 $\mu$l RNA polymerase and RNase-free water to a final volume of 20$\mu$l. The reaction mixtures were incubated at 37℃ in a water bath for 2-3 hours followed by adding 1 $\mu$l DNase and incubated for further one hour at 37℃. The quality and quantity of the labelling was determined by gel electrophoresis with λHind III as marker and also by ND-1000 Spectrophotometer (NanoDrop Technology). The labelled RNA probes were purified by adding 30 $\mu$l of RNase-free water to a total of 50 $\mu$l volume. The unincorperated labelling mix was removed by adding the final reaction mixture to the centre of a column and centrifuged at

2500rpm for 2 minutes. The purified probes were collected into new microcentrifuge tubes. The purified probes can be stored at -80℃ for up to 6 months.

4.3.1.6.3 Control Riboprobes

To assess the efficiency of transcription from the plasmid-generated template, the sense and pSTP18-neo (provided with the labeling kit) riboprobes were used as negative control probes. Mouse MLC2V cDNA was used as positive control.

4.3.1.7 Probe hybridization

300$\mu$g of riboprobes was used for each slide for hybridisation. Briefly, the riboprobe was mixed with DIG Easy Hyb (Roche Applied Science) to a total volume of 100$\mu$l for each slide to be hybridized. The hybridization mixture was then carefully added to the slide and the slide was covered with a clean RNase-free coverslip. The slides were kept horizontal in RNase free plastic trays and put in moist Hybridisation chamber (50% formamide/2xSSC). The hybridization was carried out in an oven at 65 ℃ overnight.

4.3.1.8 Post-hybridisation washes

The slides were rinsed in 5xSSC (pre-warmed to 65℃) and coverslips were removed from the slides carefully by gentle agitation and placed in a slide rack. The slides were then rinsed in 200ml of 50% formamide/2xSSC at 65℃ for 20 minutes in a fume hood. Afterwards, the slides were washed in a series of 2xSSC (30minutes at 50℃, twice), 0.2xSSC (30 minutes at 50℃, once) and 0.2xSSC (30 minutes at room

temperature, once) solution in a shaking water bath to remove the non-hybridized riboprobes.

4.3.1.9 Antibody Detection

The slides were rinsed in wash Buffer I (0.1M Tris pH7.6, 0.15M NaCl, 1mM of levamisole) for 5 minutes (twice) and blocked in a blocking solution (10% heat inactivated fetal calf serum/Buffer I) for 1 hour. 100 $\mu$l of anti-DIG mix (anti-DIG Ab: 2% fetal calf serum/Buffer I=1:1000) was then carefully applied to each slide and the slide was covered with a strip of parafilm to prevent from drying out. The slides were then put in a large moistened tray at 4℃ overnight. The slides were subsequently washed in Buffer I for 30 minutes (6 times), and equilibrated in Buffer II (0.1M Tris pH9.5, 0.1M NaCl, 50mM $MgCl_2$, 1mM of levamisole) for 5 minutes. The slides were subsequently placed in a light proof coplin jar containing NBT/BCIP (4-Nitro blue tetrazolium chloride/5-Bromo-4-chloro-3-indolyl-phosphate 4-toluidine) solution (20$\mu$l/ml stock solution in Buffer II). The slides were then left to develop for several hours. The degree of the tissue staining was periodically checked under a light microscope. The staining was stopped by placing the slides into Buffer II and rinsed in water. To reduce background staining, the slides were quickly rinsed in 1%HCl/methol solution and washed in several changes of water. The sections were mounted in Antifade Mounting Solution (Aquamount) with coverslips.

## 4.3.2 Immunohistochemical (IHC) staining of the slides

The embryos' collection, sectioning were as previously described. As the expression patterns of NIPBL (see results) prompted us to examine the possibility that

the gene product of NIPBL might be associated with active proliferation, we examined the expression of the proliferating marker-PCNA and compared the expression with the NIPBL *in situ* staining. Due to the lack of feasible RNA probes available for PCNA, we used IHC staining of the PCNA antibody (abCam) instead.

#### 4.3.2.1 Dewaxing and Antigen retrieval

Formalin fixed tissue requires an antigen retrieval step before immunohistochemical staining can proceed. This is due to the formation of methylene bridges during fixation, which cross link proteins and therefore mask antigenic sites. The two methods of antigen retrieval are (1) enzymatic and (2) heat mediated. Both serve to break the methylene bridges and so expose the antigenic sites in order to allow the antibodies to bind.

Briefly, slides were dewaxed by immersing them in 3 changes of Xylene, for 5 minutes each and rehydrated in a series of 100%, 100%, 90%, 70%, and 50% of ethanol solution for 2 minutes each. The slides were then rinsed in PBS twice for 5 minutes each.

For antigen retrieval, the slides were placed in a microwavable vessel containing retrieving solution (0.1M Tri-sodium citrate, 0.2MHCl pH6.0) and heated in a full powered microwave for 30 minutes. To prevent drying out of the solution, more retrieving solution was added after 15 minutes in the microwave. In order for the slides to cool down, the slides were placed under running water.

4.3.2.2 Staining of mouse monoclonal primary PCNA antibody

The slides were rinsed in TBS (6.06g TrisHCl, 1.39gTrisBase, 8.77g NaCl make up to 1 liter with ultra pure water) twice for 5 minutes each. To inactivate the endogenous peroxidase activity, the slides were placed in 1.6% of hydrogen peroxide in TBS for 30 minutes and then washed in 2 changes of TBS0.025%Triton for 5 minutes each. The slides were subsequently blocked in 10% normal saline/1% BSA (Bovine serum albumin) in TBS for 2 hours at room temperature. PCNA mouse monoclonal antibody was applied (1:1000 dilution with 1%BSA in TBS) and incubated at 4℃ overnight on an orbital shaker.

4.3.2.3 Secondary Antibody staining and antibody detection

After overnight incubation with primary antibody, the slides were rinsed in TBS 0.025% Triton-X twice for 5 minutes each. Secondary biotinylated antibodies (Goat anti-mouse) made up in 1%BSA/TBS were applied to slides at room temperature and incubated for 2 hours. Meanwhile, the ABC complex (0.1% Streptavidin, 0.1% HRP (or AP)-Biotin) in TBS was made. After 2 hours of incubation, the slides were rinsed twice with TBS 0.025% Triton for 5 minutes each followed by the application of ABC complex in TBS for 30 minutes and developed with chromogen (DAB, Diaminobenzidine) for variable time and checked periodically. The slides were then counterstained in 2% methyl green solution (2% (w/v) methyl green in 0.1 M nodium acetate, pH 4.2) and followed by a series of 50%, 70%, 90%, and 100% of ethanol dehydration and cleared in 2 changes of xylene. The slides were mounted histomount (Invitrogen).

### 4.3.3 Image analysis

The mounted slides were allowed to dry for at 24 hours and then the sections were visualised using a Zeiss Axiophot microscope under daylight. Images were captured with the Image Axioversion Rel 4.1 version software and photographed using cooled CCD camera.

## 4.4 RESULTS

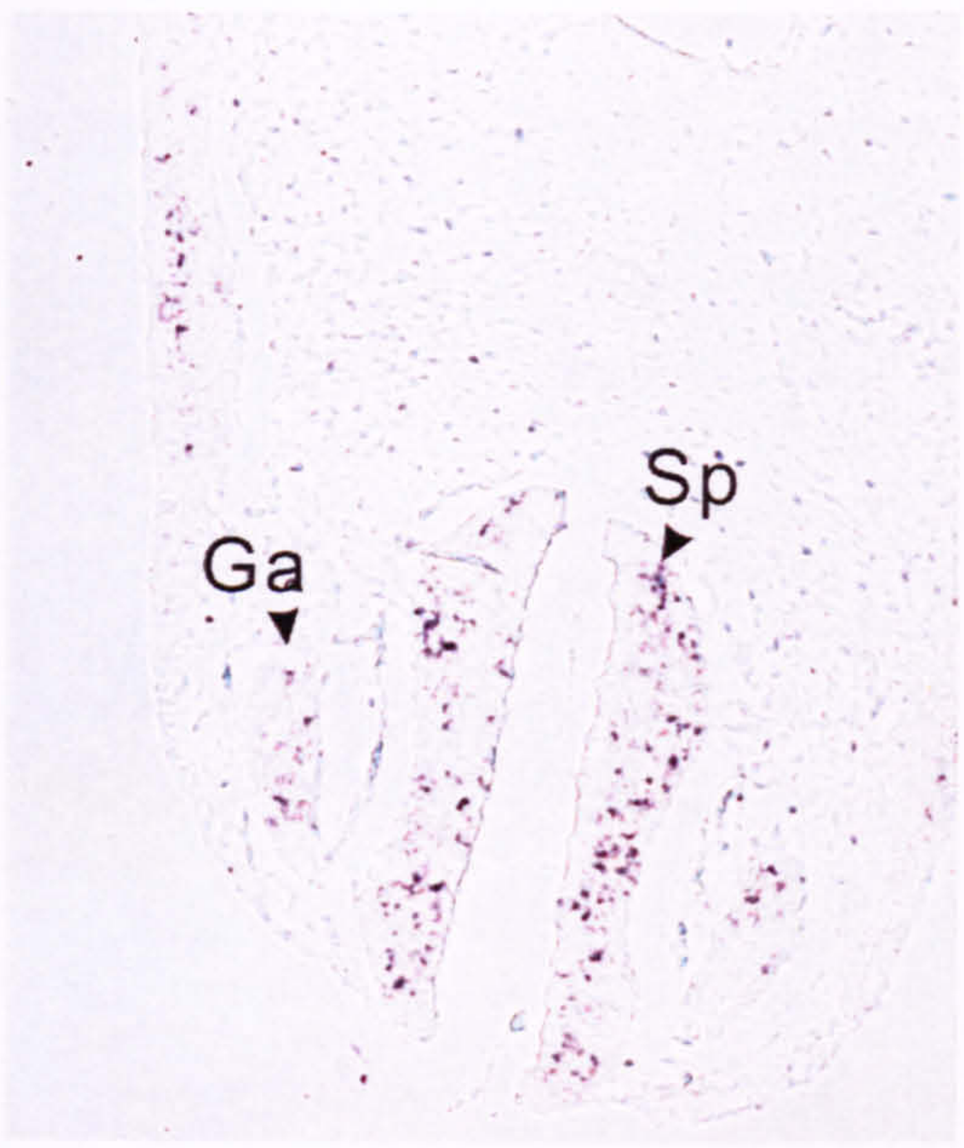
The expression in different tissues showed different intensities by the result of Northern blot. However, precise location of expression can not be detected by Northern analysis alone. Thus, I decided to perform tissue *in situ* hybridization by using riboprobes on parafilm embedded human embryos section and tried to correlate the expression pattern and the phenotypes associated with CdLS.

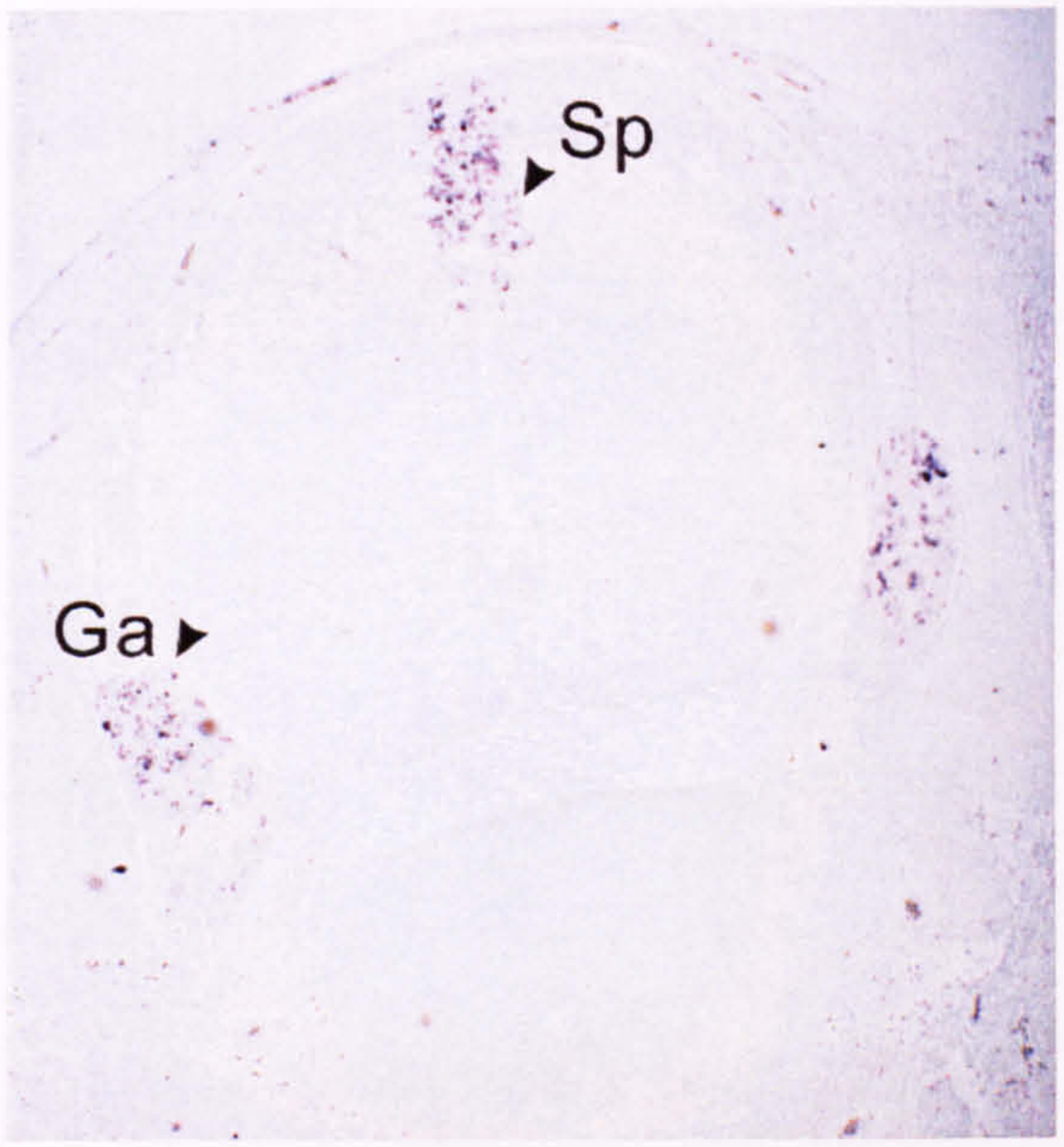
The expression of *NIPBL* was studied in human embryos ranging from 28-60 post ovulation days, which corresponds to Carnegie Stages 13-23 (see Appendix B for an approximate equivalence of Carnegie stages and postovulatory days). The expression of the human *NIPBL* gene during all these stages was seen in many different organ systems. The expression patterns by using riboprobes in different organ systems are described in detail below:

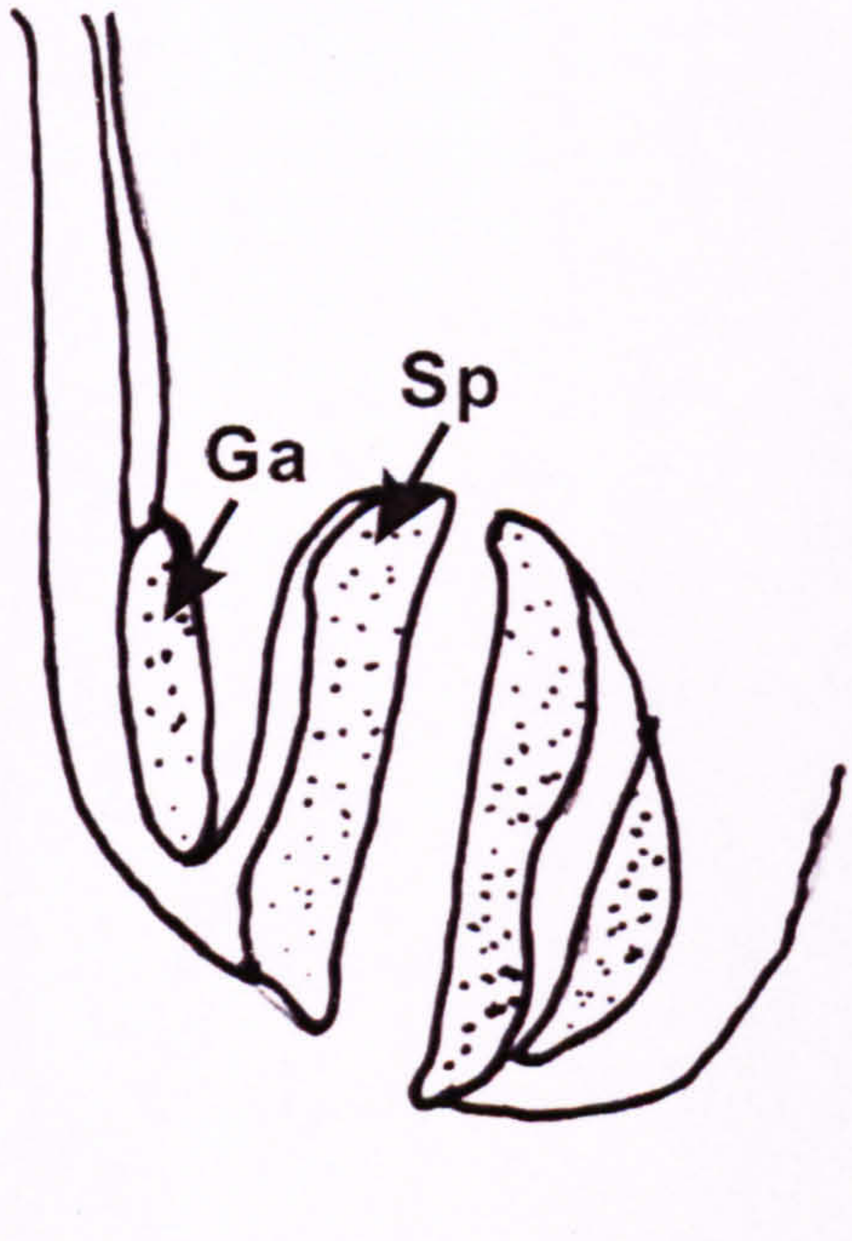
### 4.4.1 Expression of *NIPBL* in human embryonic nervous system

At stage 13 (28-30 postovulatory days), *NIPBL* transcripts were detected in the spinal cord and primordium spinal ganglion (see Fig. 4.4). At Carnegie stage 20 and 21 (52-53 postovulatory days), the expression was widely detected on cerebral hemisphere, midbrain, cerebellum, thalamus, hypothalamus and brain stem. The expression was mainly seen in the inner most ventricular zone (see Fig.4.5) where active proliferation of matrix cells (multipotential stem cells) occurs. To test whether *NIPBL* expression involved the active proliferating cell zone, I used proliferating cell nuclear antigen (PCNA) antibody as a probe for immunohistochemistry staining near the embryo section that we used for RNA tissue *in situ* hybridization and compared

the pattern of expression. The sites of tissue expression were very similar in the brain for both *NIPBL* and PCNA (see Fig. 4.6). Initially, I also intended to compare the expression pattern of PCNA and *NIPBL* on CS 13 embryo paraffin sections. Unfortunately, lack of CS13 embryo sections hindered me for further investigation.



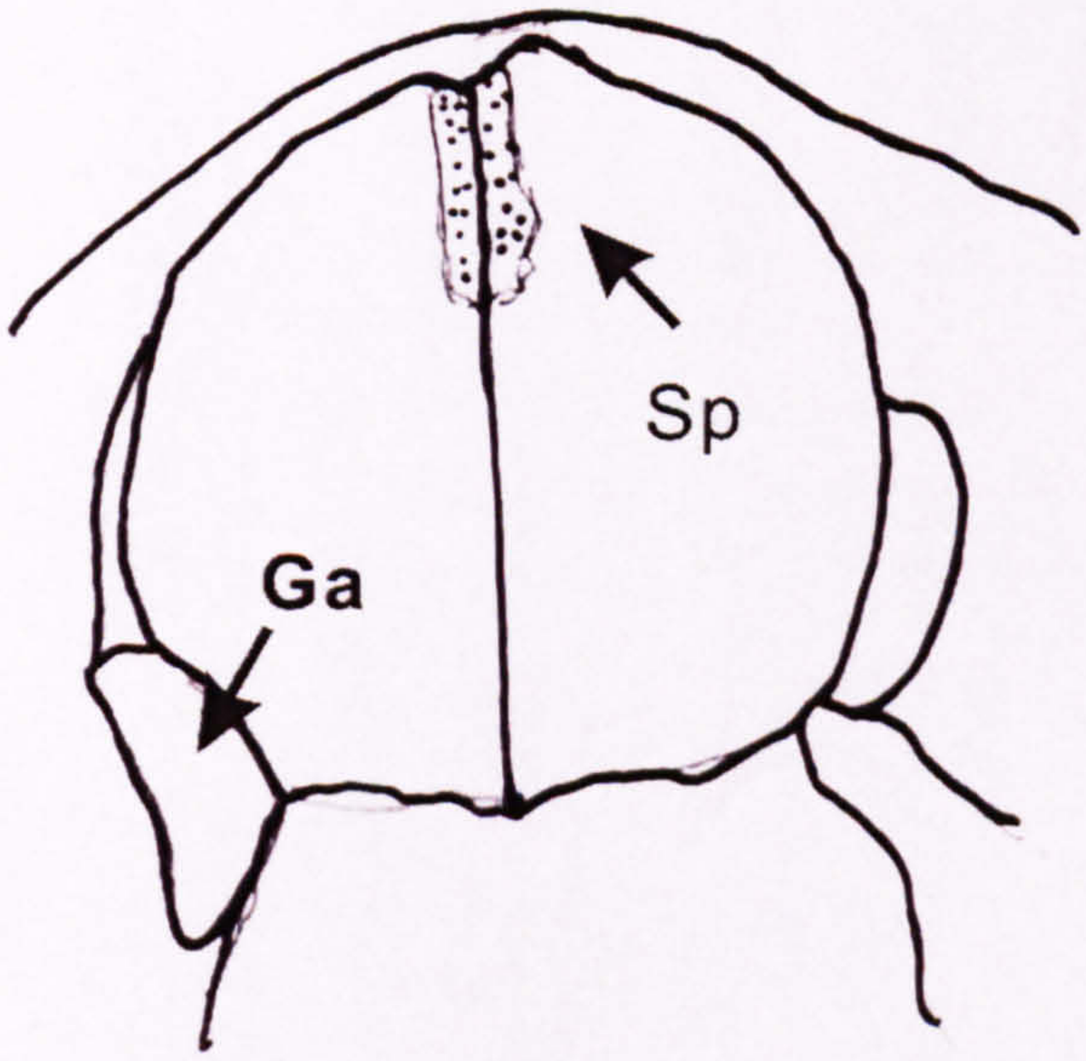




Fig. 4.4 Expression pattern of *NIPBL* transcript on CS13 (A) and CS20 (B) embryo sections. The expression of NIPBL transcript was seen on pseudostraified cylindrical epithelium of the developing spinal cord and spinal ganglion. Sp:spinal cord and Ga: spinal ganglion

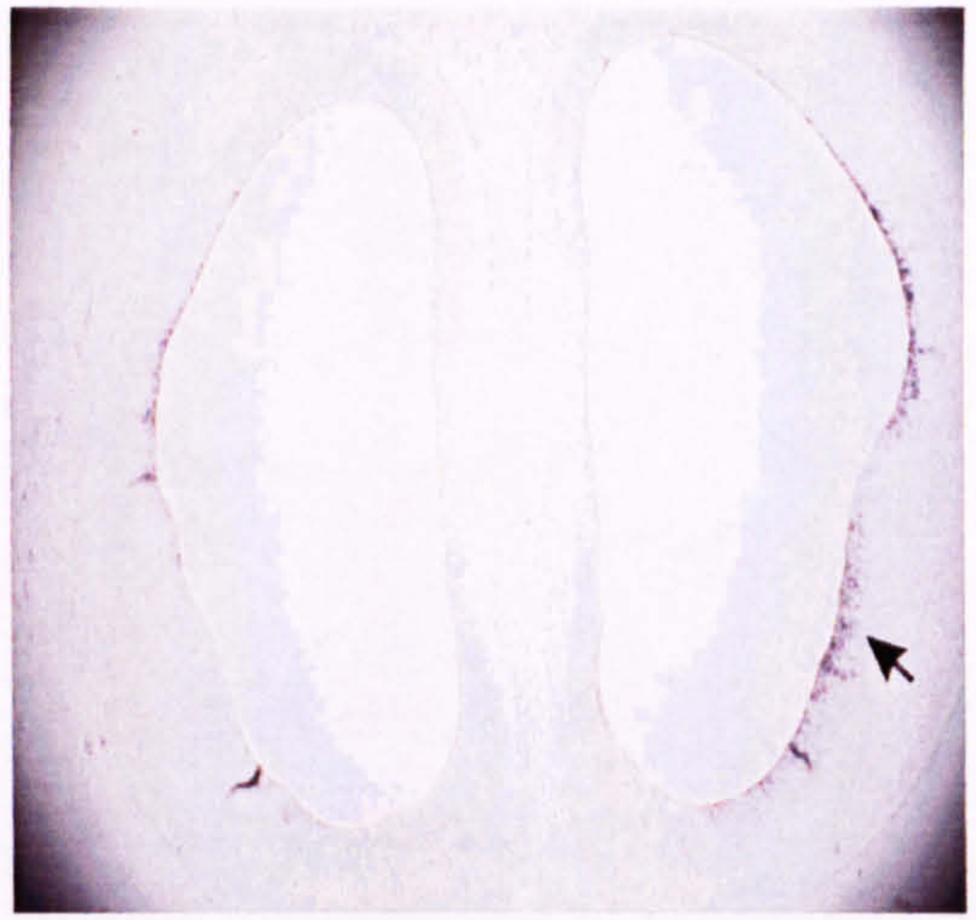
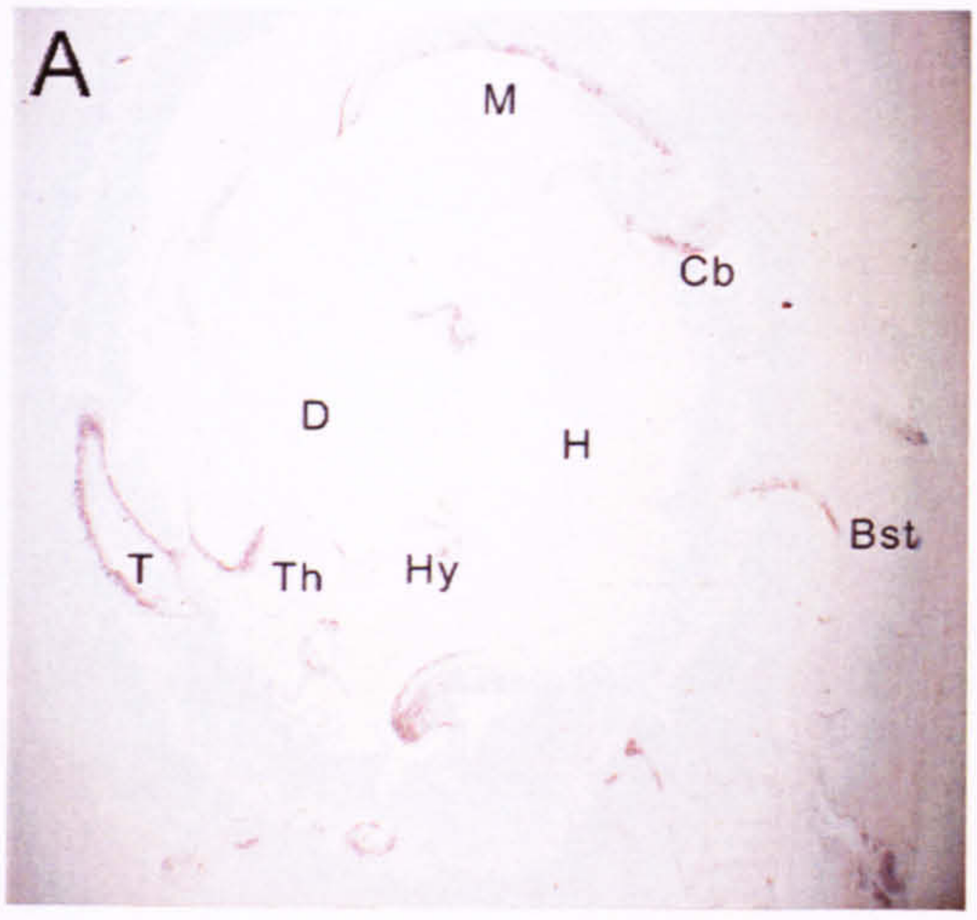
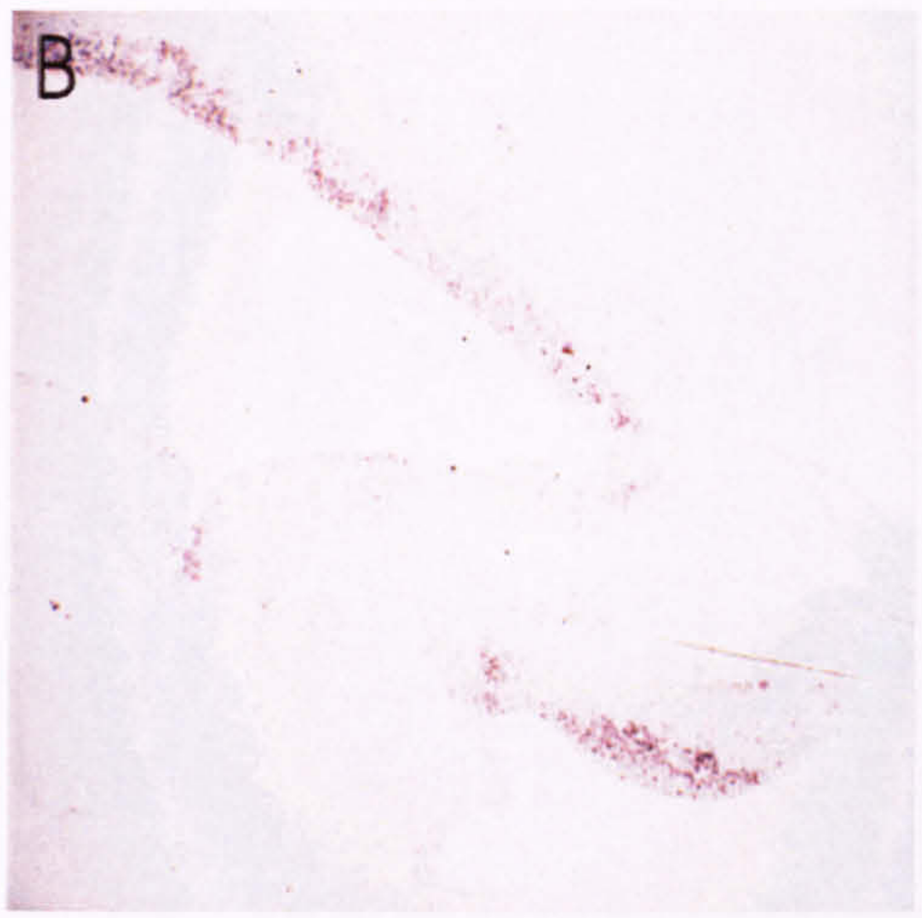
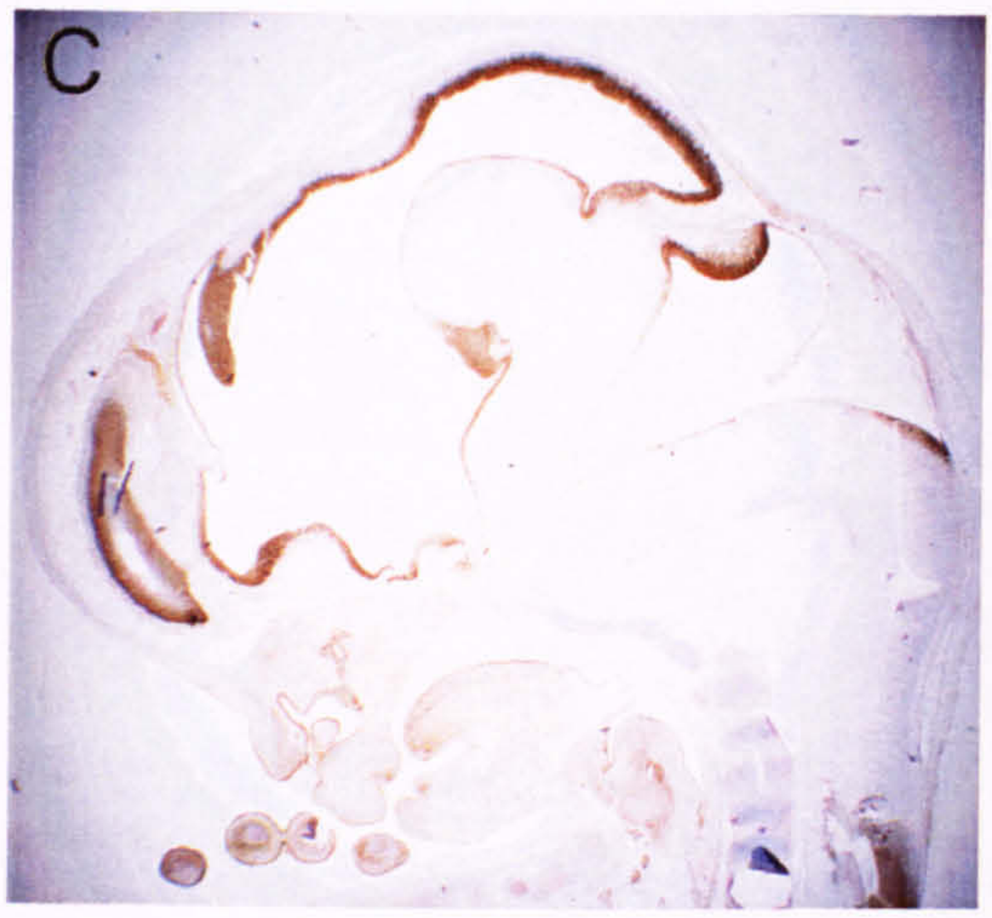



