 Human Molecular Genetics, 1993, Vol. 2, No. 12 2099–2107

Isolation of a putative transcriptional regulator from the region of 22q11 deleted in DiGeorge syndrome, Shprintzen syndrome and familial congenital heart disease

Stephanie Halford+, Roy Wadey+, Catherine Roberts, Sara C.M.Daw, Jennifer A.Whiting, Hilary O'Donnell, Ian Dunham[1], David Bentley[1], Elizabeth Lindsay[2], Antonio Baldini[2], Fiona Francis[3], Hans Lehrach[3], Robert Williamson[4], David I.Wilson[5], Judith Goodship[5], Ian Cross[5], John Burn[5] and Peter J.Scambler*

Molecular Medicine Unit, Institute of Child Health, Guilford Street, London WC1N 1EH, [1]Division of Medical and Molecular Genetics, UMDS, Guy's Tower, London SE1 9RT, UK, [2]T936 Molecular Genetics Institute, Baylor College of Medicine, Houston TX 77030, USA, [3]Genome Analysis Laboratory, ICRF, Lincoln's Inn Fields, London WC2A 3PX, [4]Department of Biochemistry and Molecular Genetics, St Mary's Hospital Medical School, Imperial College, London W2 1PG and [5]Division of Human Genetics, 19, Claremont Place, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4AA, UK

Received September 6, 1993; Revised and Accepted October 21, 1993 EMBL accession numbers X75296 and X75295

A wide spectrum of birth defects are caused by deletions of the DiGeorge syndrome critical region (DGCR) at human chromosome 22q11. Over one hundred such deletions have now been examined and a minimally deleted region of 300kb defined. Within these sequences we have identified a gene expressed during human and murine embryogenesis. The gene, named TUPLE1, and its murine homologue, encodes a protein containing repeated motifs similar to the WD40 domains found in the beta-transducin/enhancer of split (TLE) family. The TUPLE1 product has several features typical of transcriptional control proteins and in particular has homology with the yeast Tup1 transcriptional regulator. We propose that haploinsufficiency for TUPLE1 is at least partly responsible for DiGeorge syndrome and related abnormalities.

INTRODUCTION

Hemizygosity for 22q11 causes a wide variety of congenital malformations that receive several diagnostic labels, the main features being covered by the acronym CATCH22(1) (Cardiac defect, Abnormal facies, Thymic hypoplasia, Cleft palate, Hypocalcaemia, 22q11 deletions). The incidence of these anomalies has been estimated as 1:5000 live births(2). DiGeorge syndrome(3) (DGS; MIM no. 188400(4)) was the first of these haploinsufficiency syndromes to be described, and has been reported in individuals with varied chromosomal defects, most frequently monosomy for the proximal long arm of 22 (22q11)(5). We have isolated DNA markers for this region of the genome using flow-sorted libraries(6,7) and microdissection and microcloning(8). These markers have been used to develop a map of the region commonly deleted in DGS (the DiGeorge critical region or DGCR)(9–11) and show that almost all DGS patients have deletions within 22q11(12,13). Similar findings have been reported by others(14).

DGS has an extensive phenotypic overlap with the Shprintzen or velo-cardio-facial syndrome (VCFS; MIM No.19243(4)) (15–17) and it is likely these disorders are aetiologically related. Recent work in our laboratory and elsewhere has shown that VCFS patients have a deletion of the same region of 22q11 as DGS patients(18–20).

Severe DGS may occur in families in which more mildly affected individuals have congenital heart disease or craniofacial malformation(21). In our study of nine families where two or more members had congenital heart defects hemizygosity within 22q11 was detected in five(22). Preliminary analyses indicate that some patients with sporadic heart defects also have hemizygosity within 22q11(23,24). Since congenital heart disease is common (7/1000 live births) and may not be recognised as having a chromosomal or familial basis(25) these deletions are likely to be important in the pathogenesis of heart malformation.

Deletions of 22q11 are therefore associated with a wide range of phenotypes. This is typical of chromosomal disorders with abnormal gene dosage eg. Down syndrome (trisomy 21), and Turner syndrome (XO). Several commentators have surmised that these seemingly disparate features could be explained by a deficient contribution of the neural crest during development (26,27).

In this work we have analysed the extent of deletion in patients with CATCH22 phenotypes and refined the molecular cytogenetic

* To whom correspondence should be addressed

+ These authors made an equal contribution to this work and are listed alphabetically.

map of the region. In no case was a deletion smaller than 300kb. The majority of the shortest region of overlap (SRO) was cloned from various genomic libraries, and a gene encoding a transcriptional regulator isolated. We propose that this gene (and possibly sequences related to it) is involved in the pathogenesis of CATCH22.

RESULTS AND DISCUSSION

The SRO of 22q11 deletions is at least 300kb in length

Most patients with DGS/VCFS pathology (approximately 85%) have a large deletion which is detected by the marker sc11.1 (eg. refs(10,28)). This probe detects two loci (sc11.1a proximally and sc11.1b distally) approximately 2Mb apart(11) (see Fig. 1). A small proportion of patients have smaller deletions which has allowed an SRO map to be established. One independent study estimates the critical region to be greater than 750kb(20). Our recent work has shown that the scF5 and sc11.1a sequences are both within the SRO(11). FISH interphase measurements indicate that these loci are approximately 300kb apart (Fig. 1): the scF5−sc11.1a interval measured 0.48 microns (standard error 0.04) and cosmids known to be 275kb apart were separated by 0.44 microns (standard error 0.09 microns).

In an attempt to refine the SRO in order that a single gene might be incriminated in CATCH22, the extent of deletion was examined in new and previously studied patients. Specifically, we wished to ascertain whether any patient was deleted for sc11.1a but not scF5, or deleted for scF5 but not sc11.1a. Patients found to be hemizygous with the cosmid sc11.1 were not examined further as these patients have a deletion of 2Mb or greater (see Fig. 1). If patients failed to show a deletion with this cosmid they were examined with one or more of the markers proximal to sc11.1b: sc4.1, D0832, scF5, (using FISH) and NB4 (using quantitative Southern analysis). The results are summarised in Table 1 and show that both scF5 and sc11.1a are hemizygous in all patients known to have a chromosome 22 deletion. This corroborates our previously published SRO(11) and indicates that it would be impossible to narrow the DGCR SRO any further using our current sample of 114 deleted patients (of any phenotype).

The distal boundary of the SRO is defined by two breakpoints found in cell lines GM00980 and GM05878(11) (see Fig. 1): all other deleted patients were hemizygous for D0832. In order to further define the minimum possible size of our SRO a cosmid (C1/79) was isolated which is 275kb distal to scF5 (see below) and used in FISH experiments with the GM00980 and GM05878 cell lines. GM00980 is cell line from a VCFS patient with a deletion 22pter−22q11. GM05878 is a cell line from the unaffected father of a DGS child. The father carries a 10;22 balanced translocation, with the region critical for DGS located on the derivative 22 chromosome. C1/79 is dizygous in GM00980, but on the derivative 22 chromosome in GM05878 (Fig. 2a,b). Therefore all bar one of the deletions examined in our series of CATCH22 patients have a deletion of at least 575 kb (275kb C1/79−scF5, plus 300kb scF5−sc11.1a).

