7. Results 5
Assessment of Chromosomal Aberrations Using Genomic Arrays.

7.1: Introduction

7.1.1: Microdeletion Syndromes

7.1.1.1. Low Copy repeats at sites of Microdeletions.

Unlike the genomes of lower organisms the human genome consists of over 50% repetitive DNA (IHGSC 2001). These repeats fall into two different categories; common repeat elements, such as short and long interspersed nuclear elements (SINES and LINES), and segmental duplications. Segmental duplications (a subclass of low copy repeats) are regions of the genome, typically 50-500Kb long with high sequence similarity (98.5-99%). Segmental duplications can occur as intrachromosomal duplications, where the duplicated regions are on the same chromosome or as interchromosomal where the two or more duplicated regions are on different chromosomes.

The completion and publication of the whole draft genome sequence in 2001 (IHGSC 2001) allowed the comparison of regions of the genome, and the identification of segmental duplications. Within the finished sequence an estimated 3.3% of the genome was involved in segmental duplications. Intrachromosomal duplications account for about 2.64% of the total genome and interchromosomal duplications for 1.44% (Cheung, Estivill et al. 2003). Gene rich chromosomes show the highest incidence of segmental duplications.

Computational analysis of the human genome sequence by Bailey et al (Bailey, Gu et al. 2002) identified 169 large regions of the genome that had an over-representation of human shotgun sequence used to sequence the human genome by Celera Genomics. These sequences were found to be rich in these segmental duplications. Of the 169 regions identified, 24 are currently associated with disease; these include Gauchers disease on chromosome 1, Prader Willi and Angelman's syndrome on chromosome

15 and the DiGeorge region on Chromosome 22. The combined incidence of childhood disease involving segmental duplications is 1:750 (Eichler 2001). The reason for the association of duplicated regions with disease is due to the misalignment of chromosomes during meiosis where recombination occurs between duplicated regions rather than allelic loci. Recombination between homologous regions may result in deletion, amplification or inversion events. This can result in the disruption of a gene resulting in disease associated phenotypes. The mechanisms for this are shown in Figure 7.1.

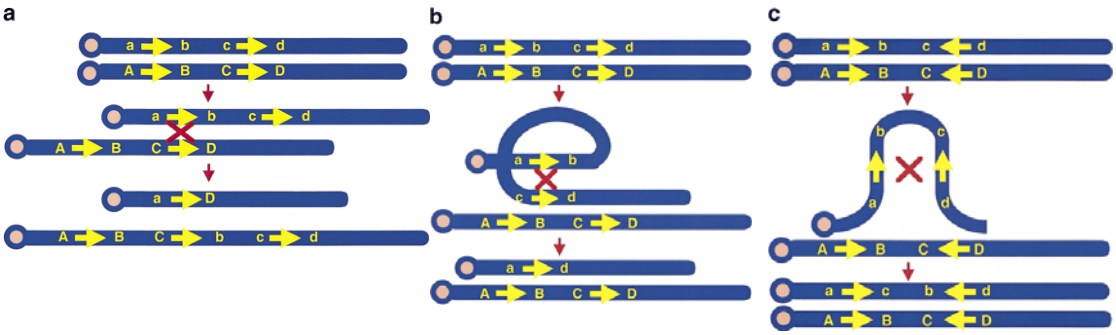


Figure 7.1: Mechanisms for segmental duplications a: recombination between repeats on two separate chromosomes leads to a deletion on one chromosome and an amplification on the other chromosome. b: recombination between repeats on the same chromosome leads to a deletion. c: recombination between repeats in an opposite orientation leads to an inversion (Adapted from (Ji, Eichler et al. 2000)).

One region of the genome particularly rich in segmental duplications is the subcentromeric region of the q arm of chromosome 22 (Dunham, Shimizu et al. 1999). In the 1.5 Mb of DNA adjacent to the centromere, which represents just 5% of the chromosome 22 sequence, 90% of sequence is duplicated on other chromosomes. 52% of the interchromosomal duplications on chromosome 22 were also located in this region (IHGSC 2001). Low copy repeats are also common within the next 7.5Mb of chromosome 22, with most of the sequence clones in the first 9Mb of the q arm containing some form of segmental duplication. The duplicated regions of chromosome 22 are represented in Figure 7.2. The highly duplicated region at the beginning of 22q includes the DiGeorge critical region (DGCR) that is deleted in patients suffering from DiGeorge syndrome. The same region involved in segmental

duplication is involved in velocardiofacial syndrome (VCFS) and CATCH22. The diversity in names is due to the variability of phenotypes observed.

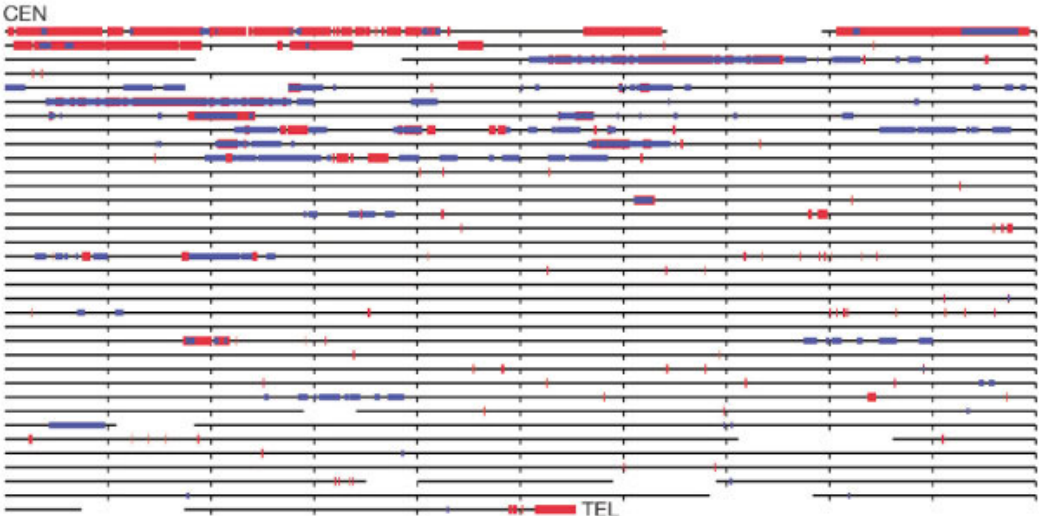


Figure 7.2: Segmental duplications on the sequenced q arm of chromosome 22. Each line represents 1Mb, with each intersection at 100Kb. Intrachromosomal repeats are shown in blue and interchromosomal repeats in red. Duplication alignments with > 90% nucleotide identity and > 1 kb long are shown (IHGSC 2001), (Bailey, Gu et al. 2002).

7.1.1.2. DiGeorge Syndrome & Conventional Diagnosis

DiGeorge is the most common microdeletion syndrome and is present in 1:4000 live births (Devriendt, Fryns et al. 1998). The syndrome is characterised by a variety of clinical features. These include a variety of heart defects, mainly affecting the aortic arch, immunodeficiencies due to a hypoplasic/absent thymus, hypocalcaemia – owing to hypoplasia of the parathyroids, and distinct facial features including low set ears and a cleft pallet. Cases presenting later in childhood tend to have a milder phenotype encompassing heart defects (OMIM entry 188400). The 3Mb deletion and the 1.5Mb deletion have indistinguishable phenotypes (Maynard, Haskell et al. 2002).

The DiGeorge critical region has been localised to chromosome 22 approx. 3966000 – 7888000bp along the q arm. The region is flanked by accession numbers AC008079-D86996 and the boundaries were defined by screening using high density genetic markers. Over 150 patients were screened with their unaffected parents used as

controls. A detectable deletion was found in 83% of the patients examined (Carlson, Sirotkin et al. 1997). Most deletions (approx. 90%) encompass a 3Mb deletion between two duplicated regions (Lindsay 2001). A further 8% of deletions have the same proximal boundary, but are smaller, encompassing 1.5Mb of DNA, between another low copy repeat (Figure 7.3). It has also been observed that rearrangement within the DiGeorge region may be associated with balanced translocations with 11q23 (Spiteri, Babcock et al. 2003). Non-22q11 deletions resulting in the DiGeorge phenotype may be due to deletions on other chromosomes. In these patients deletions have been detected on 10p13, 18q21.33 and 4q21.3-q25 (Greenberg, Elder et al. 1988 McDermid and Morrow 2002).

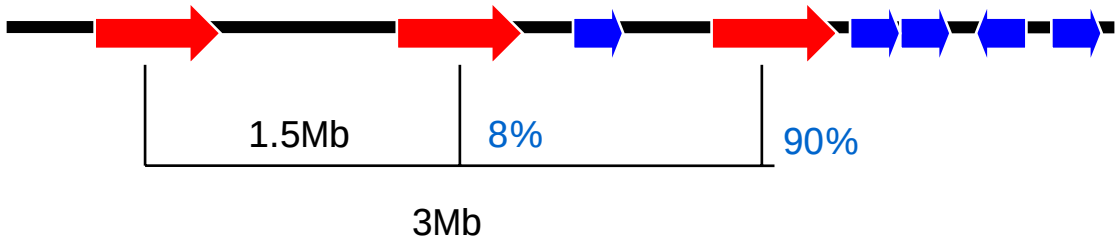


Figure 7.3: Patterns of deletion in DiGeorge patients. Red arrows: Duplications involved in DiGeorge deletions. Blue arrows: Other segmental duplications in the vicinity of the DiGeorge critical region.

The 3Mb deletion that is responsible for most patient phenotypes incorporates 30 genes. The smaller nested deletion encompasses 24 genes, with a variety of functions. However, no single gene has been identified as being solely responsible for DiGeorge syndrome. A low copy repeat (LCR 22) ranging in size from 40-350Kb and of 97-98% identity has been found at all three 22q11 breakpoint regions, as well as six adjacent locations over a 6.5Mb region. The mechanism of the chromosomal rearrangement at 22q11 is shown in Figure 7.4 (Maynard, Haskell et al. 2002).

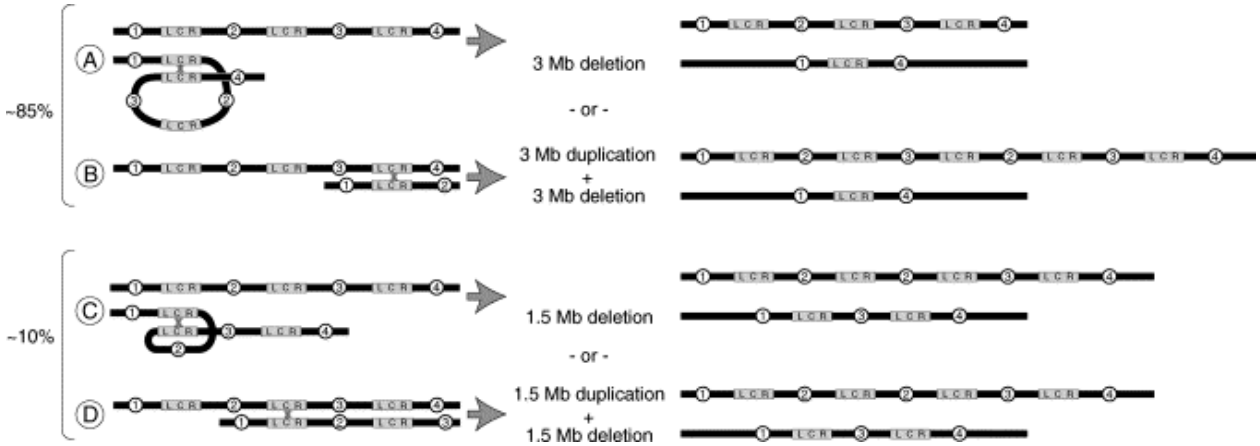
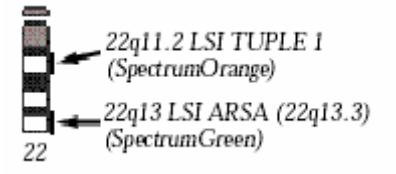


Figure 7.4: Mechanisms for deletion in DiGeorge patients (Maynard, Haskell et al. 2002).