Fig. 4.5 *NIPBL* expression on the developing brain of CS20 embryonic section. *NIPBL* transcript was detected on the ventricular zone of CS20 embryonic brain section (arrow indicates site of expression by using riboprobe). Vz: ventricular zone of the developing brain



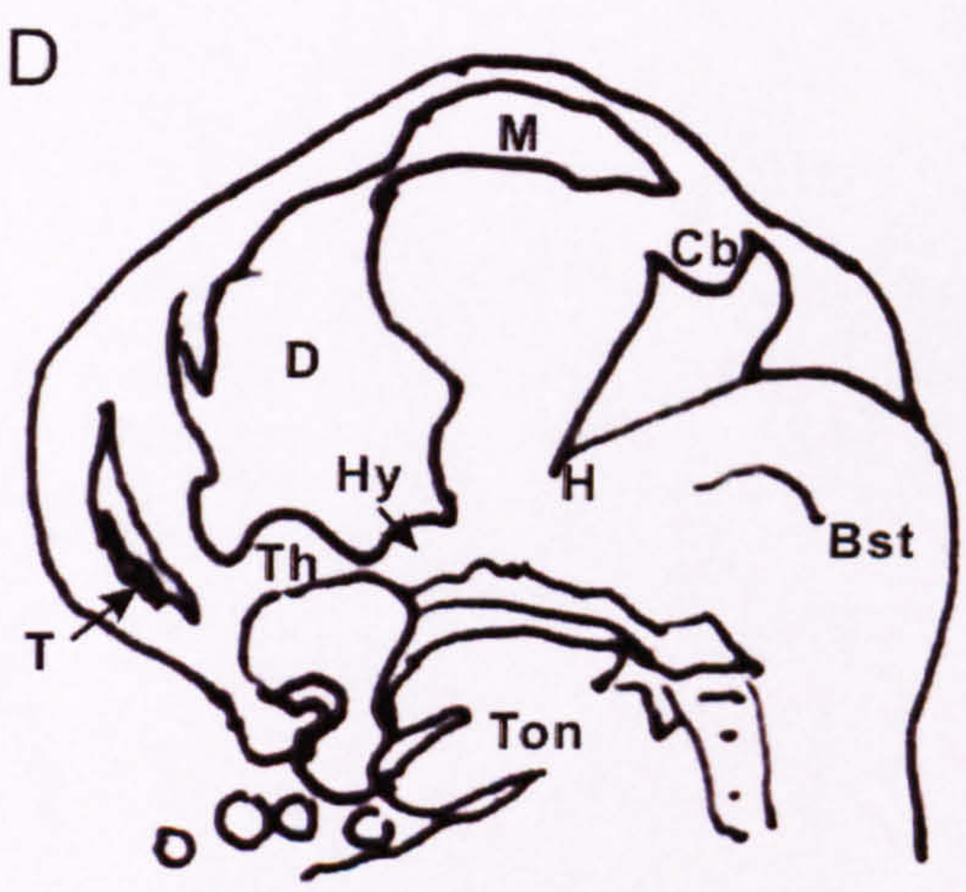




Fig 4.6 *NIPBL* expression on the developing brain of CS23 human embryonic section. Expression patterns of *NIPBL* transcript using riboprobes and PCNA antibody on CS23 embryo sections showed (A) CS23, 1.25X. The *NIPBL* expression was seen on ventricular zone of the developing brain tissue. (B) CS23, 5X of section (A); the expression of *NIPBL* transcript was mainly observed in the innermost ventricular zone. (C) Expression of NIPBL transcript is parallel to that of the expression pattern of PCNA using immunohistochemistry which indicate zone of active proliferation. Bst: brain stem; Cb: cerebellum; D: diencephalon; H: hindbrain; Hy: hypothalamus; M: midbrain; T: telencephalon ; Th: thalamus; Ton: tongue

### 4.4.2 Expression of *NIPBL* in human embryonic limbs

The expression of *NIPBL* in early limb buds (CS13, Fig. 4.7) was found mainly in the multilayered epithelial structure. As the limbs elongate, the expression on CS21 embryo is mainly traced in the chondroblasts in diaphyseal ossification centres and the surrounding developing muscles (Fig. 4.7 & 4.8).

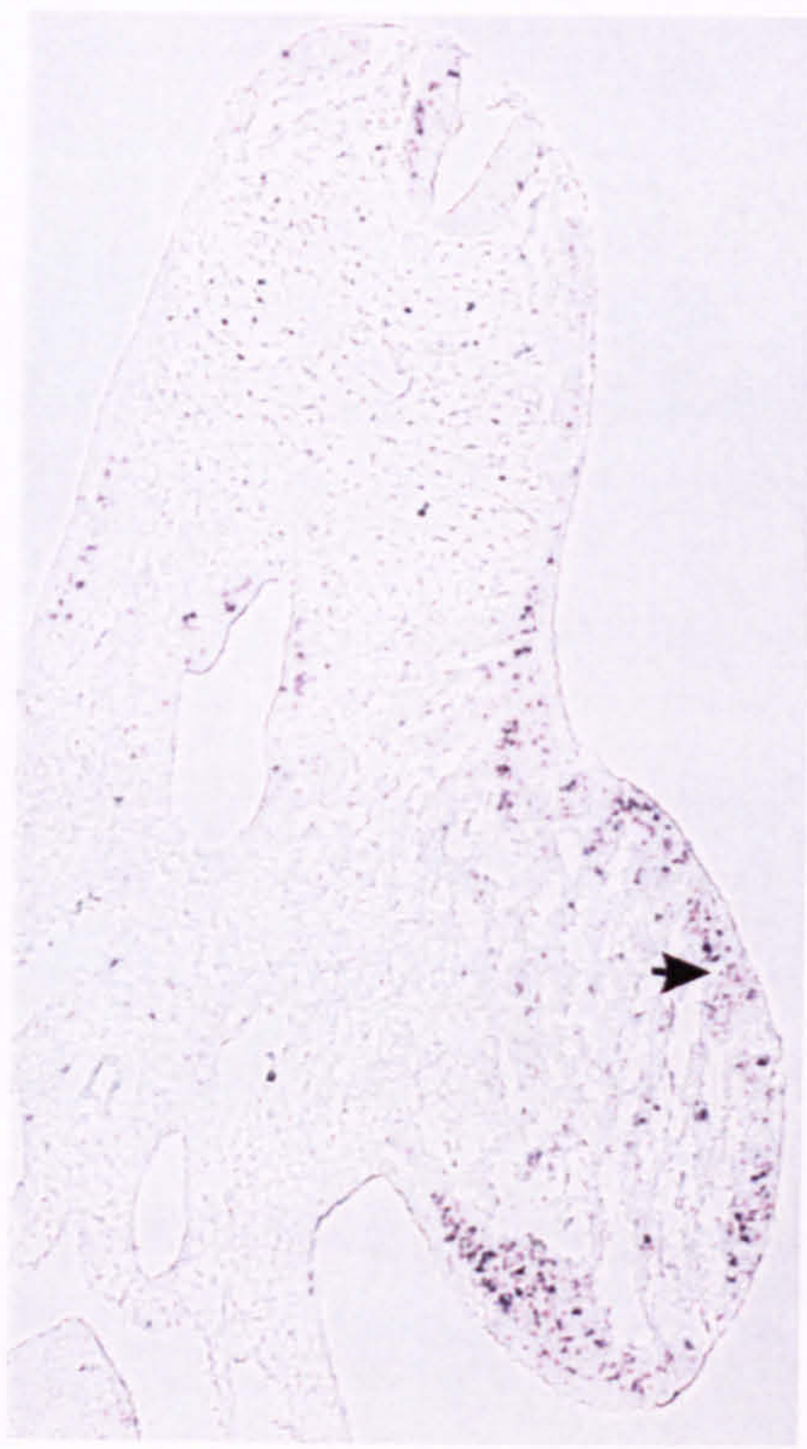
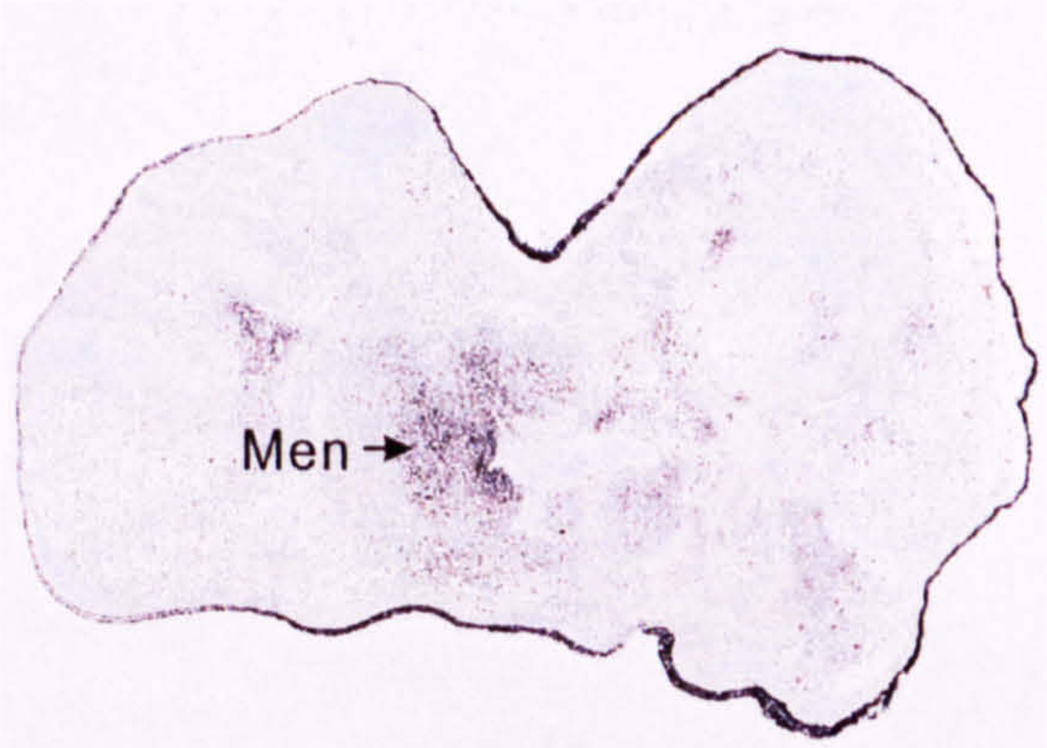



Fig. 4.7 *NIPBL* expression the early limb bud (CS13). The expression was observed mainly in the surface layers, the limb placode (arrow head). Lb: limb bud; LbP: limb placode.



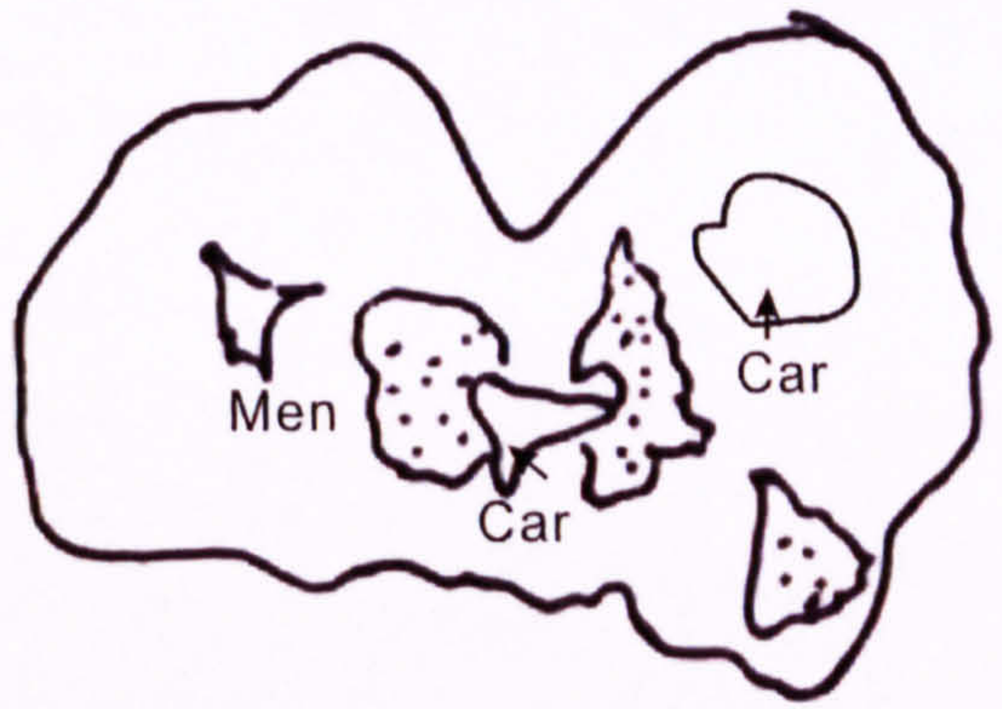
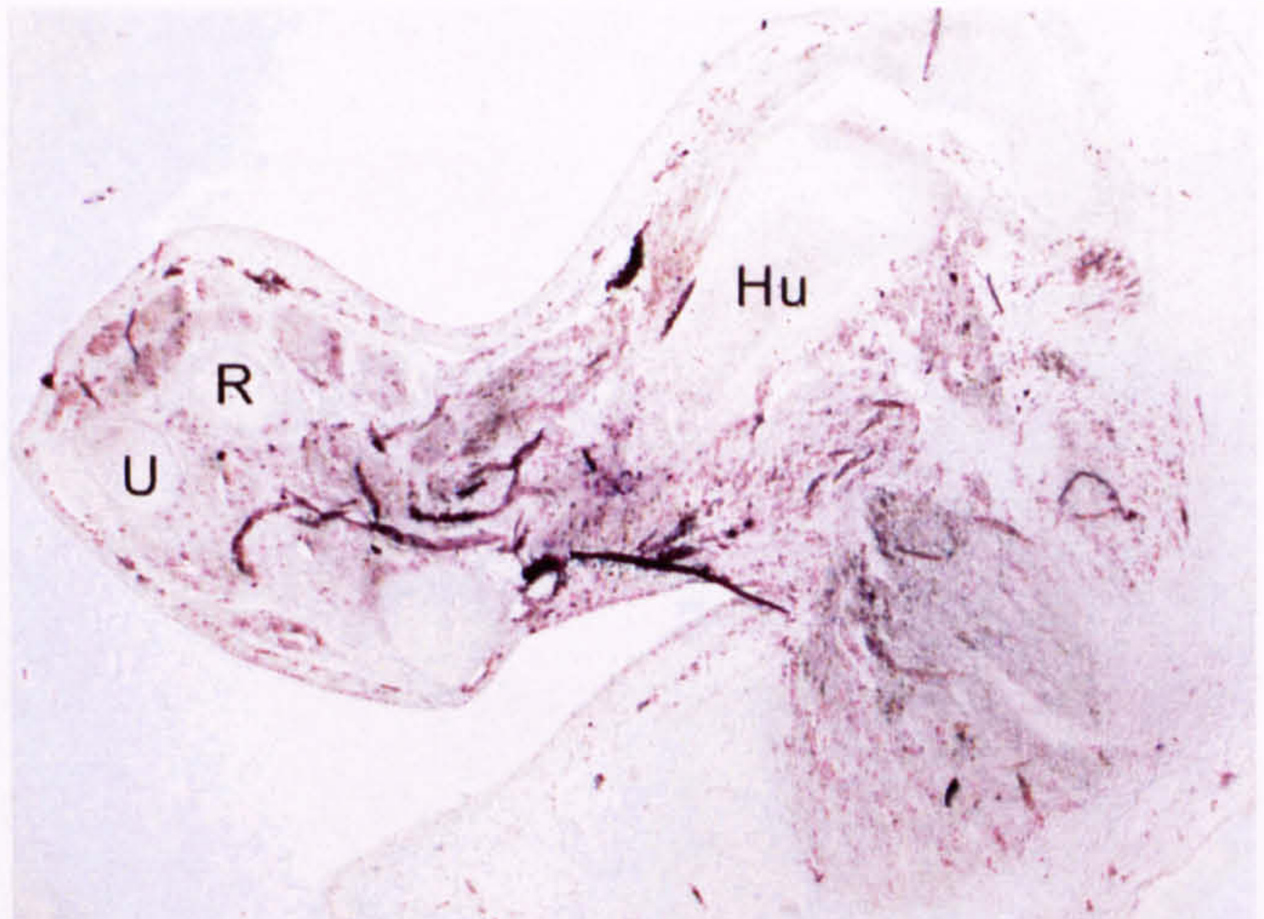




Fig. 4.8 *NIPBL* expression in developing hand plate of CS18 human embryo. Note *NIPBL* expression in multiple foci of mesenchymal condensation (arrow head). Condrocytes and mesenchyme condensation can also be observed (arrow). Men: mesenchyme; Car: developing cartilage



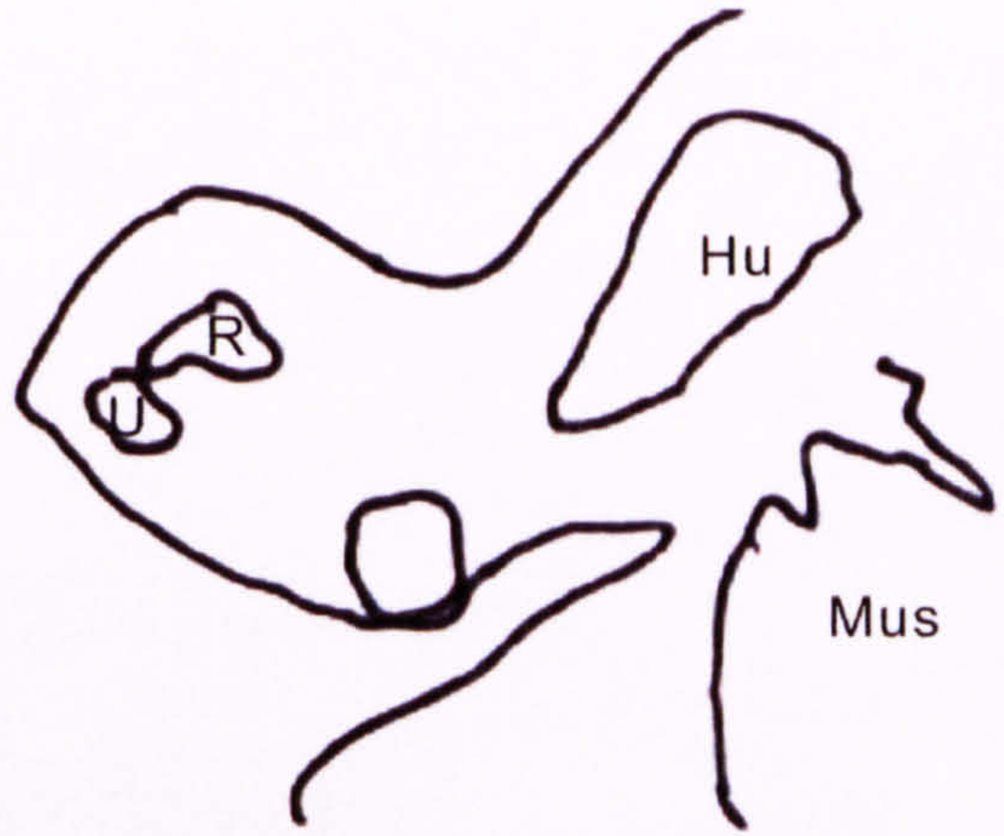




Fig. 4.9 *NIPBL* expression in upper limb of CS23 human embryo. Note the expression was seen on the developing skeleton and musculature. Hu: developing humerus; Mus: developing muscle; R: developing radius; U: developing ulna.

### 4.4.3 Craniofacial and ophthalmic expressions of *NIPBL* in human embryos

#### 4.4.3.1 Craniofacial expression of *NIPBL*

*NIPBL* expression was only observed in glabella/the nasal root area, certain facial musculatures and musculature of the tongue (see Fig. 4.10). The expressions of *NIPBL* in the developing craniofacial skeleton, however, are not obvious. I have tried several experiments, which resulted in similar expression patterns. *NIPBL* expression was seen in glabellas near the frontonasal prominence. *NIPBL* expression was also seen in the nasal septum, olfactory epithelium and pharyngeal endoderm (see Fig. 4.10 & 4.11).

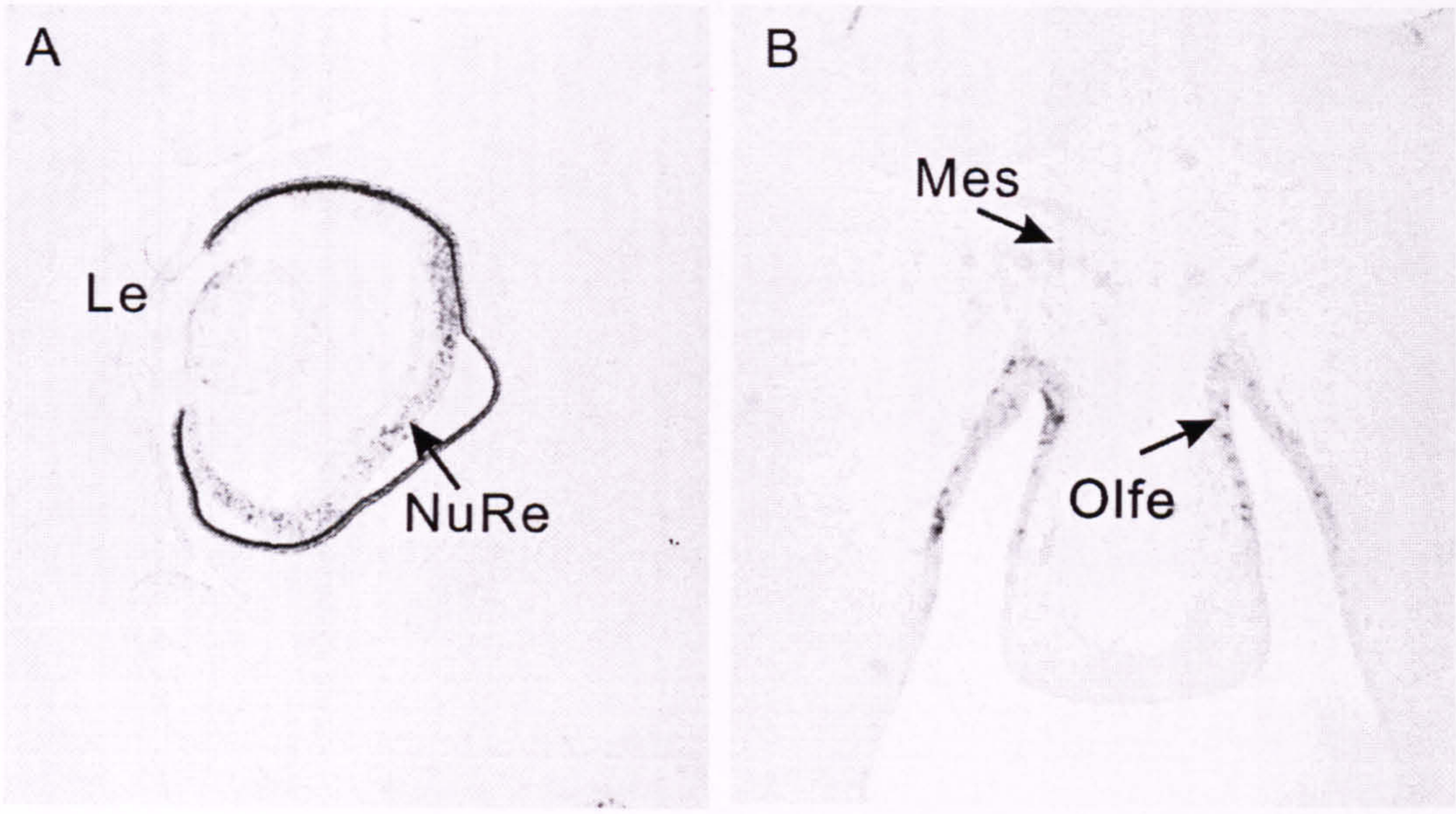
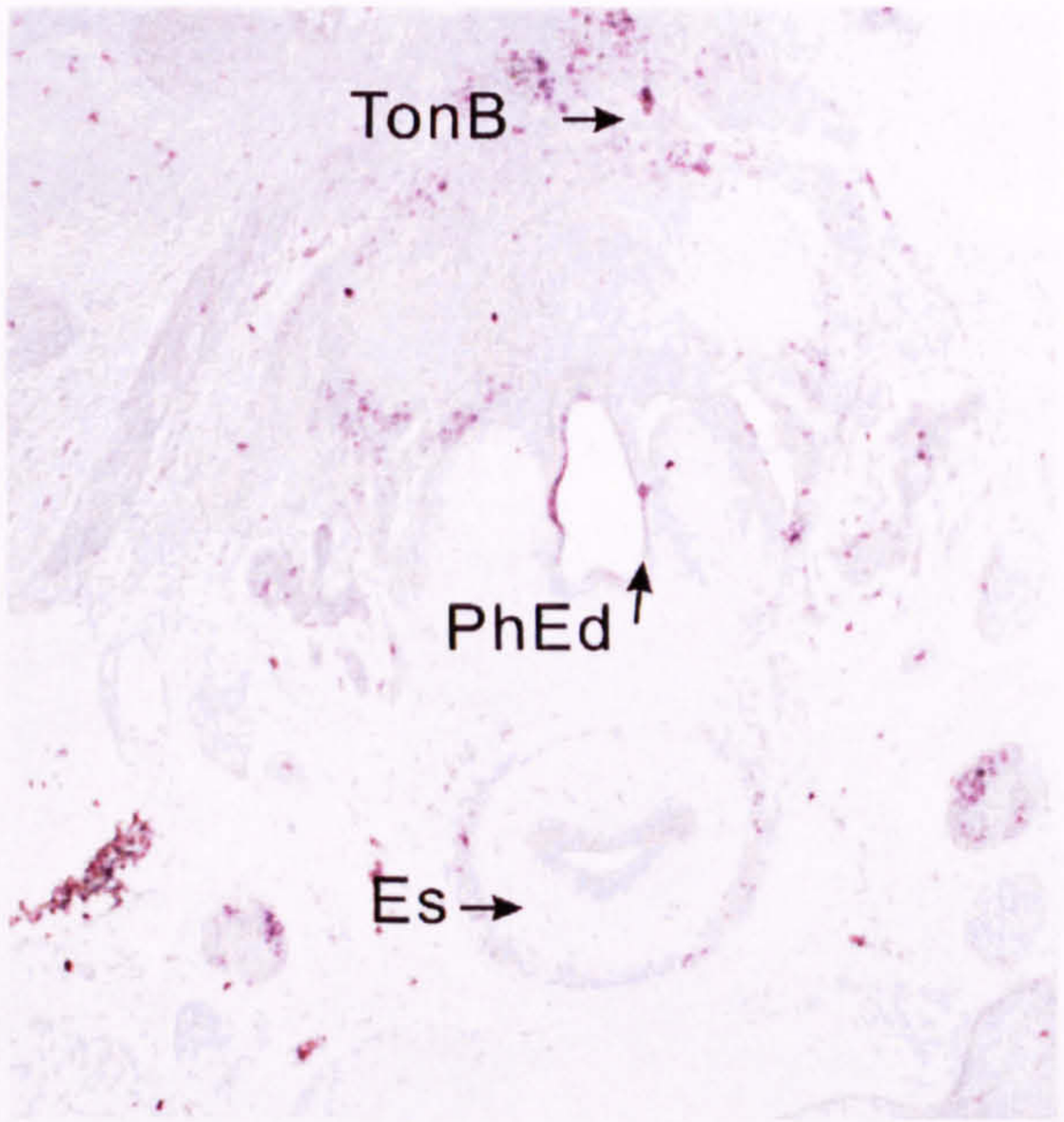



Fig. 4.10. *NIPBL* expression in developing face. A: *NIPBL* expression (CS20) was seen on nural retina and lens epithelium. B: expression of *NIPBL* (CS20) was seen on mesenchyme of glabellas and olfactory epithelium. C: expression of *NIPBL* (CS22-23) was seen on nasal septum, extraocular muscles, olfactory epithelium, midline of palate, and musculature of developing tongue. ExO: extraocular muscle; Le: lens epithelium; Mes: mesenchyme; Mi Pal: midline of palate; Ns: nasal septum; NuRe: neural retina; Olfe: olfactory epithelium; Ton: tongue; Ton Mu: tongue musculature.