While realising that DGS/VCFS might be caused by haploinsufficiency for more than one gene we reasoned that, since the scF5−sc11.1a interval is hemizygous in all patients with a known chromosome 22q11 deletion, it would contain a gene of major effect. The scF5−sc11.1a interval also contains a balanced 2;22 translocation breakpoint that occurs in a family with mild features of DGS (patient ADU, Fig. 1)(11,29), which is discussed further below. Cloning attempts were therefore concentrated in this region.

Sequences from the DGCR are under-represented in most genomic libraries

Attempts were made to derive a contiguous group (contig) of genomic clones across the scF5−NB4 interval. NB4 was used to screen the ICI(30), Washington University (WashU)(31), ICRF(32) and CEPH megaYAC(33) libraries which had been prepared as gridded arrays for hybridisation(34). The CEPH libraries(35) were also screened using PCR. In total one positive sc11.1a YAC was obtained and shown to map to 22q11 by FISH,

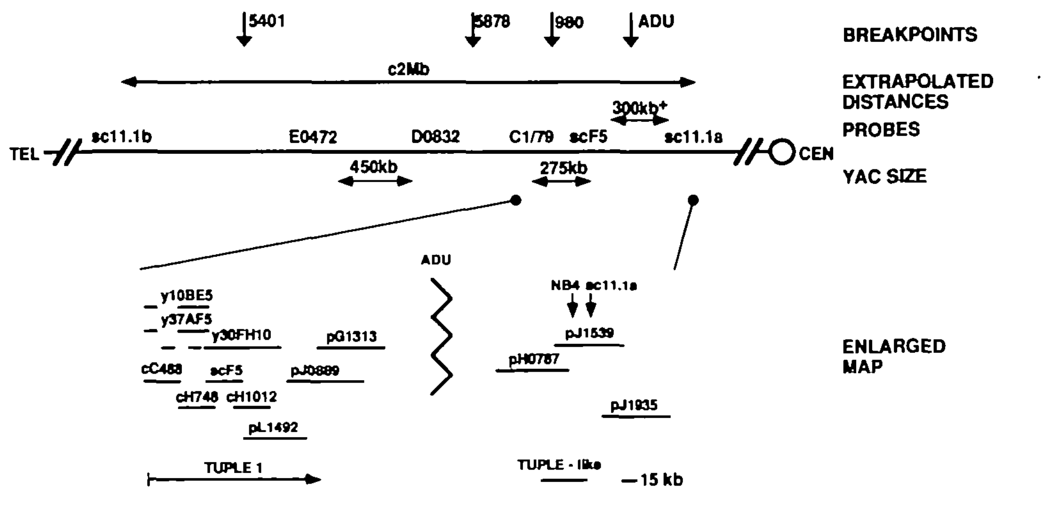


Figure 1. Map of the DGCR. The location of the probes used in this study is given above a line representing the chromosome. Important chromosome breakpoints are shown by downward pointing arrows. Distances between markers are determined from pulse-field gels or interphase FISH measurements (+). The bottom panel shows a map of the scF5−sc11.1 interval showing the YAC, cosmids and P1 contigs generated using gridded arrays. For clarity only representative clones with minimal overlap are shown. Where known, the direction of transcription of genes is given by the arrows. Cosmid coordinates are prefixed by c (all from the Livermore library apart from D0832, and E0472 (ICRF)), YACS by y (origin described in the text), and P1s by p. sc11.1a and NB4 are indicated by arrows.