The DiGeorge deletion in patients is usually clinically confirmed by FISH analysis of patient chromosomes using commercially available probes.

A



B

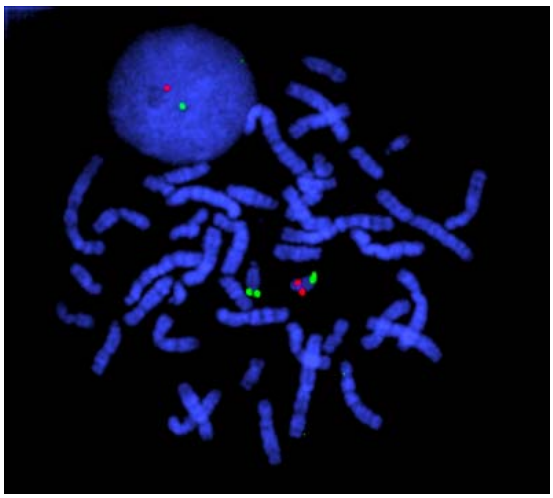


Figure 7.5: Detection of the DiGeorge deletion on patient metaphases using the commercially available probe set from Vysis. A: This is a two colour probe mix containing a spectrum orange probe mapping to the non-coding region of TUPLE 1 in the DiGeorge region and spectrum green labelled control probe hybridised to a region on 22q13.3. B: The Red signal can be seen as present on the normal chromosome 22 but absent on the copy of chromosome 22 containing the DiGeorge deletion. Reproduced from (Gribble 2003).

This commercially available Vysis probe illustrated in Figure 7.5 contains the markers D22S553, D22S609, D22S942, within the accession numbers AC000085, AC000092, AC000079. The probe is located within the first 1.5Mb of the DiGeorge critical region so will detect both the 1.5Mb and the 3Mb deletions. These commercial probes are used within clinical labs to detect the presence and absence of a deletion in the DiGeorge region of patients displaying a DiGeorge phenotype. They do not give any information about the size of the deletion. As the candidate gene(s) for DiGeorge deletion has not yet been identified, information about the deleted genes and their effect on phenotype is still critical in the understanding of DiGeorge syndrome.

The extent of the deletion in DiGeorge patients is conventionally sized using FISH probes covering contiguous regions along 22q11 (Lindsay 2001). Recently the production of a 22q tile path array has allowed the sizing of the deletion in a DiGeorge patient in a single hybridisation experiment (Buckley, Mantripragada et al. 2002). The hybridisation of DNA derived from a transformed DiGeorge lymphoblastoid cell line to the chromosome 22 tile path array showed a deletion spanning the DiGeorge critical region in the one patient examined. Some of the loci within the DiGeorge region gave ratios that are difficult to interpret on the CGH array. It was concluded that the reason for this is the high amount of common repeat elements within the cosmids located within this region. This preliminary study showed that a DiGeorge deletion could be detected on a DNA array. Further investigation of DiGeorge patients may confirm which genes are involved in the phenotype.

7.1.2: Immunoglobulin Rearrangements

7.1.2.1: Genome wide Immunoglobulin rearrangements

The human immune system relies on antibodies and T cell receptors to fight the large range of infectious agents that invade the body. As there is such a plethora of antigens the body has to deal with, the immune system has to produce a wide assortment of antibodies and T cell receptor (TCR) genes.

An individual produces more different types of antibody than all other proteins in the body put together. There are many more types of antibody in the body than there are genes in the genome, and therefore the conventional idea of one gene encoding one mRNA and one protein will not produce enough antibody diversity. A unique way of producing diversity has been observed in regions of the genome encoding antibodies.

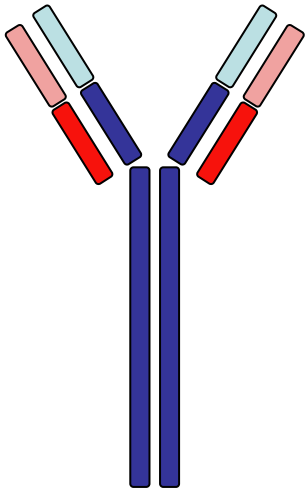


Figure 7.6: Basic four chain structure of an immunoglobulin protein. Blue: Two heavy chains, comprising of a constant region (C_H - dark blue) and a variable region (V_H – light blue). Red: Two light chains comprising of a constant region (C_L - red) and a variable region (V_L – pink).

Immunoglobulin gamma proteins have the same basic four chain structure (Figure 7.6). Antibiotic diversity is generated at the variable regions on the heavy (IgH) and light chains (IgL). The light chains are of two different types, either kappa (κ) or lambda (λ). The immunoglobulin heavy chain is encoded by a cluster of genes on chromosome 14q32.33 whilst the κ chain is encoded by genes on chromosome 2p11.2 and the λ chain is encoded on 22q11.22. The use of two different types of chain (heavy and light) immediately increases the antibody variability as any light chain can combine with any heavy chain.

In addition to the diversity generated by heavy and light chain association, immunoglobulin diversity is increased due to the rearrangement of germ-line variable

(V), diversity (D) and joining (J) gene segments. The IgH variable region exons are assembled from V, D and J gene segments. The IgL variable regions are assembled from just V and J gene segments. During B cell development the IgH variable region undergoes rearrangement first. Only 1 in 3 IgH rearrangements are in-frame and therefore successful, provided the IgH rearrangement is successful the IgL chain undergoes rearrangement. This V(D)J recombination is vital to produce diverse antibodies and only occurs in developing lymphocytes between immunoglobulin or TCR genes.

7.1.2.2: The Mechanism of λ Chain Rearrangement in Chromosome 22.

The lambda gene locus, encoded on chromosome 22q11 contains a set of variable genes and seven constant gene regions (Figure 7.7). The region was sequenced in 1997 (Kawasaki, Minoshima et al. 1997). In cells not producing antibodies, the variable genes and constant regions are found far apart. In cells that form antibodies the constant and variable genes are brought closer together, but still remain approximately 1500bp apart. The variable and constant regions are separated by a joining segment, which also contributes to diversity.

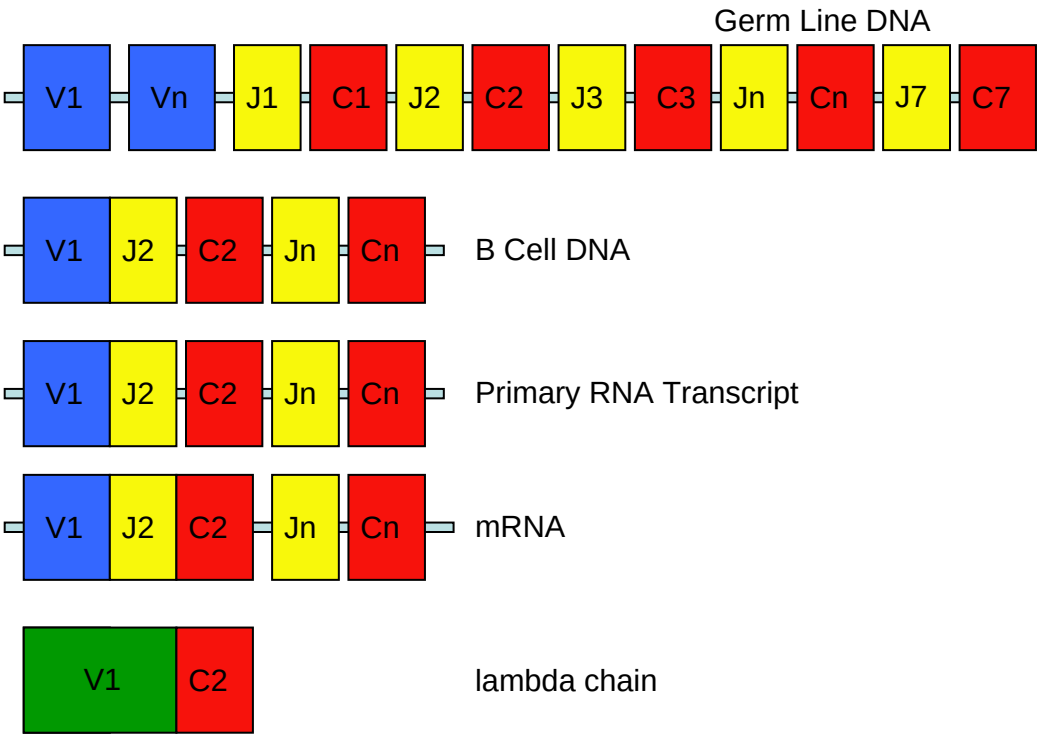


Figure 7.7: Recombination of the lambda chain of the immunoglobulin light chain. For details see text. Figure adapted from (Turner 2001).

In the germ-line of the λ chain loci there is a variety of V segments and seven different constant regions. Each constant gene region is accompanied by just one joining gene (unlike the κ chain). During B cell development the gene groups are rearranged and a region of the DNA is excised to bring variable and joining regions together. This rearrangement is mediated by a conserved recombination signal sequence which flanks the recombining regions. The recombination signal is an AT rich nonomer (ACAAAAACC) which is separated by a non-conserved 12 or 23bp spacer sequence and found upstream of the J segment. The VJ recombination occurs at sites of double stranded breaks, the recombination signal then provides complementary sequence so the ends can be precisely joined. The joining of the different segments can result in an inversion or deletion of the intervening sequences, resulting in copy number changes in these regions.

One allele will initially undergo rearrangement. If this is unsuccessful the second allele will undergo rearrangement. The IgLκ and IgLλ loci expression is under negative feedback control. The production of a functional IgL protein feeds back onto the IgLκ and IgLλ loci and prevents unnecessary rearrangement.

In this way different B cells contain different rearrangements of the constant, joining and variable genes leading to antibody diversity. The newly formed variable, joining and constant arrangement is transcribed into primary RNA. The introns are then removed and the spliced mRNA is translated into the lambda chain protein.

Expression of the successfully rearranged IgL chain is enhanced by epigenetic modification of the chromatin associated with active chromatin. The CpG islands are demethylated, histones are acetylated and transcriptional activators are also recruited to the chromatin (Blackwood and Kadonaga 1998). The active allele is switched to become early replicating whilst the inactive allele is late replicating (Goren and Cedar 2003).

Deletions in this region due to excision of DNA during VJ recombination can be detected, both by FISH and by using a CGH microarray (Buckley, Mantripragada et al. 2002). Assessment of the VJ recombination on CGH arrays will allow the sizing of rearrangements. The arrays may be able to detect incomplete VJ recombination, which leads to an imbalance in B and T cells in the immune system by reporting copy number change. They may also be used to detect aberrant VJ recombination which can lead to chromosomal translocations (Bassing, Swat et al. 2002).

7.1.3: Assessment of DiGeorge and IgL λ copy number change on genomic arrays.

The two previous chapters have described how genomic clone microarrays have been used to assess DNA replication timing in a human cell line. However the microarray sampling the genome at a 1Mb resolution and the 22q tile path array described in Chapter 4 also have other uses, such as detecting DNA copy number changes.