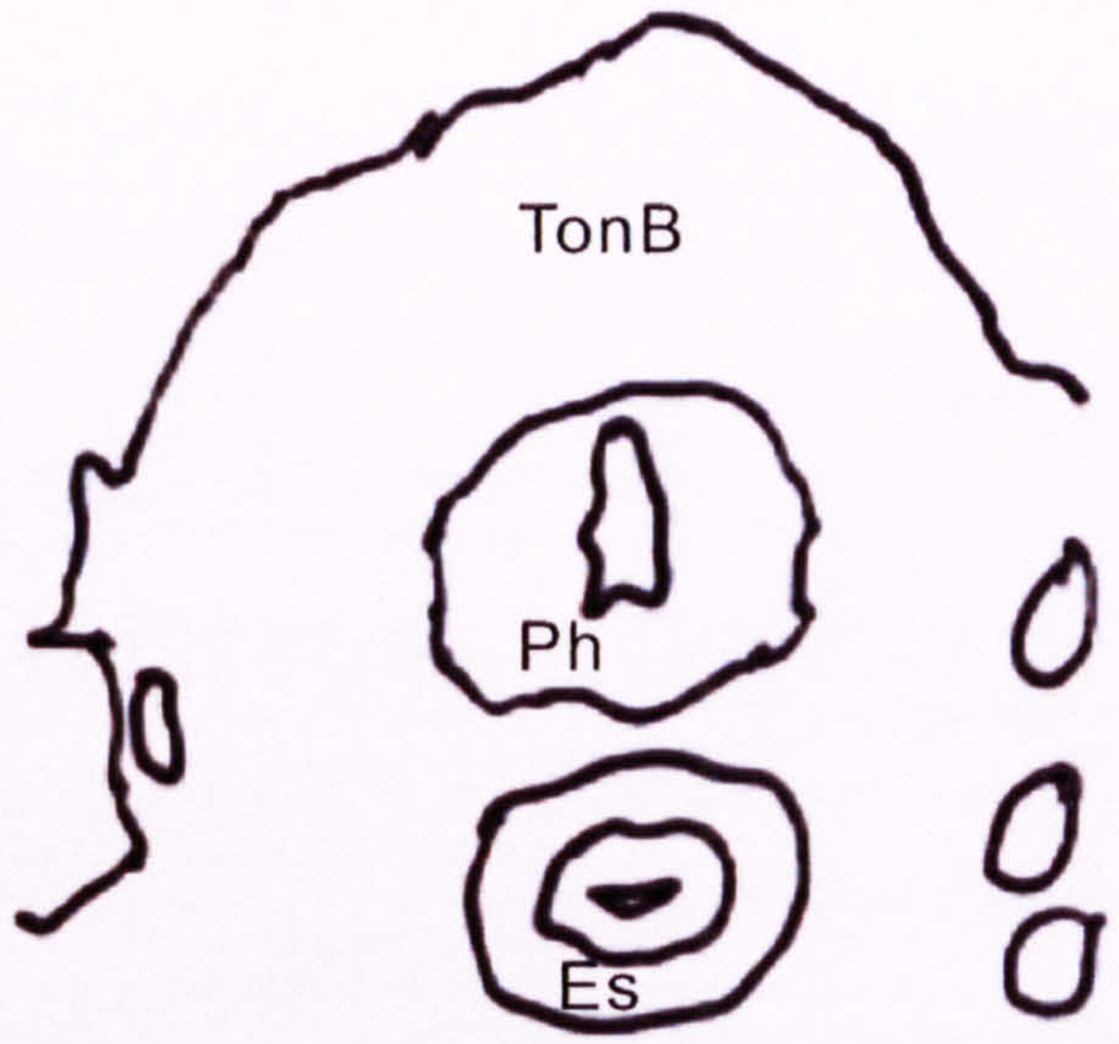




Fig. 4.11. *NIPBL* expression in developing pharynx and esophagus. The expression of *NIPBL* transcript was seen on pharyngeal endoderm, esophageal endodermal epithelium and mesothelium and musculature of tongue on CS 20 embryo section. Es: esophagus; Ph: pharynx; PhEd: pharyngeal endoderm; TonB: tongue base.

### 4.4.3.2 Expression of NIPBL in developing eyes

The expression of the NIPBL transcript in the developing eyes was mainly seen on the retina and lens epithelia, sees (Fig. 4.10A).

### 4.4.3.3 Expression of *NIPBL* in developing ears

The expression of *NIPBL* transcript is mainly seen on the surrounding otic mesenchyme, which will form the future otic canal (see Fig. 4.12).

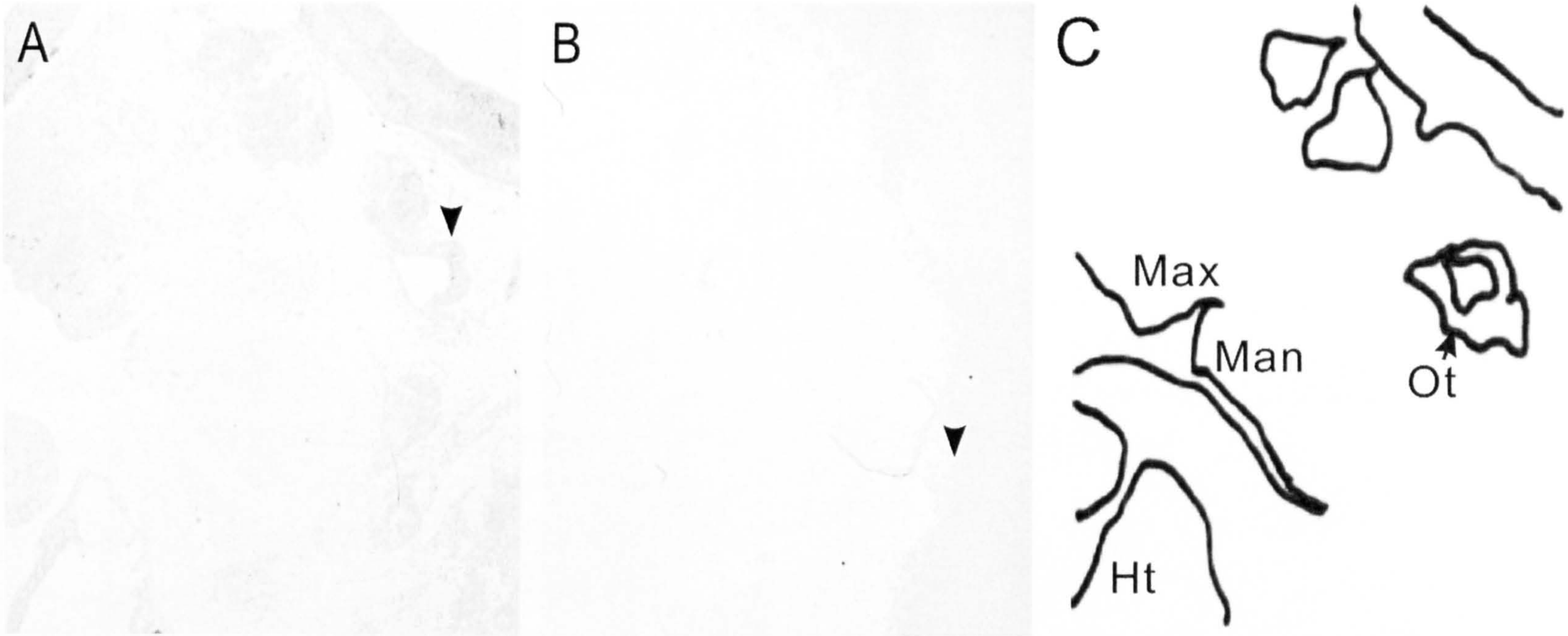



Fig. 4.12. *NIPBL* expression in the developing otic canal of CS23 embryo sections. (A) expression was seen on surrounding otic mesenchyme (arrowhead). No expression was seen on control (B) *NIPBL* expression was not seen on nearby section by using mRNA (sense strand) as control probe. (C) Graph of the facial structure presented in (A). Ht: developing heart; Man: developing mandible; Max: developing maxilla; Ot: developing otic placode.

### 4.4.4 Expression of *NIPBL* in developing heart

After finishing the preparation of the manuscript published in Nature Genetics (Tonkin et al., 2004), I decided to perform more detailed, serial expression data on human embryos of different stages. After several experiments, I accidentally found out that my gene expression pattern is very similar to that of genes such as *Wnt 8B* where heart expression should not exist. This prompted me to examine whether the expression data that I used in the published Journal were correct or not. I then performed radioactive labeled tissue *in situ* hybridization, which showed similar expression data as compared to the previous data except that the *NIPBL* expressions are stronger in the liver and weaker in the heart (data not shown). However, due to high cellularity and the presence of a large amount of red blood cells in the liver, I did not consider it significant. I also considered the use of *NIPBL* antibodies for histoimmunocytostaining. Unfortunately, no feasible antibodies were present until recently. After several months of trying, with the help from staffs in the developmental genetics group, the optimal protocol for tissue *in situ* hybridization that will yield the least background and tissue specific staining was determined. The major difference between the new protocol and old protocol include: (1) acetylation of the slide sections, for which I incubated slides with 0.1M triethanolamine (TEA) buffer, pH 8.0, containing 0.25% (v/v) acetic anhydride to acetylate the slides first; (2) quantification of the DIG-labeled probes by using direct blotting procedures with the test strip provided by the manufacturer (Roche), spectrometry quantification by using nanodrop spectrometer, and by direct electrophoresis of the probes. 300 $\eta$ g of labeled probes was used per slide instead of 500 $\eta$ g. (3) use of the hybridization buffer directly purchased from Roche Lab. (with the concentration of salmon sperm being 5X more, Denhardt's solution is 50% less, and SSC are twice more than previously

used. (4) more postbybridization washes than previously used (10X compared to 3X). In comparison to the previously used protocol, the tropism to myocardium was diminished. The new protocol yielded less background and less myocardium tropism in several different probes we used. The expression result of *NIPBL* by using the new protocol is still pretty similar to that of previous protocol except in the myocardium. No *NIPBL* expression was observed in myocardium by using the modified protocol. I also suspected that if the new protocol will lead to the decreased myocardium staining. I used mouse myocardium light chain 2V (*MLC2V*) antisense and sense strands as controls since MLC2V is exclusively expressed in embryonic mouse myocardium. Expression on myocardium was seen in MLC2V antisense probe but not with sense probe in mouse myocardium. Thus I excluded the possibility of non-staining of the myocardium by the modified protocol (see Fig. 4.13).

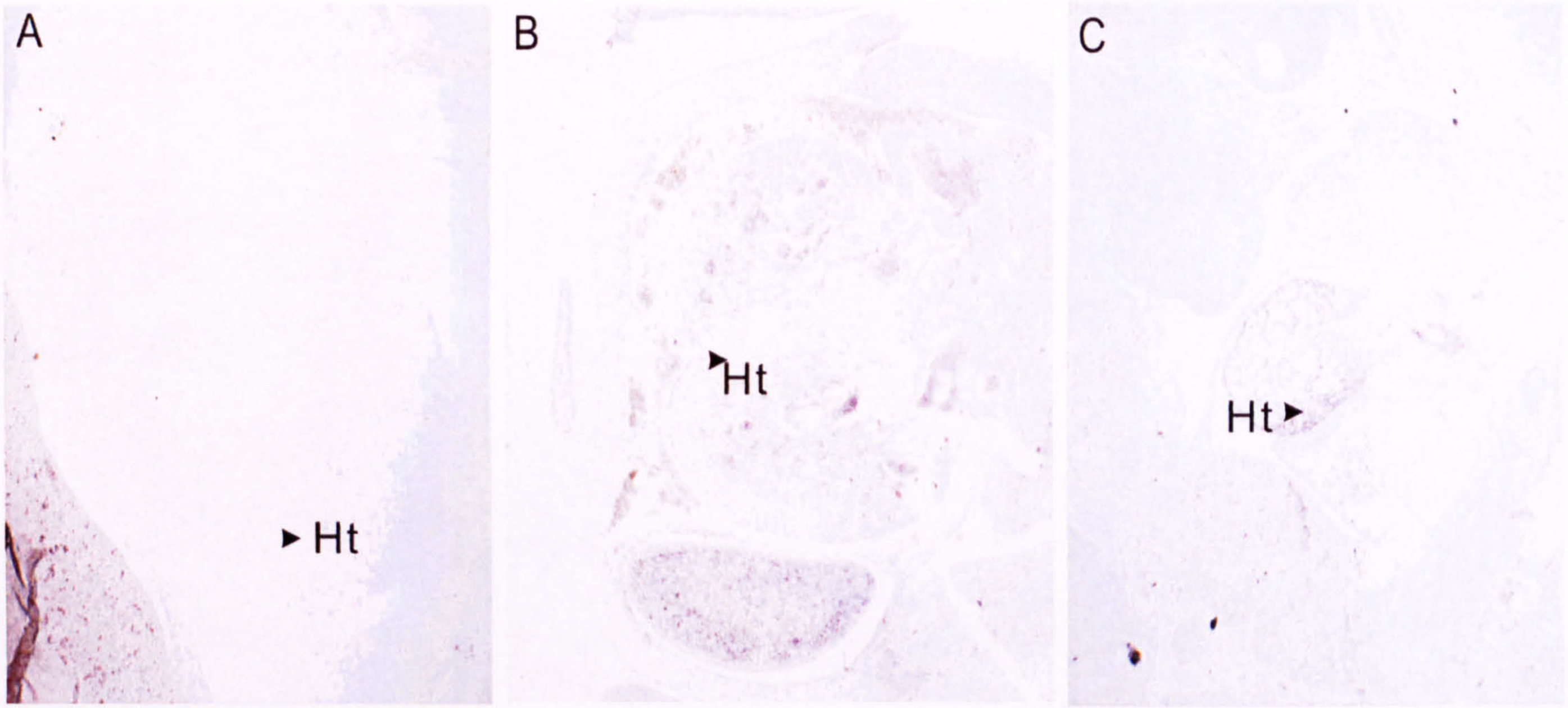
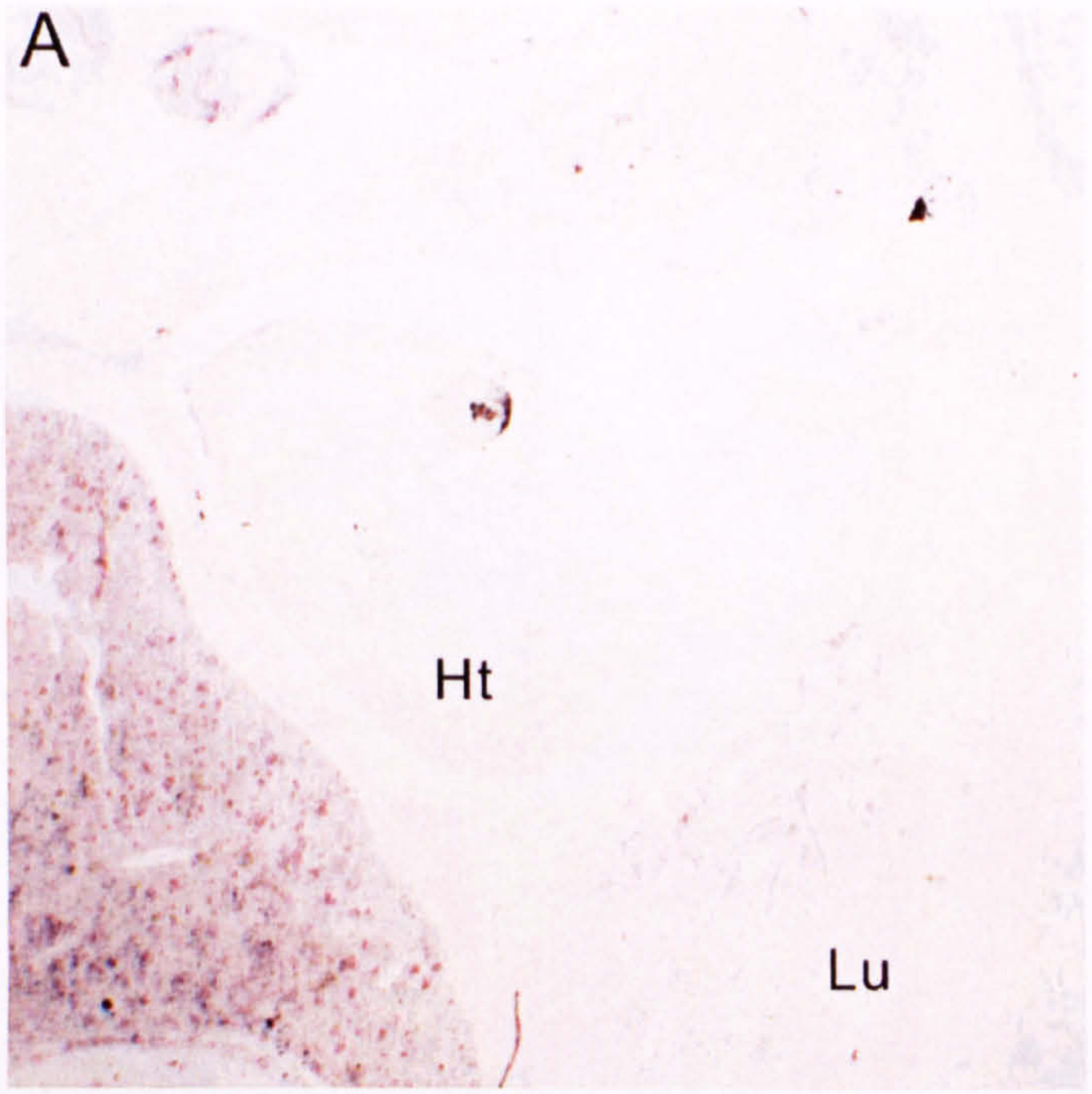



Fig. 4.13. *NIPBL* expression in developing heart. (A&B) NIPBL expression was not seen in myocardium of CS19 embryonic section. (C) Expression was seen in mouse embryonic section by using *MLC2V* as probe. Arrowhead indicates sites of developing myocardium.

### 4.4.5 Expression of *NIPBL* in developing lung, kidney and pancreas

#### 4.4.5.1 Expression of *NIPBL* in developing lung

In my experiments, *NIPBL* expression was seen on tubule epithelium of the developing lung (see Fig. 5.14), which initiates branching morphogenesis.



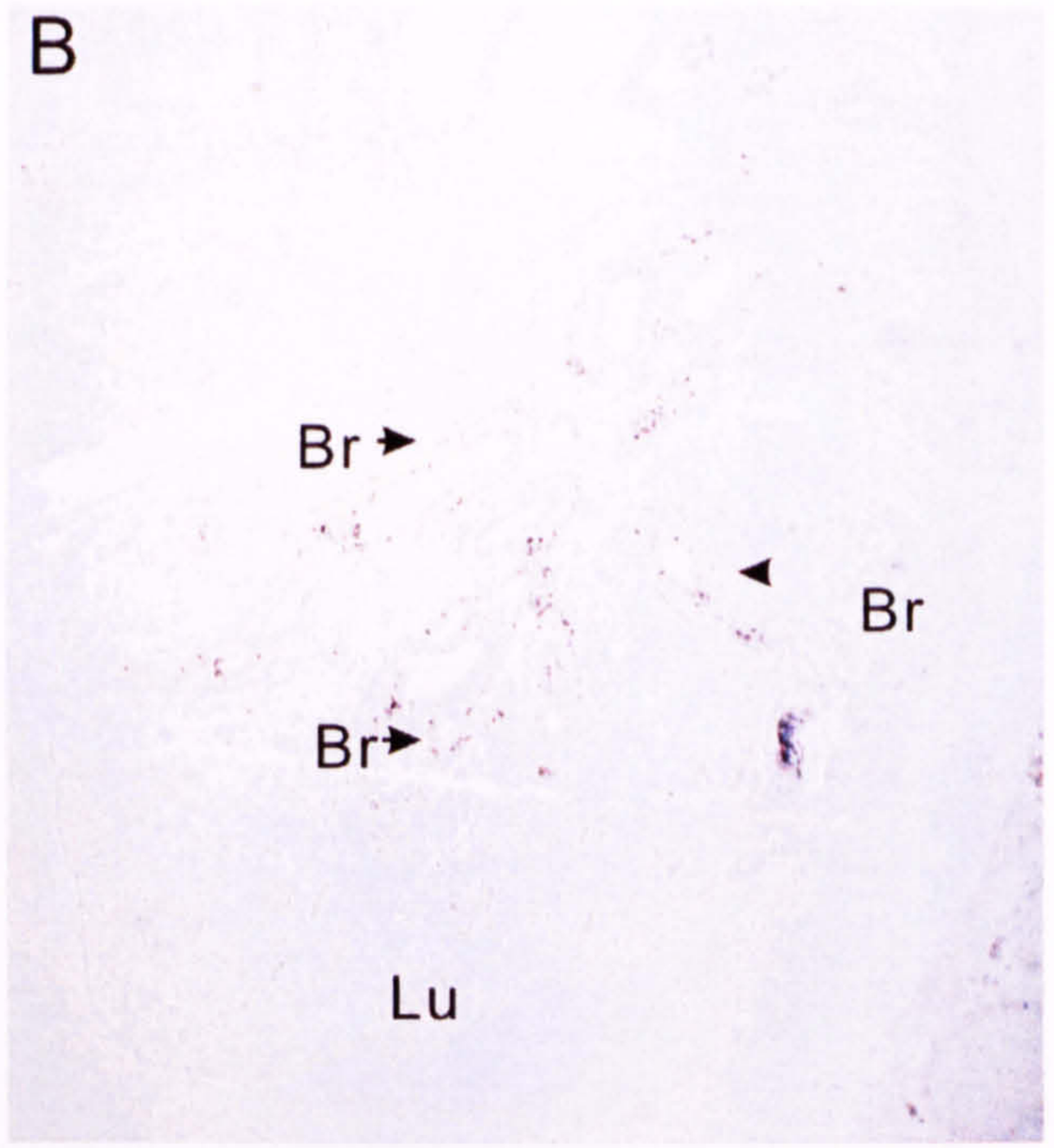
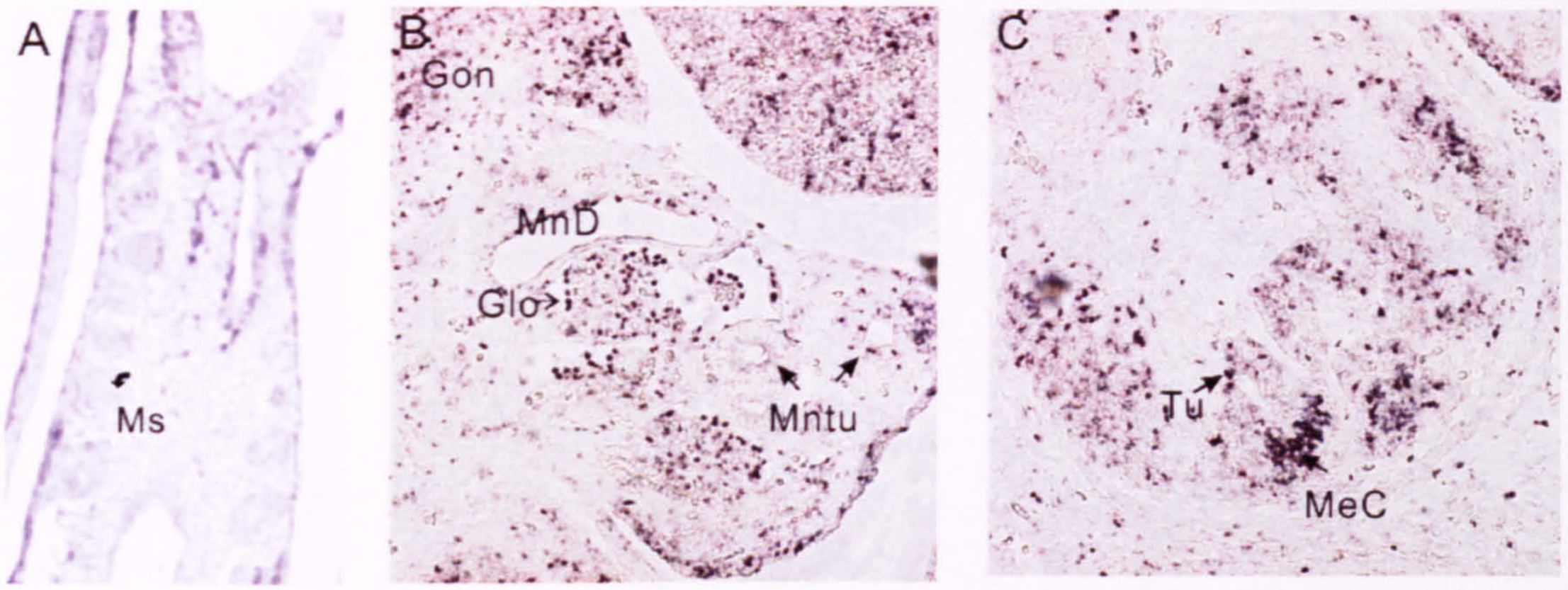




Fig. 4.14 *NIPBL* expression of the developing lung. (A&B) *NIPBL* expression was seen on epithelium of the developing lung tubules on CS23 embryo section. No expression was seen on myocardium of the developing heart. Br: developing bronchial tubules; Ht: developing heart; Lu: developing lung.

### 4.4.5.2 Expression of *NIPBL* in developing kidney

In my expression studies, *NIPBL* expression was seen in metanephric vesicles of CS 14 (around 35 postovulartory days) human embryo and in both epithelia and surrounding mesenchyme of CS18 mesonephros and metanephros tubules (see Fig.4.15.).