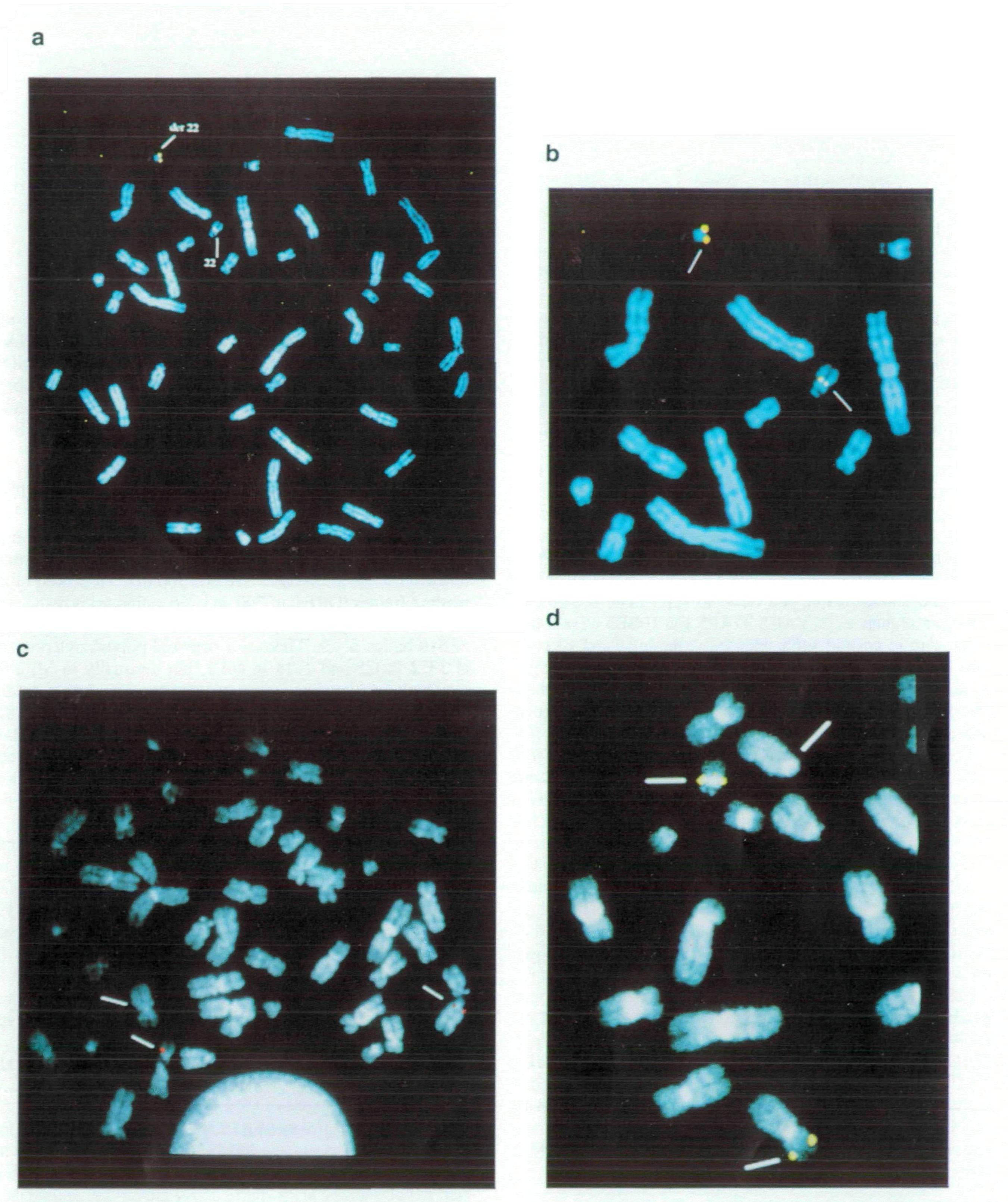


Figure 2. (a) C1/79 maps proximal to the translocation breakpoint in cell line GM05878. The bars indicate the signal which is detected on the normal chromosome 22 and the derivative 22 chromosome. A full metaphase spread is shown on the left (a), and a magnified section on the right (b). This localises C1/79 proximal to the breakpoint, and therefore C1/79 is hemizygous in 113/114 CATCH22 patients. (c) FISH mapping of the sc11.1a YAC in the balanced translocation cell line ADU. The YAC hybridises to the proximal long arm of chromosome 22 and to the derivative 22 and therefore maps proximal to the balanced translocation breakpoint as expected. Alu PCR products were labelled and hybridised to ADU metaphases, and signal detected using a CCD as described previously(11). (d) FISH mapping of TUPLE1 to proximal 22q. The white bars indicate the normal chromosome 22 and the two translocation derivatives. Signal can be seen (false coloured yellow) on the normal 22 and the derivative 2 chromosome.

Table 1. Extent of deletion in catch22

	Diagnosis DGS	VCFS	CHD	Charge	Total
Patients deleted for interval:					
sc11.1b–sc11.1a	84	10	5	1	100
sc4.1–sc11.1a	6	–	1	–	7
D0832–sc11.1a	–	5	–	–	5
C1/79–sc11.1a	–	1	–	–	1
scF5–sc11.1a	1	–	–	–	1
no detectable deletion	2	6	25	2	

Some of this data has already been published in references 6,8,9,11, 13,19,21,23–25,28,62,66,70.

proximal to the breakpoint in ADU (Fig. 2c); no evidence for non-contiguity of insert was found. However, PFGE indicated that this YAC was just 47kb in size, and was, therefore, probably deleted for some of the insert DNA. End fragments were obtained by Alu-vector PCR but no recombinants were obtained in several screenings of the YAC arrays.

scF5 was similarly used to screen the YAC libraries. Three positives were obtained which mapped to 22q11 by FISH. End fragment cosmid isolation and FISH allowed the orientation and overlap of these YACs to be assessed (Fig. 1), and a cosmid contig was established around scF5. Cosmid C1/79 was isolated from the 275kb YAC 10BE5, employing vector-Alu PCR to obtain end fragment probes, and used in FISH analysis as described above; it mapped distal to scF5. The proximal end fragment sequences of YACS 37AF5 and 10BE5 were found to be present in cosmid scF5. However, the left hand vector-Alu end fragment (LHE) of YAC 30FH10, which is the more proximal end, was not represented in the Livermore cosmid grids, and 30FH10 showed a marked tendency to undergo deletion in culture.

The difficulty in obtaining clones within the 30FH10LHE-sc11.1a interval despite the complexity of the YAC and cosmid libraries available to us, coupled with the instability of the YACs in culture, suggested that the relevant genomic sequences might not be represented in standard human libraries. We therefore screened a gridded array of P1(36) recombinants which contained a 1.2-fold representation of the human genome (FF and HL manuscript in preparation). Positives were obtained and further screens conducted with the cDNAs described below and with the first round P1 positives (Fig. 3a); the resulting contigs are shown in Fig. 1.

FISH was used to map the ADU balanced translocation breakpoint with respect to the contigs. The karyotype of ADU is 46,XX,t(2;22)(2pter–2q14.1::22q11.1–22qter; 22pter–22q11.1::2q14.1–2qter)(8). As mentioned above the sc11.1a YAC detects sequences proximally to the breakpoint i.e. on the derivative 22 chromosome. The breakpoint was not detected using P1 clone G1313 or P1 clone H0787 and therefore lies in the region we have been unable to clone even using P1 and fosmid libraries.

Isolation of coding sequences from the DGCR

A probe from scF5 detects conserved sequences in a range of mammals (not shown) and was used to screen a human fetal brain library and a mouse embryo library. Several independent positives were obtained after tertiary screening of the human fetal brain library and one clone was isolated from a 10.5 day (pc) whole mouse embryo library. The largest human clone, C5, was analysed further; it has an insert of 3.1kb. Human fetal brain clones were confirmed as mapping to chromosome 22 by somatic cell hybrid analysis and/or FISH (Fig. 2d). The C5 DNA was hybridised to a Northern blot containing human fetal mRNAs. One transcript of 3.4kb was detected in all tissues (Fig. 3b), and one smaller 3.2kb transcript was observed in liver. The missing 3′ untranslated sequences were obtained by RACE PCR. The murine clone, MF2, has a 1.89kb insert.

The C5 and MF2 sequences encode a putative transcriptional regulator

The C5 and MF2 inserts were sequenced by primer walking and the results given in Fig. 4. There is a good Kozak consensus initiation sequence at bases 228 and 360 in C5 followed by open reading frames (ORFs) of 790 and 766 amino acids respectively. MF2 has an initiation sequence at 381 and an open reading frame of 519 amino acids. There is a potential polyadenylation signal at 3372 in C5 and 1918 in MF2, but unusually in MF2 there is a 3′ untranslated sequence of just 2bp. The larger predicted C5 protein has a molecular weight of 89kd and a pI of 7.8; the shorter ORF has a predicted molecular weight of 84kd. The shorter ORF is favoured because there is a termination codon between the two potential initiation codons at the corresponding position in the MF2 clone. The MF2-encoded protein has a predicted molecular weight of 57kd and a pI of 7.7.

The predicted proteins of MF2 and C5 are remarkably similar, amino acids 1–511 MF2 being 96% identical to amino acids 1–512 of C5 (C5 has a one amino acid insertion). Remarkably, the DNA sequence is highly conserved over this region (91% identity). The two predicted proteins have different carboxy-termini, with MF2 having a eight amino acid terminus from 512–519 unrelated to the much longer 254 amino region from 513–766 in C5. These data are consistent with MF2 being the murine homologue of C5, but representing an alternative splice variant.

Several features of both ORFs suggest they encode a transcriptional regulator. Firstly, in C5 there are five and in MF2 there are six occurences of the amino acid sequence SP; SP is relatively over represented in many such genes (37). Secondly, there is a potential nuclear localisation sequence at amino acids 235–242. Thirdly, the carboxy-terminal half of the protein is rich in polar amino acids, particularly glutamine, and serine and threonine (38% S and T in the carboxy terminal 89 amino acids of MF2). Fourthly, and most compelling, there are a series of sequence similarities to previously isolated gene products.

A significant similarity to a gene of known function is found with the Saccharomyces cerevisae Tup1 regulatory gene product(38). Similarity is also seen with several other proteins which contain a so called WD40 repeat, a motif first described in the beta subunit of transducin, the GTP-binding protein(39).