Section 7.2.1 describes how the chromosome 22q tiling path array can be used to detect a deletion in the DiGeorge region of 22q11. Patient DNA was obtained by collaboration with Charles Shaw-Smith, from the Department of Medical Genetics at Addenbrookes Hospital, Cambridge. Patients exhibiting features of the DiGeorge phenotype, but showing no 22q11 deletion by conventional FISH analysis were applied to the chromosome 22q array, and to the array sampling the genome at a 1Mb resolution. This is described in section 7.2.2. The DNA was obtained from collaboration with Katrina Prescott, from the Institute of Child Health, University College London. The arrays detected a deletion in one patient. Follow-up work, including additional FISH analysis and microsatallite analysis was performed at the Institute of Child Health.

The genome of B cells can undergo a rearrangement during development of the immunoglobulin light chain λ locus located in 22q11. Section 7.3 describes deletions observed in lymphoblastoid cell lines due to this rearrangement.

7.2: Array analysis of DiGeorge syndrome patients

7.2.1: Assessment of DiGeorge Patient DNA samples on the Chromosome 22q Tile path array.

DiGeorge Syndrome is a congenital defect caused by a deletion on chromosome 22q11. Five individuals who displayed a DiGeorge phenotype and who had demonstrated a deletion at 22q11 using a commercial diagnostic fluorescence *in situ* hybridisation (FISH) probe set were selected for analysis. The size of the 22q11 deletion was assessed by the hybridisation of DNA from these patients onto the 22q tile path array.

Initially, DNA from five different patients was hybridised to individual arrays using DNA from a male lymphoblastoid cell line (HRC 575) as a reference. A typical result is shown in Figure 7.8.

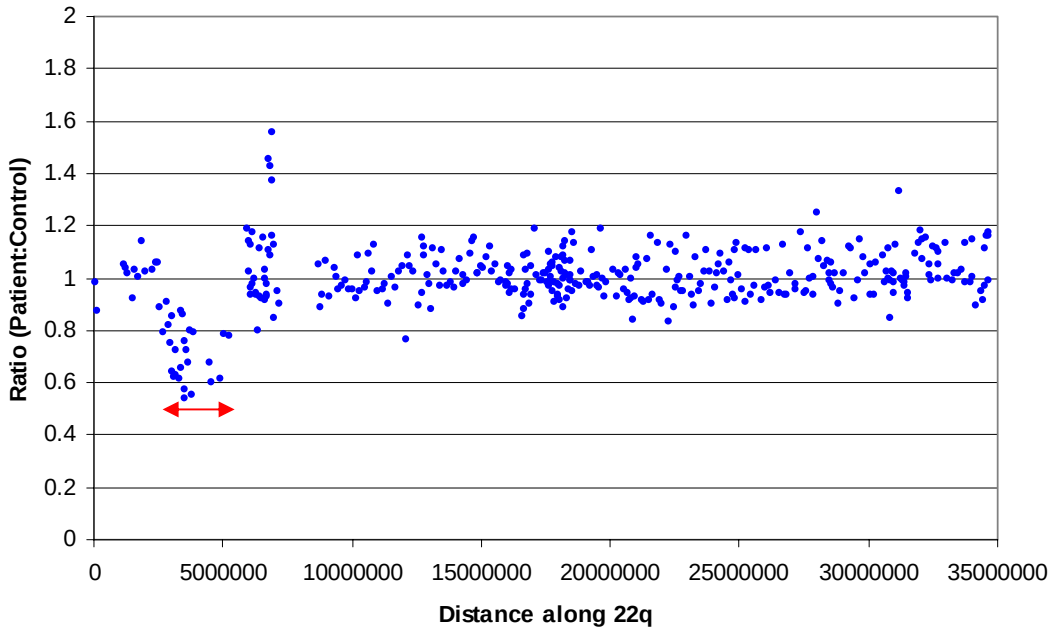


Figure 7.8: Hybridisation of DNA from a DiGeorge patient onto the 22q tile path array. DNA from a male lymphoblastoid cell line was used as a reference. Red arrow shows the region of the DiGeorge deletion.

In Figure 7.8, a deletion was can be seen between 2577096-5227316bp along chromosome 22q (between clones bac 519d21 – pac 52f6; international clone names

and accession numbers can be found in Appendix 9). A copy number change was defined as 5 times the standard deviation of the self:self hybridisation performed in 4.3.2 (0.2) to be highly statistically significant. A deletion was defined as clones that report a ratio below 0.8. A copy number gain was defined as clones with a ratio above 1.2. An amplification can also be seen in four of the clones between 6789448-6935464bp along chromosome 22q (between clones cN24A12 and cN9G6 – Table 7.1). However comparison with other experiments (See 4.3.3 and 7.3) showed that the amplified ratios seen were actually due to a deletion in the reference cell line at the immunoglobulin light chain λ locus. Consequently all the DiGeorge experiments were repeated using a pool of DNA extracted from units of donated blood from twenty anonymous individuals.

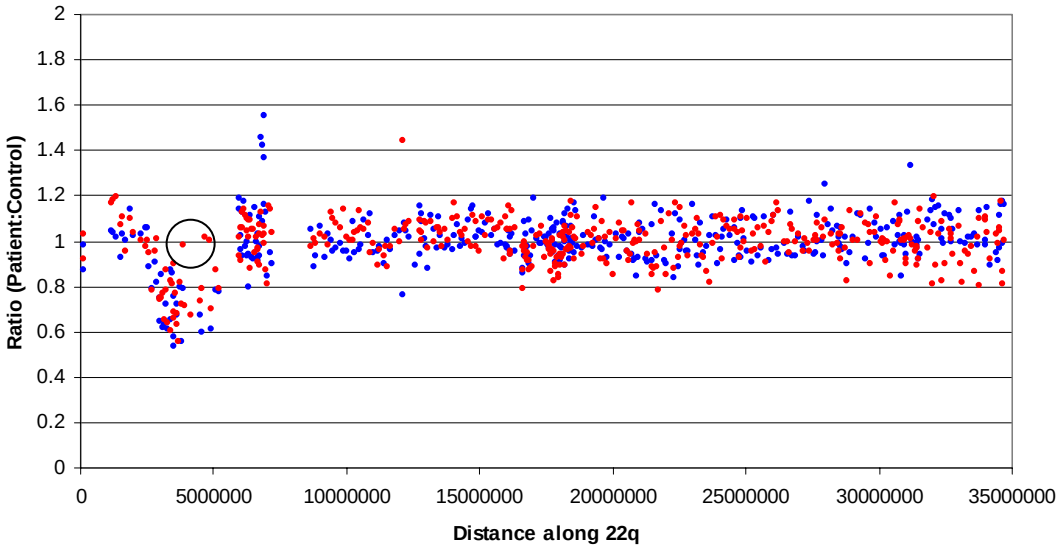


Figure 7.9: DNA from the same DiGeorge patient hybridised to two different arrays using two different control samples. Blue data points: Patient DNA hybridised against a single lymphoblastoid reference cell line (HRC 575). Red data points: Patient DNA hybridised against a pool of blood DNA from twenty different individuals. The three red points in the DiGeorge deletion region reporting normal ratios (circled) are discussed below.

When the patient DNA was hybridised against pooled DNA from twenty different individuals the four clones previously identified as having an elevated patient:control

ratio when hybridised against the lymphoblastoid cell line showed normal 1:1 ratios (Table 7.1).

Table 7.1: The region of 22q which exhibited a gain when the cell line HRC 575 DNA was used, and yet showed normal ratios when hybridised against a pool DNA control. Clones marked with an * exhibited the same copy number as the control DNA when hybridised against both control DNA's.

Clone	Midpoint of clone	Ratio V. cell line DNA	Ratio V. pool DNA
cN24A12	6789448.5	1.45	1.13
cN86D6*	6823353	1.09	1.02
cN92H4	6859162	1.42	Not Available
cN84E4	6894369	1.37	0.99
cN9C5*	6915218.5	1.16	1.03
cN9G6	6935464.5	1.56	1.06

DNA samples from five different DiGeorge patients were hybridised each to separate arrays using the pool DNA as a control. The results are shown in Figure 7.10 and Table 7.2.

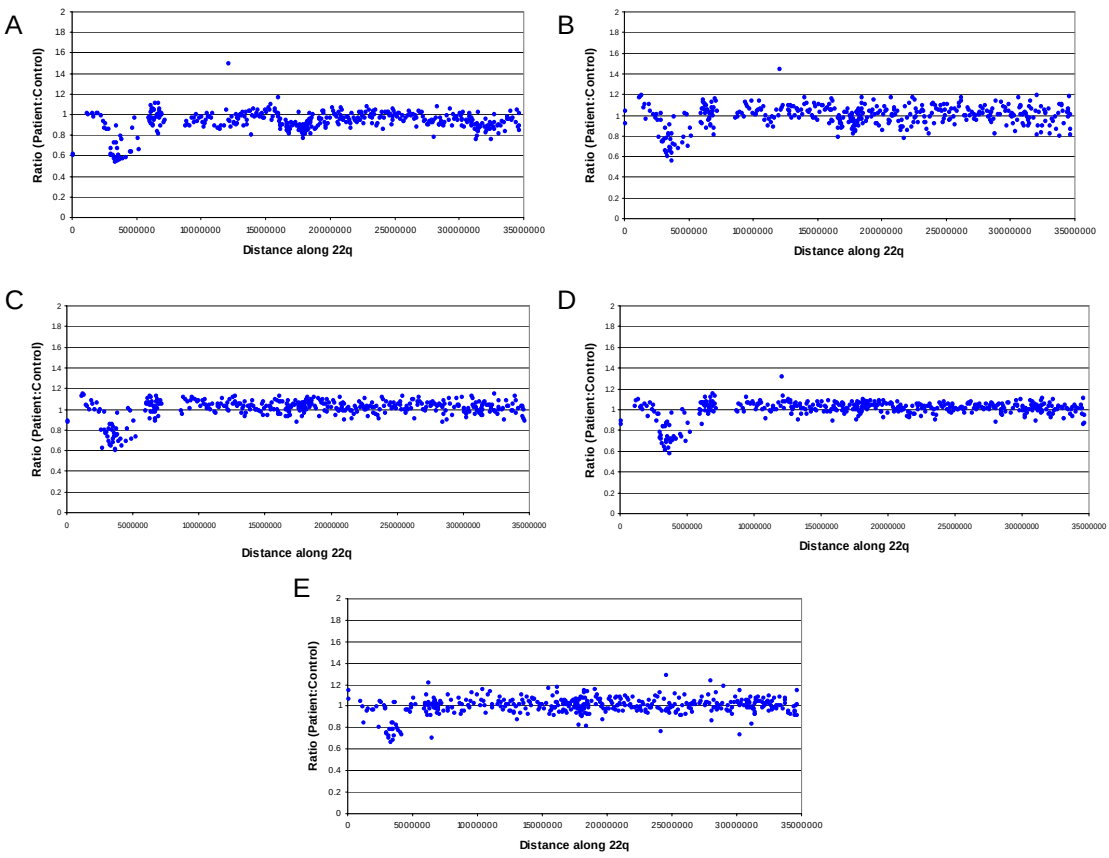


Figure 7.10: (Previous page) Patient:Control ratio profiles for five separate DiGeorge Patient DNA samples. A: Patient 1, B: Patient 2, C: Patient 3, D: Patient 4, E: Patient 5.

Table 7.2: Patient:Control ratios of clones in the DiGeorge region of chromosome 22. Shaded ratios indicated deletions in the loci represented on the array. Ratios shaded with pale grey only show a slightly reduced but not significant ratio, but are within the deleted region. NA: Clone ratio is not available. (Loci that did not pass the analysis criteria described in 2.5.2 and were rejected.) NP: clone was not present on early arrays.