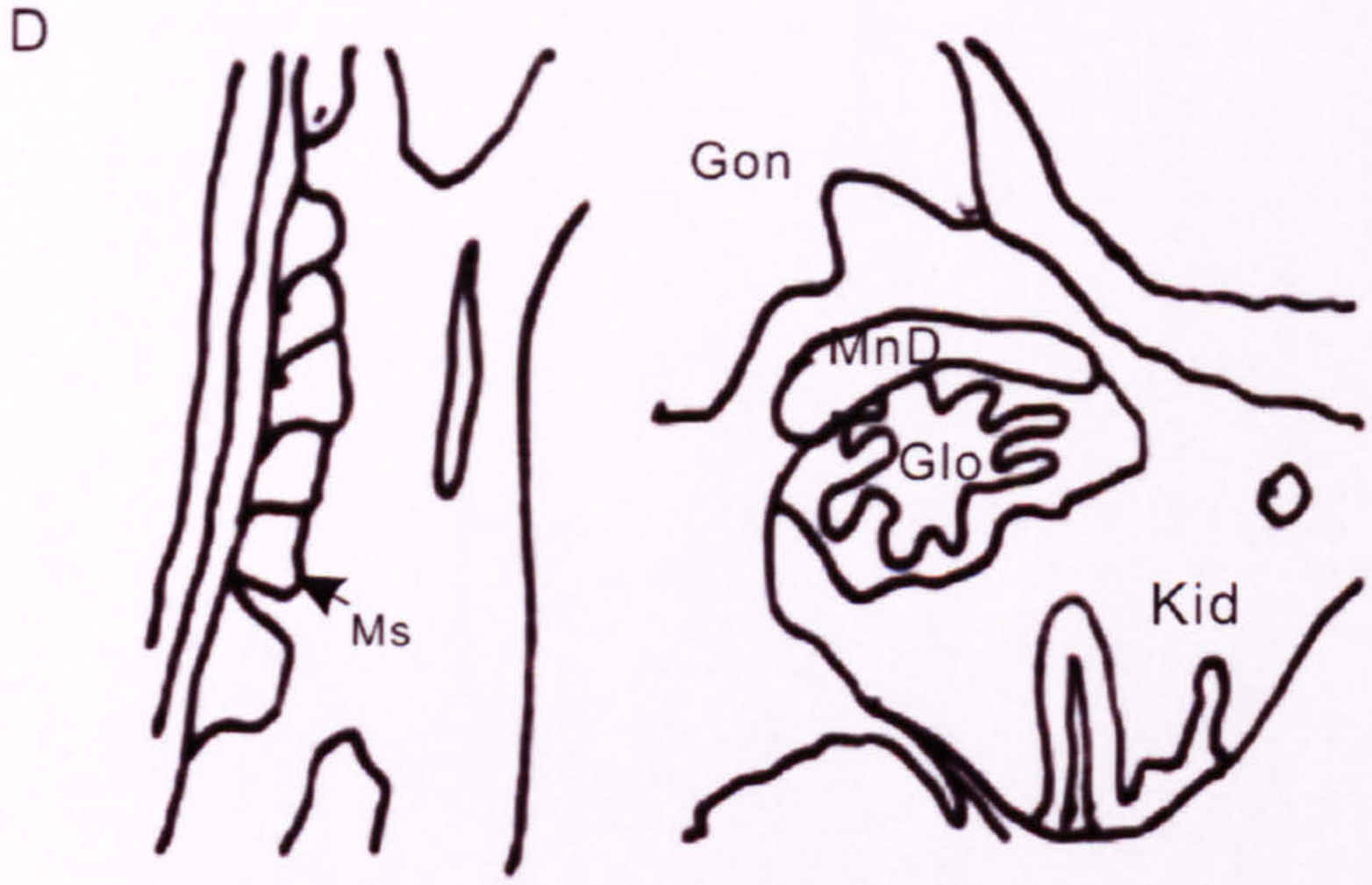




Fig. 4.15 *NIPBL* expression in developing embryonic kidney. (A) expression was seen on mesonephros of CS 14 embryonic section. (B) Expression was seen in mesonephric duct, tubules and the developing mesonephric glomerulus at Carnegie stage 18 embryo. (C) Expression was seen in metanephric tubules and the surrounding mesenchyme of the same embryonic section as (B). Glo: developing glomerulus; Gon: gonad; Kid: developing kidney (mesonephro); MeC: mesenchyme; MnD: mesonephric duct; Mntu: mesonephric tubule; Ms: mesonephron; Tu: metanephric tubule.

#### 4.4.5.3 Expression of *NIPBL* in developing pancreas

In my studies, *NIPBL* expression was seen in developing pancreatic ductal epithelium and foregut endoderm (Fig. 4.16).

### 4.4.6 Expression of *NIPBL* in developing gastrointestinal tract

In my experiment, *NIPBL* was seen both in the endoderm and mesenchyme of the developing foregut and midgut (see Fig. 4.16).

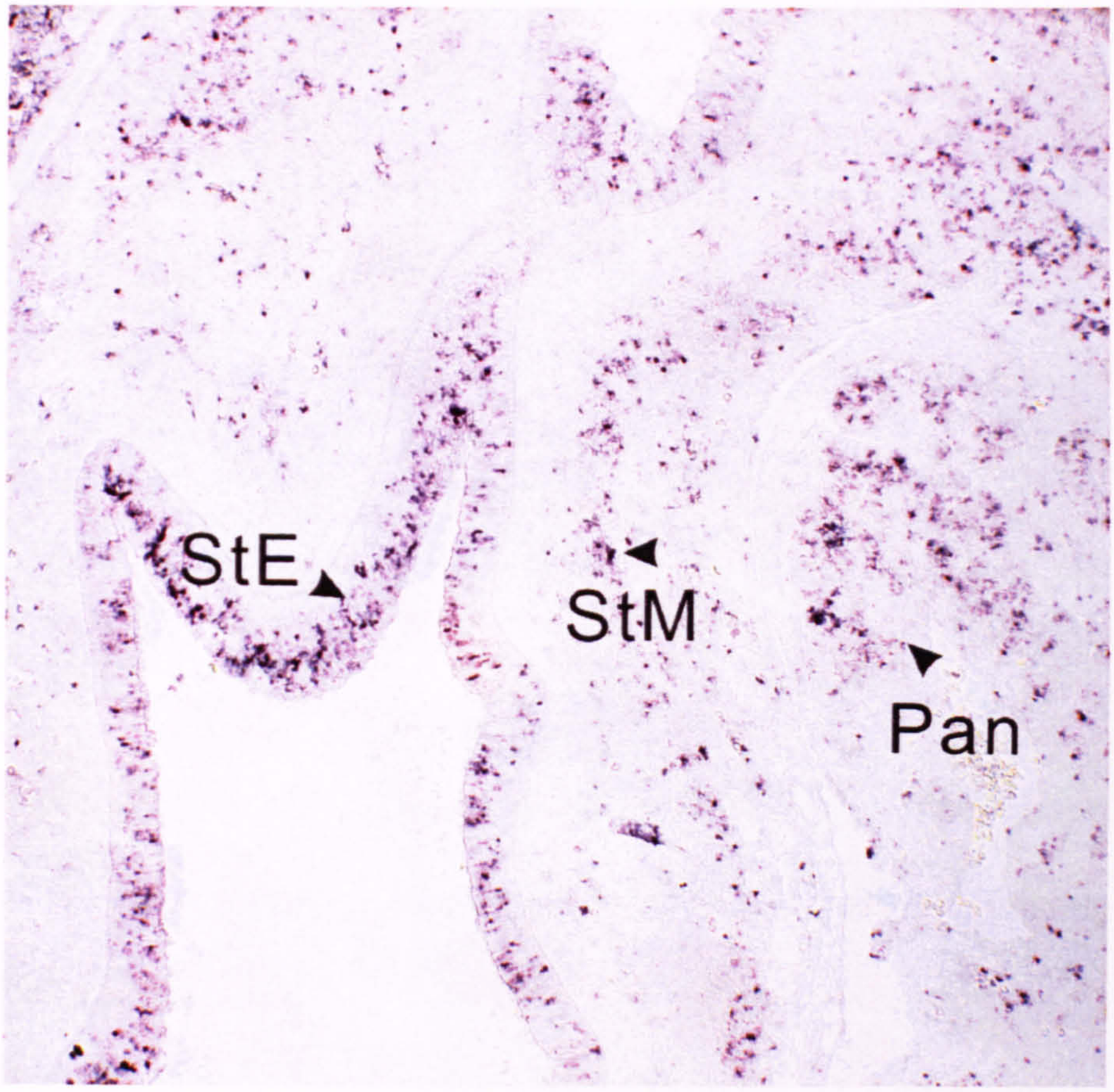



Fig. 4.16 *NIPBL* expression in the developing stomach and pancreas. Expression was seen on stomach epithelium and mesenchyme and pancreatic ductal epithelium in CS18 human embryo.

### 4.4.7 Expression of *NIPBL* in developing gonads.

*NIPBL* expression was seen on the premodium germ cells of the urogenital ridge in CS 13 embryonic sections. No expression was seen on control slide by using sense RNA as control probe.

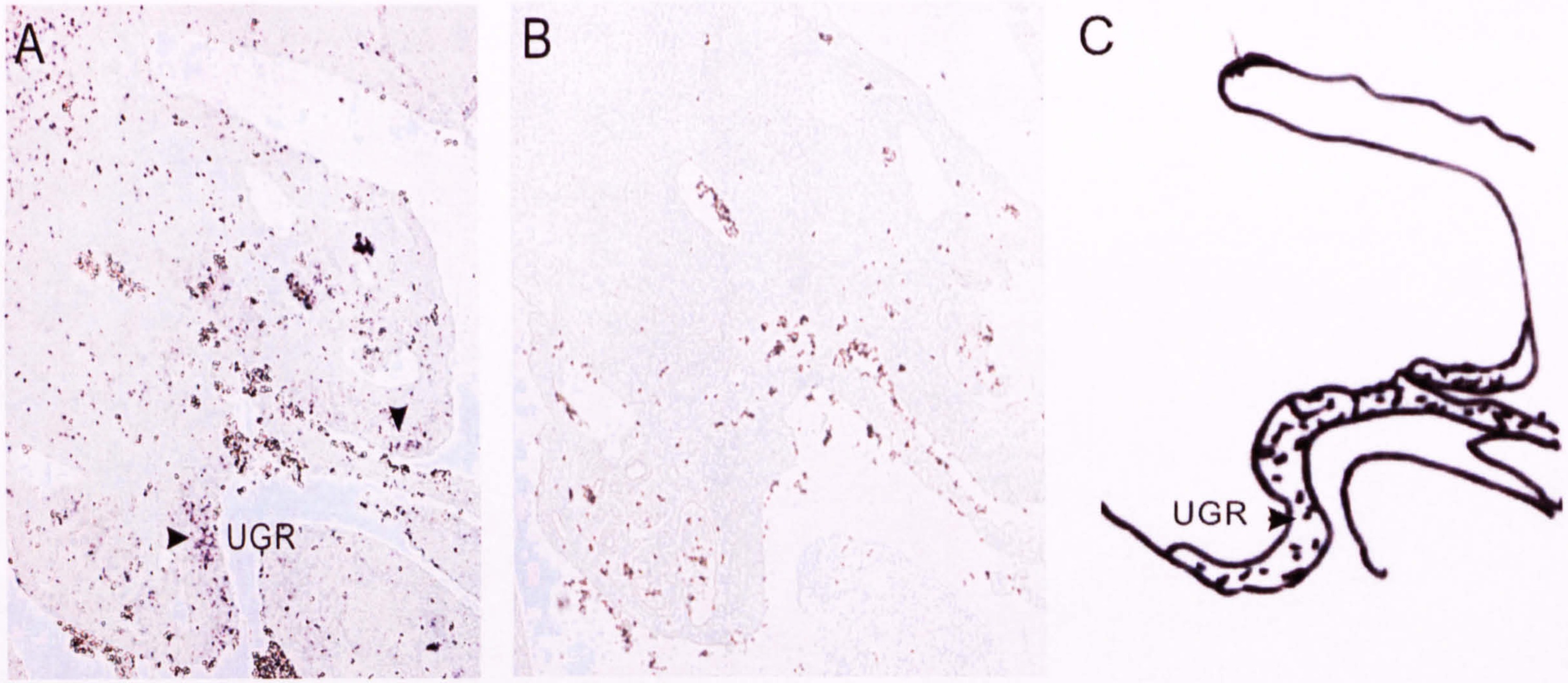



Fig. 4.17 *NIPBL* expression in the developing gonads. (A) *NIPBL* expression was seen on urogenital ridge (arrowhead) on CS13 embryonic section. (B) expression was not seen in nearby section by using mRNA as control probe. (C) Graph of the embryonic section at the level of urogenital ridge. UGR: urogenital ridge.

## 4.5 DISCUSSION

### 4.5.1 Expression of *NIPBL* in nervous system

In my experiments, I used PCNA as a proliferating marker instead $^{3}$H-thymidine incorporation. Homogeneous staining was seen exclusively on the ventricular zone of the developing brain (Fig. 4.6). The expression of the *NIPBL* transcript is parallel to that of PCNA. Thus, the *NIPBL* expression should also exist on matrix cells that are in their active proliferation state at stage I of cytogenesis. Defects in matrix cells proliferation will result in defective neuronogenesis and defective production of glial progenitors. Microcephaly, decreased cognitive function and behavior problems may result, as seen in patients with CdLS.

The yeast homologue of human *NIPBL*, Scc2, is found to be involved in cohesion loading which is important for cell replication (Ciosk et al., 2000; Neuwald and Hirano, 2000) and postreplicative double strand DNA repair (Sjogren and Nasmyth, 2001). By interacting with Scc4, Scc2/Scc4 complexes help in the establishing and loading of cohesin onto chromatin during the G1/S phase (Ciosk et al., 2000). Thus, it is plausible that defects in Human Scc2, *NIPBL*, will result in aberrant proliferation of the matrix cells and differentiation and thus affect the size and function of the developing brain. However, most of the cells undergoing stage II or III of cytogenesis may only appear later in the fetal or even postnatal stage. To examine the effect of NIPBL protein on neuronogenesis and glial progenitors differentiation, further investigation by using mouse or rat at the later developmental stage, such as the fetal and postnatal stage, is necessary.

### 4.5.2 Expression of *NIPBL* in developing limb

In mild CdLS patients with minimal limb defects, proportionately small long bones, hands and feet are observed in 90% of patients with CdLS (Jackson et al., 1993). My expression data showed that the *NIPBL* expressions were seen in the limb placode close to apical ectodermal ridge in CS 13 (28-32 postovulatory days) early limb bud where rapid cell proliferation occurs, as seen in Fig. 4.7. At the later stage, expression was seen in regions of increased cell density in the CS 18 embryo (44-48 postovulatory days) when ossification commences. Multiple foci of mesenchymal cell aggregations (arrow heads) and some areas of chondrogenic ossification centre (arrow) can be observed (see Fig. 4.8). The expression pattern seen in Fig. 4.8 was similar to that seen in PCNA immunocytostaining (data not shown). This suggests that *NIPBL* expression was mainly involved in cells under active proliferation. If *NIPBL* gene product, delangin, is involved in cell proliferation as suggested by my finding, it is plausible to suspect that the mutations that lead to truncation of delangin found in severe CdLS patients with limb defects may result from a decrease in chondrogenic progenitors at the early stage of limb development. If the mutations lead to severe reduction in progenitor cells under certain thresholds, amelia and severe limb reduction defects may occur.

However, the patterning process is complex and involves lots of regulating genes (Tickle, 2003). Thus, it is not clear to us whether the minor limb defects such as syndactyly observed in CdLS patients result from direct impaired joint formation or from the impaired apoptosis of the interdigital mesenchyme. In CS 23 embryo sections, expression was seen in mesenchyme surrounding the developing finger skeletons (data not shown). Whether this expression is associated with

cutaneous/skeletal syndactyly observed in patients with CdLS (Jackson et al., 1993) is not known. It has been reported that drosophila homologue, *nipped-B,* might facilitate the communication between the remote enhacer and the *cut* promoter (Dorsett et al., 2005; Rollins et al., 1999). Moreover, I do not know the level of *NIPBL* expression in the chondrocyte at the time when the condensations differentiated into cartilage templates. At that stage, the chondrocytes proliferation was decreased with low mitotic indexes. To investigate whether the level of *NIPBL* expression shows an inverse relationship as the chondrocyte enter a more differentiated stage, further serial studies under a more stringent quantitative control are mandatory. They would involve the use of a stringent amount of labeled riboprobes and hybridization to sections from the embryos/fetuses of mice/rats at different stages of development in one experiment.

Although observed with minor statistical significance, mutations that lead to truncation of the *NIPBL* gene product were more likely to be seen in patients with severe CdLS. Thus, it is reasonable to speculate that truncation mutations of *NIPBL* might result in a severe reduction of the gene dosage. Such a reduction would lead to a decreased progenitor cells population during early embryogenesis. These events will consequently lead to the severe limb reduction defect and misshaped limbs. For mutations that lead to altered protein quality rather than quantity, milder forms of limb defects such as small hands and feet may result. Whether this is via the direct effect of the change in delangin function or through the indirect effect by interaction with other interacting proteins is not known. *In vivo* and *in vitro* studies using the site-directed mutagenesis of the full length cDNA containing *NIPBL* (homologues) sequences in other animal models are necessary to clarify the role of *NIPBL* protein during limb development.

### 4.5.3 Expression of *NIPBL* in developing craniofacial skeletons, eye and ear

In addition to microcephaly, long, bushy and neat pencil like eyebrows as well as a peculiar face with a long philtrum and a short and anteverted nose are facial hallmarks for patients with CdLS (Ireland and Burn, 1993; Ireland et al., 1993). To our surprise, *NIPBL* expression is not obvious in craniofacial skeletons. Although *NIPBL* expression was not directly seen in the neural crest cells of the developing craniofacial skeletons, the *NIPBL* gene product may exert its influence on craniofacial patterning through the interaction of pharyngeal endoderm and neural ectoderm. Factors that lengthen the developmental window related to neural crest induction may lead to facial dysmorphism. Defects in *NIPBL* protein may lead to the decreased number of cyclindrical epithelium cells overlying/underlying neural crests, and thus affect the epithelial-mesenchyme interactions. This will result in hypoplasia of the facial skeleton seen in patients with de Lange syndrome.

A cleft palate was also observed in patients with CdLS (Kim et al., 2005; Kousseff et al., 1994). As seen in Fig. 5.10. *NIPBL* expression was seen in the midline where the two sides of the palate meet. Failure of the unification of two palatal halves might lead to the clefting seen in patients with CdLS.

Furthermore, the *NIPBL* expression was also seen in the musculature of the developing tongue and pharyngeal endoderm. As described above, the pharyngeal endoderm may play a role in lower face and neck skeleton patterning, while involvement of the developing tongue musculature might have an impact on speech delay and feeding difficulties as seen in patients. Though mental retardation is a characteristic in classic CdLS patients, we do not know whether the existence of

difficulty in communication due to tongue muscle weakness will further impair the I.Q. score in patients with CdLS. *NIPBL* protein may also exert its effect through interaction with other molecules such as *Wnt*, *Bmp*, *Fgfs* and *Hedgedog* proteins during early craniofacial development. However, to clarify this, gene knock down in feasible animal models is necessary and such experiments are underway now.

As described in Chapter 1, the ophthalmic findings associated with CdLS include synophrys, long accurate eyelashes, hypertelorism, telecanthus, myopia, any degree of nasal lacrimal duct obstruction, ptosis, strabismus and nystagmus (Levin et al., 1990). The expression of *NIPBL* was seen in the nasal root area. If the defective expression of *NIPBL* results in the flat nasal root as observed in patients with CdLS, then obstruction of the nasal lacrimal duct may occur since most obstructions appear where the nasolacrimal duct enters the nasal cavity (Robb, 2001). Ptosis of the eye is due to failure of normal development of the levator palpebrae superioris muscle or abnormal development of the superior division of oculomotor nerve, which supplies this muscle. Moreover, extraocular muscle involvement defects are also associated with strabismus. The retina and optic nerve fibers are derived from neuroectoderm of the forebrain. Up to now, only one CdLS patient with idiopathic retinal telangiectasia has been reported to be associated with retinopathy (Folk et al., 1981). However, this finding is not compatible with the finding we have. Nevertheless, the vertebrate retina is derived from paired evaginations from the neural tube in embryonic development and is initially produced by progenitor cells similar to those that generate the neurons and glia of other areas of the central nervous system (Reh and Levine, 1998). Furthermore, the central nervous system, which is consistent with that *NIPBL* expression, is mainly on sites with active proliferation.

In patients with CdLS, hearing defects may be present in 60-100% of children (Jackson et al., 1993). Comprehensive studies of the ear and hearing examinations in 45 children revealed stenosis of the external auditory canals in 80% and hearing loss from a mild to a severe form in all patients with CdLS (Sataloff et al., 1990). Thus, defects in otic mesenchymal progenitors might lead to conductive hearing impairment.

### 4.5.4 Expression of *NIPBL* in developing heart

Expression of *NIPBL* was not seen by using the modified protocols for tissue *in situ* hybridization. However, congenital heart disease was reported to be observed in one fourth of patients with CdLS. (Greenwood et al., 1977; Tsukahara et al., 1998) Right-sided obstruction is most common. Recently, a 32-year-old woman with CdLS and a severe calcified aortic valve was described. Inactivation of the Notch signaling causing differentiation of the valvular cells into osteoblast-like cells was presumed to be the cause of calcified aortic valve seen in this patient (Oudit et al., 2006).

In *Drosophila*, Notch signaling pathway was suggested to have a genetic association with Drosophila *Nipped-B* (Dorsett et al., 2005; Rollins et al., 2004). Notch, which is involved in a wide range of developmental contexts, but is best known for its role in lateral inhibition, is activated by its transmembrane ligands Delta (Drosophila and vertebrates) as well as Serrate (Drosophila) and its vertebrate homologue Jagged. The glycosyl transferase Fringe differentially modulates the responsiveness of Notch to Delta versus Serrate/Jagged ligands. In addition, the transmission of the signal to target genes involves the nuclear factor Su(H)/RBP-J and its co-repressor(Bray and Furriols, 2001; Mumm and Kopan, 2000). In *Drosophila*,

the Su(H) insulator protein encoded by the suppressor of the Hairy-wing [su(H)] gene interferes with enhancer-promoter communication. Su(H) binds a DNA sequence in the gypsy transposon. When gypsy inserts into a gene, enhancer-promoter interactions are blocked in a Su(H)-dependent manner (Rollins et al., 1999). During *Drosophila* cardiogenesis, Notch functions throughout two distinct developmental events, namely the process of lateral inhibition and the control of lineage decisions during asymmetric cell divisions of heart progenitors. Activation of Notch signaling will prevent the cells from differentiating into cardiac progenitors. On the contrary, inactivation of the pathway causes ectopic expression of heart markers and expansion of the heart (Zaffran et al., 1995). After specification of the cardiac progenitor cells, differentiatial activity is responsible for the acquisition of asymmetric cardiac cell fates in the two daughter cells. *Drosophila Nipped-B* has been suggested to facilitate promoter-enhancer communication through antagonistic to insulation by Su(H)(Rollins et al., 1999).

The occurrence of congenital heart defect in patients with CdLS may be through interaction of *NIPBL* with other signaling pathways, such as Notch signaling, that affect cardiac development. We can not exclude the possibility that locations of *NIPBL* transcripts are in locations outside of the developing heart and secrete its protein product to interact with other major signaling pathways involved in cardiac development. It is also possible that the expression of *NIPBL* in the developing heart is a very early event (earlier than CS 17) such as in the early induction phase of cardiac progenitors. Thus, it is important to for us to perform immunohistochemistry staining to verify the expression pattern of *NIPBL* in developing embryonic heart from very early stage of development to late embryonic stages when feasible delangin antibodies are available.

### 4.5.5 Expression of *NIPBL* in developing lung, kidney and pancreas

Considering that the most frequent cause of death in patients with CdLS is recurrent pulmonary infections due to chronic lung disease (Beck and Fenger, 1985; Gupta and Goyal, 2005), it is reasonable that the *NIPBL* gene product might play a crucial role in lung morphogenesis.

Anomalies of the kidney and urinary tract occur in approximately 40% of the patients with de Lange syndrome. Renal manifestation of patients with CdLS includes: absent or poor corticomedullary differentiation, pelvic dilatation, small kidneys, isolated renal cyst, vesicourethral reflux, decreased glomerular filtration rate and renal ectopia. All is associated with branching defects as described in section 4.2.3.5. Renal function impairment was found in approximately one third of CdLS patients with structural anomalies. Anomalies of the kidney and urinary tract and renal functional impairment more frequently occur in patients with the classic rather than the mild clinical phenotype (Selicorni et al., 2005).

Aside from complete renal agenesis, very few knockout mutations of genes that operate during early branching result in moderate to severe branching defects. On the contrary, these phenotypes tend to be slight or undetectable (Qiao et al., 1999b). The lack of moderate to severe branching defects may arise from a requirement for an essential number of nephrons for organ survival. In addition, mechanisms such as redundancy may allow for activation of compensatory processes that are able to buffer the effects of genetic mutations (Wagner, 2000a; Wagner, 2000b). Furthermore, gene dose may also play a role in the lack of moderate to severe branching defects. For example, although *Gdnf* null mice lack kidneys and die, heterozygotes show an array

of renal phenotypes ranging from small kidneys to unilateral agenesis with no intermediate phenotypes (Cullen-McEwen et al., 2001).

Defects during early or late branching may predispose to disorders of reduced nephron number and results in reduced filtration surface area (Cullen-McEwen et al., 2003). Finally, in the developing kidney, branching must cease and the caliber of the tubule lumen must be regulated. Defects in branching termination may result in generation of cysts with corresponding clinical manifestations ranging from asymptomatic simple cysts to end-stage polycystic kidney disease (Selicorni et al., 2005). Defects during initial ureteric bud outgrowth may result in renal agenesis, as ureteric budding requires the presence of GDNF as a primary inducer. Collecting system duplication which may result in bifid ureter and duplex kidney is another common renal malformation during the initial ureteric bud outgrowth stage (Pohl et al., 2002).

Annular pancreas, a malformation in which a band of pancreatic tissue completely encircles the duodenum associated with formation of a second ventral bud, has been reported in patients with CdLS (Lagunas Flores et al., 1981; Wick et al., 1982).

In conclusion, it seems that *NIPBL* is expressed on sites that are associated with the branching morphogenesis in human embryos. However, in more than 50% of patients with CdLS phenotypic presentation shows no defects associated with branching morphogenesis. It is possible that *NIPBL* gene haploinsufficiency may not be sufficient to cause the urinary abnormality due to variable penetrance and/or a dose-dependent genotype-phenotype relationship. In total, that *NIPBL* functional

redundancy and/or dose effect might exist during early branching morphogenesis. Finally, the number of human disorders involving defective branching morphogenesis, for which a molecular basis has been identified, is far smaller than the number predicted by the number and diversity of identified phenotypes. Thus, from the expression data and the associated phenotypes observed in patients with CdLS, I believe that *NIPBL* gene product might play an important role during early branching morphogenesis in a dosage-dependent manner. To verify this, establishing an animal model by using conditional knockout of *NIPBL* gene in mouse is crucial and is now underway.

### 4.5.6 Expression of *NIPBL* in developing gut

Few gastrointestinal malformations were reported to be associated with CdLS except for cecal volvulus, hiatal hernia and gastroesophgeal reflux (Holthusen and Rottingen, 1998; Husain et al., 1994; Jackson et al., 1993; Lachman et al., 1981; Masumoto et al., 2001; Pei et al., 2000). The cecal vovulus is associated with non-fixation of the cecum and ascending colon (Husain et al., 1994; Masumoto et al., 2001). The gastroesophageal reflux is observed in around 65% of CdLS patients (Luzzani et al., 2003). The main cause of gastroesophageal reflux disease (GERD) is incompetence of the antireflux barriers at the esophagogastric junction. The antireflux barriers include two sphincters: the lower esophageal sphincter, and the crural diaphragm that functions as an external sphincter.

The *NIPBL* expression was seen in most the developing skeletal muscles including the developing diaphragmatic musculature during human embryogenesis (data not shown). This may contribute to the high prevalence of gastroesophageal

reflux disease (GERD) seen in CdLS patients. Furthermore, congenital diaphragmatic hernia has also been reported in CdLS patients (Cunniff et al., 1993; Jelsema et al., 1993; Pankau and Janig, 1993). Thus, the external sphincter malfunctions might contribute, at least partly, to the pathogenesis of GERD seen in patients with CdLS. Also, a set of *Hoxd* genes contributes to the formation of a major anatomical and physiological subdivision of the gut, the ileocaecal sphincter, which forms the boundary between the small intestine and the large bowel. Inactivation of several Hoxd genes results in abnormal ileocecal sphincters. Is it possible that *NIPBL* gene product may influence the sphincter architect in the gastrointestinal tract though direct or indirect pathway? Is *NIPBL* involved in foregut positional patterning and does it lead to the posterior homeotic shift in which esophageal epithelium has the character of the more caudal gastric phenotypes? Why are no severe gut phenotypes reported in patients with CdLS if *NIPBL* expression was seen both in endodermal epithelium and mesenchyme? It is possible that haploinsufficiency of the *NIPBL* product alone is not enough to cause the phenotypes. Moreover, genetic redundancy makes it necessary for several functionally similar genes to be inactivated to generate a recognizable gut phenotype (Beck, 2002).

### 4.5.7 Expression of *NIPBL* in developing gonads

Patients with a severe form of CdLS seldom reproduce and genital abnormality such as hypospadia has been reported in CdLS patients (Elhassani and Moyd, 1987). My expression study in the early human embryo revealed the expression of *NIPBL* transcript in the gonadal ridge. Thus, it is plausible that mutations in *NIPBL* lead to severe phenotypes of patients with CdLS and result in infertility in these patients.

## 4.6 CONCLUSION:

In this study, I found that the expression pattern in tissue *in situ* hybridisation is similar with that of northern blot except that (1) Expression was not visible in adult brain and weak in foetal brain. However, this is not surprising if the *NIPBL* expression only involves cells under active proliferation during early development. (2) The expression on heart is strong. However, more experiments need to evaluate this. (3) Weak expression of adult and foetal lung seen in Northern blot may reflect that *NIPBL* expressions are only restricted to early generation of branching tubules.

During human development, similar mitotic pattern as seen in neuroepithelium was observed in a wide variety of proliferating tissues. Specifically, the so-called germinal cells are rounded and always lie directly (1) beneath the luminal surface showing mitotic chromosomes; (2) between the mitotic zone (M) and the deep nuclear zone, an intermediate zone with sparse nuclei which are at the latter half of telophase and early prophase. The proliferating tissues with these characteristics include: Adult: the epithelium of the efferent duct of the epididymis; the gastric mucosa; the intestinal mucosa; the epithelium of the endometrial gland. Embryo: the epithelium of the branchial clefts; the tracheal and lung tubular epithelium; the gastric mucosa; the intestinal epithelium; the epithelium of the bile duct system including the gall-bladder; the epithelium of the pancreas (the above six embryonic tissues are of endodermal origin); the mesonephric tubules; the mesonephric duct; the metanephric tubule; the thickened part of the mesothelium (these four tissues are of mesodermal origin); the epithelium of the nasal pit; Rathke's cleft; the anlange of the retina; the neural tube (these four tissues are of ectodermal origin) (Fujita, 2003). The sites of *NIPBL* expression are parallel to the embryonic tissues described above. What is common

to these tissues is the fact that they consist of a single layer of cylindrical epithelium. Thus, it is quite possible that the *NIPBL* expression is mainly on tissues undergoing active proliferation. Defect in *NIPBL* gene products will lead to decrease in these totipotent progenitors. Generalized proportionate growth deficiency may result as seen in patients with CdLS.