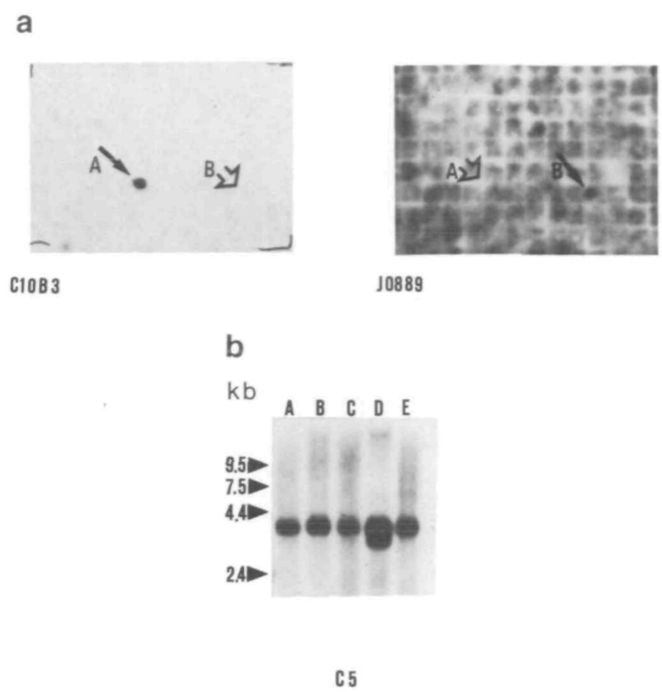


Figure 3. (a) Examples of P1 gridded array hybridisations. The left panel shows hybridisation of the whole sc11.1a YAC DNA to a section of a gridded array, with a positive colony indicated by the closed arrow A. The right panel shows hybridisation of P1 clone J0889 to the same section of the array, with closed arrow B indicating a positive colony. (b) Northern analysis of C5 expression in human fetal tissues. C5 C5 detects a 3.4kb transcript in all tissues tested, and a 3.2kb transcript in liver. Lanes A=heart; B=brain; C=lung; D=liver; E=kidney.

The varied effects of Tup1 alleles(40) suggested a defect in a transcriptional regulator, a hypothesis that has received strong support from transcriptional studies of test gene constructs in yeast (41).

Tup1 forms a complex with the transcriptional regulator Ssn6, which then further complexes with α2-Mcm1 and a1-α2 to form a general repressor of transcription(41). The α2 and **a**1 proteins contain homeodomains; Mcm1 contains a domain similar to that found in mammalian serum response factor. The pleiotropism of the Tup1 mutants and its interactions with homeodomain proteins is of interest given the variability of CATCH22 and the phenocopy of DGS produced in the HoxA3 knockout mouse (42). However, the similarity with Tup1 does not extend over the whole length of the protein, and the glutamine-rich domain of C5 and MF2 is at the carboxy- rather than the amino-terminus of the protein. The sequence similarity is essentially confined to the WD40 region and homology is particularly marked here because amino acids not present in the WD40 consensus are matched (Fig. 4c). Like Tup1, the C5/MF2-encoded protein is rich in serine and threonine (19% overall); potential phosphorylation sites within the ORFs were predicted using the Prosearch programme. There is some genetic evidence that Tup1 activity may be controlled via phosphorylation by proteins such as the SNF1 kinase (43).

Fig. 4d displays an alignment of the WD40 region of MF2 with the WD40 consensus sequence (taken from(38)). In the predicted ORFs there are three full WD40 domains, and there appears to be a fourth 'half' domain. Two WD40 proteins of unknown function were in fact the strongest database matches: YCR27C from Saccharomyces cerevisiae, and the AAC3 developmental regulator from Dictyostelium discoideum(44). The amino terminal WD40 repeat unit is most strongly related to that seen in the Drosophila gene E(spl) product (using the MPsrch programme), but is also similar to the Tup repeats. Human homologues of E(spl) have been isolated and contain WD40 repeats; they are called TLE proteins for Transducin-like/Enhancer of split (45). Their role in any human genetic defects is unknown. The C5/MF2 sequences are much more strongly related to Tup1 than to the TLE proteins or beta-transducin so, in the absence of any functional data, we propose to call the C5/MF2 gene Tuple1 for Tup-like/Enhancer of split gene 1.

A second, weaker, series of sequence similarities are detected in the region flanking the penta-Q stretch at amino acids 364–369 of MF2 and includes regulators such as the mastermind(mam)(46) and grainy head(47) proteins found in Drosophila. Assuming Tuple1 is a transcriptional regulator, the carboxy terminus of the protein presumably contains an activation or repression domain. In this region there is a stretch of weak similarity (detected by MPsrch) to two Drosophila repressors, the gap gene hunchback and the suppressor of zeste product (amino acids 462–507 and 437–511 respectively).

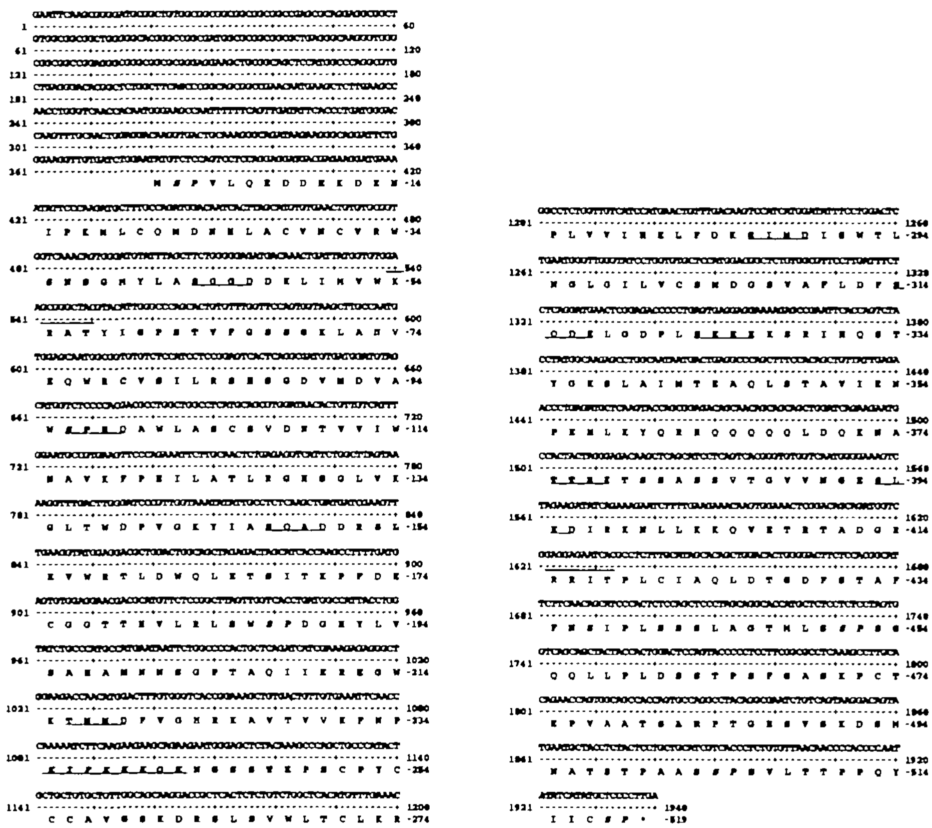
Mapping the TUPLE transcripts within the DGCR

Subfragments of the cDNAs were used to orientate the direction of transcription, which is towards the centromere (see Fig. 1). The 3′ end of C5 maps within the P1 clone G1313. Therefore, C5 sequences are not disrupted by the ADU balanced translocation breakpoint, and lie distal to it. Low stringency hybridisations of the fetal brain cDNAs to Southern blots of digested human genomic DNA gave a more complex set of fragments than seen at high stringency and we reasoned that sequences related to TUPLE1 might be present in the genome. Low stringency hybridisation of C5 subfragments to the gridded P1 arrays detected cross-hybridising clones from the proximal (NB4) contig as well as the scF5 contig and we are attempting to isolate cDNAs encoded within these sequences.