Clone	Location	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
b677f7	2523529	0.867599	1.00634	0.997294	1.041255	1.043958
bac519d21	2577097	0.910715	0.95083	0.799467	0.976757	1.036773
pac995o6	2710128	0.856811	0.782641	0.626426	0.893028	1.007301
p423	2881004	0.857811	1.012522	0.969525	0.958052	0.98568
fos41c7	2950903	NP	NP	NP	NP	0.979491
cN119F4	2979519	0.667419	0.746817	0.76231	0.723682	0.758473
18c3	3013238	0.616361	0.743239	0.772776	0.777589	0.747465
111f11	3052391	0.609105	0.750534	NA	0.753778	0.719907
b72f8	3192997	0.607635	0.873848	0.804798	0.841273	0.704972
fF39E1	3236756	0.727743	0.78112	0.852192	0.834134	0.785443
Cos98c4	3306156	0.568455	0.637356	0.656379	0.638873	NA
Cos49c12	3369955	0.534524	0.603945	0.647299	0.603859	0.664103
Cos 83c5	3406693	0.721168	0.825051	0.801886	0.836507	0.784862
Cos83e8	3443825	0.888218	0.813	0.847148	0.869298	0.84075
Cos 59f	3467897	0.578088	NA	NA	NA	0.779052
Cos105a	3493862	0.546522	0.658493	0.678754	0.679448	0.685424
Cos81h	3532078	0.67372	0.90295	0.828353	0.882227	0.84065
Cos31e	3569626	0.609019	0.768846	0.765578	0.737782	0.724199
Cos100h	3612315	0.600973	0.679369	0.729034	0.71584	1.036948
Cos91c	3652472	0.557729	0.63308	0.60015	0.63035	1.03047
Cos 89h	3731131	0.79355	0.818929	0.81366	0.842539	0.827075
c2h	3804446	0.571592	0.721842	0.75199	0.683078	0.794157
c56c	3878158	0.752982	0.981838	0.957891	0.969274	NA
p888c9	3925566	0.568191	0.712153	0.687371	0.705867	0.774482
p158I19	4056427	NA	NA	0.70935	0.740995	0.74914
b444p24	4165628	0.581054	0.676969	0.648217	0.719954	0.73486
b562F10	4491054	0.634745	0.734261	0.686802	0.760658	0.967303
p_m11	4591569	0.639606	0.787911	0.811635	0.740377	0.965098
bac32	4686897	0.86681	1.01556	1.005373	0.998268	0.967807
pac408	4854070	0.967018	1.00157	0.984653	0.969228	0.949261
b135h6	4935029	NA	0.69915	0.709847	0.697637	1.012156
p_n5	5082378	0.763858	0.872079	0.886715	0.864976	1.075634
p52f6	5227317	0.662279	0.793727	0.737265	0.778774	1.019238
cN109G12	5976695	0.964023	1.014324	0.981454	1.003975	1.02227

DiGeorge deletions at 22q11 can be seen in all five patients tested. An additional gain in clone dJ477H23 can also be seen in patients 1, 2 and 4, 12.1Mb along chromosome 22q.

Detailed plots within the DiGeorge region, for all five patients are shown superimposed in Figure 7.11.

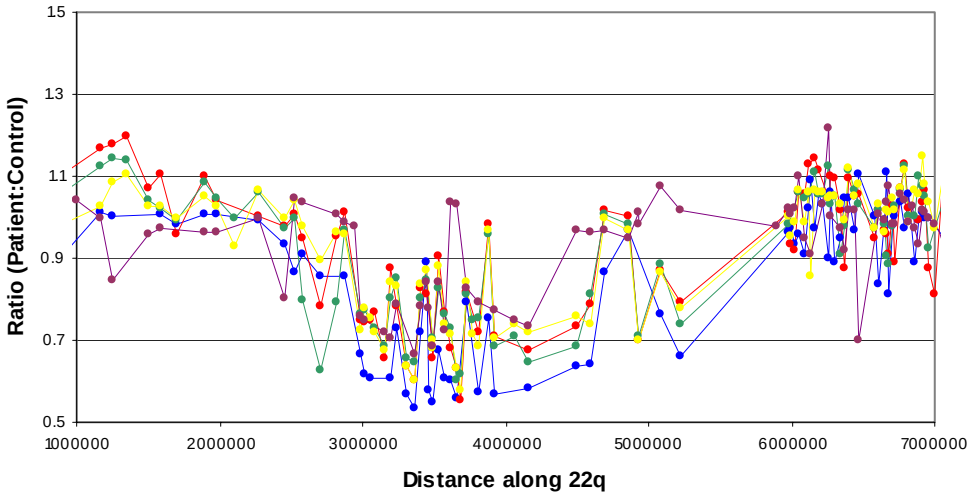


Figure 7.11: Patient:Control ratios obtained when five different patients are plotted on the same axis. Blue: Patient 1. Red: Patient 2. Green: Patient 3. Yellow: Patient 4. Purple: Patient 5.

A full single copy loss (ratios approximating 0.5:1) can only be seen in a few of the clones in the deleted region. The clones showing a full single copy loss are interspersed with clones reporting an intermediate ratio. There are also clones in the centre of the deleted region that report ratios that are modal as would be expected for non-deleted regions. The four clones that have this characteristic on multiple arrays are; Pac423 (average ratio = 0.95), Bac32 (0.91), Cos56c (0.97) and Pac408 (0.97).

These intermediate ratios make defining the boundaries of the deletion uncertain. At the edges of the deletion it is difficult to distinguish deleted ratios from the background variation. Because of this the shaded deleted regions in Table 7.2 do not have their boundaries defined exactly. It is however possible to determine that the deletion in patient 5 is smaller than the other deletions. This was confirmed by FISH

analysis of the patient chromosomes using clones bK562F10, p_n5 and p52F6 as probes (Figure 7.12).

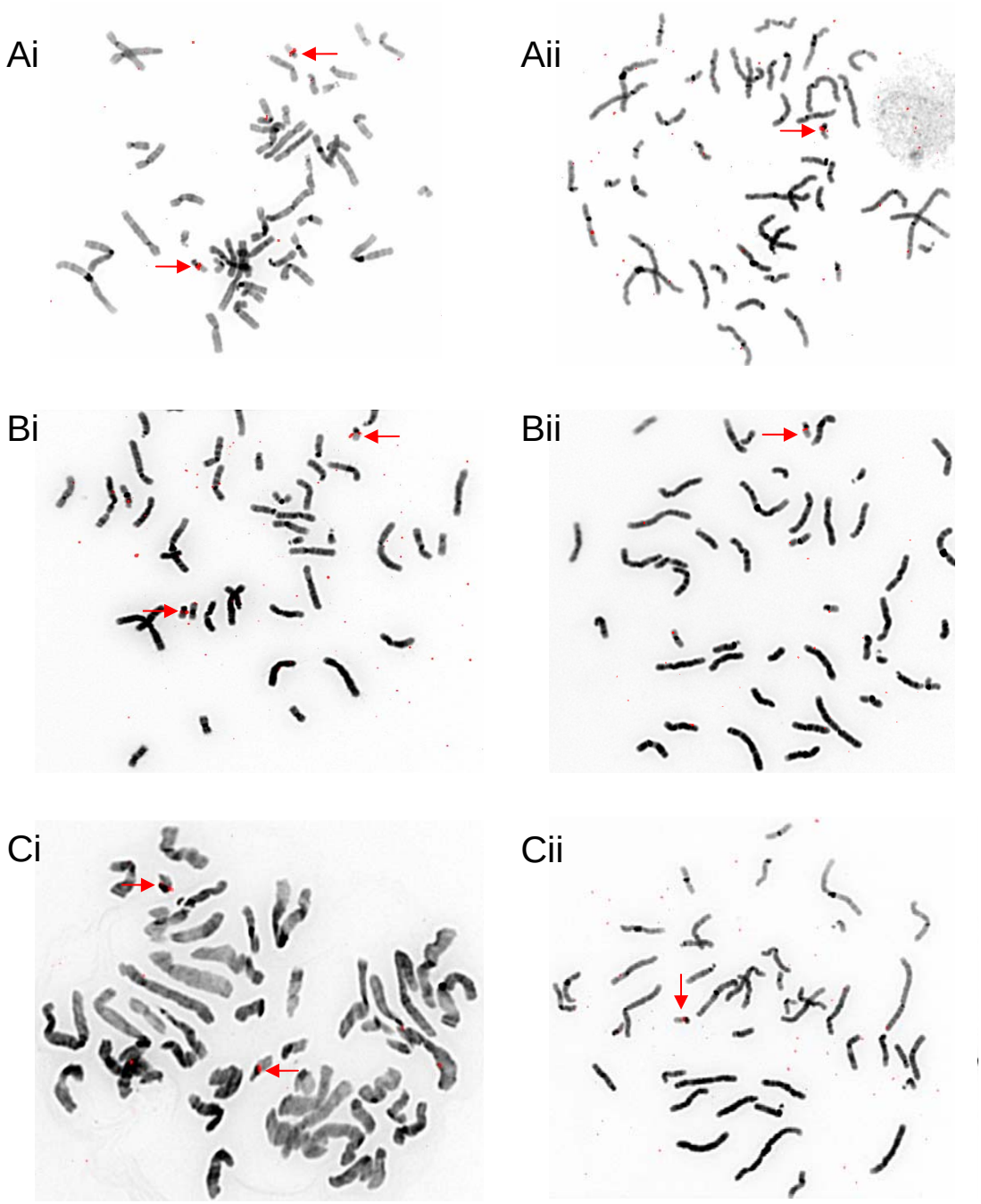
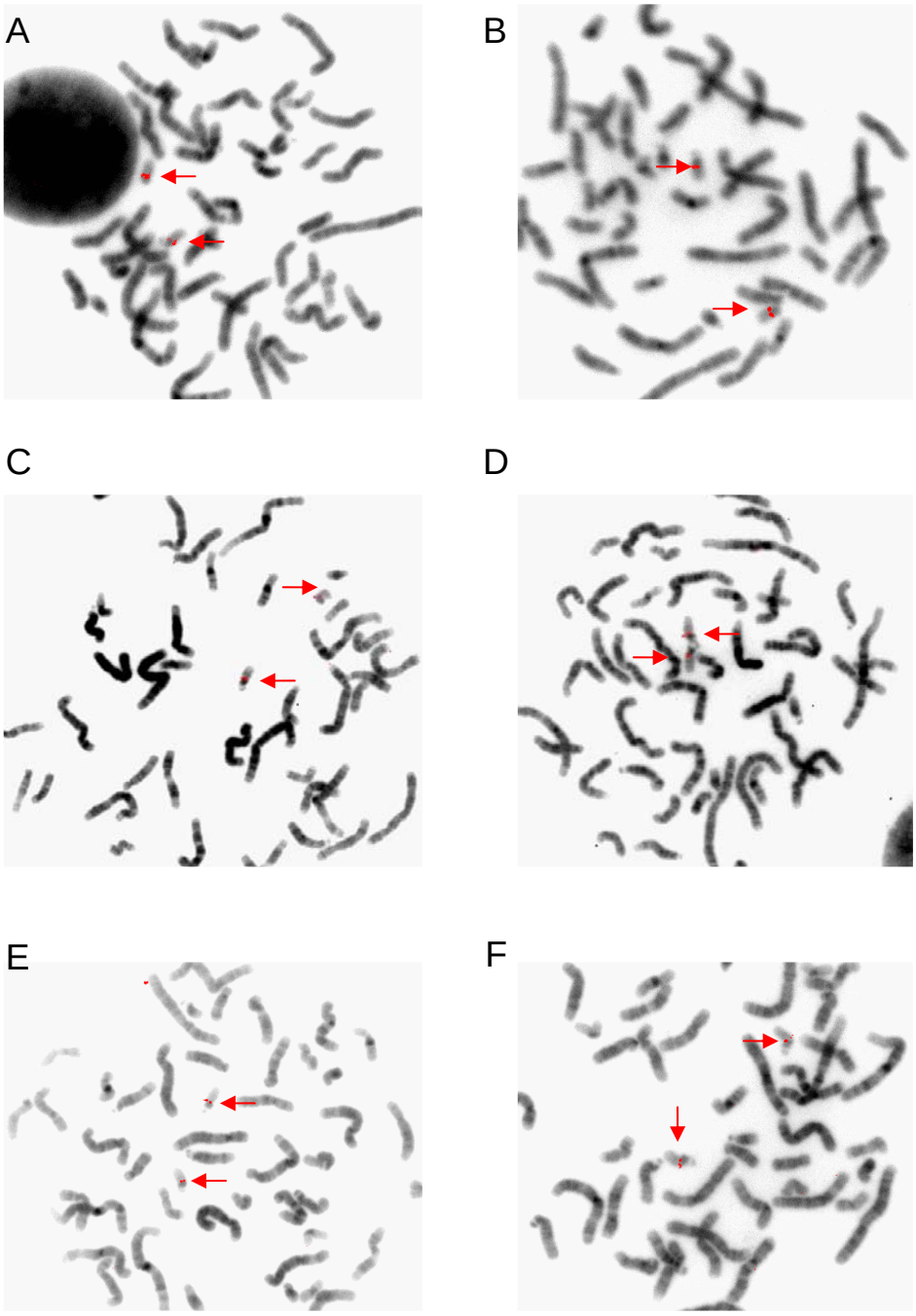