In most proliferating somatic tissues, growth is tightly coupled to the cell cycle, which makes the increase in the cell number a reliable indicator of tissue growth and size during development. There are several ways in which cell numbers in a tissue can be modulated. These include alterations in patterning which determines the initial size of the stem or founder cell population, the rate and mode of proliferation of founder and progenitor cells, cell death and cell migration. In plants, cell proliferation and growth primarily contribute to organ size; the removal of overproliferated cells and cell migration does not occur (Shpak et al., 2004).

In *Drosophila*, the coordination of cell proliferation and cell death is essential for determining the correct numbers of cells in a tissue. In mammals, the contributions of proliferation, migration and death to the regulation of cell number vary greatly between tissues. One example of this is retina, in which proliferation has a significant effect on cell number during development and slowing down of the cell cycle causes a major deficit in the cell number (Green et al., 2003).

Cell growth is tightly coupled to cell cycle progression. However, it is not clear how coupling occurs. Several mechanisms may be involved, which include nutritional availability, protein synthesis and RNA transcription (Britton et al., 2002; Orian et al., 2003). In mammals, cell size at cell division appears to depend on the speed of the

cells progress through the cell cycle and the speed of the growing process, which in turn depends on extracellular signals that control cell cycle progression, cell growth, or both.

In addition to cell cycle regulation, patterning and morphogenesis may be directly regulated by cell-cycle related events in early embryogenesis. In *Xenopus*, progression through a limited number of cell cycles in the midblastula transition stage is necessary for *HoxB* gene cluster activation in neuroectoderm. More cell cycles are required for the correct spatial expression of *HoxB* genes along the AP axis (Fisher and Mechali, 2003).

In my experiments, *NIPBL* expression patterns are similar to that of the proliferating marker, PCNA. Their expressions are mainly on sites of active proliferation such as early limb bud, gonads and ventricular zone of developing nervous system, and most of the cylindrical epithelium as described above. If mutations occuring in *NIPBL* lead to the prolongation for each cell cycle and result in a decreased cell number during organogenesis, growth deficiency may occur as seen in patients with CdLS. Although severe patterning defects (except in limbs) seldom reported to be associated with CdLS, *NIPBL* mutations seen in patients with CdLS are due to hemizygous rather than homozygous defects and thus lead to less severe phenotypes. It is possible that homozygous mutation of *NIPBL* will severely impair cell cycle progression and is non-viable for the developing early human embryos. Thus, it is important to establish a gene knockout model to understand the effect of *NIPBL* gene product and the impact they have in embryonic development. Such an experiment is underway now.

## 4.7 REFERENCES


Angerer, L. M., Yang, Q., Liesveld, J., Kingsley, P. D., and Angerer, R. C. (1992). Tissue-restricted accumulation of a ribosomal protein mRNA is not coordinated with rRNA transcription and precedes growth of the sea urchin pluteus larva. Dev Biol *149*, 27-40.

Angle, B., Yen, F., Hersh, J. H., and Gowans, G. (2003). Patient with terminal duplication 3q and terminal deletion 5q: comparison with the 3q duplication syndrome and distal 5q deletion syndrome. Am J Med Genet A *116*, 376-380.

Aoki, T. O., David, N. B., Minchiotti, G., Saint-Etienne, L., Dickmeis, T., Persico, G. M., Strahle, U., Mourrain, P., and Rosa, F. M. (2002). Molecular integration of casanova in the Nodal signalling pathway controlling endoderm formation. Development *129*, 275-286.

Ausubel, F. M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (2002). Short protocols in molecular biology.

Bartram, U., and Speer, C. P. (2004). The role of transforming growth factor beta in lung development and disease. Chest *125*, 754-765.

Baumgart, E., Schad, A., Volkl, A., and Fahimi, H. D. (1997). Detection of mRNAs encoding peroxisomal proteins by non-radioactive in situ hybridization with digoxigenin-labelled cRNAs. Histochem Cell Biol *108*, 371-379.

Beck, B., and Fenger, K. (1985). Mortality, pathological findings and causes of death in the de Lange syndrome. Acta Paediatr Scand *74*, 765-769.

Beck, F. (2002). Homeobox genes in gut development. Gut *51*, 450-454.

Braddock, S. R., Lachman, R. S., Stoppenhagen, C. C., Carey, J. C., Ireland, M., Moeschler, J. B., Cunniff, C., and Graham, J. M., Jr. (1993). Radiological features in Brachmann-de Lange syndrome. Am J Med Genet *47*, 1006-1013.

Bray, S., and Furriols, M. (2001). Notch pathway: making sense of suppressor of hairless. Curr Biol *11*, R217-221.

Britton, J. S., Lockwood, W. K., Li, L., Cohen, S. M., and Edgar, B. A. (2002). Drosophila's insulin/PI3-kinase pathway coordinates cellular metabolism with nutritional conditions. Dev Cell *2*, 239-249.

Buongiorno-Nardelli, M., and Amaldi, F. (1970). Autoradiographic detection of molecular hybrids between RNA and DNA in tissue sections. Nature *225*, 946-948.

Burstyn-Cohen, T., Stanleigh, J., Sela-Donenfeld, D., and Kalcheim, C. (2004). Canonical Wnt activity regulates trunk neural crest delamination linking BMP/noggin signaling with G1/S transition. Development *131*, 5327-5339.

Calegari, F., Haubensak, W., Haffner, C., and Huttner, W. B. (2005). Selective lengthening of the cell cycle in the neurogenic subpopulation of neural progenitor cells during mouse brain development. J Neurosci *25*, 6533-6538.

Cerny, R., Lwigale, P., Ericsson, R., Meulemans, D., Epperlein, H. H., and Bronner-Fraser, M. (2004). Developmental origins and evolution of jaws: new interpretation of "maxillary" and "mandibular". Dev Biol *276*, 225-236.

Chien, K. R. (2000). Genomic circuits and the integrative biology of cardiac diseases. Nature *407*, 227-232.

Chinoy, M. R. (2003). Lung growth and development. Front Biosci *8*, d392-415.

Christoffels, V. M., Habets, P. E., Franco, D., Campione, M., de Jong, F., Lamers, W. H., Bao, Z. Z., Palmer, S., Biben, C., Harvey, R. P., and Moorman, A. F. (2000). Chamber formation and morphogenesis in the developing mammalian heart. Dev Biol *223*, 266-278.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol Cell *5*, 243-254.

Cohn, M. J., Lovejoy, C. O., Wolpert, L., and Coates, M. I. (2002). Branching, segmentation and the metapterygial axis: pattern versus process in the vertebrate limb. Bioessays *24*, 460-465.

Cordero, D., Marcucio, R., Hu, D., Gaffield, W., Tapadia, M., and Helms, J. A. (2004).

Temporal perturbations in sonic hedgehog signaling elicit the spectrum of holoprosencephaly phenotypes. J Clin Invest *114*, 485-494.

Couly, G., Grapin-Botton, A., Coltey, P., and Le Douarin, N. M. (1996). The regeneration of the cephalic neural crest, a problem revisited: the regenerating cells originate from the contralateral or from the anterior and posterior neural fold. Development *122*, 3393-3407.

Cox, K. H., DeLeon, D. V., Angerer, L. M., and Angerer, R. C. (1984). Detection of mrnas in sea urchin embryos by in situ hybridization using asymmetric RNA probes. Dev Biol *101*, 485-502.

Crump, J. G., Maves, L., Lawson, N. D., Weinstein, B. M., and Kimmel, C. B. (2004a). An essential role for Fgfs in endodermal pouch formation influences later craniofacial skeletal patterning. Development *131*, 5703-5716.

Crump, J. G., Swartz, M. E., and Kimmel, C. B. (2004b). An integrin-dependent role of pouch endoderm in hyoid cartilage development. PLoS Biol *2*, E244.

Cullen-McEwen, L. A., Drago, J., and Bertram, J. F. (2001). Nephron endowment in glial cell line-derived neurotrophic factor (GDNF) heterozygous mice. Kidney Int *60*, 31-36.

Cullen-McEwen, L. A., Kett, M. M., Dowling, J., Anderson, W. P., and Bertram, J. F. (2003). Nephron number, renal function, and arterial pressure in aged GDNF heterozygous mice. Hypertension *41*, 335-340.

Cunniff, C., Curry, C. J., Carey, J. C., Graham, J. M., Jr., Williams, C. A., Stengel-Rutkowski, S., Luttgen, S., and Meinecke, P. (1993). Congenital diaphragmatic hernia in the Brachmann-de Lange syndrome. Am J Med Genet *47*, 1018-1021.

DeLise, A. M., Fischer, L., and Tuan, R. S. (2000a). Cellular interactions and signaling in cartilage development. Osteoarthritis Cartilage *8*, 309-334.

DeLise, A. M., Stringa, E., Woodward, W. A., Mello, M. A., and Tuan, R. S. (2000b). Embryonic limb mesenchyme micromass culture as an in vitro model for chondrogenesis and cartilage maturation. Methods Mol Biol *137*, 359-375.

Dorsett, D., Eissenberg, J. C., Misulovin, Z., Martens, A., Redding, B., and McKim, K. (2005). Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila. Development.

Dudley, A. T., Ros, M. A., and Tabin, C. J. (2002). A re-examination of proximodistal patterning during vertebrate limb development. Nature *418*, 539-544.

Elhassani, S. B., and Moyd, P. K. (1987). Topics in perinatal genetics. Cornelia de Lange syndrome. J Perinatol *7*, 156-157.

Fisher, D., and Mechali, M. (2003). Vertebrate HoxB gene expression requires DNA replication. Embo J *22*, 3737-3748.

Folk, J. C., Genovese, F. N., and Biglan, A. W. (1981). Coats' disease in a patient with Cornelia de Lange syndrome. Am J Ophthalmol *91*, 607-610.

Francis-West, P. H., Abdelfattah, A., Chen, P., Allen, C., Parish, J., Ladher, R., Allen, S., MacPherson, S., Luyten, F. P., and Archer, C. W. (1999). Mechanisms of GDF-5 action during skeletal development. Development *126*, 1305-1315.

Fujita, S. (2003). The discovery of the matrix cell, the identification of the multipotent neural stem cell and the development of the central nervous system. Cell Struct Funct *28*, 205-228.

Garcia-Castro, M. I., Marcelle, C., and Bronner-Fraser, M. (2002). Ectodermal Wnt function as a neural crest inducer. Science *297*, 848-851.

Gomperts, M., Wylie, C., and Heasman, J. (1994). Primordial germ cell migration. Ciba Found Symp *182*, 121-134; discussion 134-129.

Green, E. S., Stubbs, J. L., and Levine, E. M. (2003). Genetic rescue of cell number in a mouse model of microphthalmia: interactions between Chx10 and G1-phase cell cycle regulators. Development *130*, 539-552.

Greenwood, R. D., Sommer, A., Craenen, J., Waldman, J. D., and Rosenthal, A. (1977). Congenital heart disease in de Lange's syndrome. South Med J *70*, 80-81.

Gupta, D., and Goyal, S. (2005). Cornelia de-Lange syndrome. J Indian Soc Pedod

Prev Dent *23*, 38-41.

Gurrieri, F., Kjaer, K. W., Sangiorgi, E., and Neri, G. (2002). Limb anomalies: Developmental and evolutionary aspects. Am J Med Genet *115*, 231-244.

Higuchi, R., Krummel, B., and Saiki, R. K. (1988). A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res *16*, 7351-7367.

Holthusen, J., and Rottingen, J. A. (1998). [Cecal volvulus as a complication in Cornelia de Lange syndrome. A case report and literature review]. Tidsskr Nor Laegeforen *118*, 1559-1560.

Hu, D., Marcucio, R. S., and Helms, J. A. (2003). A zone of frontonasal ectoderm regulates patterning and growth in the face. Development *130*, 1749-1758.

Husain, K., Fitzgerald, P., and Lau, G. (1994). Cecal volvulus in the Cornelia de Lange syndrome. J Pediatr Surg *29*, 1245-1247.

Ireland, M., and Burn, J. (1993). Cornelia de Lange syndrome--photo essay. Clin Dysmorphol *2*, 151-160.

Ireland, M., Donnai, D., and Burn, J. (1993). Brachmann-de Lange syndrome. Delineation of the clinical phenotype. Am J Med Genet *47*, 959-964.

Jackson, L., Kline, A. D., Barr, M. A., and Koch, S. (1993). de Lange syndrome: a clinical review of 310 individuals. Am J Med Genet *47*, 940-946.

Janners, M. Y., and Searls, R. L. (1970). Changes in rate of cellular proliferation during the differentiation of cartilage and muscle in the mesenchyme of the embryonic chick wing. Dev Biol *23*, 136-165.

Jelsema, R. D., Isada, N. B., Kazzi, N. J., Sargent, K., Harrison, M. R., Johnson, M. P., and Evans, M. I. (1993). Prenatal diagnosis of congenital diaphragmatic hernia not amenable to prenatal or neonatal repair: Brachmann-de Lange syndrome. Am J Med Genet *47*, 1022-1023.

Johansson, C. B., Momma, S., Clarke, D. L., Risling, M., Lendahl, U., and Frisen, J.

(1999). Identification of a neural stem cell in the adult mammalian central nervous system. Cell *96*, 25-34.

Jorgensen, E. D., Durbin, R. K., Risman, S. S., and McAllister, W. T. (1991). Specific contacts between the bacteriophage T3, T7, and SP6 RNA polymerases and their promoters. J Biol Chem *266*, 645-651.

Kardon, G., Campbell, J. K., and Tabin, C. J. (2002). Local extrinsic signals determine muscle and endothelial cell fate and patterning in the vertebrate limb. Dev Cell *3*, 533-545.

Kelly, R. G., and Buckingham, M. E. (2002). The anterior heart-forming field: voyage to the arterial pole of the heart. Trends Genet *18*, 210-216.

Kim, I. T., Park, J. W., and Choi, W. C. (2005). A Korean case of Cornelia de Lange syndrome. Korean J Ophthalmol *19*, 153-155.

Kousseff, B. G., Newkirk, P., and Root, A. W. (1994). Brachmann-de Lange syndrome. 1994 update. Arch Pediatr Adolesc Med *148*, 749-755.

Lachman, R., Funamura, J., and Szalay, G. (1981). Gastrointestinal abnormalities in the Cornelia de Lange syndrome. Mt Sinai J Med *48*, 236-240.

Lagunas Flores, A., Duran Pinales, C., Aguirre Anorve, R., Rosales Peimbert, E., Hernandez Sarmiento, J. L., and Cuanalo Vida, M. (1981). [Brachmann-De Lange syndrome: Typus degenerativus amstelodamensis. Report of 5 cases and review of the literature]. Bol Med Hosp Infant Mex *38*, 331-345.

Le Douarin, N. M., Creuzet, S., Couly, G., and Dupin, E. (2004). Neural crest cell plasticity and its limits. Development *131*, 4637-4650.

Lee, S. H., Bedard, O., Buchtova, M., Fu, K., and Richman, J. M. (2004). A new origin for the maxillary jaw. Dev Biol *276*, 207-224.

Levin, A. V., Seidman, D. J., Nelson, L. B., and Jackson, L. G. (1990). Ophthalmologic findings in the Cornelia de Lange syndrome. J Pediatr Ophthalmol Strabismus *27*, 94-102.

Luzzani, S., Macchini, F., Valade, A., Milani, D., and Selicorni, A. (2003). Gastroesophageal reflux and Cornelia de Lange syndrome: typical and atypical symptoms. Am J Med Genet A *119*, 283-287.

Lyons, D. A., Guy, A. T., and Clarke, J. D. (2003). Monitoring neural progenitor fate through multiple rounds of division in an intact vertebrate brain. Development *130*, 3427-3436.

Mariani, F. V., and Martin, G. R. (2003). Deciphering skeletal patterning: clues from the limb. Nature *423*, 319-325.

Masumoto, K., Izaki, T., and Arima, T. (2001). Cornelia de Lange syndrome associated with cecal volvulus: report of a case. Acta Paediatr *90*, 701-703.

Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., and Green, M. R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res *12*, 7035-7056.

Moore, K. L., Persaud, T.V.N. (2003). The Developing Human. Clinically oriented embryology, 7th edn: Saunders).

Moorman, A. F., De Boer, P. A., Vermeulen, J. L., and Lamers, W. H. (1993). Practical aspects of radio-isotopic in situ hybridization on RNA. Histochem J *25*, 251-266.

Mumm, J. S., and Kopan, R. (2000). Notch signaling: from the outside in. Dev Biol *228*, 151-165.

Neuwald, A. F., and Hirano, T. (2000). HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. Genome Res *10*, 1445-1452.

O'Rahilly R, M. F. (1999). The embryonic human brain. An atlas of developmental stages 2nd edition Wiley-Liss

Orian, A., van Steensel, B., Delrow, J., Bussemaker, H. J., Li, L., Sawado, T., Williams, E., Loo, L. W., Cowley, S. M., Yost, C., *et al.* (2003). Genomic binding by the Drosophila Myc, Max, Mad/Mnt transcription factor network. Genes Dev *17*,

1101-1114.

Oudit, G. Y., Chow, C. M., and Cantor, W. J. (2006). Calcific bicuspid aortic valve disease in a patient with Cornelia de Lange syndrome: linking altered Notch signaling to aortic valve disease. Cardiovasc Pathol *15*, 165-167.

Pankau, R., and Janig, U. (1993). Diaphragmatic defect in Brachmann-de Lange syndrome: a further observation. Am J Med Genet *47*, 1024-1025.

Pei, R. S., Lin, C. C., Mak, S. C., Chi, C. S., and Chou, G. (2000). Barrett's esophagus in a child with de Lange syndrome: report of one case. Acta Paediatr Taiwan *41*, 155-157.

Piscione, T. D., and Rosenblum, N. D. (1999). The malformed kidney: disruption of glomerular and tubular development. Clin Genet *56*, 341-356.

Pohl, M., Bhatnagar, V., Mendoza, S. A., and Nigam, S. K. (2002). Toward an etiological classification of developmental disorders of the kidney and upper urinary tract. Kidney Int *61*, 10-19.

Polkinghorne, J., Hoffenberg, R., Kennedy, I., Macintyre, S. (1989). Review of the quidance on the research use of fetuses and fetal material. Cm, 762.

Qiao, J., Sakurai, H., and Nigam, S. K. (1999a). Branching morphogenesis independent of mesenchymal-epithelial contact in the developing kidney. Proc Natl Acad Sci U S A *96*, 7330-7335.

Qiao, J., Uzzo, R., Obara-Ishihara, T., Degenstein, L., Fuchs, E., and Herzlinger, D. (1999b). FGF-7 modulates ureteric bud growth and nephron number in the developing kidney. Development *126*, 547-554.

Reh, T. A., and Levine, E. M. (1998). Multipotential stem cells and progenitors in the vertebrate retina. J Neurobiol *36*, 206-220.

Reue, K. (1998). mRNA quantitation techniques: considerations for experimental design and application. J Nutr *128*, 2038-2044.

Robb, R. M. (2001). Congenital nasolacrimal duct obstruction. Ophthalmol Clin

North Am *14*, 443-446, viii.

Roberts, D. J. (2000). Molecular mechanisms of development of the gastrointestinal tract. Dev Dyn *219*, 109-120.

Rollins, R. A., Korom, M., Aulner, N., Martens, A., and Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. Mol Cell Biol *24*, 3100-3111.

Rollins, R. A., Morcillo, P., and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics *152*, 577-593.

Ruhin, B., Creuzet, S., Vincent, C., Benouaiche, L., Le Douarin, N. M., and Couly, G. (2003). Patterning of the hyoid cartilage depends upon signals arising from the ventral foregut endoderm. Dev Dyn *228*, 239-246.

Sataloff, R. T., Spiegel, J. R., Hawkshaw, M., Epstein, J. M., and Jackson, L. (1990). Cornelia de Lange syndrome. Otolaryngologic manifestations. Arch Otolaryngol Head Neck Surg *116*, 1044-1046.

Selicorni, A., Sforzini, C., Milani, D., Cagnoli, G., Fossali, E., and Bianchetti, M. G. (2005). Anomalies of the kidney and urinary tract are common in de Lange syndrome. Am J Med Genet A *132*, 395-397.

Shpak, E. D., Berthiaume, C. T., Hill, E. J., and Torii, K. U. (2004). Synergistic interaction of three ERECTA-family receptor-like kinases controls Arabidopsis organ growth and flower development by promoting cell proliferation. Development *131*, 1491-1501.

Sjogren, C., and Nasmyth, K. (2001). Sister chromatid cohesion is required for postreplicative double-strand break repair in Saccharomyces cerevisiae. Curr Biol *11*, 991-995.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol *98*, 503-517.

Tapadia, M. D., Cordero, D. R., and Helms, J. A. (2005). It's all in your head: new insights into craniofacial development and deformation. J Anat *207*, 461-477.

Tickle, C. (2003). Patterning systems--from one end of the limb to the other. Dev Cell *4*, 449-458.

Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J., and Strachan, T. (2004). NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. Nat Genet *36*, 636-641.

Tsukahara, M., Okamoto, N., Ohashi, H., Kuwajima, K., Kondo, I., Sugie, H., Nagai, T., Naritomi, K., Hasegawa, T., Fukushima, Y., *et al.* (1998). Brachmann-de Lange syndrome and congenital heart disease. Am J Med Genet *75*, 441-442.

Wagner, A. (2000a). Robustness against mutations in genetic networks of yeast. Nat Genet *24*, 355-361.

Wagner, A. (2000b). The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. Genetics *154*, 1389-1401.

Weier, H. U., and Rosette, C. (1988). Generation of labeled RNA probes from enzymatically amplified DNA templates. Nucleic Acids Res *16*, 11836.

Wick, M. R., Simmons, P. S., Ludwig, J., and Kleinberg, F. (1982). Duodenal obstruction, annular pancreas, and horseshoe kidney in an infant with Cornelia de Lange syndrome. Minn Med *65*, 539-541.

Wilcox, J. N. (1993). Fundamental principles of in situ hybridization. J Histochem Cytochem *41*, 1725-1733.

Wolpert, L., Tickle, C., and Sampford, M. (1979). The effect of cell killing by x-irradiation on pattern formation in the chick limb. J Embryol Exp Morphol *50*, 175-193.

Young, I. D., Ailles, L., Deugau, K., and Kisilevsky, R. (1991). Transcription of cRNA for in situ hybridization from polymerase chain reaction-amplified DNA. Lab Invest *64*, 709-712.

Zaffran, S., Astier, M., Gratecos, D., Guillen, A., and Semeriva, M. (1995). Cellular interactions during heart morphogenesis in the Drosophila embryo. Biol Cell *84*, 13-24.

Zaffran, S., and Frasch, M. (2002). Early signals in cardiac development. Circ Res *91*, 457-469.

Zitnik, G., and Martin, G. M. (2002). Age-related decline in neurogenesis: old cells or old environment? J Neurosci Res *70*, 258-263.

# CHAPTER 5 THE EXPRESSION PATTERNS AND MUTATION SCREENING OF HUMAN *MAU-2*

## 5.1 INTRODUCTION

The proper transmission of genetic material is crucial for the growth and survival of all organisms. Each time a eukaryotic cell divides; it must faithfully produce two copies of each chromosome, known as sister chromatids, and accurately distribute them between two daughter cells. Missegregation produces cells bearing too few or too many chromosomes, with often fetal consequences. Sister chromatids are moved to opposite poles of the cells by forces generated by microtubules, which attach to specialized structures at centromeric regions of the chromatids called kinetochores. The process can only deliver a complete set of cell's chromosomes to each daughter cell if the kinetochore of each chromatid is attached to microtubules with opposite orientations to those attached to its sister, which is known as amphitelic attachment. Amphitelic attachments result in a competition between microtubules attempting to pull sister chromatids apart, and cohesion complexes between sisters glue them together to resist the splitting force. It is sister chromatid cohesin complex that make it possible for cells to segregate their chromosomes long after duplication. Only when all chromosomes are bi-oriented on the metaphase plate, do cells start to destroy the connection between sisters and triggers their segregation to opposite poles during anaphase. The discovery of two multiprotein complexes, cohesin and condesin, has provided insight s into the molecular mechanisms of chromosome segregation (Ciosk et al., 2000; Toyoda et al., 2002)

Cohesin is a four-subunit protein complex, in which a heterodimer of SMC proteins (SMC1/SMC3) associates with two other non-SMC subunits, the Scc1/RAD21/Mcd1 and Scc3 proteins. In vertebrates there are two variants of Scc3, called SA1 and SA2. The cohesin forms a tripartite ring in which the open-V structure of the SMC heterodimer is closed by the binding of Scc1, a member of kleisins superfmaily proteins, to the head domains of SMC1 and SMC3 (Gruber et al., 2003; Haering and Nasmyth, 2003). It is proposed that the cohesin complex may hold sister chromatids together by embracing two DNA duplexes within its coiled-coil arms (Gruber et al., 2003). Proteolytic cleavages of the Scc1 subunit of the complex might open the ring and thereby trigger sister chromatid separation at the onset of anaphase (Uhlmann et al., 1999).

Loading of cohesin onto chromatin requires Scc2/Scc4 complexes during G1 phase (Ciosk et al., 2000). Scc2 are HEAT-repeat proteins that physically interact with cohesin. It has been suggested that Scc2 may promote hydrolysis of ATP bound to SMC's heads and thereby stimulate the opening of the ring to allow loading onto chromatin (Arumugam et al., 2003). In *Xenopus egg* extracts, loading of Scc2 on chromatin depends on the assembly of prereplication licensing complex but not on the initiation of DNA replication (Gillespie and Hirano, 2004). Cytological studies suggested that cohesin does not colocalize with Scc2/Scc4 on chromosome. However, Scc2 and Scc4 showed an identical pattern of localization and were distinct from the site occupied by cohesin (Ciosk et al., 2000; Lengronne et al., 2004) and their binding to chromatin coincided with strong transcription (Lengronne et al., 2004). High-resolution mapping of cohesion-binding sites in yeast revealed a preference of cohesin for sites of convergent transcription (Lengronne et al., 2004). It is proposed that after DNA is transported into cohesion ring, cohesin may be able to slide along

chromatin, responding to steric requirements of transcription. Cohesin is loaded onto chromosomes next to its loading factor Scc2/Scc4, from where it relocates to places of convergent transcription. Once loaded, prevention of further Scc2/Scc4 action on cohesin would be important to secure stability of the ring (Arumugam et al., 2003). Pushing cohesin away from its loadings sites by the transcriptional machinery until they reach the sites where RNA polymerase traveling in opposite direction meet (Lengronne et al., 2004). In vertebrate, most cohesin dissociates from chromatin at prophase, and only a small population in pericentrimeric region remains on chromosome by metaphase (Waizenegger et al., 2000). This step is regulated by two mitotic kinases, polo and aurora B (Haering and Nasmyth, 2003).


Fig.5.1 The cohesin cycles. The licensing complexes and cohesin are loaded during G1 and SCC2/SCC4 complexes help in the establishment of the cohesin loading onto chromatin. During S phase, the licensing complexe (Mcm2-7 complex) is displaced from DNA as it replicates, and cohesin ring established. During anaphase, cohesin is cleaved, thereby allowing segregation of sister chromatids.

Aside from their canonical role for the proper segregation in yeast chromosomes, recent studies have shown that cohesin might also contribute to gene regulation by regulating enhancer-promoter communication (Rollins et al., 2004; Rollins et al., 1999). It is proposed that the mechanism used by cohesin to attach two sister chromatids *in trans* might also be used *in cis* on the same sister chromatid to help enhancer contact promoters over long distance (Rollins et al., 2004). *gypsy*, a *Drosophila* transposon, functions as an insulator to block enhancer-promoter communication. Insertion of *gypsy* causes loss-of-function "cut wing" phenotypes in *Drosophila* when located between a wing enhancer and the *cut* promoter. In a genetic screen for mutations that exacerbate the *gypsy* induced cut wing phenotype identified *Nipped-B*, a drosophila homologue of yeast cohesin regulatory factor *Scc2* (Rollins et al., 1999). It was suggested that *Nipped-B* normally facilitates long-range enhancer-promoter communication and thus, mutation of *Nipped-B* would reduce their ability to overcome the *gypsy* insulator. Also, *Nipped-B* mutations cause weak cut-wing phenotypes even in the absence of gypsy, which supports a normal role in gene regulation at the *cut* locus. They also observed that Nipped-B protein and cohesin have opposite effects on this long-range regulation supports the model that *Scc2/Nipped-B* acts both as a loader and unloader of cohesin and certain developmental genes could be particularly sensitive to the silencing effect of cohesin. Reduction levels of Scc2 will result in the unopenable cohesin ring and cause transcriptional silencing during development. This might also explain that mutation of one copy of human *Nipped-B like (NIPBL)* gene cause Cornelia de Lange syndrome, a developmental defects characterized by growth deficiency, limb defects and mental retardation (Krantz et al., 2004; Tonkin et al., 2004). However, it remains unclear whether these developmental defects are derived from misregulation of cohesin or whether *Scc2/Nipped-B/NIPBL* may have a cohesin-independent functions that affects

the dynamics of other transcriptional regulators.