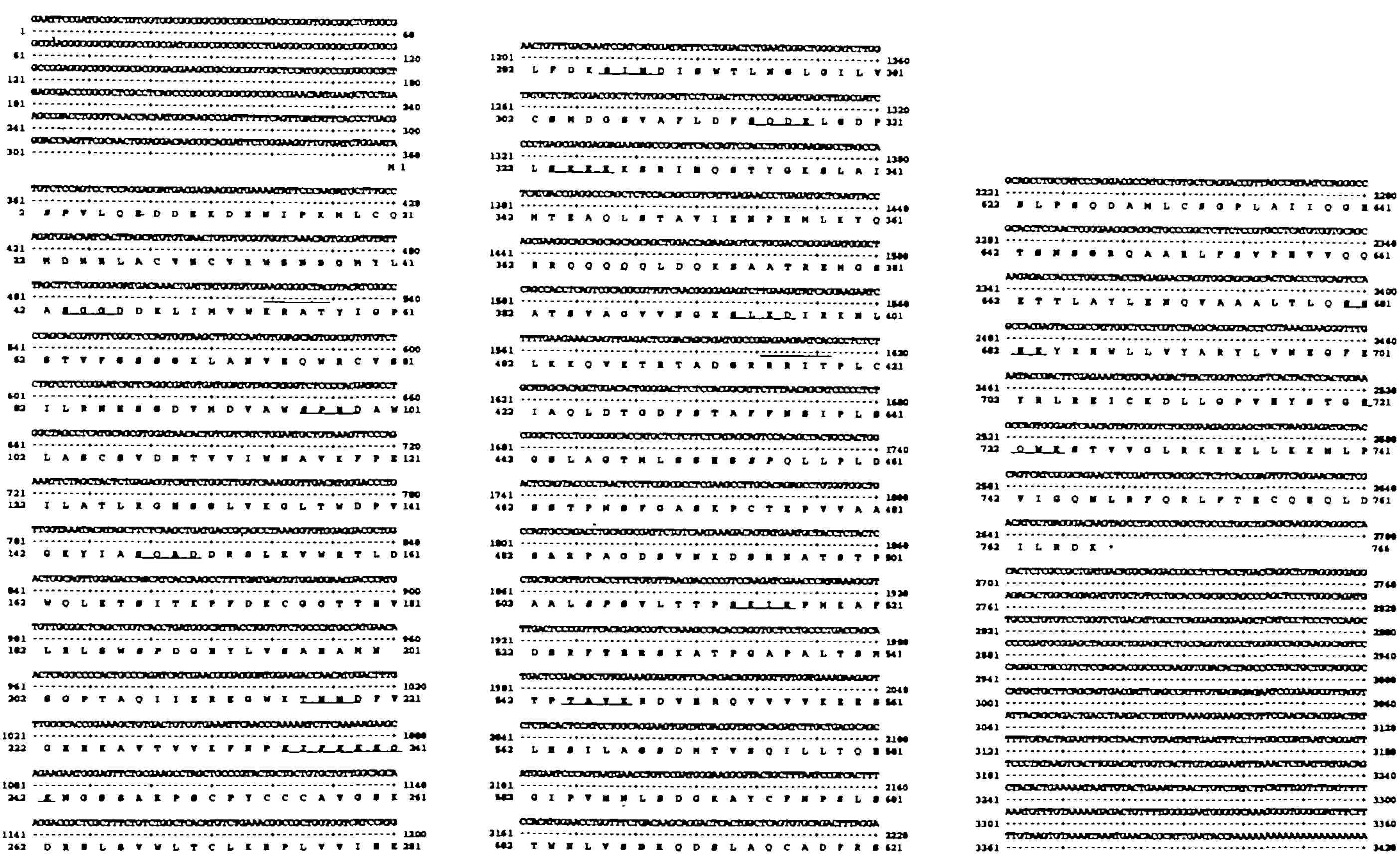
We have two patients with DGS, six patients with VCFS and twenty-five patients with conotruncal heart defects, who have no detectable deletion within 22q11. C5 and a subclone of the P1 G1313 were used to screen these patients for gross rearrangements or deletions, but none were found. However, there is a second DGS locus mapping to 10p13(48,49), and 17p(50) and 4q(51) have been proposed as further loci. Submicroscopic deletions at any of these loci, point mutations within TUPLE genes, or any teratogenic phenocopy(52–54) may be responsible for the birth defects in these cases.

TUPLE1 as a candidate gene in CATCH22

TUPLE1 is hemizygous in all patients known to have a deletion within 22q11 and encodes a protein similar to known transcriptional regulators. It is therefore feasible that haploinsufficiency of this gene is involved in DiGeorge syndrome and related abnormalities. However, although the balanced translocation of patient ADU falls within the scF5–sc11.1a interval it does not disrupt TUPLE1 sequences, at least those present in clone C5. It is therefore possible that another gene just proximal to TUPLE1 is disrupted by this breakpoint and that this is sufficient to cause DGS. Alternatively, the balanced translocation could exert a position effect on TUPLE1 (a similar distance exists between the PAX6 aniridia locus and a balanced translocation found in an a



b



c

```
     29                        53   73                                        115 142            157
Tuple1 VNCVRWSNSGMYLASGGDDKLIMVW     NVEQWRCVSILRSHSGDVMDVAWSPHDAWLASCSVDNTVVIWN     GKYIASQADDRSLKVW
       +  V  S  G +LA+G++D+LI +W    ++E ++ V IL++H  D+  +    P    L S S D TV IW+    GKYIA+ + DR+++VW
Tup1   IRSVCFSPDGKFLATGAEDRLIRIW     DIENRKIVMILQGHEQDIYSLDYFPSGDKLVSGSGDRTVRIWD     GKYIAAGSLDRAVRVW
     402                          426 427                                       469 495            510

     24                   44
Tuple1 NHLACVNCVRWSNSGMYLASG
       +H + V CV+ SN G YLA+G
Tup1   DHTSVVCCVKFSNDGEYLATG
     298                   318
```