Figure 7.12: FISH analysis of the region that the array indicated is not deleted on patient 5. Two categories of metaphase spread from the patient are observed; (i) those that show a signal on both copies of the chromosome 22s, and (ii) those that show a signal on just one copy of chromosome 22. The probes are A: bK562F10, B: p_n5 and C: p52F6.

One hypothesis that would explain intermediate ratios observed in the DiGeorge region is the abundance of segmental duplications in 22q11. Duplications would mask the single copy deletion that is characteristic of DiGeorge syndrome, and would explain the intermediate ratios exhibited. A clone with a duplication at just one other loci within the genome would report a 3:4 ratio opposed to a 1:2 ratio. Because of this, selected clones from the DiGeorge region were mapped by fluorescence *in situ* hybridisation to normal chromosomes, and chromosomes isolated from two of the DiGeorge patients (Table 7.13 and Figure 7.13-7.15). Due to the limitation on the number of patient metaphases, experiments that were unsuccessful could not be repeated.

Table 7.3: Clones chosen for FISH analysis, and results on the patient metaphases. (A & B: clones on the edge of the DiGeorge deletion not reporting any copy number loss. C & D clones in the middle of the DiGeorge deletion not reporting any copy number loss. E-G: Clones reporting a single copy number loss on the arrays. H & I: clones reporting intermediate ratios on the array)

	Clone	**Accession no**	**FISH – Normal cell line**	**FISH – Patient 1**	**Array Ratio - 1**	**FISH – Patient 4**	**Array Ratio - 4**
A	519d21	AC008079	2x22	2x22	0.91	2x22	0.98
B	995o6	AC008132	2x22	2x22	0.85	2x22	0.89
C	Cos56c	Ac000080	2x22	NA	0.75	1x22	0.97
D	Bac32	Ac007050	2x22	1x22	0.87	2x22	1.00
E	49c12	Ac000079	2x22	1x22	0.53	1x22	0.60
F	98c4	Ac000092	2x22	NA	0.56	NA	0.64
G	52f6	Ac005500	2x22	NA	0.66	1x22	0.78
H	Pn_5	Ac002472	2x22	1x22	0.76	1x22	0.86
I	83c5	Ac000087	2x22	1x22	0.72	1x22	0.84


A
B
C
D
E
F

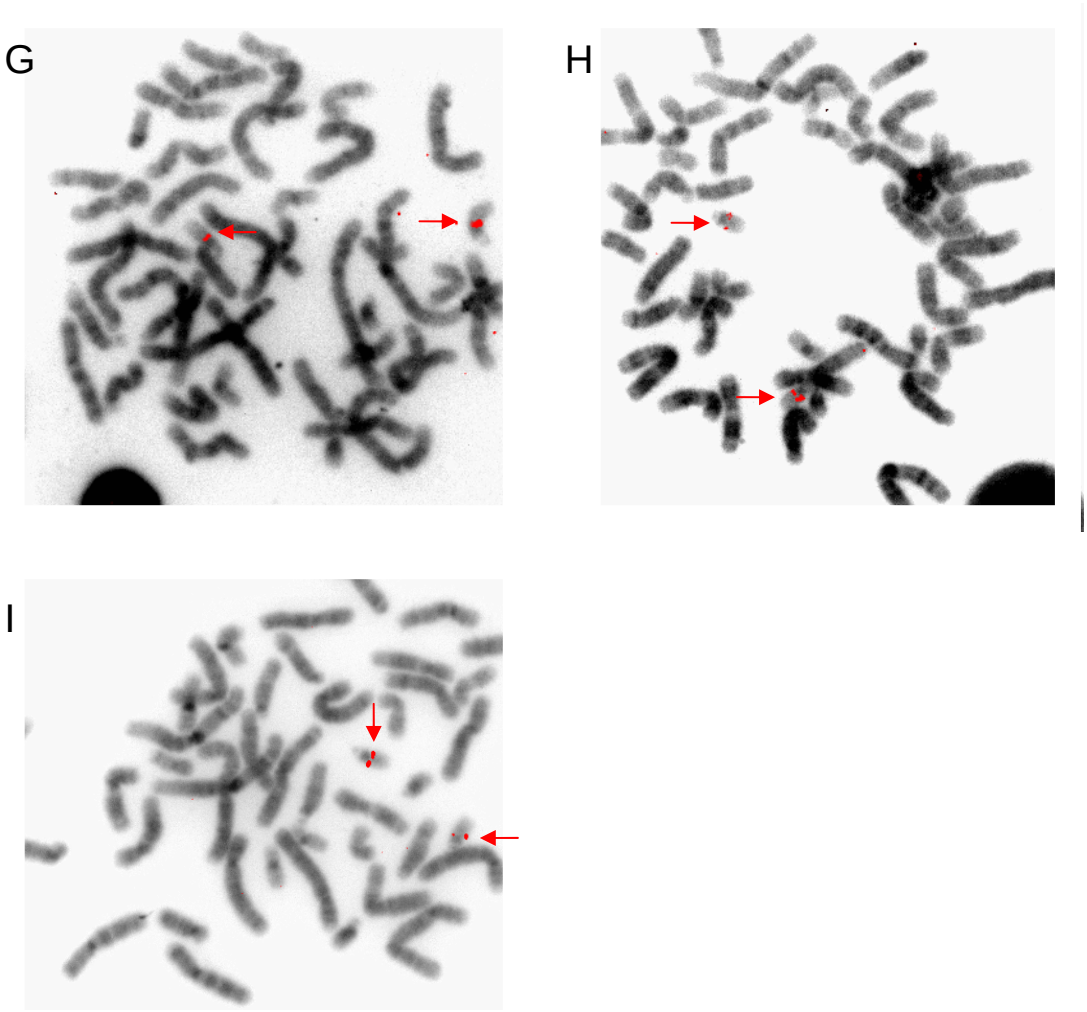


Figure 7.13: Probes hybridised to chromosomes prepared from a normal (46, XY) lymphoblastoid cell line. Lettered images relate to the probes described in Table 7.3.

The hybridisation of the probes to normal metaphase chromosomes showed no secondary signals that may indicate segmental duplications, although all the clones that were examined by FISH analysis have previously been shown to contain segmental duplications elsewhere on 22q (Buckley, Mantripragada et al. 2002). However the resolution of metaphase FISH would not enable intrachromosomal repeats elsewhere on chromosome 22q11 closer than 2 Mb from the FISHed clone to be resolved.

The results in the DiGeorge region indicate segmental duplications may affect the ratio reported by the arrays. However either the stringency of the FISH or the inability to resolve intrachromosomal repeats may produce disparate array and FISH results.