The amino acids sequences of the Human *Nipped-B like (NIPBL)* gene product, delangin, are well conserved among different species ranging from Scc2 of *S. cerevisiae,* Mis4 of *S. pombe* and Rad9 of *C. cinereus*, and Nipped-B of *D. melanogaster.* In many cases, the ortholog proteins are known to have a major role in sister chromatid cohesin. The C-terminal domain, spanning around 1300-1500 amino acids that is especially evolutionary strongly converved, contains at least 5 HEAT repeats. HEAT repeats are tandomly arranged bihelical structure and appear to function as flexible joints to wrap target substrate and as scaffolding on which other molecular components may assemble (Groves and Barford, 1999). Repeat proteins contain strings of tandem repeats of a basic structural element. Their construction is quite different from that of globular proteins, in which sequentially distant elements coalesce to form the protein. The different families of repeat proteins use their diverse scaffolds to present highly specific binding surfaces through which protein–protein interactions are mediated. HEAT repeats have been found in a variety of chromosome-associated proteins such as cohesin and condensing (Neuwald and Hirano, 2000) and may function as flexible scaffolding that recognizes and binds target substrates.

On the contrary, the N-terminal region of Scc2 orthologs is comparatively poor conserved. Notably, a consensus PXVXL motif was found to be located within the relatively poor conserved N-terminal. PXVXL motif binds directly to the chromoshadow domain of HP1 proteins. The HP1 proteins function as adaptors, bring together different proteins in multiprotein complexes through protein-protein interactions with the chromo and shadow domains. When shadow domain binds to

PXVXL consensus motif and with the recognition of Lys-9-methylated histone H3 by the chromo domain, the HP1 family proteins can mediate chromatin packaging and gene silencing (Lechner et al., 2005; Thiru et al., 2004). Thus, it is believed that delangin might also play a role in developmental regulation through the interaction with the heterochromatin proteins.

In order to elucidate the function of delangin as developmental regulators, we tried to find the proteins that interact with it. As described above, yeast Scc2 interacts with Scc4 to help in the loading of cohesin onto chromatin. In contrast to Scc2, Scc4 appeared to be poorly conserved during evolution (Ciosk et al., 2000). Standard BLAST analyses by using Scc4 protein query to screen for the protein and translated nucleotide sequence databases identify significant sequence matches in a limited group of organism such as *Candida glabrata*, *Kluyveromyces lactis* and *Eremothecium gossypii*. To identify the human homolog of yeast *Scc4*, we used PSI-BLAST to search the databases (http://www.ncbi.nlm.nih.gov/BLAST/). After 5 rounds of search identify the human *KIAA 0982* as the most possible ortholog with the highest value (the analysis was performed by Prof. Tom Strachen).

As described above, ortholgues of yeast Scc4 were not readily identified in other species. In order to search for the interacting paterner for the Drosophila Nipped-B, one group identified a 71-kDa Drosophila Nipped-B interacting protein which is co-immunoprecipitated with the FLAG-Nipped-B extract in a genetic screen (personal communication by Dale Dorsett).After query protein databases with selected value mass values, the 71-kDa protein was identified as the *Drosophila CG4203* gene (Seitan et al., 2006) and subsequent global yeast two-hybrid screen and the reciprocal yeast two-hybrid experiments in *Drosophila* confirmed that the CG4203 protein was

bound to Nipped-B protein (Seitan et al., 2006).

Recently, it has become clear that CG4203 and KIAA0892 are ortholgs of *C. elegans mau-2*, a gene broadly required for the proper guidance of cells and axons that undergo long-range migrations during development of worm (Benard et al., 2004; Takagi et al., 1997). Expression of Mau-2 in *C. elegans* was reported to be cytoplasmic (Takagi et al., 1997).

In order to determine whether KIAA 0892 protein is the human ortholog of *Drosophila* CG4203 and as the interacting protein of delangin, Northern blot analysis and tissue *in situ* hybridization were used for the expression studies of human *Mau-2* (*hMau-2*)

## 5.2. MATERIAL AND METHODS:

### 5.2.1. Northern blot analysis

#### 5.2.1.1 Probe used for Northern blot analysis and Tissue *in situ* hybridization

The human *Mau-2* cDNA probe was generated by PCR amplifying a 400bp fragment corresponding to exon 19 (3' UTR) of human *Mau-2* (*KIAA0892*). The primers used for the amplifying are: 5' TGTCTCTCCAGAGCCATCCT 3' and 5' AGGTGCAGATGTTTGGAAGC 3'. The cDNA fragment was cloned into the pGEM-T Easy Vector (Promega) as described previously. All the procedures are the same as described in Chapter 4.

#### 5.2.2.2 Probe labelling-random priming

The purified PCR amplified fragment was used as probe. Briefly, 25ng of the purified probe with distilled water to a total of 20 $\mu$l was denatured at 100℃ for 5 minutes in hot block and snapped cool on ice for 2 minutes. Then, 3 $\mu$l of 10x labelling buffer, 3 $\mu$l of mixed oligonucleotides, 3 $\mu$l of [$\alpha^{32}$ P] and 1 $\mu$l of Klenow were added to the probe mixture the reaction mixture was placed in oven at 37℃ for 1 hour. After one hour of labelling procedure, the unincorporated nucleotides was removed by applying the reaction mixture into the nick column (Amersham) as described in chapter 2. The purified labelled probe was denatured at 95℃ for 5 minutes and snapped cool on ice for 2 minutes before hybridization.

### 5.2.2.3 Prehybridization of the RNA blots

Preheat the hybridization buffer (ULTRAhyb, Ambion) at 65℃ for warming up. Then, prehybridization the blots of human adult and fetal RNA containing ~2 $\mu$ g of mRNA per lane (Clontech) with 10ml of hybridization buffer (ULTRAhyb, Ambion) in the oven at 42℃ for 30minutes.

### 5.2.2.4 Hybridization

The purified and denatured labelled probe was added to the prehybridized blots and hybridized at 42℃ for overnight.

### 5.2.2.5 Post-hybridization wash and film development

After overnight hybridization, the blots were washed with 2xSSC/0.1% SDS twice, 15 minutes each. Then the membranes were washed on 1xSSC/0.1%SDS twice, 15 minutes each followed by 0.5xSSC/0.1%SDS, if necessary. Washes generally started at room temperature, but higher temperature, lower concentration of SSC and higher concentration of SDS have been used for more stringent washings. After washing, the blots were placed between two sheets of polyethylene wrap in a cassette and X-ray film (Kodak) was placed on top of the blots. The cassette were kept overnight at -70℃. The autoradiography was performed by passing the X-ray film through Amidol developer which converts the $Ag^{+}$ ions to Ag atoms and thiol which removes the unexposed silver halide crystals (Buongiorno-Nardelli and Amaldi, 1970).

### 5.2.2 Tissue *in situ* hybridization:

The probe we used for tissue *in situ* hybridization is same as that of Northern blotting.

All the procedures are same as those described in Chapter 4.

## 5.3 RESULTS

### 5.3.1 Northern blot analysis

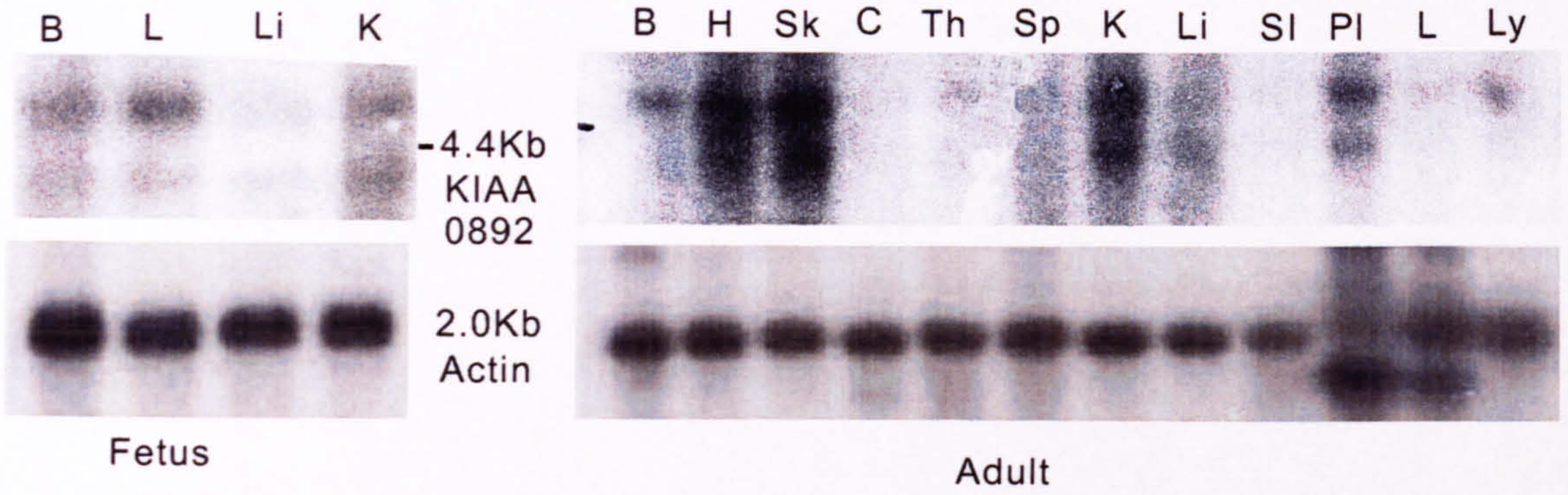



Fig. 5.2 Northern analysis of *KIAA 0892* transcript. B: brain; C: colon; H: heart; K: kidney; L: lung; Li: liver; Ly: peripheral lymphocyte; Pl: placenta; Sk: skeletal muscle; SI: small intestine; Sp: spleen; Th: thymus.

The expression of *hMau-2* transcript was seen on multiple organ system with unequal intensity. There are two isoforms and the transcript sizes are around 4.0 Kb and 6.0Kb respectively. The expression was strongest on fetal brain, lung, and kidney and also on adult brain, heart, skeletal muscle, kidney, placenta; expression was only moderate on fetal and adult liver and lymphocytes; weak expression was seen on adult colon, small intestine, thymus and lung.

### 5.3.2 Tissue *in situ* hybridization by using riboprobes

The expression of *hMau-2* on human embryonic sections showed similar pattern to that of *NIPBL*. The expression was strong in the proliferation zone of the developing embryonic brain and muscles on CS18 and CS23 embryos.

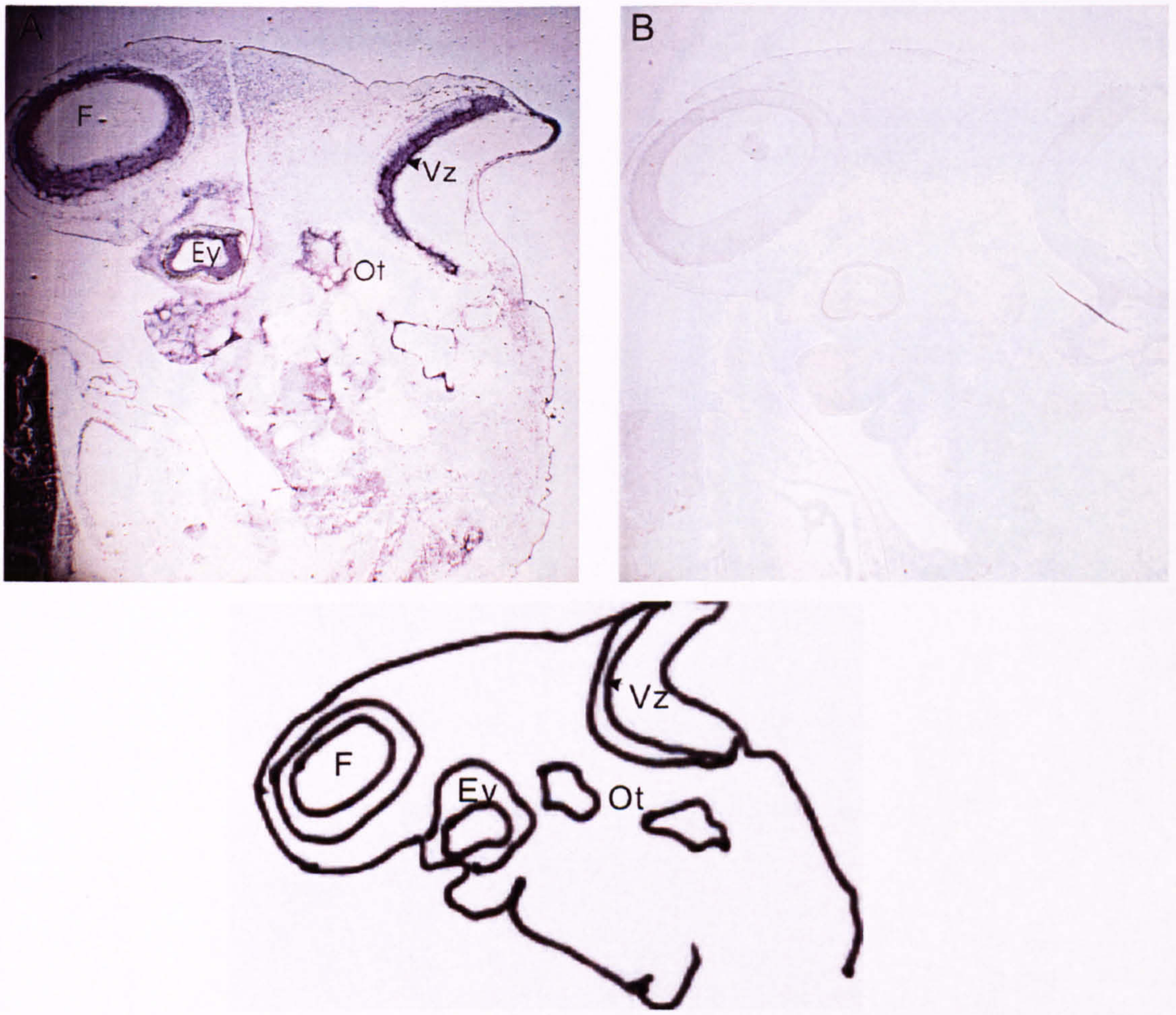



Fig.5.3 *hMau-2* expression on CS23 embryo. The expression was seen on (A) ventricular zone of the developing brain, retina of the developing eye and mesenchyme of the developing ear. (B) no expressison was seen on control slide by using sense strand as the riboprobe. F: forebrain; Ey: developing eye; Ot: otic placode. Vz: ventricular zone of the developing brain.

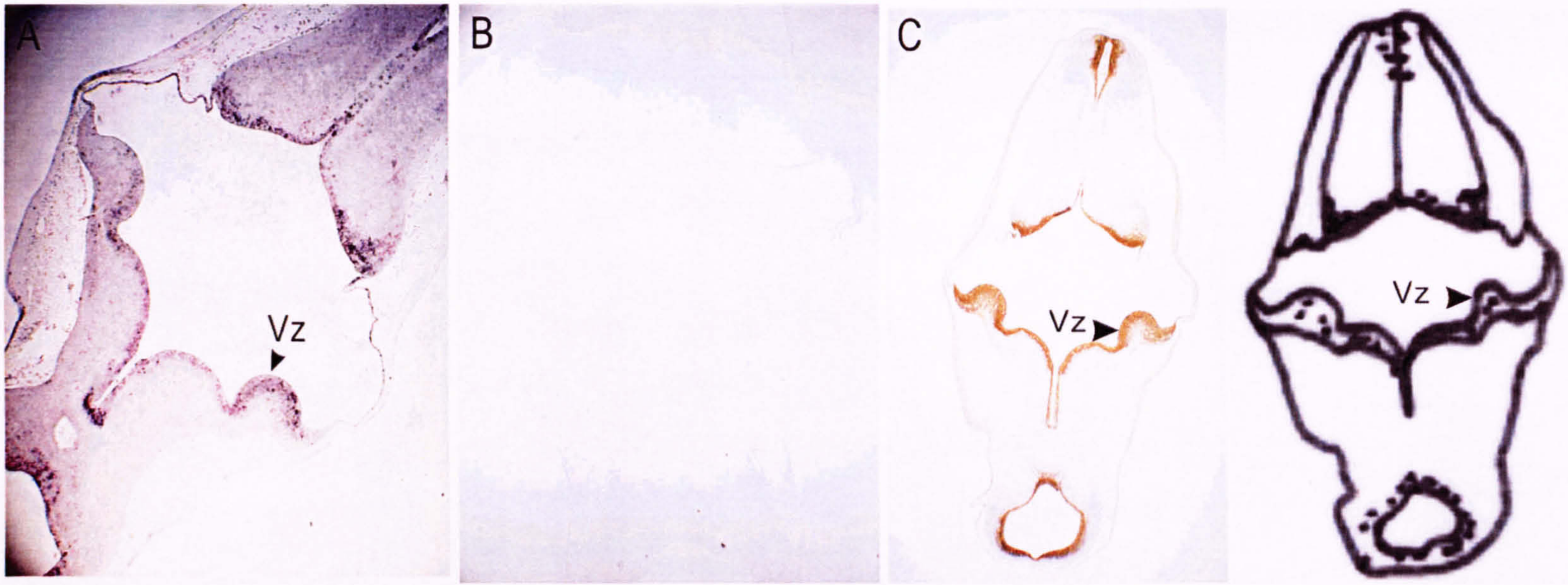



Fig. 5.4 *hMau-2* expression on transverse section of CS18 embryo. (A) expression was seen on ventricular zone of the developing brain. (B) no expression was seen on control slide by using sense strand as riboprobe. (C) immunohistochemistry of the nearby slide by using PCNA as probe.

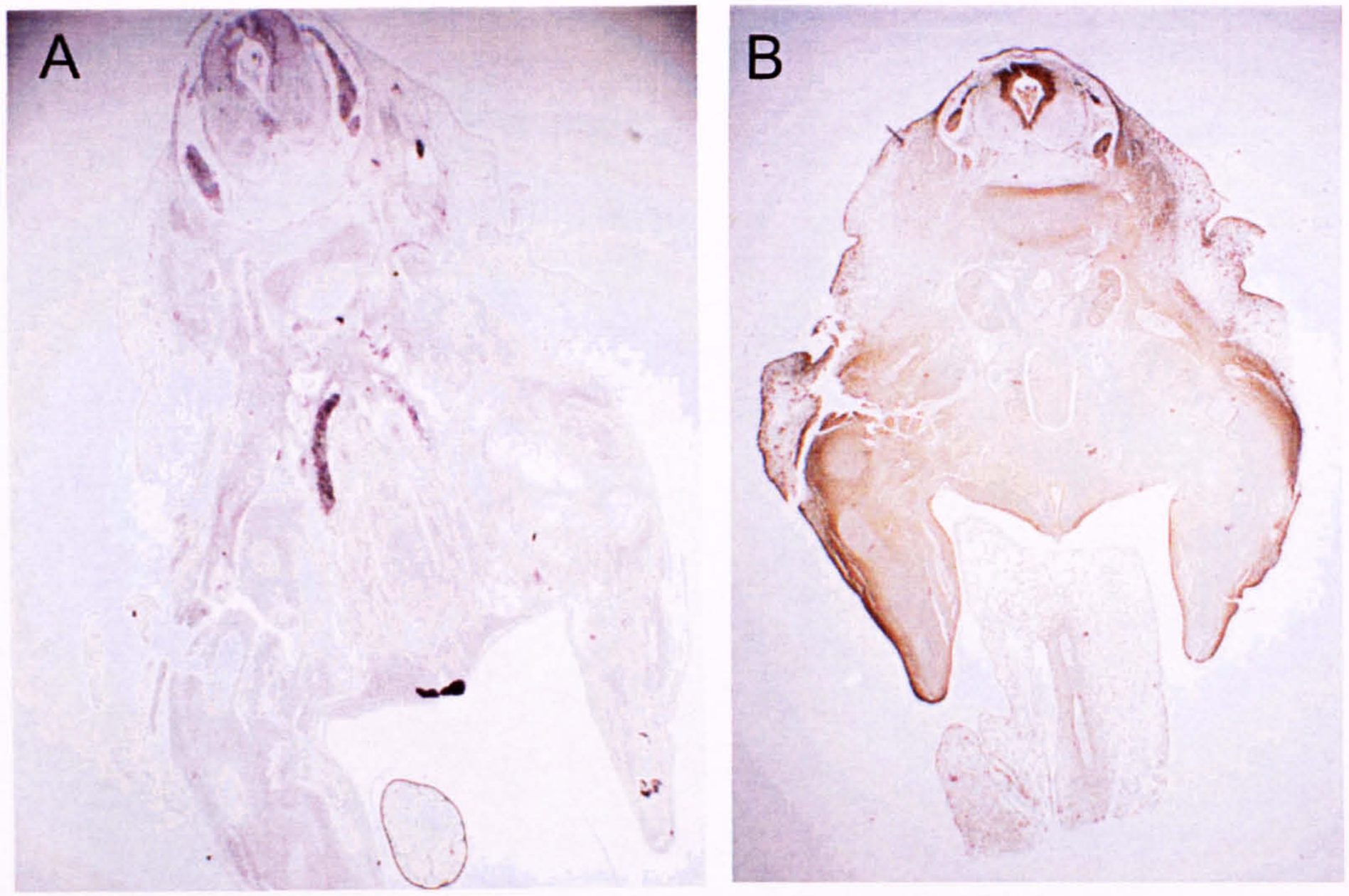



Fig 5.5 *hMau-2* expression on CS18 embryonic section. (A) Human *Mau-2* as a riboprobe. (B) Immunohistochemical staining by using PCNA as a probe. The expression pattern of human *Mau-2* is similar to that of PCNA.

Generally speaking, the expression pattern of *hMau-2* in developing human embryo is less tissue specific in sagital sections (Fig. 5.4) as compared to *NIPBL* and is more similar to the expression pattern of PCNA. However, expression pattern of *hMau-2*, *NIPBL* and PCNA in human embryonic brain are very similar. Thus, it is also expressed on the active proliferative zone of the developing brain.

### 5.3.3 Mutation screening

The similar expression pattern seen in *NIPBL* and *hMau-2* transcripts prompted us to further investigate if *hMau-2* is a Human Scc4 homologue, the interactor protein of delangin. Human delangin was found to be co-immunoprecipitated with *hMau-2* gene product (Seitan et al. 2006) in human hela cells. Thus, it is highly likely that hMau-2 is the interacting partner of the human delangin. Therefore, we decided to perform mutation screens in *NIPBL* mutation-negative CdLS patients. The *hMau-2* contains 4161bps, 218 amino acids and 19exons. Primers were designed on exon-intron flanking regions (see appendix C) for the primer information and amplification conditions).

No mutation was found in patients with CdLS up to now. This is compatible with the phenotypes observed in worms with *Mau-2* mutation (homologue of *hMau-2*). Heterozygous worms produced by the homologous mutant mother appear completely wild type which suggested autosomal recessive inheritance. As autosomal dominant inheritance is suggested as mode of transmission for CdLD, this makes *hMau-2* an unlikely candidate for CdLS.

However, considering that Roberts syndrome, an autosomal recessive disorder

characterized by craniofacial anomalies, tetraphocomelia and loss of cohesion at heterochromatic regions of centromeres and the Y chromosome, is caused by mutations in a novel human gene, *ESCO2*. The *ESCO2* gene product is required for the establishment of sister chromatid cohesion during S phase (Vega et al., 2005). Moreover, the highest mutation detection rate of CdLS patients in NIPBL gene was around 50%. Thus, it is possible that other gene products involved in the establishment of cohesin complexes might be candidates for CdLS.

Recently, another gene, *SMC1*, was found to be mutated in X-linked CdLS patients (Musio et al., 2006). SMC1 constitutes part of the cohesin complexes. Thus, it is necessary for us to perform mutation screening of the human *SMC1* gene in *NIPBL* mutation negative CdLS patients, especially for patients with suspected X-linked inheritance within families. Other genes involved in the establishment of cohesin may need further investigation for the possibility of being other CdLS disease causing genes.

## 5.4 DISCUSSION:

In comparison of *NIPBL*, the expression of *hMau-2* transcript on northern blot analysis revealed that the intensity of expression of *hMau-2* is higher in both fetal and adult brain and fetal lung than those of *NIPBL*. The expression of *hMau-2* is also higher in adult lymphocyte than that of *NIPBL*. In a serial of experiments in tissue hybridization studies, *hMau-2* tends to has more background and less tissue specific expression than *NIPBL* (Tonkin et al., 2004). To elucidate this, more serial expression studies by using fetal and postnatal mouse sections may be necessary.

In metazoan, Mau-2 was found to be associated with cellular and axonal migration (Benard et al., 2004; Takagi et al., 1997). Defects in Mau-2 will lead to uncoordinated movements, early demise in worms. However, the larval lethality can be rescued by the maternal effects. That is, homozygous mutants derive from heterozygous mothers; the uncoordinated phenotypes could be fully rescued. However, phenotype which develops post-embryonically such as egg-laying can not be rescued. Thus, mau-2 is necessary for *C. elegans* both in early developmental and post-embryonic stages (Takagi et al., 1997). Up to now, no vertebrate phenotypes are reported to be associated with defects in vertebrate mau-2 (KIAA0892) proteins. However, this may be due to the fact that vertebrate has just recently been recognized (Seitan et al., 2006; Watrin et al., 2006) and thus not much is known about it. Or according to the phenotypes observed in worm may suggest that the phenotypes in vertebrate null mutants of vertebrate *Mau-2* may be rescued by the wild-type maternal *Mau-2* existed on the cytoplasm of the fertilized ovum and the surviving null mutants may not reproduce. Heterozygous mutant may not result in any defective phenotypes. Nevertheless, to further elucidate the development regulatory role of vertebrate Scc4

and Scc2 during different developmental stages even after birth, knock out experiments using animal models such as mouse is necessary for our future work.

## 5.5 REFERENCES


Arumugam, P., Gruber, S., Tanaka, K., Haering, C. H., Mechtler, K., and Nasmyth, K. (2003). ATP hydrolysis is required for cohesin's association with chromosomes. Curr Biol *13*, 1941-1953.

Benard, C. Y., Kebir, H., Takagi, S., and Hekimi, S. (2004). mau-2 acts cell-autonomously to guide axonal migrations in Caenorhabditis elegans. Development *131*, 5947-5958.

Buongiorno-Nardelli, M., and Amaldi, F. (1970). Autoradiographic detection of molecular hybrids between RNA and DNA in tissue sections. Nature *225*, 946-948.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol Cell *5*, 243-254.

Gillespie, P. J., and Hirano, T. (2004). Scc2 couples replication licensing to sister chromatid cohesion in Xenopus egg extracts. Curr Biol *14*, 1598-1603.

Groves, M. R., and Barford, D. (1999). Topological characteristics of helical repeat proteins. Curr Opin Struct Biol *9*, 383-389.

Gruber, S., Haering, C. H., and Nasmyth, K. (2003). Chromosomal cohesin forms a ring. Cell *112*, 765-777.

Haering, C. H., and Nasmyth, K. (2003). Building and breaking bridges between sister chromatids. Bioessays *25*, 1178-1191.

Krantz, I. D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L. A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C. A., *et al.* (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nat Genet *36*, 631-635.

Lechner, M. S., Schultz, D. C., Negorev, D., Maul, G. G., and Rauscher, F. J., 3rd (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. Biochem Biophys

Res Commun *331*, 929-937.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G. P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature *430*, 573-578.

Musio, A., Selicorni, A., Focarelli, M. L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P., and Larizza, L. (2006). X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat Genet *38*, 528-530.

Neuwald, A. F., and Hirano, T. (2000). HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. Genome Res *10*, 1445-1452.

Rollins, R. A., Korom, M., Aulner, N., Martens, A., and Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. Mol Cell Biol *24*, 3100-3111.

Rollins, R. A., Morcillo, P., and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics *152*, 577-593.

Seitan, V. C., Banks, P., Laval, S., Majid, N. A., Dorsett, D., Rana, A., Smith, J., Bateman, A., Krpic, S., Hostert, A., *et al.* (2006). Metazoan Scc4 Homologs Link Sister Chromatid Cohesion to Cell and Axon Migration Guidance. PLoS Biol *4*, e242.

Takagi, S., Benard, C., Pak, J., Livingstone, D., and Hekimi, S. (1997). Cellular and axonal migrations are misguided along both body axes in the maternal-effect mau-2 mutants of Caenorhabditis elegans. Development *124*, 5115-5126.

Thiru, A., Nietlispach, D., Mott, H. R., Okuwaki, M., Lyon, D., Nielsen, P. R., Hirshberg, M., Verreault, A., Murzina, N. V., and Laue, E. D. (2004). Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. Embo J *23*, 489-499.

Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J., and Strachan, T. (2004). NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly

Nipped-B, is mutated in Cornelia de Lange syndrome. Nat Genet *36*, 636-641.

Toyoda, Y., Furuya, K., Goshima, G., Nagao, K., Takahashi, K., and Yanagida, M. (2002). Requirement of chromatid cohesion proteins rad21/scc1 and mis4/scc2 for normal spindle-kinetochore interaction in fission yeast. Curr Biol *12*, 347-358.

Uhlmann, F., Lottspeich, F., and Nasmyth, K. (1999). Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. Nature *400*, 37-42.

Vega, H., Waisfisz, Q., Gordillo, M., Sakai, N., Yanagihara, I., Yamada, M., van Gosliga, D., Kayserili, H., Xu, C., Ozono, K., *et al.* (2005). Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nat Genet *37*, 468-470.

Waizenegger, I. C., Hauf, S., Meinke, A., and Peters, J. M. (2000). Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. Cell *103*, 399-410.

Watrin, E., Schleiffer, A., Tanaka, K., Eisenhaber, F., Nasmyth, K., and Peters, J. M. (2006). Human Scc4 is required for cohesin binding to chromatin, sister-chromatid cohesion, and mitotic progression. Curr Biol *16*, 863-874.

# CHAPTER 6: CONCLUSION & FUTURE WORKS

In 2004, we identified a novel gene, *NIPBL*, which is found to be mutated in around 45% of CdLS patients (Tonkin et al., 2004). Standard point mutational screening of the *NIPBL* exons and sequences flanking the exon-intron boundaries revealed a wide variety of pathogenic mutations. More than half of the mutations resulted in premature termination. Loss of function causing haploinsufficiency is supported by patients with heterozygous deletion involving the whole *NIPBL* gene. No genotype-phenotype correlation was observed.