d

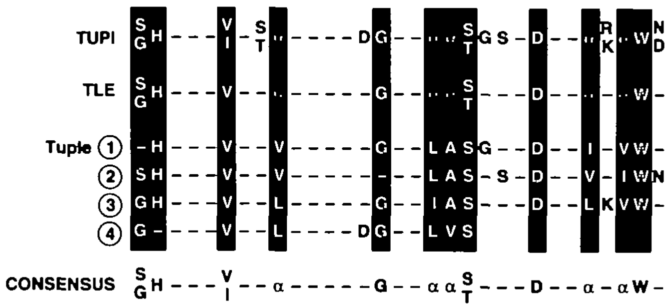


Figure 4. Sequence of C5 and MF2. (a) Complete nucleotide sequence of C5 and the derived protein product. (b) Complete nucleotide sequence of MF2 and the derived protein product. The polyadenylation signals are italicised. The amino acids SP are italicised. The putative nuclear localisation signal is bold, italicised and underlined. Potential sites of protein phosphorylation are given as follows: cAMP-dependent protein kinase phosphorylation sites are overlined; the CKII sites are underlined and the PKC sites are shown in bold. (c) Protein sequence match of the derived MF2 sequence (Tuple1) and Tup1 as detected with BLAST; the Poisson obtained using the BLAST program(68) is 5.4×10^{-9}. + indicates a conservative amino acid change. (d) Comparison of the WD40 domains of Tuple1 with the consensus sequences derived from the Tup1, cdc4 and beta transducin(39) proteins, and the TLE family (45). A new consensus is given below the matches. α = aliphatic amino acid.

aniridia family(55)). The deletion data taken as a whole are consistent with an additional mechanism. All interstitial deletions detected to date have hemizygosity for markers scF5–NB4 i.e. a deletion of at least 300kb (the proximal boundary of the SRO has yet to be accurately mapped), in contrast to other deletion syndromes where shortest region of deletion overlap analysis implicates a smaller critical region. This raises the possibility that hemizygosity of more than one gene is required for the development of CATCH22, for instance a second TUPLE-related sequence. In this model the balanced translocation in patient ADU could disrupt the co-ordinate regulation of a gene cluster by removal of a gene or genes from the action of a locus control region. 'Leakiness' of position effects or disruption of transcriptional regulation would explain the mild phenotypes seen in association with the 2;22 balanced translocation(29)(Dr P.Jouk, pers. comm.).

Given that TUPLE1 is deleted in all patients known to have a deletion of 22q11 we propose that haploinsufficieny of this gene (and possibly other related genes) is at least partially responsible for the phenotype of DGS and related birth defects. A WD40 protein has been identified as being haploinsufficient in the Miller–Dieker lissencephaly syndrome(56). Although this protein is unlikely to be a transcriptional regulator this demonstrates that dosage of these repeat subunit proteins can be important(56). Since the majority of patients have a 2Mb deletion it is conceivable that hemizygosity of other genes outside the critical region might contribute to the developmental defects in some patients. We have recently described a gene of unknown function, T10 (see Fig.1), which is expressed in early murine embryogenesis (57), and others have described zinc finger sequences (58,59).

Proof that TUPLE1 is haploinsufficient in DGS and related conditions will require creation of mice hemizygous for Tuple homologue(s). It is also possible that point mutations within individual TUPLE genes may occur in non-deletion cases of this series of congenital anomalies.

Possible mechanism of action of TUPLE1 haploinsufficeincy

Analysis of the Tuple1 sequence leads us to predict that the encoded proteins will, like other WD40 proteins, form multimeric complexes that function as transcription factors. Whatever the role of the TUPLE genes in human birth defects the identification of a WD40 transcriptional regulator expressed during mammalian embryogenesis is of interest given the role of such genes in cell fate determination in yeast and Drosophila. Likely candidates for proteins that interact with TUPLE would be proteins which, like Ssn6, contain a TPR snap helix motif. This is a 34 amino acid repeat defined by a degenerate consensus sequence with conserved spacing(60). One model predicts that the repeats form interlocking helical structures in which the 'knob' of one protein associates with the 'hole' of another(61). Several WD40/TPR protein interactions have been described, based on genetic(60) and functional(41) data. Such a TUPLE-TPR complex might then interact with DNA binding proteins and alter the transcription of target genes. Altered levels of the TUPLE subunits secondary to hemizygosity at 22q11 would alter the stoichiometric relationship with other subunits of the complex, which could conceivably disrupt function in a manner analagous to altered ratios of haemoglobin subunits in the thalassaemias. Whatever the exact mechanism, hemizygosity of TUPLE genes is associated with a set of abnormalities which, taken together, is a significant cause of human birth defects.

MATERIALS AND METHODS

Probes

sc4.1(62), sc11.1, D0832, and scF5(11) have been described previously. NB4 is 1.6kb HindIII/PstI single copy fragment detecting deletions involving sc11.1a alone.

Patients

Previously unreported patients were ascertained through the British Pediatric Association and assessed by the Newcastle group. Cell lines GM00980 and GM05878 were obtained from the NIGMS repository. ADU was a kind gift of Dr P.Jouk.

Genomic library screening

The chromosome-specific gene library LL22NC03 used in this work was constructed at the Biomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550 under the auspices of the National Laboratory Gene Library Project sponsored by the US Department of Energy. The library no. 106 (L4/FS22) was constructed by D.Nizetic from digests of DNA from 4× cell line LCL 127, ligated into Lawrist 4 and propagated in DH5 alpha.

Whole insert was isolated from cosmids, and whole YAC DNAs were isolated from pulse field gels. P1s were used as template for Alu PCR and the resulting products purified on Magic-Prep™ columns (Promega) or Amicon30 spun columns. YAC end fragments were isolated by Alu-vector and/or vectorette PCR(63,64). These various DNAs were labeled by random priming, competed with an excess of total human competitor and hybridised to gridded arrays as previously(57,65). Gridded arrays were manufactured using a Biomek 1000 robot(34); genomic clone libraries were stored at −70°C in microtitre dishes. Subfragments of the cDNAs, as described in the text, were similarly hybridised to the arrays, but without competition. Each filter set was stripped and exposed to check for absence of signal before re-use. Positive clones were picked and DNA purified for confirmation of cross-hybridisation by Southern analysis.

FISH

Alu PCR products from YAC and P1 clones were labelled and hybridised to metaphase spreads as described previously(11). Cosmids were nick-translated whole(11). FISH interphase measurements were conducted as we have described previously on 40 interphase nuclei(11). The scF5−sc11.1a distance of 300kb was extrapolated using the C1/79−scF5 measurement; C1/79 and scF5 are at the opposite ends of YAC 10BE5, 275kb in length.

Northern and Southern analysis

A northern blot filter was purchased from Clontech, and probed according to the manufacturer's instructions with oligolabeled C5 probe; the filter was washed at a final stringency of 0.1×SSC 0.1% SDS at 65°C. Quantitative Southern analysis and phosphorimaging were performed as previously described(25,66).

Sequencing and sequence analysis

cDNAs were sequenced using the dideoxy chain termination method(67) and primer walking. The DNA sequence was translated using the GCG package and the derived protein sequence used to search the NBRF and Swissprot databases using BLAST(68), BLOCKS(69) and MPsrch programs (developed by S Sturrock and JF Collins). Phosphorylation sites were predicted using the Prosearch programme developed by F.Kolakowski. All programmes are available at the HGMP resource center, Northwick Park, Harrow, UK. Consensus matching was conducted by hand.