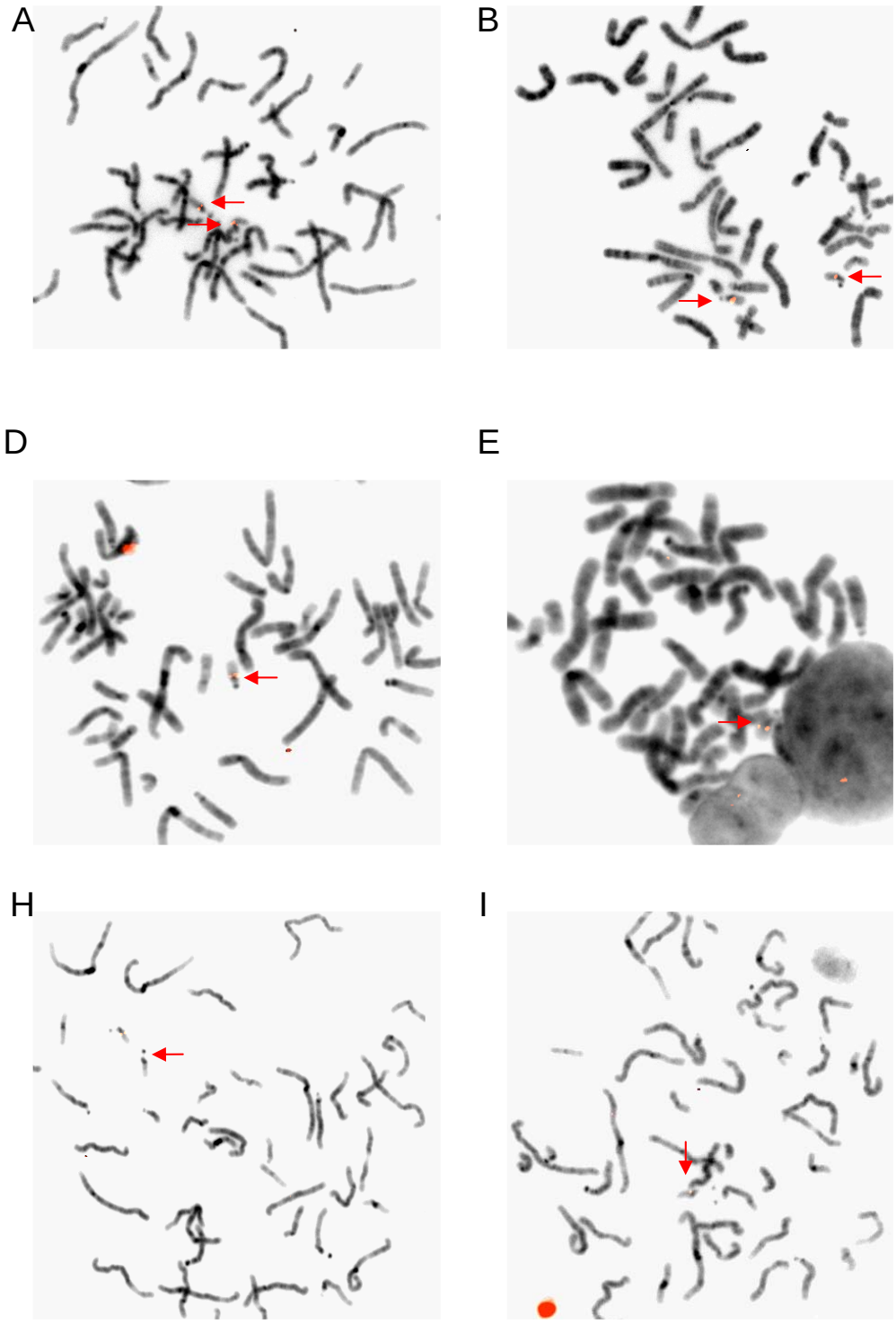
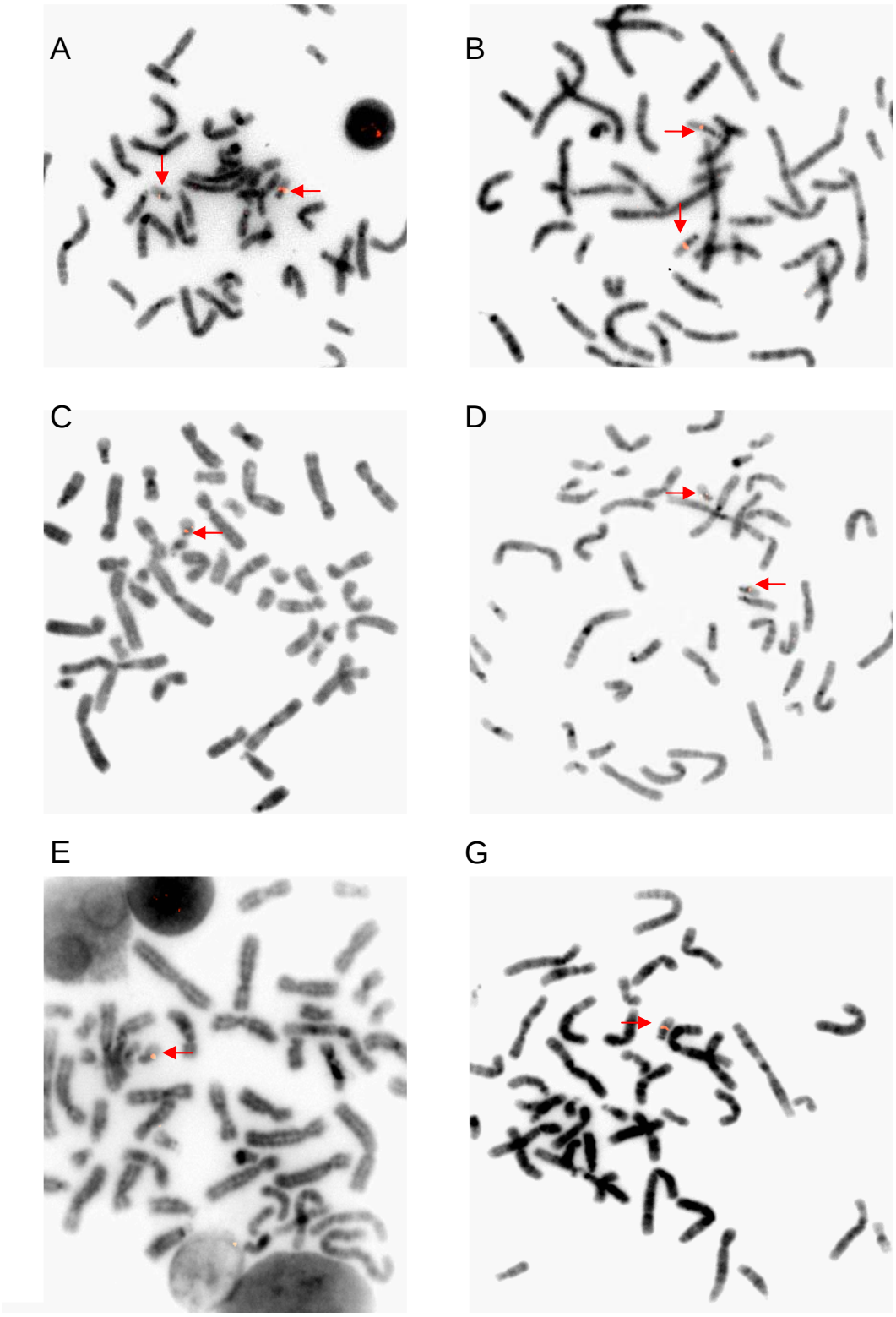


Figure 7.14: DiGeorge region probes hybridised to chromosomes isolated from patient 1. For clones used as the probe see Table 6.3 (letters correspond to clones used for hybridisation experiment.


A
B
C
D
E
G

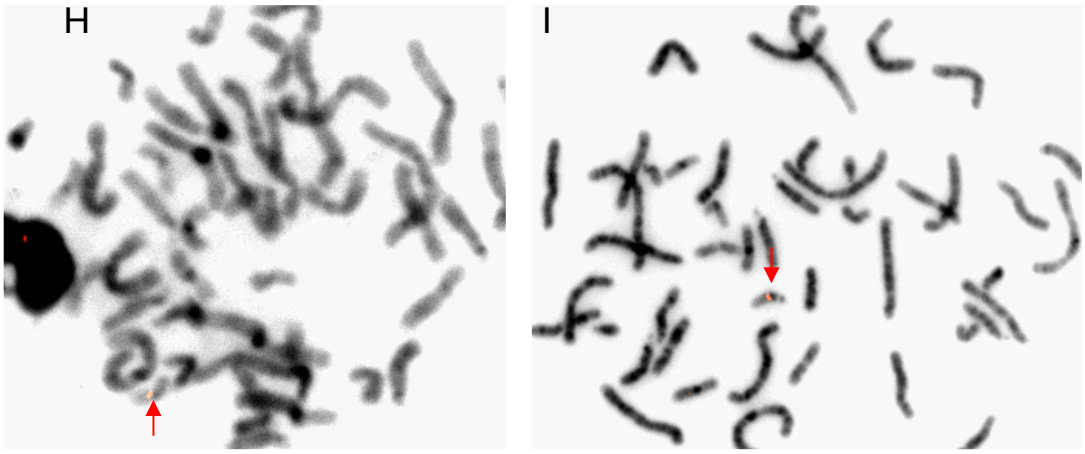


Figure 7.15: DiGeorge Region Probes hybridised to chromosomes isolated from patient 4

For clones used as the probe, see Table 7.3 (letters correspond to clones used for hybridisation experiment.

7.2.2: Assessment of patients with the DiGeorge phenotype that do not show a deletion in 22q by FISH analysis.

DNA from six patients that have aspects of the DiGeorge phenotype, but that had no deletion detected by conventional FISH, were applied to microarrays to characterise the patient DNA. Initially, patient DNA was hybridised to the 22q array to see if a deletion could be detected in the DiGeorge critical region that could not be detected by FISH. The DNA from the patient's blood was then hybridised to the 1Mb array for genome wide analysis to detect copy number changes elsewhere in the genome. The patient phenotypes are described in Table 7.4. The patient DNA samples were hybridised against pool DNA.

Table 7.4: The phenotype characteristics of patients showing some characteristics of DiGeorge syndrome, but with no 22q11 deletion when analysed by FISH.

Patient	Phenotype
1	Absent Thymus, bilateral cleft lip and pallet, tetralogy of fallot (heart defect), malformed ears, tracheoesophageal fistula, anomalous right subclavian artery, small testes, abnormal renal arteries, Arrinencphaly (absent optic tracts)
2	Facial dysmorphism, Coloboma (defect of the iris), Interrupted aortic arch, ventricular septal defect, atrial septal defect.
3	Ventricular septal defect, pulmonary atresia (obstruction of the pulmonary artery), cleft lip and palate, micropenis, undescended testes, hypoplastic scrotum, facial defects, thymus hypoplasia, deafness,
4	Nasal speech, nasal regurgitation, tetralogy of fallot, facial dysmorphism
5	Hypocalcaemia, aortic coarctation (heart defect), facial dysmorphism
6	Hypocalcaemia, interrupted aortic arch, low set ears, small mouth, interrupted aortic arch type B (heart defect)

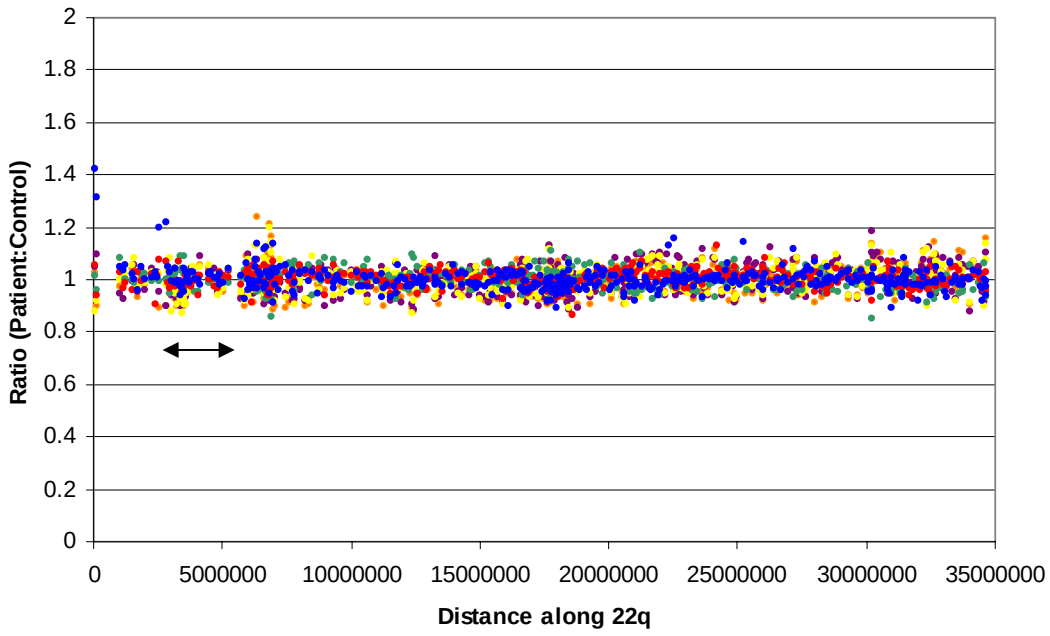


Figure 7.16: Patient:Control ratios obtained when six different patients are plotted against position on chromosome 22. Blue: Patient 1. Red: Patient 6. Green: Patient 3. Yellow: Patient 2. Purple: Patient 4, Orange: Patient 5. The DiGeorge region is indicated with a black arrow.

Patient 1 shows a copy number gain in the 2 clones adjacent to the centromere. However these two clones often show abnormal ratios (also reported in sections 4.3.3 and 7.3) so their elevation was not of note. The data was seen as being noisier at

around 6.5Mb along the q arm of chromosome 22. The standard deviation at this locus is 0.065 oppose to 0.043 along the rest of the chromosome arm. This coincides with the VJ recombination region of the immunoglobulin light chain λ region (Section 7.3). Clones bac519d21 and pac699j1 (located within the DiGeorge region) are also slightly elevated in one patient with ratios of 1.195 and 1.217 respectively.

CGH profiles of the six patients on the 1Mb array are shown in Appendix 10. Most loci on the graph that had been identified as containing gains or losses had already been identified as clones consistently reporting atypical ratios, as reported in Appendix 11. However some clones that did not consistently report atypical ratios were elevated or deleted. These are indicated in Table 7.5. No deletions were seen in the chromosome 22q clones on the 1Mb array.

Table 7.5: Clones showing amplification or deletion on the DiGeorge phenotype patients when analysed on the 1Mb array.