Our preliminary data obtained from tissue *in situ* hybridisation on early human embryos by using riboprobes suggest that the expression patterns of *NIPBL* are similar to that of the proliferating marker, PCNA. Thus, it is plausible that defects in Human Scc2, *NIPBL*, will result in aberrant proliferation of the matrix cells and differentiation and thus affect the size and function of the developing brain. However, most of the cells undergoing stage II or III of cytogenesis may only appear later in the fetal or even postnatal stage. To examine the effect of *NIPBL* on neuronogenesis, glial progenitors differentiation and axonal migration, further investigation by using mouse or rat at the later developmental stage, such as the fetal and even the postnatal stage, is necessary.

The major role of Scc2 in yeast and Xenopus is to interact with Scc4 to help in loading the cohesin ring onto chromatin in the early S phase during cell replication. We also identified the sequence of the human Scc4 homologue, which we call human *Mau-2*. The cohesin ring can help to hold the sister chromatids together during cell replication. Defects occurring in the cohesion can cause premature sister chromatids

separation (PSCS). However, no significant PSCS in CdLS patients' chromosomes was observed in our series (personal communication by Tom Strachan). We speculated that haploinsufficiency of delangin is not enough to cause PSCS. Thus, it is necessary to test if homozygous defects in *NIPBL* will cause PSCS.

In Drosophila, Nipped-B was also found to participate in remote activation of the *cut* and *Ultrabithorax* genes. Reducing the *Nipped-B* dosage reduces activation of the wild-type *cut* gene by the remote wing margin enhancer and causes the wing margin defect in Drosophila. Thus, to test if mouse delangin and Mau-2 help in cohesion loading onto the chromatin and participate in the remote control of the mammalian *cut* and *ultrabithorax* homologues, we need to examine the levels of gene expression in *NIPBL/hMau-2/hSMC1/or hSCC1* by siRNA knockdown in selected cell lines and try to elucidate the relationship between these genes. Moreover, to overcome the possibility of cell death after siRNA knockdown, conditional siRNA knockdown technology and phenotypic rescue by cDNA transfection need to be used.

Furthermore, to examine if the highly conserved sequence within intron 1 contains a *cis*-regulatory element, experiments need to be designed to test the transactivating ability of this sequence.

As mutations in *NIPBL* may lead to the prolongation for each cell cycle and result in a decreased cell number during organogenesis, growth deficiency may occur as seen in patients with CdLS. Although severe patterning defects (except in limbs) are seldom reported to be associated with CdLS, *NIPBL* mutations seen in patients with CdLS are due to hemizygous rather than homozygous defects and thus lead to less severe phenotypes. It is possible that homozygous mutation of *NIPBL* will severely

impair cell cycle progression and is non-viable for the developing early human embryos. Thus, it is important to establish an animal gene knockout model to understand the effect of *NIPBL* gene product and the impact it has on embryonic development. This work is now underway.

# Appendix A: *NIPBL* specific primers

| Exon | Forward primer | Reverse primer | Size |
|---|---|---|---|
| 2 | AGGTTGAACAAACCAAAGCAG | AAAGGGCAGTTTCAGTTGCT | 383 |
| 3 | GGAAGAGGAGGAATGCCTTT | TGAAATAAAACCAGGAATACGG | 383 |
| 4 | TTGGCCATACCAGTGTGATT | GCTTGTTAAAGAAGACACCTGGA | 299 |
| 5 | AAAATCTCTGGAATGTTTGAAAGAA | TCCACCATTACCTAAATGTATGAAC | 291 |
| 6 | GAGGACTTTGTGACAGTCAGATTT | GCCATATCTTTAAAAAGGTTTACTGAC | 377 |
| 7 | GGCAGAGTGTTAAGATTTGATGA | CAAATTTTGAAATGGAAATACTAGG | 400 |
| 8 | GAAAACAGGAAAGTGCAGAAGAA | CATGATCCAGACCTGCTTGT | 394 |
| 9A | GTGAAACCACCACAACTG | TGAGCAGCATTTAGTGGGC | 429 |
| 9B | CAGGACAGACTTCAAAAACACC | CCAAATCTCATATAGTTGTTTCAG | 512 |
| 10A | TTGCATTTGCATTTTACTCCA | GTGTCTCAGGATGGTTTTCTGG | 428 |
| 10B | TACGGGAAATGGGGTCAAGGC | AGGCTCAACTATGGTGCTCTCG | 424 |
| 10C | TGAGAGCAGAACAACTGAATGC | TGGCTTTCCAGGAATCCCTCC | 352 |
| 10D | AGGTGAGAGCCGCCCTGAAACTC | CACGAGGACTGTCAGGTCTTGA | 467 |
| 10E | TGAATCAGGGGACTCAAGGG | AGGGAACTTCTTGATTTGTCCTC | 468 |
| 10F | AGGAGCTAAGCCTGTAGTTGTG | CTTGAGTAGTGGGTGGGGAAGA | 349 |
| 11 | ATTGTCACTTTAGGGTTAAGAGTATT | AAGGTGACTGTGCTTTTGCT | 446 |
| 12 | TTCCCCATGTGATTCATTTGT | CGATAATCACTGCACATAGAAACT | 432 |
| 13-14 | AGTTTCTATGTGCAGTGATTATCG | TGATTTCAAGGTAGGACACATCA | 485 |
| 15 | TGTCCTTTTTCATTCAGGGTTT | TCTCAGAAGTCAACAATGAGTTTTC | 400 |
| 16 | CCTCCATAGCTCAAAGGGAAT | TGAGAAAATGGAAAATGAAACAA | 279 |
| 17 | TCCACCAGTGAAAATCAAATCA | GGTGCCATTTTAAGTCCTATTTTG | 400 |
| 18 | GCTTTATCTTCCAGGTTCTGTAGC | GGAAAAATTCTTGATAATATAGTGCTG | 388 |
| 19 | CCAGTAAGCTTATGAATGTATTGGAA | TTGAAATGAATGTACTGAAAACCA | 292 |
| 20 | CATTTTCATTCTAAATGGCAGGTA | AAATCAGACAGAAATGAAGAATAAATG | 293 |
| 21 | TGGCAAACACAGTATCGTGAA | AGCGAGACTCCGTCTCAAAA | 357 |
| 22 | CCCCTCTTCAGTAATAACATGACAAT | GAAACCAAGTGTCAGGAAGATG | 299 |
| 23 | GCTAACAATTTCAATCATGTTGGT | GCACATGCAAAAAGAGATGC | 352 |
| 24 | TGGGACAATATCACAGGAAAAA | TTGAATCCACACAAAATGAAATG | 375 |
| 25 | CAGTTTTTCCTTCAGATTTGTGTTT | TGAAAATCAATATTCCAAACAACTT | 300 |
| 26-27 | TTTGTATTCCTGTAATGTGAGCA | GATTACAGGCGTGAGCCACT | 682 |
| IntF1 | CACACCTTCTCAGTTTAGCACA | GATTACAGGCGTGAGCCACT | 400 |
| IntF2 | TTTGTATTCCTGTAATGTGAGCA | GGGAAATAGATATCAAGGAGATAAATG | 362 |
| 28-29 | TTGTTTTCTTTTGCATGTTTTCA | TTGTTTAGAAAGTGGCAATAGTTCA | 576 |
| 30 | GCAGAAAGCATGTAAAAAGCAA | CGATTCATTGTAAAGTGTTTGGA | 392 |

| | | | |
|---|---|---|---|
| 31 | TCCTGGCAGTTTGTGTTTTG | GGCAGCAGTACGAAATGTCA | 345 |
| 32 | TTTTGAATTGAGAAAATTGAAACAT | TCCCTACCAAAAGAACTAAAAAGG | 283 |
| 33 | ACCTTAGGTCTTACACAGCAA | TGTGCTCAACTAGGTTATCAAC | 362 |
| 34 | TTGAGGCCTATACTGGACCTATTT | GTGGTTGACGCATGTGAACT | 335 |
| 35-26 | TATTTCTGCCCCCAAATACG | GAGAGGACCACGGTGGATAA | 598 |
| 37 | TGGTGGCACACGACTGTAATCC | TCATCCTGGGTCACTACTGTCAT | 467 |
| 38 | CACACAAATTCTCATTGTTTTAAGG | TGACAAGAATGCAACCCAGT | 362 |
| 39 | CTAGGTAAGGCCACCAGCAT | CCCAGCTATAGACCTCAGCA | 472 |
| 40 | TGTAACGTTTTGTGATTTTGGTT | AGAAGAGTAAGAGGAACGAAGAACT | 400 |
| 41 | AGGAAGGCCTATAAGGTTAAATTC | AACTTAGTTGCATGTTTAGAAGATAGG | 364 |
| 42 | GCTAGCCTCAGAATGTAATGCTC | TCCCCTTCACTTCTGACTGTG | 459 |
| 43 | TGAGGTGAAAGTGCCCTGTA | CCAAGTCAAGTATTGCCCAGA | 398 |
| 44 | CAAGCTGTTGAATGGAGCATAC | CATGAGCCACCACACCCAGC | 434 |
| 45 | TGAAGCTGTCCTAGGATCACAA | TGCAATGGGAAAGACAAGC | 500 |
| 46 | CACCCACACCAAACTACTGC | CTTGGTGTACATTTAATAGACTGACAT | 382 |
| 47 | GCGTCAAGGGATTAAAAGCA | CCTGTAAATTTGGCTGCATGT | 488 |

## Appendix B

| Carnegie stage | Days (approx) | Size (mm) | Events |
|---|---|---|---|
| 1 | 1 | 0.1-0.15 | fertilized oocyte, pronuclei |
| 2 | 2-3 | 0.1-0.2 | cell division with reduction in cytoplasmic volume, formation of inner and outer cell mass |
| 3 | 4-5 | 0.1-0.2 | loss of zona pellucida, free blastocyst |
| 4 | 5-6 | 0.1-0.2 | attaching blastocyst |
| 5 | 7-12 | 0.1-0.2 | implantation |
| 6 | 13-15 | 0.2 | extraembryonic mesoderm, primitive streak |
| 7 | 15-17 | 0.4 | gastrulation, notochordal process |
| 8 | 17-19 | 1.0-1.5 | primitive pit, notochordal canal |
| 9 | 19-21 | 1.5-2.5 | Somite Number 1 - 3<br>neural folds, cardiac primordium, head fold |
| 10 | 22-23 | 2.0-3.5 | Somite Number 4 - 12<br>neural fold fuses |
| 11 | 23-26 | 2.5-4.5 | Somite Number 13 - 20<br>rostral neuropore closes |
| 12 | 26-30 | 3.0-5.0 | Somite Number 21 - 29<br>caudal neuropore closes |
| 13 | 28-32 | 4.0-6.0 | Somite Number 30<br>leg buds, lens placode, pharyngeal arches |
| 14 | 31-35 | 5.0-7.0 | lens pit, optic cup |
| 15 | 35-38 | 7.0-9.0 | lens vesicle, nasal pit, hand plate |
| 16 | 37-42 | 8.0-11 | nasal pits moved ventrally, auricular hillocks, foot plate |
| 17 | 42-44 | 11-14 | finger rays |
| 18 | 44-48 | 13-17 | ossification commences |
| 19 | 48-51 | 16-18 | straightening of trunk |
| 20 | 51-53 | 18-22 | upper limbs longer and bent at elbow |

| | | | |
|---|---|---|---|
| 21 | 53-54 | 22-24 | hands and feet turned inward |
| 22 | 54-56 | 23-28 | eyelids, external ears |
| 23 | 56-60 | 27-31 | rounded head, body and limbs |
| Following this stage **Fetal Development** occurs until birth (approx 40 weeks) | | | |

Sources: UNSW Embryology,Version 5.3.-An educational resource for learning concepts in embryological development (http://embryology.med.unsw.edu.au/)

# Appendix C

## *hMau-2* specific primers

| Exon | Forward primer | Reverse primer | Size |
|---|---|---|---|
| 1 | AGC CTA TCA GCG ATG CCT AA | GCT TTC GTG TCC CTT GAG AG | 679 |
| 2 | AGT AAG CCT ACC CTC TCA GGC | GGA CAC TCC AGG TGC TCT GTG | 312 |
| 3 | GCT GTG CTC CTC TTT GTG CT | ATT CCA ACT CCC GGA ACC | 384 |
| 4 | ACT TTT AAT AAA GTA AAT TGG CAA CC | TGT CCC CAC CTA CTC CAT GT | 389 |
| 5 | GCC AAC TGG AGC TGT GTA CC | TGC CAT TTC TGT TAT TTG TCC T | 371 |
| 6 | GCA GTC TTG TAA GGG GAC TGT T | GCT CTT GCC TCA GCA TGA AT | 396 |
| 7 | CAC CCT ATG GGA TTC ATG CT | GAC CCT GTT CCC AGA TAC CA | 539 |
| 8-9 | TGG CTG ATT CAT GGG AGT GCG | TCT GAC GTT GGA CAC CAG CAT | 670 |
| 10 | CAC TGT GGG AAC CAA CCT CT | GAT CAG TGG TGC CCA TGC | 483 |
| 11 | CCT CAT GGT GGA CGT GTC T | ATC GTG CCT GGG AAG GTT AT | 392 |
| 12 | GCT CCC AGT GAG CAT CTT GT | GCC CCA TAG CTG TCC CTT AT | 472 |
| 13 | CCG TAA CGC AGA TGA TGG TT | CAT AGC CCG ATT CAC CTC TG | 466 |
| 14 | CGG CCC TTC TCA CCA GAG | GGT TTA CAA AAC GGC AGT GG | 385 |
| 15 | CTC CCC GTG CTC TTT GGT | CGG TTC AGG TCC TCA GCA T | 399 |
| 16 | TCC AGG GAC GCT ACA ACG AGG | TCA GGG CAT CCA CCT AGA CGT | 368 |
| 17 | GAT GGG CTC TGA GGT AAG CA | TCG CAG CCT CTG TGT CCT AT | 493 |
| 18 | GAC AGG GCT CAG GTC CAG T | CCG GGA CTC TGT TCT CCA T | 739 |
| 19 | CTC AGG TGA TCC CTT GAT GG | CAC CTG TGA GCC TCT GCA C | 493 |

PAGE

MISSING

IN

ORIGINAL

nature genetics

# *NIPBL*, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome

Emma T Tonkin[1,3], Tzu-Jou Wang[1,3], Steven Lisgo[1], Michael J Bamshad[2] & Tom Strachan[1]

**Cornelia de Lange syndrome (CdLS) is a multiple malformation disorder characterized by dysmorphic facial features, mental retardation, growth delay and limb reduction defects[1,2]. We indentified and characterized a new gene, *NIPBL*, that is mutated in individuals with CdLS and determined its structure and the structures of mouse, rat and zebrafish homologs. We named its protein product delangin. Vertebrate delangins have substantial homology to orthologs in flies, worms, plants and fungi, including Scc2-type sister chromatid cohesion proteins, and *D. melanogaster* Nipped-B. We propose that perturbed delangin function may inappropriately activate *DLX* genes, thereby contributing to the proximodistal limb patterning defects in CdLS. Genome analyses typically identify individual delangin or Nipped-B-like orthologs in diploid animal and plant genomes. The evolution of an ancestral sister chromatid cohesion protein to acquire an additional role in developmental gene regulation suggests that there are parallels between CdLS and Roberts syndrome.**

The multisystem nature of the CdLS phenotype suggests that it is caused by a microdeletion or microduplication affecting several genes or by a single gene that regulates various target genes. A high-density BAC microarray comparative genome hybridization screen found no evidence for a consistent pattern of microdeletion or microduplication[3]. Because CdLS is rare and most cases are sporadic, genome-wide linkage screens are problematic. As an alternative, we analyzed chromosomal breakpoints associated with CdLS, focusing first on three classical cases with *de novo* balanced translocations, including the previously described translocations t(3;17)(q26.3;q23.1)[4] and t(14;21)(q32;q11)[5]. We first analyzed the 3q26.3 breakpoint because of

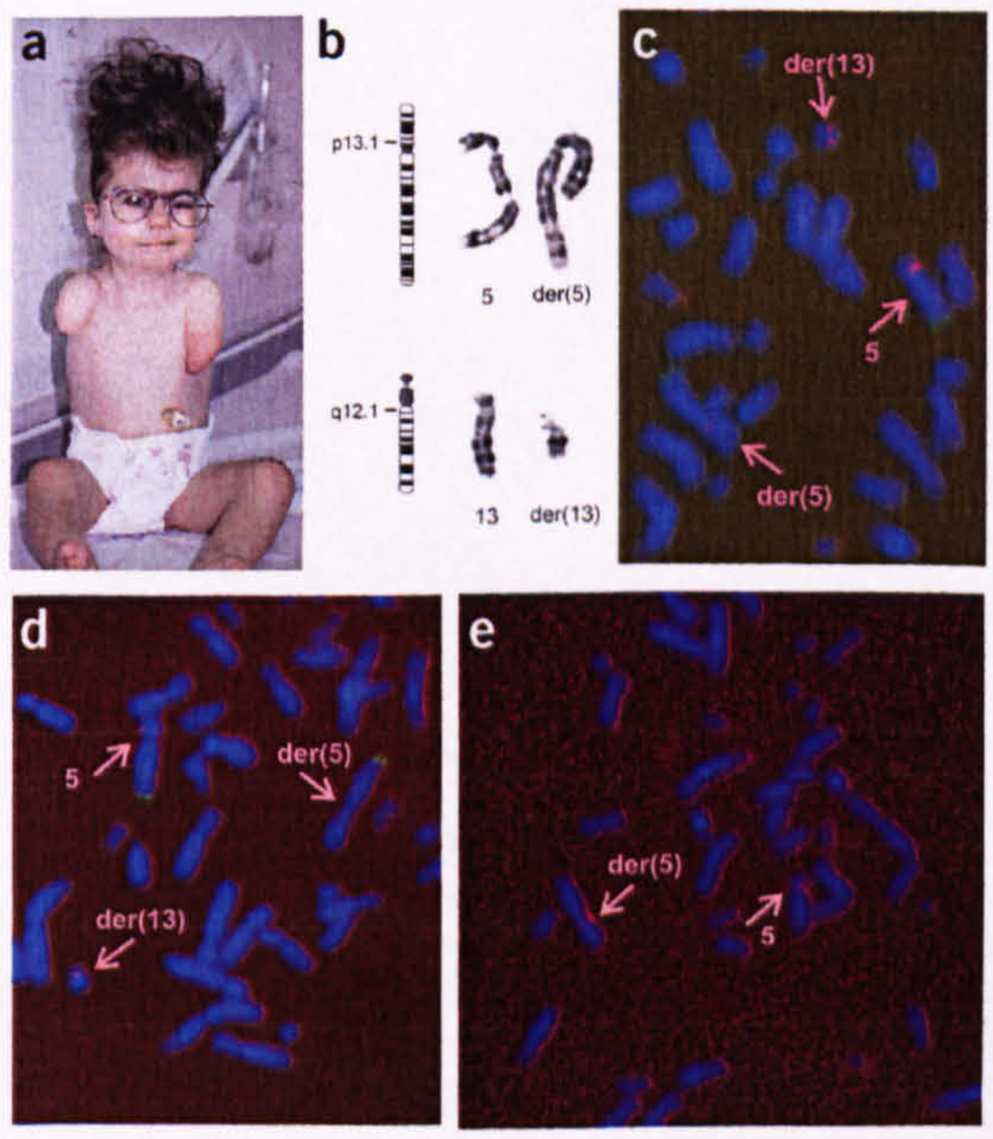



**Figure 1** FISH mapping of a 5p13 translocation breakpoint in an individual with classical CdLS. (**a**) Individual with classical CdLS with characteristic limb and facial abnormalities (including an upturned triangular nose, long philtrum, thin upper lip, downturned corners of the mouth; see fuller description for individual P46 in **Table 1**). (**b**) Giemsa chromosome banding showing a balanced *de novo* t(5;13)(p13.1;q12.1) translocation. (**c–e**) Metaphase chromosome FISH with the breakpoint-spanning BAC clone CTD-2653m23 (**c**) and overlapping fosmid clones G248P84262B4 (**d**) and G248P8840C10 (**e**), all labeled with Spectrum Red. Labeled in green is a chromosome 5q telomere-specific probe. Arrows indicate the normal chromosome 5 and the der(5) t(5;13)(p13.1;q12.1) and der(13) t(5;13)(p13.1;q12.1) chromosomes. In occasional metaphases a weak G248P84262B4 signal can be detected on the der(5) chromosome as well as a strong signal on the der(13). The combined data suggest that the most likely location for the breakpoint is close to the proximal end of the region of overlap for inserts of G248P84262B4 and G248P8840C10 (**Fig. 2a**).

[1]Institute of Human Genetics, University of Newcastle, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK. [2]Department of Pediatrics and Department of Human Genetics University of Utah Salt Lake City, Utah 84112, USA. [3]These authors contributed equally to this work. Correspondence should be addressed to T.S. (tom.strachan@ncl.ac.uk).

Published online 16 May 2004; doi:10.1038/ng1363

**Table 1 Phenotypes of selected individuals with CdLS with identified mutations**

| Individual | Category[a] | Nature of mutation | Location of mutation[b] | Inherited or *de novo* | Amino acid change | Phenotype |
|---|---|---|---|---|---|---|
| P2 | Classical | 1-bp insertion | 7306_7307insG, exon 43 | *De novo* | S2462X | Growth retardation, microbrachycephaly, flat facial profile, long philtrum, thin lips, crescent-shaped mouth, low-set ears, synophrys, bushy eyebrows, general hirsutism, hearing impairment, myopia, micromelia, clindactyly, proximally placed thumbs, fixed flexion of the elbows, syndactyly of the feet, recurrent pulmonary infection, hypoplastic umbilicus, bilateral inguinal hernias, undescended testes |
| P3 | Classical | Missense | 7289A→G, exon 43 | *De novo* | Y2430C[c] | Severe growth retardation, lobster limb defect, characteristic face, feeding difficulty, gastroesophageal reflux |
| P11 | Classical | Splice acceptor AG dinucleotide | 5575–2A→G, intron 29 | *De novo* | Undetermined | Growth retardation, characteristic facial features including long philtrum, anteverted nostrils, thin lips, crescent-shaped mouth, left postaxial polydactyly, bilateral hearing impairment, heart murmur |
| P13 | Mild | 3-bp deletion | 3616–3618 delATA, exon 14 | Change absent in maternal DNA; no paternal DNA available | I1206del[c] | Growth retardation, small hands, microcephaly, speech delay, inguinal hernia |
| P16 | Classical | Missense | 4043T→G, exon 17 | *De novo* | L1348R[c] | Growth retardation, synophrys, long philtrum, thin upper lip, speech delay, high arched palate, congenital heart defect, short digits, feeding difficulty with gastroesophageal reflux |
| P18 | Mild | Missense | 3931T→C, exon 17 | *De novo* | C1311R[c] | Small for gestational age, growth retardation, microcephaly, mild synophrys, neat arched eyebrows, depressed nasal bridge, smooth philtrum, thin lips, micromelia with short first and fifth digits of hands |
| P27 | Classical | Nonsense | 7903G→T, exon46 | No parental DNA samples available | E2635X | Severe growth retardation, microcephaly, flat facial profile, synophrys, bushy eyelashes, short nose, anteverted nostrils, thin lips, downturned corners of mouth, micrognathia, small hands, hypoplastic and overlapping toes, bilateral syndactyly of second and third toes, delayed speech |
| P29 | Classical | 1-bp insertion | 347_348insA, exon 4 | No parental DNA samples available | Y116X | Severe growth retardation, microcephaly, micromelia of the hands, feeding difficulty, atrioseptal defect type II, cryptochidism |
| P37 | Classical | Splice acceptor AG dinucleotide | 7686–2A→G, intron 44 | No parental DNA samples available | Undetermined | Growth retardation, characteristic facial features including long philtrum, anteverted nostrils, thin lips, crescent-shaped mouth, bilateral upper limb reduction defect |
| P46 | Classical | t(5;13) | intron 1 (see **Fig. 2a**) | *De novo* | N/A | Severe growth retardation, characteristic facial abnormalities and bilateral transverse upper limb reduction defects (see **Fig. 1a**), ventricular septal defect, cleft palate, sacral myelomeningocele |

[a]Classical and mild CdLS are primarily distinguished according to the severity of growth retardation and the extent of developmental delay[29,30]. [b]Nucleotide numbering refers to the cDNA sequence specifying the 2,804–amino acid isoform and commencing at the +1 position of the initiation codon. [c]Known amino acid conservation data are as follows: Ile1206 is conserved in mouse, rat and zebrafish, changed to methionine in fly and valine in *C. elegans*; Cys1311 is conserved in mouse, rat, zebrafish, fly and *C. elegans*; Leu1348 is conserved in mouse, rat and zebrafish, changed to isoleucine in fly and *C. elegans*; Tyr2430 is conserved in mouse, rat, zebrafish and fly, changed to phenylalanine in *C. elegans*.

perceived phenotypic overlap between duplication 3q syndrome and mild CdLS[6,7]. The 3q breakpoint disrupts a large gene undergoing unusual alternative splicing, but we found no additional mutations specific to any individuals with CdLS[3]. Molecular analyses of regions spanning the 17q23, 14q32 and 21q11 breakpoint regions also did not identify a gene likely to underlie CdLS (data not shown).

We localized the breakpoints in a third translocation case to 5p13.1 and 13q12.1 (**Fig. 1a,b** and **Table 1**). Fluorescent *in situ* hybridization (FISH) mapping identified BACs crossing the 5p breakpoint (**Fig. 1c**) and the 13q breakpoint. CdLS was recently reported to be associated with a 5p13.1-5p14.2 deletion (D. Viskochil, personal communication), and so we focused on the 5p breakpoint. We continued FISH mapping with fosmids until we identified two clones with overlapping inserts that mapped to either side of the 5p13 breakpoint. G248P84262B4 gave a clear hybridization signal on the normal chromosome 5 and the der(13), indicating preferential binding to the region telomeric to the 5p breakpoint (**Fig. 1d**). In contrast, G248P88840C10 hybridized clearly to the der(5) chromosome but was not visible on the der(13) chromosome (**Fig. 1e**).

The 5p breakpoint mapped close to a previously predicted gene-like sequence of unknown function, called *IDN3*. Using *in silico* analyses and in-house cDNA sequencing, we determined that *IDN3* was a gene fragment and that it comprises 90 kb of a new 190-kb gene, which we named *NIPBL* (*Nipped-B*-like; **Fig. 2a,b**). *NIPBL* contains 47 exons and is predicted to generate isoforms of 2,804 or 2,697 amino acids (**Fig. 2c**). Northern-blot analysis confirmed the predicted 9.8-kb transcript size and showed that *NIPBL* was strongly expressed in fetal and adult kidney, fetal liver, adult placenta, heart, skeletal muscle and thymus, but weakly or almost undetectably expressed in fetal and adult brain and lung and in adult liver, colon, small intestine and leukocytes (**Supplementary Fig. 1** online).

We screened other individuals with CdLS for mutations in *NIPBL* and identified nine plausible point mutations, at least five of which arose *de novo* (**Table 1** and **Supplementary Fig. 2** online). As we found *NIPBL* mutations in individuals with severe and mild CdLS, phenotype variation can be explained, at least in part, by allelic heterogeneity. The spectrum and distribution of mutations imply that pathogenesis arises from loss or altered function of a single *NIPBL* allele. Our mutation detection rate was ~50%. Locus heterogeneity could be a factor, but limitations of the screening methods is a plausible explanation for the comparatively low mutation detection rate. Considerable intrafamilial variation in phenotype, even between siblings with CdLS[8], has been reported, suggesting that additional factors may be important.

Using BLAST and BLAT analyses we determined the full-length sequence of the mouse, rat and zebrafish *NIPBL* homologs (data not shown). The exon structure is very well conserved in vertebrates and specifies a protein of ~2,800 amino acids. Sequence identities between human delangin and vertebrate orthologs are 96% (for mouse and rat) and 63% (for zebrafish; data not shown). TBLASTN searches against expressed-sequence tag databases showed that the two C-terminal isoforms are conserved in cow, pig, mouse, rat and chick (data not shown).

We also identified delangin homologs in flies (*Drosophila melanogaster* Nipped-B, *Anopheles gambiae* XM_320088), worms (*Caenorhabditis elegans* PQN-85, *C. briggsae* CBG0727), plants (*Arabidopsis thaliana* NM_121558, *Oryza sativa* NM_186173) and fungi (Scc2 family of sister chromatid cohesion proteins). In each case, homology is largely confined to a segment of ~1,500 amino acids spanning most of the delangin C-terminal half (**Fig. 2c**). Discounting small polyglutamine- and lysine-rich segments, most of the homologs do not share significant homology with any other protein sequence predicted from the relevant genome sequence. Because they are expected to be essential and most of their sequence shows homology to delangins, they may be viewed as orthologs. The pufferfish may be an exception: BLAT analyses suggest that there are two related *NIPBL*-like gene sequences.

Many of the fungal homologs have crucial chromosomal roles: *Saccharomyces cerevisiae* Scc2 and *Schizosaccharomyces pombe* Mis4 in sister chromatid cohesion[9,10] and *Coprinus cinereus* Rad9 in meiotic chromosome pairing and DNA repair[11]. Some metazoan orthologs, however, are known or likely to be developmental regulators. By facilitating activation of remote enhancers, the *D. melanogaster* Nipped-B protein regulates the Ultrabithorax (*Ubx*) and *Cut* homeobox genes[12,13]. RNA-interference knock-down of the gene encoding *C. elegans* PQN-85 results in a high level of embryonic lethality; survivors have a paralyzed uncoordinated phenotype, body morphology defects and sometimes a vulval defect (J. Ahringer, personal communication).