ACKNOWLEDGEMENTS

We would like to thank patients, families and their clinicians for their support. We are grateful to Drs J.A.Garnes, M.Batzer and P.de Jonge (Lawrence Livermore Laboratories, CA) and Dr D.Nizetic (ICRF, London) for provision of 22 gridded cosmid arrays; Dr M.Simon's lab. (Caltech, CA) for fosmid clones; Dr A.Markham (Leeds University, UK), Dr D.le Paslier (CEPH, Paris), and Dr D.Schlessinger (Washington University) for supplying YAC libraries. The MRC, MRC Human Genome Mapping Programme, the British Heart Foundation, Action Research, EEC, Dunhill Medical Trust, the Borwick trust, and the Trustees of Great Ormond St. Hospital generously funded this work. The sequence data has accession numbers X75296 (HSTUPLE) and X75295 (MMTUPLE1) in the EMBL database.

REFERENCES

1. Wilson,D.I., Burn,J., Scambler,P. and Goodship,J. (1993) J. Med. Genet., **30**, 852−856.
2. Scambler,P.J. (1993) In Yacoub,M. and Pepper,J. (eds.), Annual of Cardiac Surgery. Current Science., London.
3. DiGeorge,A.M. (1965) J. Pediatr., **67**, 907.
4. McKusick,V.A. (1990) Mendelian Inheritance in Man. Johns Hopkins University Press, Baltimore.
5. de la Chapelle,A., Herva,R., Koivisto,M. and Aula,P. (1981) Hum Genet, **57**, 253−256.
6. Carey,A.H., Roach,S., Williamson,R., Dumamanski,J.P., Nordenskjold,M., Collins,V.P., Rouleau,G., Blin,N., Jalbert,P. and Scambler,P.J. (1990) Genomics, **7**, 299−306.
7. Sharkey,A.M., McLaren,L., Carroll,C., Fantes,J., Green,D., Wilson,D., Scambler,P.J. and Evans,H.J.E. (1992) Hum Genet, **89**, 73−78.
8. Carey,A.H., Claussen,U., Ludecke,H.-J., Horsthemke,B., Ellis,D., Oakey,H., Wilson,D., Burn,J., Williamson,R. and Scambler,P.J. (1992) Mammalian Genome, **3**, 101−105.
9. Scambler,P.J., Carey,A.H., Wyse,R.K.H., Roach,S., Dumanski,J.P., Nordenskjold,M. and Williamson,R. (1991) Genomics, **10**, 201−206.
10. Halford,S., Lindsay,E., Nayudu,M., Carey,A.H., Baldini,A. and Scambler,P.J. (1993) Hum Mol Genet, **2**, 191−196.
11. Lindsay,E.A., Halford,S., Wadey,R., Scambler,P.J. and Baldini,A. (1993) Genomics, **17**, 403−407.
12. Carey,A.H., Kelly,D., Halford,S., Wadey,R., Wilson,D., Goodship,J., Burn,J., Paul,T., Sharkey,A., Dumanski,J., Nordenskjold,M., Williamson,R. and Scambler,P. (1992) Am J Hum Genet, **51**, 964−970.
13. Wilson,D.I., Cross,I., Goodship,J.A., Scambler,P.J., Taylor,J.F.N., Walsh,K. and Burn,J. (1992) Am J Hum Genet, **51**, 957−963.
14. Driscoll,D.A., Budarf,M.L. and Emanuel,B. (1992) Am. J. Hum. Genet., **50**, 924−933.
15. Stevens,C.A., Carey,J.C. and Shigeoka,A.O (1990) Pediatrics, **85**, 526−530.
16. Shprintzen,R.J. (1990) In Buyse,M.A. (ed.), Birth Defects Encyclopaedia. Dover, MA, pp. 1744−1745.
17. Goldberg,R., Motzkin,B., Marion,R., Scambler,P.J. and Shprintzen,R. (1993) Am J Med Genet, **45**, 313−319.
18. Scambler,P.J., Kelly,D., Williamson,R., Goldberg,R. and Shprintzen,R. (1992) Lancet, **339**, 1138−1139.
19. Kelly,D., Goldberg,R., Wilson,D., Lindsay,E., Carey,A.H., Goodship,J., Burn,J., Cross,I., Shprintzen,R. and Scambler,P.J. (1993) Am J Med Genet, **45**, 308−312.
20. Driscoll,D.A., Spinner,N.B., Budarf,M.L., McDonald-McGinn,P.M., Zackai,E.H., Goldberg,R.B., Shprintzen,R.J., Saal,H.M., Zunana,J., Jones,M.C., Mascarello,J.T. and Emanuel,B.S. (1992) Am. J. Med. Genet., **44**, 261−268.
21. Wilson,D.I., Cross,I.E., Goodship,J.A., Coulthard,S., Carey,A.H., Scambler,P.J., Bain,H.H., Hunter,A.S., Carter,P.E. and Burn,J. (1991) Br. Heart J., **66**, 308−312.
22. Wilson,D.I., Goodship,J.A., Burn,J., Cross,I.E. and Scambler,P.J. (1992) Lancet, **340**, 573−575.
23. Wilson,D.I., Scambler,P.J., Goodship,J.A. and Burn,J. (1992) Proceedings of the British Paediatric Association Meeting., **April 7-10.**, Warwick UK.(Abstract)
24. Burn,J., Takao,A., Wilson,D.I., Cross,I., Momma,K., Wadey,R., Scambler,P.J. and Goodship,J. (1993) J. Med. Genet., **30**, 822−824.
25. Holder,S.E., Winter,R.M., Kamath,S. and Scambler,P.J. (1993) J. Med. Genet., **30**, 825−827.
26. Van Mierop,L.H.S. and Kutsche,L.M. (1986) Am. J. Cardiol., **58**, 133−137.
27. Siebert,J.R., Graham,J.M. and MacDonald,C. (1985) Teratology, **31**, 331−336.
28. Desmaze,C., Scambler,P., Prieur,M., Halford,S., Sidi,D., Le Deist,F. and Aurias,A. (1993) Hum Genet, **90**, 663−665.
29. Augusseau,S., Jouk,S., Jalbert,P. and Priur,M. (1986) Hum. Genet., **74**, 206.
30. Anand,R., Riley,J.H., Butler,R., Smith,J.C. and Markham,A.F. (1990) Nucleic Acids Res., **18**, 1951−1956.
31. Brownstein,B.H., Silverman,G.A., Little,R.D., Burke,D.T., Korsmeyer,S.J., Schlessinger,D. and Olsen,M.V. (1989) Science, **244**, 1348−1351.
32. Larin,Z., Monaco,A.P. and Lehrach,H. (1991) Proc. Natl. Acad. Sci. (USA), **88**, 4123−4127.