Patient	**Clones with a ratio >1.2**	**Clones with Ratio <0.8**
1	RP11-537N4 (Chr 19)	None
2	None	None
3	None	None
4	RP11-537N4 (Chr 19) RP11-383B4 (Chr 10)	CTD-2022G9 (Chr 5) RP11-412L4 (Chr 5) RP11-506H20 (Chr 5)
5	RP3-432E18 (Chr 12)	RP1-24K19 (Chr 21)
6	RP4-679K16 (Chr 1)	None

The chromosome 19 clone RP11-537N4 shows ratios elevated above 1.2 in two of the patients when hybridized against a female pool control. However this clone was also identified to contain segmental duplications, with interchromsomal duplications on chromosomes 11, 6 and 2 (Bailey, Gu et al. 2002), making the results obtained for this clone difficult to interpret.

Patient 4 shows a deletion across 3-4Mb of chromosome 5 (Figure 7.17). Three clones clearly showed deleted ratios (Table 7.6). One clone, RP11-92M7, proximal to the three deleted clones also shows a slightly reduced. This deletion was investigated further by our collaborators as discussed in section 7.4

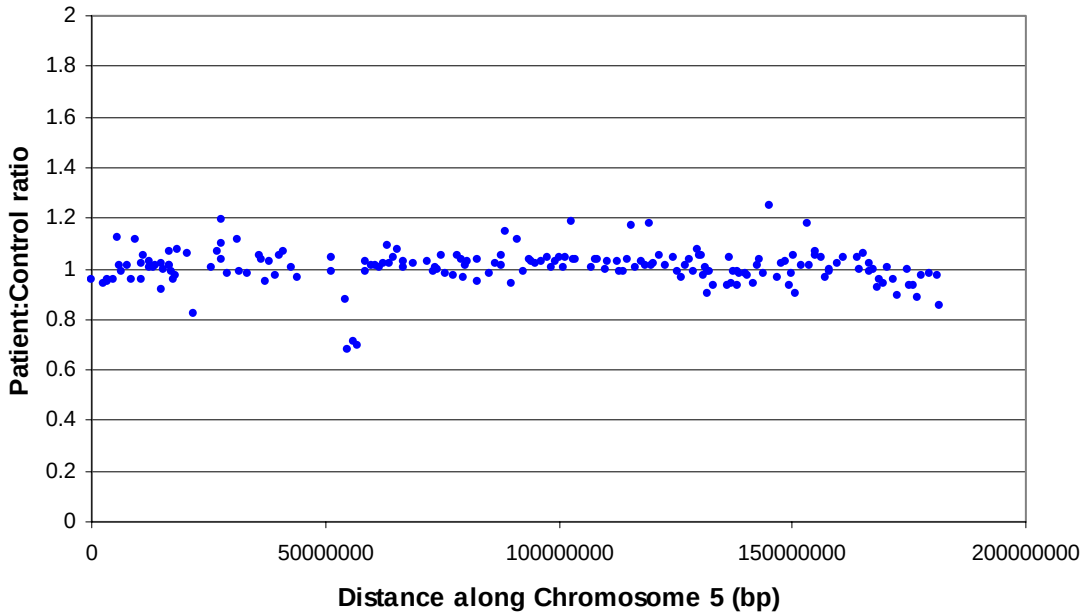


Figure 7.17: Deletion detected in chromosome 5 of patient 4 on the 1Mb array.

Table 7.6: The chromosome 5 clones deleted in patient 4

Clone	Chromosome	Position	Patient: Control
CTD-2022G9	5	54753069	0.68
RP11-506H20	5	56074899	0.71
RP11-412L4	5	57066879	0.70

The other gains and losses are also being investigated by our collaborators. Two of the six patients showed no gains or losses or deletions. This could be due to the fact that any deletions are not detected using an array of a 1Mb resolution or that the phenotype is not due to a DNA copy number change. Epigenetic changes in the genome may lead to the phenotypic effects observed. These would not be detected by the arrays.

7.3: Assessment of VJ recombination of the Immunoglobulin light chain λ using the 22q tile path array

The immunoglobulin light chain λ genes are located at approximately 6.5Mb along the q arm of chromosome 22. As lymphoblastoid cell lines are derived from differentiated B cells and are generally clonal, the immunoglobulin light chain λ genes will have undergone VJ recombination in these cells. As this leads to the

excision of DNA between the variable and the joining regions, clones on the arrays in this region will report these losses. Such changes were detected on the male: female control experiments reported in section 4.3.3. The VJ recombination in five different lymphoblastoid cell lines was assessed by hybridising DNA from the cell line against the pooled DNA from twenty anonymous blood donor samples (Fig 7.18).

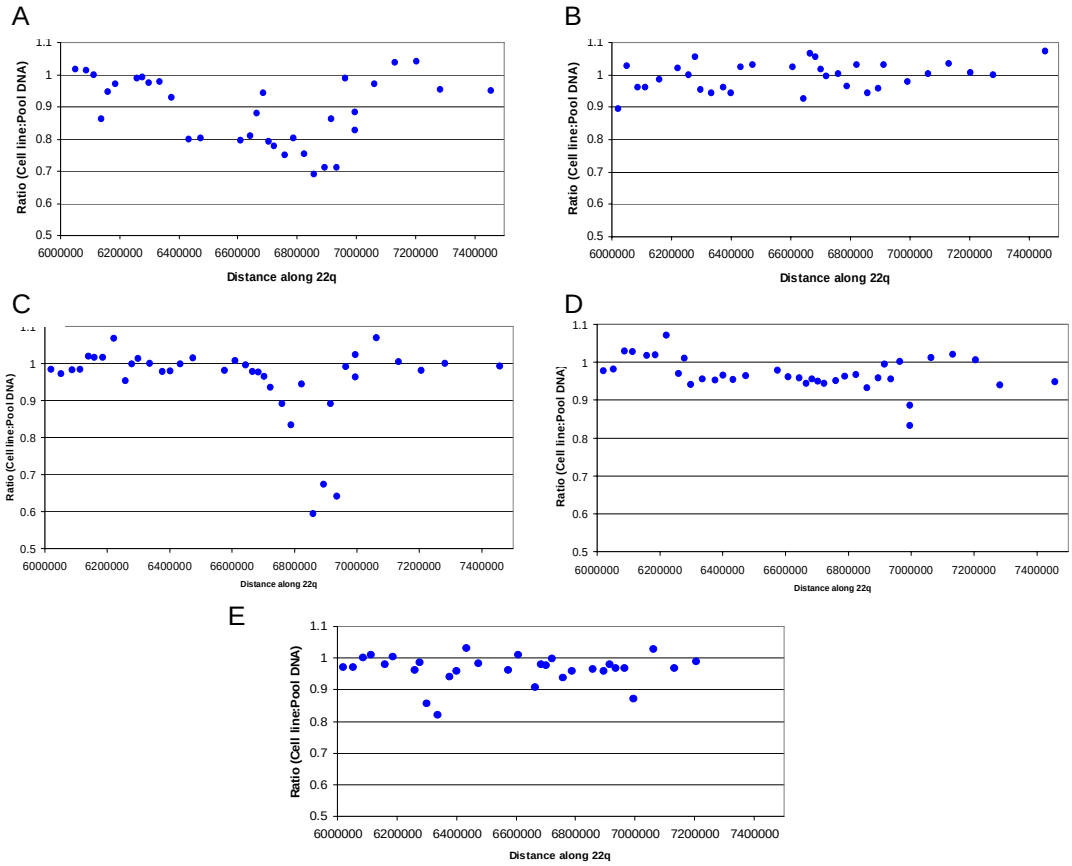


Figure 7.18: DNA from five lymphoblastoid cell lines with a normal karyotype were hybridised against DNA from a pool of 20 individuals A: Cell line HRC 575, B: Cell line HRC 146, C: Cell line HRC 159, D: Cell line HRC 160, E: Cell line HRC 196.

A deletion, defined using the <0.8 criteria described in 6.2.1, was seen in two out of five normal cell lines. The boundary of the deletion cannot be accurately defined. This is due to the large amount of segmental duplication within this region. The deletion in the HRC 575 cell line included clones cN22A12 – cN75C12 (midpoints 6433944-6995343) and represents a deletion of approximately 561Kb.

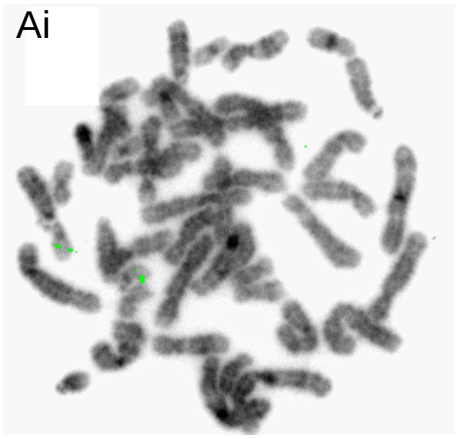
Comparison with the published map of the immunoglobulin light chain λ region allows determination of which constant and variable regions may be involved in the

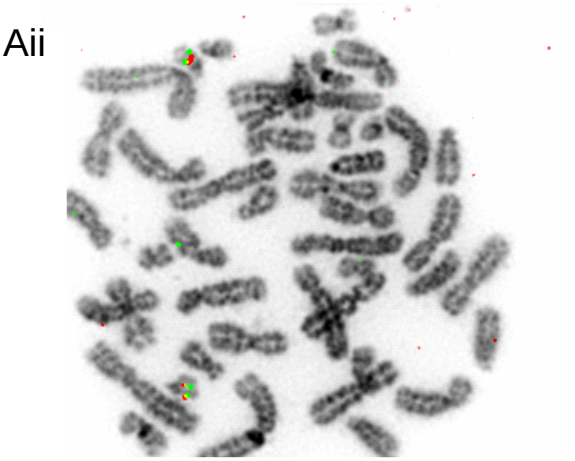
rearrangement (Kawasaki, Minoshima et al. 1997). The clone cN75C12 contains half the genes that encode the constant region of the IgL λ locus. It can therefore be determined that the constant region used is IgLC 3-7 (although IgLC 4,5, and 6 are known to be pseudogenes). The clone cN22A12 contains the IgLV genes 7-46, 5-45, 1-44 and 7-43. All IgLV genes more telomeric than this have been deleted. There was also a deletion in the cell line HRC 159. This deletion was smaller and covered approximately 76Kb between clones cN92H4 and cN9G6 (midpoints 6859162-6935464). The deletion mapped to between IgLC1-3 and IgLV 3-7. Again these clones contain a significant amount of segmental duplications so defining the exact size of the deletion was difficult. The other three cell lines show no deletion at this region, although there is slightly more background variation at the immunoglobulin light chain λ locus.

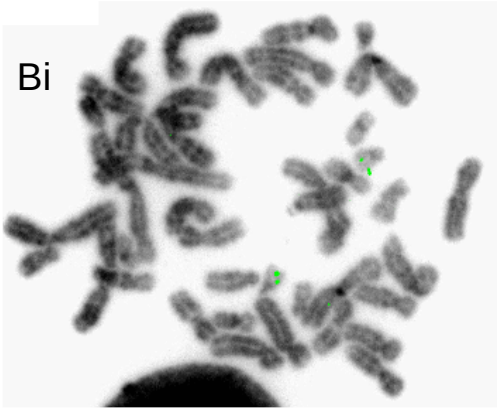
The deletion status at the immunoglobulin light chain λ locus was confirmed for two cell lines by FISH. HRC 575 showed a deletion, whereas the cell line HRC 160 did not. Biotin labelled FISH probes were made from the DNA from the same clones that were spotted onto the array and are shown in Table 7.7. A digoxygenin labelled control probe (bK57G9) from the non-deleted region of chromosome 22 was used to aid identification of chromosome 22. Selected images can be seen in Figure 7.19.