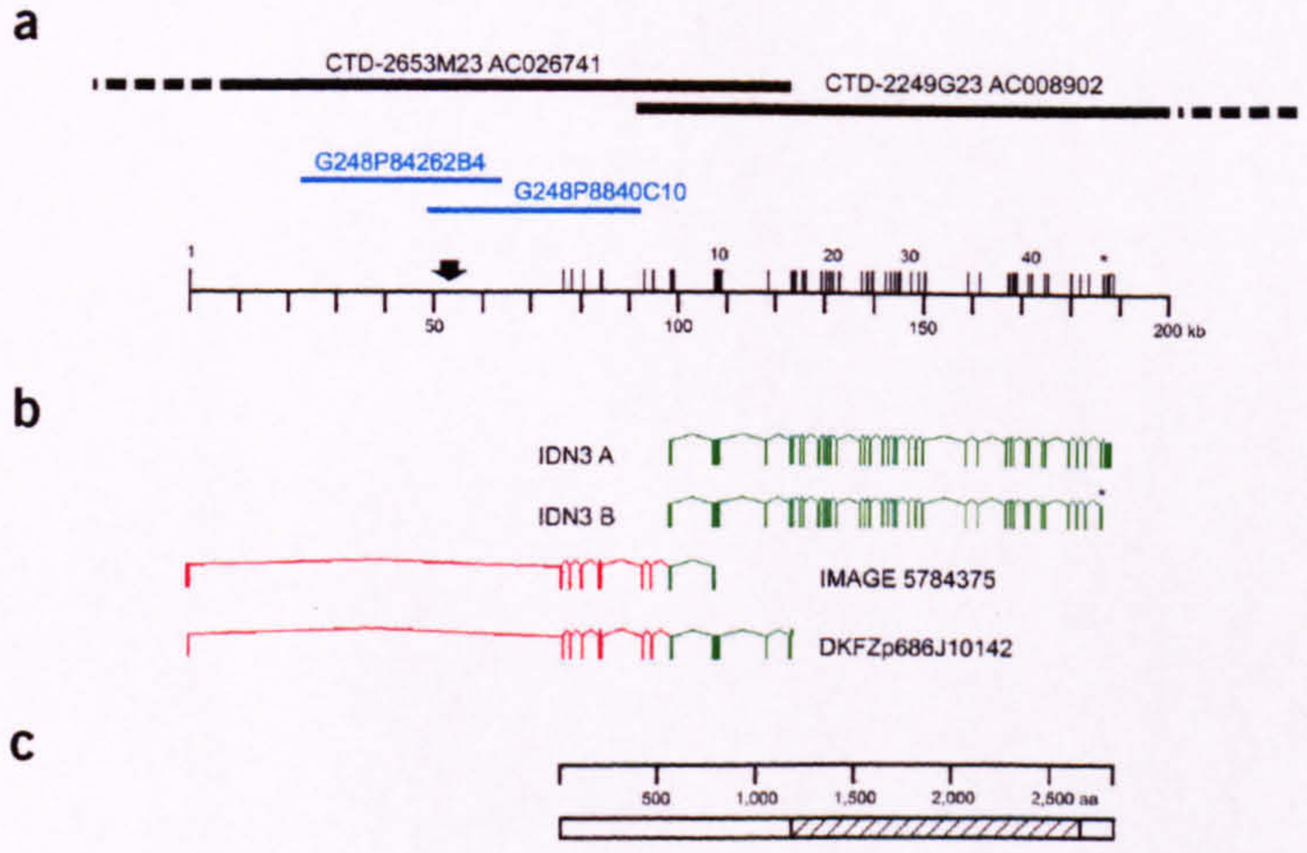



**Figure 2** Normal exon-intron organization and expressed products of the gene *NIPBL* severed by the 5p13 breakpoint. (**a**) Genomic organization and FISH clones. A scale diagram of the exon-intron organization of *NIPBL* is shown at the bottom. Thick black horizontal bars at top show relevant BAC clones spanning and extending beyond the gene (solid and dotted black lines, respectively). Two fosmid clones that immediately flank the breakpoint (marked by the vertical arrow; see also **Fig. 1d,e**) are shown in blue. (**b**) Cognate cDNAs. Linked vertical bars illustrate exon selection in representative cDNAs used to determine the full exon complement, including the original IDN3 A and IDN3 B clones and overlapping clones that defined new 5′ exons (in red). The 5′ end of the IMAGE 5784375 clone sequence defines the start of exon 1. The asterisk denotes alternative 3′ end truncation distinguishing the short isoform. (**c**) Protein product. The coding sequence commences in exon 2 and continues either to exon 47, generating a long isoform, or to an expanded variant of exon 46, generating a slightly shorter isoform. The open bar represents the long (2,804 amino acids) isoform, with the internal striped box representing a region that is highly conserved during evolution and significantly related to fungal Scc2-type proteins. The short isoform contains 2,697 amino acids, of which residues 1–2,683 are identical to those of the long isoform, leaving a 14–amino acid C-terminal end that is unrelated to the 121–amino acid C-terminal end of the long isoform.

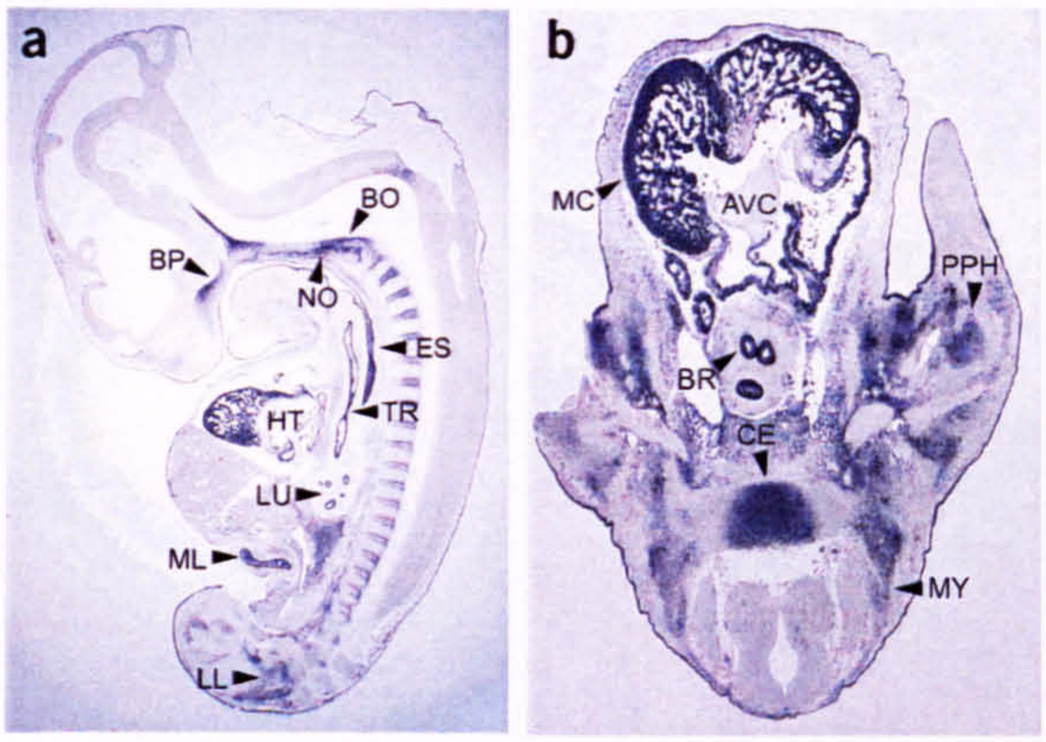



**Figure 3** Specific embryonic expression of *NIPBL* transcripts shown by *in situ* hybridization. Bright-field microscopy images after hybridization to embryonic sections of a 424-bp antisense riboprobe corresponding to sequences in exons 10–12 of *NIPBL*. Images represent a midline section through a sagittally sectioned Carnegie stage 18 embryo (**a**) and a transverse section of a Carnegie stage 17 embryo (**b**). BP, cartilage primordium of the basisphenoid bone; BO, cartilage primordium of the basioccipital bone; NO, notochord; ES, esophagus; HT, heart; TR, trachea; LU, lung; ML, midgut loop; LL, lower limb; MC, heart myocardium; AVC, A-V cushion; BR, bronchus; PPH, precartilage primordium of the humerus; CE, centrum (vertebral body); MY, migrating myoblasts.

Because embryonic expression can differ substantially between some human-mouse orthologs[14], we carried out *NIPBL in situ* hybridization analyses on human embryonic tissue sections. The observed expression pattern was largely consistent with the CdLS phenotype (**Figs. 3** and **4**). *NIPBL* was expressed in developing limbs (**Fig. 3a,b**) and later in cartilage primordia of the ulna and of various hand bones (**Fig. 4c**). Sites of craniofacial expression included the cartilage primordium of the basioccipital and basisphenoid skull bones (**Figs. 3a** and **4f**) and elsewhere in the head and face, including a region encompassing the mesenchyme adjacent to the cochlear canal (**Fig. 4e,f**).

*NIPBL* was also expressed in the spinal column, notochord and surface ectoderm (**Figs. 3a,b** and **4b**), sclerotome and what seem to be migrating myoblasts (**Fig. 3b**). Expression in the developing heart was pronounced in the atrial and ventricular myocardium and in the ventricular tubeculae but absent in the endocardial cushions (**Fig. 3a,b**). *NIPBL* was also expressed in the developing esophagus, trachea and midgut loops (**Fig. 3a**), in the bronchi of the lung (**Figs. 3b** and **4d**) and in the tubules of the metanephros (**Fig. 4a**).

Expression in organs and tissues not typically affected in CdLS (*e.g.*, the developing trachea, bronchi, esophagus, heart and kidney) may reflect a bias towards underreporting of more subtle aspects of the phenotype or problems that typically present later in life. Respiratory and feeding difficulties and gastroesophageal reflux are known CdLS complications[1,15], individuals with CdLS have a greater incidence of congenital heart abnormalities[1,16] and renal abnormalities can be found in >50% of classical cases (A. Selicorni, personal communication). Expression in the mesenchyme surrounding the cochlear canal (**Fig. 4e**) may be related to the hearing impairment commonly found in CdLS[17]. Expression of *NIPBL* in embryonic brain was not evident, but the main neurodevelopmental deficits in CdLS are thought to occur during late gestation[18,19].

The involvement of Nipped-B in activating the *Ubx* and *Cut* homeobox genes[12] may provide insights into the molecular basis of CdLS pathogenesis. *Ubx* suppresses limb formation in the fly abdomen by repressing Distalless (*Dll*), a gene required for distal limb development[20]. The Dlx family of mammalian *Dll* homologs are involved in multiple developmental processes, including limb and branchial arch patterning, neurogenesis and hematopoiesis[21]. They are expressed in the apical ectodermal ridge of the developing limb bud, which partly coordinates limb outgrowth, and also in facial primordia. Therefore, the proximodistal limb patterning defect that underlies limb reduction in CdLS and possibly the associated facial abnormalities could largely be explained by inappropriate activation of *DLX* genes. Mutations in fly *Cut* cause leg and wing abnormalities. The mouse homolog *Cutl2* (*Cux2*) is dynamically expressed in branchial arch and limb bud progress zones[22], and so reduced expression of a human homolog in CdLS could also contribute to facial and limb abnormalities. The other mouse homolog, *Cutl1*, is widely expressed but important in lung development[23].

A dual role for Nipped-B in sister chromatid cohesion and developmental regulation was recently confirmed[24]. Similar dual roles

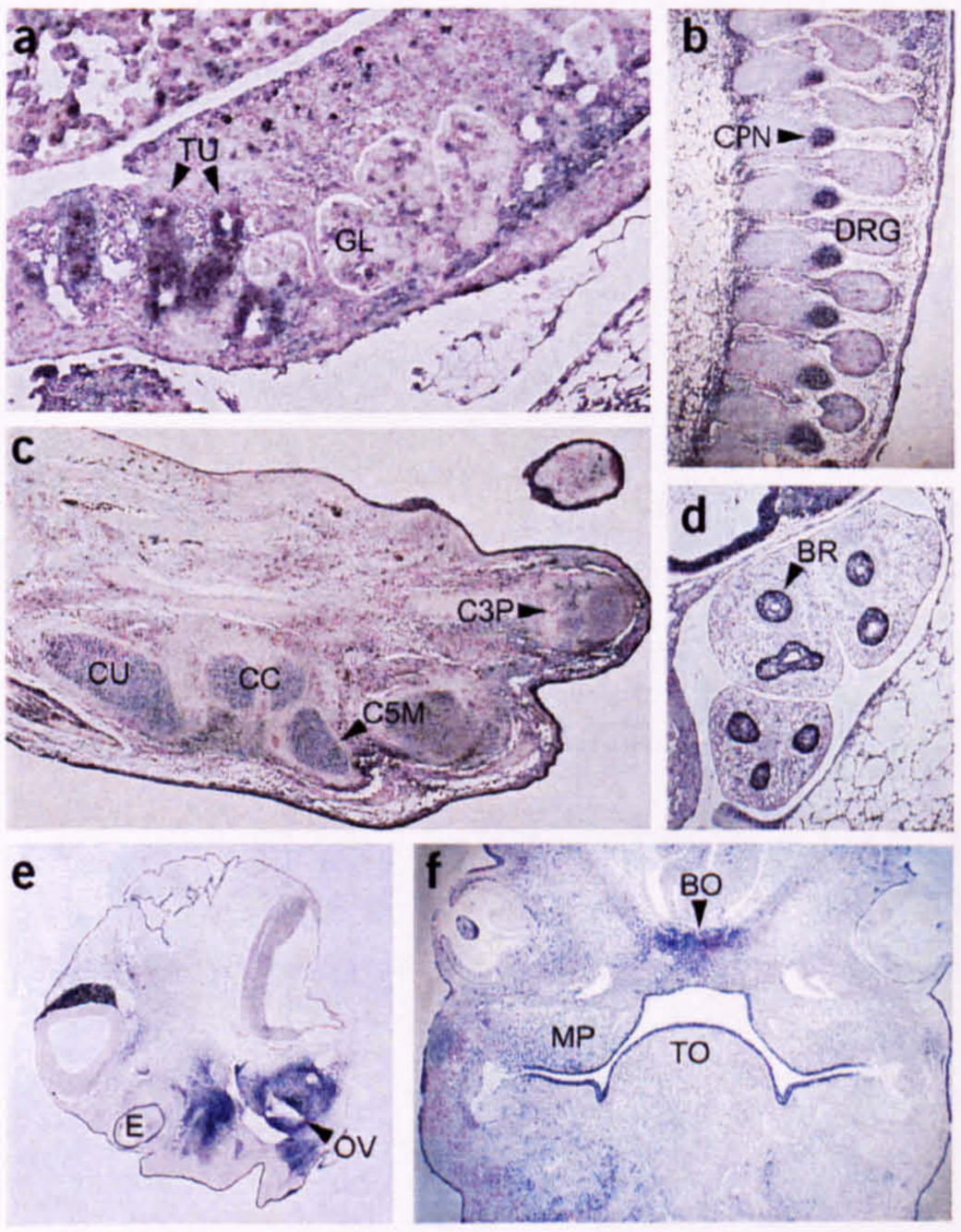



**Figure 4** Detail of *NIPBL* expression in human embryonic development. Expression of *NIPBL* transcripts in specific structures of the metanephros (**a**), spinal column (**b**) and lung (**d**) of a single Carnegie stage 18 embryo, and in the hand plate of a Carnegie stage 21 embryo (**c**). Lower panels illustrate head and craniofacial structures in sagittal (**e**) and transverse (**f**) sections of a Carnegie stage 18 embryo. TU, metanephric tubule; GL, metanephric glomerulus; CPN, cartilage primordium of the neural arch; DRG, dorsal root ganglia; CU, cartilage primordium of the ulna; CC, cartilage primordium of the carpal bones; CM5, cartilage primordium of the metacarpal of the fifth digit; CP3, cartilage primordium of the phalangeal of the third digit; BR, bronchi of the lung; E, eye; OV, otic vesicle; BO, cartilage primordium of the basioccipital bone; MP, maxillary process; TO, tongue.

can be expected for vertebrate delangins, suggesting a possible parallel between CdLS and Roberts syndrome (OMIM 268300), which is characterized by growth retardation, limb reduction defects, craniofacial abnormalities and premature centromere separation. We assayed C-banded samples from individuals with CdLS for premature centromere separation but, perhaps unsurprisingly, detected no abnormalities; targeted knock-down of both alleles might be more informative. If delangin does have a dual functional role, the housekeeping role in facilitating *trans* interactions between sequences on sister chromatids could be satisfied with a basal level of expression. An additional role in enabling long-distance *cis* interactions (between promoter plus remote enhancer) for select target genes could require strong expression in tissues and organs where the target genes are active. The Scc2–Nipped-B–delangin family provides a model system for investigating evolutionary diversification of protein function.

## METHODS

**Chromosome analyses.** We used thymidine to synchronize phytohemagglutinin-stimulated blood cultures and carried out G-banding according to standard protocols[25]. For premature centromere separation assays, we carried out standard C-banding[25] on fresh slides of samples from 12 individuals with CdLS. We scored 25 cells from each individual sample for premature centromere separation. We carried out chromosome FISH analysis by nick translation labeling of assorted genomic clones with SpectrumRed (Vysis) according to the manufacturer's instructions. Genomic clones included YAC, BAC and fosmid clones. We hybridized labeled probes along with a prelabeled chromosome 5p telomere-specific probe (Qbiogene) to metaphase chromosomes using standard methodology.

**DNA sequencing and mutation screening.** We obtained Image cDNA clones from the MRC Geneservice (see URL below). We sequenced all inserts using a combination of vector-specific and insert-specific primers and the MegaBACE ET system (Amersham). We screened mutations by direct sequencing and, when exons were suitably small, by SSCP-heteroduplex analysis using standard protocols. In the latter case, we denatured PCR products, size-fractionated them in 1× MDE gels (BioWhittaker) containing 5% glycerol and 0.6× TBE buffer at 300 V for ~20 h (depending on fragment size) and visualized them by silver staining. Any samples that had band differences relative to an unaffected control were sequenced using the MegaBACE ET system (Amersham). If the primary chromatogram suggested the presence of a deletion or insertion, we cloned the PCR product and sequenced a number of transformants to confirm the change. We reamplified and resequenced all mutations to confirm that the change observed was not the result of base misincorporation by the DNA polymerase. We used a panel of genomic DNA samples from 45 individuals of European descent (mostly from the UK, some from Poland and Ireland) with CdLS. For each mutation, we also screened 200 normal chromosomes. We screened parental DNA (when available) to confirm that the observed mutation had occurred *de novo*.

Because of the very long coding sequence (8,412 nucleotides), our mutation screening protocol surveyed a subset (26) of the 46 coding exons, namely exons 2–8, 13–20, 23, 24, 30, 34–36, 38, 40, 43, 45 and 46. The coding sequence sampled in these 26 exons corresponds to ~31% of the total. This means that approximately one-third of the coding sequence was sampled in 45 individuals and more than one-half of the proximal intronic sequence was also sampled for splicing mutations. On the basis of observed relative frequencies of splice site mutations and other mutations in large multiexon genes, and assuming that all affected individuals are heterozygotes (as expected from the strong evidence for autosomal dominant transmission[26]) and that there are no strong mutational hot spots in the coding sequence, the identification of 9 mutations in a panel of 45 individuals with CdLS (**Table 1**) equates roughly to a detection rate >50%. The relatively low mutation detection rate could reflect limitations of the mutation screening protocol: only coding sequences and proximal intronic sequences were analyzed, and the gene is large and possibly prone to undetected large-scale mutations.

**Gene expression analyses.** We designed PCR primers to amplify a 424-bp cDNA probe spanning exons 10–12 of *NIPBL*, which should hybridize to transcripts encoding both the long and short isoforms (primer sequences are available on our website; see URL below). For northern-blot analyses, we labeled the probe to high specific activity with [$\alpha^{32}$P]-dCTP by random priming. After removing unincorporated nucleotides (NICK column, Amersham), we hybridized the probe against blots of human adult and fetal RNA (Clontech) containing ~2 μg of mRNA per lane at 42 °C overnight in ULTRAhyb (Ambion). We washed the blots in 0.1× SSC in 0.1SDS at 65 °C before exposing them to film. After removing the test probe, we rehybridized the blots with random-primed labeled control cDNA for human β-actin.

For tissue *in situ* hybridization, we cloned the 424-bp cDNA fragment into the pGEM-T Easy vector (Promega) and transcribed it with T7 and SP6 RNA polymerases incorporating DIG-11-UTP to generate labeled sense and antisense riboprobes, respectively. We generated additional isoform-specific probes to correspond to the long isoform of 2,804 amino acids and the short isoform of 2,697 amino acids (primer sequences are available on our website; see URL below). We hybridized the probes to sections of human embryonic tissue as described[27]. The isoform-specific probes generated similar expression patterns as the non-isoform-specific probe. We collected and used human embryonic tissue samples with ethical permission from the joint Ethics Committee of the Newcastle Health Authority and with appropriate signed consents. Samples were staged by microscopic examination. We fixed and processed tissue samples as previously described[28]. We selected the material we studied to have normal karyotypes and to be unrelated to disease.

**URLs.** Sequences of primers used for expression are available at our Newcastle CdLS research website at http://www.ncl.ac.uk/ihg/cdls. Servers used for nucleotide sequence analysis were the US National Center for Biotechnology Information's BLAST server (http://www.ncbi.nih.gov/BLAST/), the University of California at Santa Cruz's BLAT genome search server (http://genome.ucsc.edu/cgi-bin/hgBlat), the Ensembl genome browser (http://www.ensembl.org/), the University of California at Santa Cruz genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway), the NIX suite of nucleotide sequence analysis programs (http://www.hgmp.mrc.ac.uk/Registered/Webapp/nix/) and Baylor College of Medicine's sequence utilities programs (http://searchlauncher.bcm.tmc.edu/seq-util/seq-util.html). Servers for protein sequence analysis included PSORTII (http://psort.nibb.ac.jp) and the DomPred program for predicting protein domains (http://bioinf.cs.ucl.ac.uk/dompred/). The Rep program (http://www.embl-heidelberg.de/~andrade/papers/rep/search.html) allowed us to identify five HEAT repeats in the conserved C-terminal domain. Alignment of multiple orthologous sequences was aided by using the ClustalW program at http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html. IMAGE cDNA clones were obtained from the MRC Geneservice, available at http://www.hgmp.mrc.ac.uk/geneservice/index.shtml.

**GenBank accession numbers.** Human *NIPBL* mRNA encoding the long delangin isoform, AJ627032; homologous mouse mRNA encoding the long delangin isoform, AJ627033; human *NIPBL* mRNA encoding the short delangin isoform, AJ640137; homologous mouse mRNA encoding the short delangin isoform, AJ640138; Image clone 5784375, AJ627564.

*Note: Supplementary information is available on the Nature Genetics website.*

ACKNOWLEDGMENTS

This paper is dedicated to the memory of F. Strachan (1921–2004). We thank many of our current Newcastle colleagues, especially S. Zwolinsky and J. Wolstenholme, for discussions, carrying out chromosome-banding analyses and conducting the premature centromere separation assay; H. Peters and D. Henderson for contributions to analysis of our expression data; S. Humphray for supplying fosmid clones; L. Jackson for facilitating the collaboration between M.B. and the Newcastle group; I. Krantz for discussions; A. Peaford and colleagues for their support; M. Walasek, M. Ireland and many other clinical geneticists for providing blood samples from individuals with CdLS and access to phenotype data; many individuals with CdLS and their families for their generosity; and previous members of the Newcastle CdLS research team, notably M. Smith, P. J. A. Eichhorn and B. Imamwerdi, for their earlier contributions. We thank the UK Community Fund and previously Action Research for providing funding for this project and the MRC-Wellcome Human Developmental Biology Resource for supplying human embryonic tissue samples.

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

Received 19 February; accepted 28 April 2004
Published online at http://www.nature.com/naturegenetics/

1. Jackson, L., Kline, A.D., Barr, M.A. & Koch, S. de Lange syndrome: a clinical review of 310 individuals. *Am. J. Med. Genet.* **47**, 940–946 (1993).
2. Ireland, M., Donnai, D. & Burn, J. Brachmann-de Lange syndrome Delineation of the clinical phenotype. *Am. J. Med. Genet.* **47**, 959–964 (1993).
3. Tonkin, E.T. *et al.* A giant novel gene undergoing extensive alternative processing is severed by a Cornelia de Lange-associated translocation breakpoint at 3q26.3. *Hum. Genet.* (in the press).
4. Ireland, M., English, C., Cross, I., Houlsby, W.T. & Burn, J. A de novo translocation t(3;17)(q26.3;q23.1) in a child with Cornelia de Lange syndrome. *J. Med. Genet.* **28**, 639–640 (1991).
5. Wilson, W.G., Kennaugh, J.M., Kugler, J.P. & Wyandt, H.E. Reciprocal translocation 14q;21q in a patient with the Brachmann-de Lange syndrome. *J. Med. Genet.* **20**, 469–471 (1983).
6. Steinbach, P. *et al.* The dup(3q) syndrome: report of eight cases and review of the literature. *Am. J. Med. Genet.* **10**, 159–177 (1981).
7. Wilson, G.N., Dasouki, M. & Barr, M. Jr. Further delineation of the dup(3q) syndrome. *Am. J. Med. Genet.* **22**, 117–123 (1985).
8. Krajewska-Walasek, M., Chrzanowska, K., Tylki-Szymanska, A. & Bialecka, M. A further report of Brachmann-de Lange syndrome in two sibs with normal parents. *Clin. Genet.* **47**, 324–327 (1995).
9. Ciosk, R. *et al.* Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol. Cell* **5**, 243–254 (2000).
10. Furuya, K., Takahashi, K. & Yanagida, M. Faithful anaphase is ensured by Mis4, a sister chromatid cohesion molecule required in S phase and not destroyed in G1 phase. *Genes Dev.* **12**, 3408–3418 (1998).
11. Seitz, L.C., Tang, K., Cummings, W.J. & Zolan, M.E. The rad9 gene of Coprinus cinereus encodes a proline-rich protein required for meiotic chromosome condensation and synapsis. *Genetics* **142**, 1105–1117 (1996).
12. Rollins, R.A., Morcillo, P. & Dorsett, D. Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* **152**, 577–593 (1999).
13. Dorsett, D. Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Curr. Opin. Genet. Dev.* **9**, 505–514 (1999).
14. Fougerousse, F. *et al.* Human-mouse differences in the embryonic expression patterns of developmental control genes and disease genes. *Hum. Mol. Genet.* **9**, 165–173 (2000).
15. Luzzani, S., Macchini, F., Valade, A., Milani, D. & Selicorni, A. Gastroesophageal reflux and Cornelia de Lange syndrome: typical and atypical syndromes. *Am. J. Med. Genet.* **119A**, 283–287 (2003).
16. Mehta, A.V. & Ambalavanan, S.K. Occurrence of congenital heart disease in children with Brachmann-de Lange syndrome. *Am. J. Med. Genet.* **71**, 434–435 (1997).
17. Sataloff, R.T., Spiegel, J.R., Hawkshaw, M., Epstein, J.M. & Jackson, L. Cornelia de Lange syndrome. Otolaryngologic manifestations. *Arch. Otolaryngol. Head Neck Surg.* **116**, 1044–1046 (1990).
18. Yamaguchi, K. & Ishitobi, F. Brain dysgenesis in Cornelia de Lange syndrome. *Clin. Neuropathol.* **18**, 99–105 (1999).
19. Vuilleumier, N. *et al.* Neuropathological analysis of an adult case of the Cornelia de Lange syndrome. *Acta Neuropathol. (Berl.)* **104**, 327–332 (2002).
20. Gebelein, B., Culi, J., Ryoo, H.D., Zhang, W. & Mann, R.S. Specificity of Distalless repression and limb primordia development by abdominal Hox proteins. *Dev. Cell* **3**, 487–498 (2002).
21. Panganiban, G. & Rubinstein, J.L.R. Developmental functions of the *Distal-less*/Dlx homeobox genes. *Development* **129**, 4371–4386 (2002).
22. Iulianella, A., Vanden Heuvel, G. & Trainor, P. Dynamic expression of murine Cux2 in craniofacial, limb, urogenital and neuronal primordia. *Gene Expr. Patterns* **3**, 571–577 (2003).
23. Ellis, T. *et al.* The transcriptional repressor CDP (Cutl1) is essential for epithelial cell differentiation of the lung and the hair follicle. *Genes Dev.* **15**, 2307–2319 (2001).
24. Rollins, R.A., Korom, M., Aulner, N., Martens, A. & Dorsett, D. Drosophila Nipped-B protein supports sister chromatid cohesion and opposes the Stromalin/Scc3 Cohesion factor to facilitate long-range activation of the cut gene. *Mol. Cell. Biol.* **24**, 3100–3111 (2004).
25. Rooney D.E. (ed.) *Human Cytogenetics: Constitutional Analysis. A Practical Approach.* 3rd. edn. (Oxford University Press, Oxford, 2001).
26. Russell, K.L. *et al.* Dominant paternal transmission of Cornelia de Lange syndrome: a new case and review of 25 previously reported familial recurrences. *Am. J. Med. Genet.* **104**, 267–276 (2001).
27. Breitschopf, H., Suchanek, G., Gould, R.M., Colman, D.R. & Lassmann, H. *In situ* hybridization with digoxigenin-labeled probes: sensitive and reliable detection method applied to myelinating rat brain. *Acta Neuropathol.*, **84**, 581–587 (1992).
28. Lako, M. *et al.* A novel mammalian Wnt gene, *WNT8B*, shows brain-restricted expression in early development, with sharply delimited expression boundaries in the developing forebrain. *Hum. Mol. Genet.* **7**, 813–822 (1998).
29. Van Allen, M.I. *et al.* Clinical variability within Brachmann-de Lange syndrome: a proposed classification system. *Am. J. Med. Genet.* **47**, 947–958 (1993).
30. Allanson, J.E., Hennekam, R.C. & Ireland, M. De Lange syndrome: subjective and objective comparison of the classical and mild phenotypes. *J. Med. Genet.* **34**, 645–650 (1997).