33. Chumakov,I.M., Rigalt,P., Guillou,S., Ougen,P., Billault,A., Guasconi,G., Gervy,P., Le Gall,I., Soularue,P., Grinas,L., Bougueleret,L., Bellanne-Chantelot,C., Lacroix,B., Gesnouin,P., Pook,S., Vaysseix,G., Frelat,G., Schmitz,A., Sambucy,J.-L., Bosch,A., Estivill,X., Weissenbach,J., Vignall,A., Riethman,H.C., Cox,D., Patterson,D., Gardiner,K., Hattori,M., Sakaki,Y., Ichikawa,H., Ohki,M., Le Paslier,D., Helig,R., Antonarakis,S.E. and Cohen,D. (1992) Nature, **359**, 380–387.
34. Bentley,D.R., Todd,C., Collins,J.E., Holland,J., Dunham,I., Hassock,S., Bankier,A. and Giannelli,F. (1992) Genomics, **12**, 334–341.
35. Albertson,H.M., Abderrahim,H., Cann,H.M., Dausset,J., Le Paslier,D. and Cohen,D. (1990) Proc. Natl. Acad. Sci. (USA), **87**, 4256–4260.
36. Sternberg,N. (1990) Proc. Natl. Acad. Sci. (USA), **87**, 103–107.
37. Suzuki,M. (1989) J. Mol. Biol., **207**, 61–84.
38. Williams,F.E. and Trumbly,R.J. (1990) Mol. Cell. Biol., **10**, 6500–6511.
39. Fong,H.K.W., Hurley,J.B., Hopkins,R., Miake-Lye,M.S., Johnson,R.F., Doolittle,R.F. and Simon,M.I. (1986) Proc. Natl. Acad. Sci. (USA), **83**, 2162–2166.
40. Williams,F.E., Varanasi,U. and Trumbly,R.J. (1991) Mol. Cell. Biol., **11**, 3307–3316.
41. Keleher,C.A., Redd,M.J., Schultz,J., Carlson,M. and Johnson,A.D. (1992) Cell, **68**, 709–719.
42. Chisaka,O. and Capecchi,M.R. (1991) Nature, **350**, 473–479.
43. Carlson,M., Osmond,B.C., Neigeborn,L. and Botstein,D. (1984) Genetics, **107**, 19–32.
44. Shaw,D.R., Richter,H., Giorda,R., Ohmachi,T. and Ennis,H.L. (1989) Mol. Gen. Genet., **218**, 453–459.
45. Stifani,S., Blaumueller,C.M., Redhead,N.J., Hill,R.E. and Artavanis-Tsakonas,S. (1992) Nature Genetics, **2**, 119–127.
46. Smoller,D., Friedel,C., Schmid,A., Bettler,D., Lam,L. and Yedvobnick,B. (1990) Genes Dev., **4**, 1688–1700.
47. Dynlacht,B.D., Attardi,L.D., Admon,A., Freeman,M. and Tjian,R. (1989) Genes Dev., **3**, 1677–1688.
48. Greenberg,F., Elder,F.F.B., Haffner,P., Nothrup,H. and Ledbetter,D. (1988) Am. J. Hum. Genet., **43**, 605–611.
49. Lai,M.M.R., Scriven,P.N., Ball,C. and Berry,A.C. (1992) J. Med. Genet., **29**, 586–588.
50. Greenberg,F., Courtney,K.B., Wessels,R.A., Huhta,J., Carpenter,R.J., Rich,D.C. and Ledbetter,D.H. (1988) Am. J. Med. Genet., **31**, 1–4.
51. Fukushima,Y., Ohashi,H., Wakui,T., Nishida,T., Nakamura,Y., Hoshino,K., Ogawa,K. and Oh-ishi,T. (1992) Am. J. Hum. Genet., **51**, 306.(Abstract)
52. Lammar,E.J., Chen,D.T., Hoar,N.D., Agnish,P.J., Benke,J.T., Braun,C.J., Curry,P.M., Fernhoff,A.W., Grix,A.T., Lott,J.M., Richard,J.M. and Sun,S.C. (1986) New. Eng. J. Med., **313**, 837–841.
53. Oster,G., Kilburn,K.H. and Siegal,F.P. (1983) Clin. Immunol. Immunopathol., **28**, 128–134.
54. Binder,M. (1985) Am. J. Pathol., **118**, 179–193.
55. Jordan,T., Hanson,I., Zaletayev,D., Hodgson,S., Prosser,J., Seawright,A., Hastie,N. and van Heyningen,V. (1992) Nature Genetics, **1**, 328–332.
56. Reiner,O., Carrozzo,R., Shen,Y., Wenert,M., Faustinella,F., Dobyns,W.B., Caskey,C.T. and Ledbetter,D.H. (1993) Nature, **364**, 717–721.
57. Halford,S., Wilson,D.I., Daw,S.C.M., Roberts,C., Wadey,R., Kamath,S., Wickremasinghe,A., Burn,J., Goodship,J., Mattei,M.-G., Moorman,A.F.M. and Scambler,P.J. (1993) Hum. Molec. Genet., **2**, 1577–1582.
58. Aubry,M., Marineau,C., Zhang,F.R., Zahed,L., Figlewicz,D., Delattre,O., Thomas,G., De Jong,P.J., Julien,J-P. and Rouleau,G.A. (1992) Genomics, **13**, 641–648.
59. Aubry,M., Demczuk,S., Desmaze,C., Aikem,M., Aurias,A., Julien,J.-P. and Rouleau,G.A. (1993) Hum. Molec. Genet., **2**, 1583-1587.
60. Goebl,M. and Yanagida,M. (1991) TIBS, **16**, 173–177.
61. Hirano,T., Kinoshita,N., Morikawa,K. and Yanagida,M. (1990) Cell, **60**, 319–328.
62. Carey,A.H., Kelly,D., Halford,S., Wadey,R., Wilson,D., Goodship,J., Burn,J., Paul,T., Sharkey,A., Dumanski,J., Nordenskjold,M., Williamson,R. and Scambler,P.J. (1992) Am. J. Hum. Genet., **51**, 964–970.
63. Nelson,D.L. (1990) GATA, **7**, 100–106.
64. Riley,J., Butler,R., Ogilvie,D., Finniear,R., Jenner,D., Powell,S., Anand,R., Smith,J.C. and Markham,A.F. (1990) Nucleic Acids Res., **18**, 2887–2890.
65. Scambler,P.J., Estivill,X., Bell,G., Farrall,M., McLean,C., Newman,R., Little,P.F.R., Frederick,P.A., Hawley,K.M.F., Wainwright,B.J., Williamson,R. and Lench,N.J. (1987) Nucl. Acid. Res., **15**, 3639–3652.
66. Wadey,R., Wickremasinghe,A., Roberts,C., Daw,S., Halford,S., Wilson,D., Goodship,J., Burn,J. and Scambler,P.J. (1993) J. Med. Genet., **30**, 807–812.
67. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. (USA), **74**, 5463–5467.
68. Altschul,S.F., Gish,W., Miller,W., Myer,E.W. and Lipman,D.J. (1990) J. Mol. Biol., **215**, 403–410.
69. Wallace,J.C., Agus,S., Henikoff,J.G. and Henikoff,S. (1991) Nucleic Acids Res., **19**, 6565–6572.
70. (1993) Genome Priority Reports, **1**, 574–587.