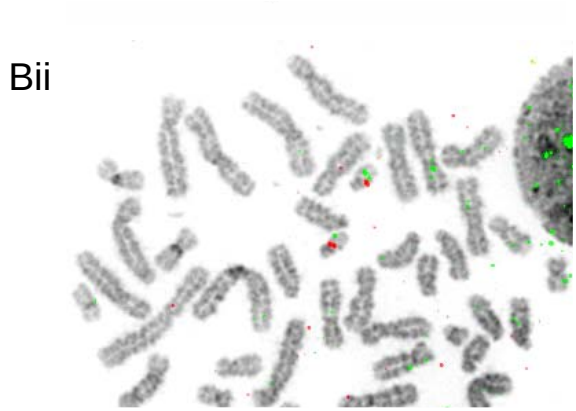
Table 7.7: Clones from the immunoglobulin light chain λ locus hybridised to metaphases from two different lymphoblastoid cell lines.

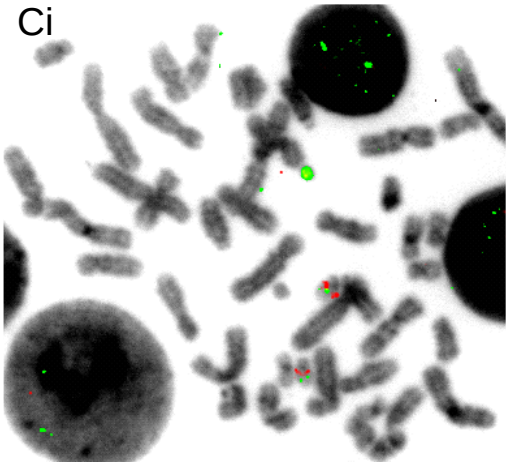
Clone	Accession no.	Signal on HRC 575	Signal on HRC 160
cN22A12	D86999	No	Yes
cN35B9	D87010	No	Yes
cN50D10	D87011	No	Yes
cN63E9	D87013	-	Yes
cN61E11	D87014	-	Yes
cN31F3	D87002	No	Yes
cN52F2	D87006	No	Yes
cN102D1	D86994	No	Yes
cN48A11	D87007	No	Yes
cN24A12	D86998	No	Yes
cN68D6	D87015	No	Yes
cN92H4	D87024	No	No
cN84E4	D87021	No	Yes
cN9C5	D87023	No	Yes
cN9G6	D87020	No	Yes
cN75C12	D87017	No	Yes
cN81C12	AP000360	Yes	Yes

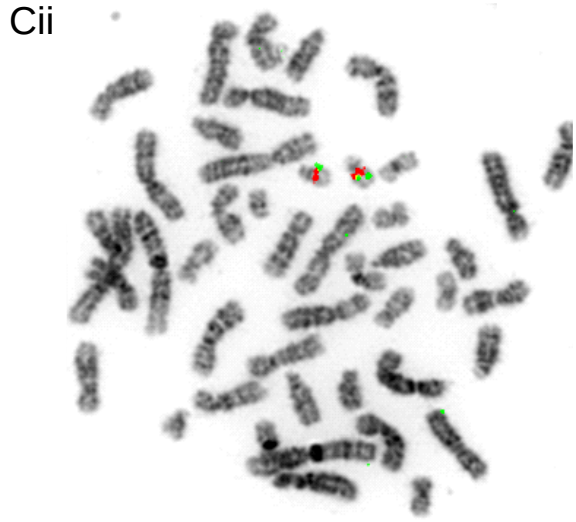

Ai


Aii


Bi


Bii


Ci


Cii

Di Dii

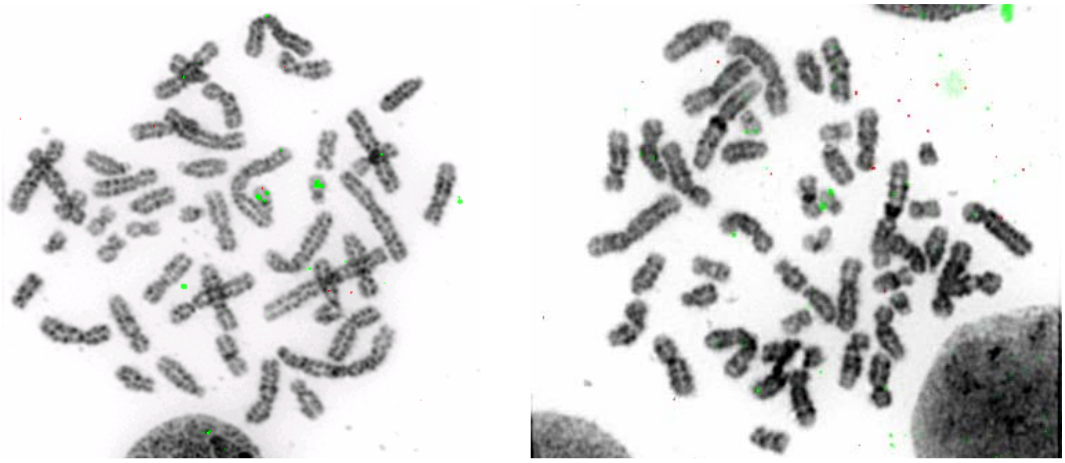


Figure 7.19: Fluorescence *in situ* hybridisation of clones from the immunoglobulin light chain λ locus (red) and a control probe (bK57G9 – green) to metaphases from the cell lines HRC 575 (i) and HRC 160 (ii). Letters relate to the clone used as the probe as reported in Table 7.8.

Table 7.8: Results from FISH experiments performed with clones from the VJ recombination region hybridised to metaphases from two different lymphoblastoid cell lines.

Hybridisation	**Clone**	**Result on HRC 575**	**Result on HRC 160**
A	cN68D6	deleted	present
B	cN75C12	deleted	present
C	cN81C12	present	present
D	cN92H4	deleted	deleted

The FISH analysis of HRC 575 cell line confirmed a deletion between cN68D6, and cN75C12, while no rearrangement was found in the HRC 160 cell line. cN81C12, was identified by arrays as being retained and distal to the HRC 575 deletion and was shown by FISH to be present in both cell lines. cN92H4 was found to be absent in both cell lines, despite showing a (non-deleted, although still reduced) ratio of 0.93 on the HRC 160 array.

7.4: Discussion

7.4.1: Segmental Duplications and the DiGeorge region.

A copy number loss was found in all the patients that had their DiGeorge status confirmed by FISH. However the reduced ratio observed for many clones rarely reached the 0.5:1 ratio that would indicate a full single copy number loss and deletion of one allele. This could be due to one of two explanations; either the clone is not fully deleted, or the clone DNA is cross hybridising with another region of the genome that is not deleted.

Examining the first possibility, the arrays have been shown to be quantitative (Fiegler, Gribble et al. 2003) and a deletion of only half a clone would report an intermediate ratio on the array. However it is unlikely that this is the reason for all the intermediate ratios seen as clones in the middle of the deleted region are affected. FISH analysis of some of these clones using metaphases from patients has shown that they are fully deleted on one copy of chromosome 22.

The second hypothesis is that non-deleted DNA from other regions of the genome are cross hybridising to the DNA on the microarray from the DiGeorge region. There is an abundance of segmental duplications in the 22q11 region of chromosome 22 (Figure 7.2) (Dunham, Shimizu et al. 1999; Bailey, Yavor et al. 2001; Bailey, Yavor et al. 2002). These repeated regions of the genome include intrachromosomal duplications, which exhibit homology to regions elsewhere on chromosome 22 and interchromosomal duplications which show homology to DNA sequences on other chromosomes. Duplications at the DiGeorge region can be seen in Figure 7.20.

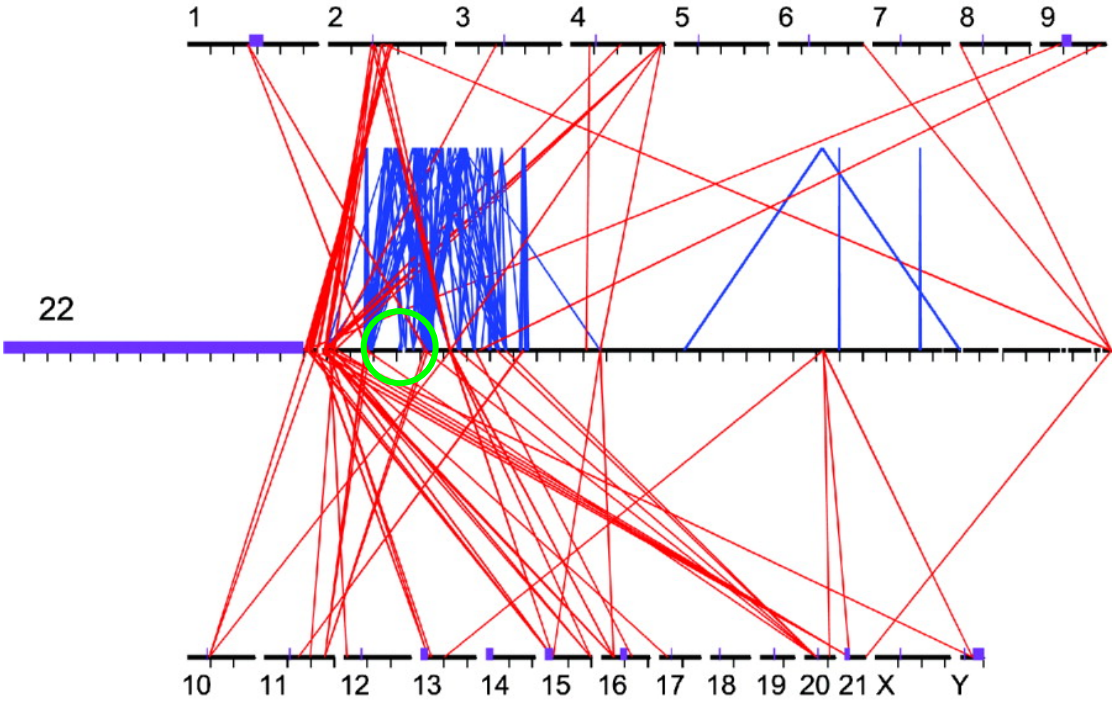


Figure 7.20: Segmental duplications on chromosome 22. Blue: Intrachromosomal deletions. Red: Interchromosomal deletions. DiGeorge region is indicated in green. Figure from (Bailey, Yavor et al. 2002).

In the experiments performed, whole genomic DNA from the DiGeorge patients is hybridised to the array. DNA that is not part of the deleted region, but has a high homology to clones within the deleted region will hybridise to these loci on the array. This would mask the single copy deletion that is characteristic of DiGeorge syndrome, and would explain the intermediate ratios exhibited.

Bioinformatic analysis of the clones in the DiGeorge region showed that many of these clones contained segmental duplications (Bailey, Yavor et al. 2002). Clones in the DiGeorge region such as pac699j1, pac995o6, bac519d21 and bac32 contain duplications in many other locations on 22q11. However, other clones in the DiGeorge region such as the cosmids 18c3, 111f11 and 119F4 contained no duplications and yet did not show a full single copy deletion on the array. However the bioinformatics approach to the detection of duplications may not be sufficient to find all regions of homology and is likely to underestimate the true amount of duplication (Eichler 2001). Most clones within the DiGeorge region contain a segmental duplication.

The chromosome 22 add-in experiment described in section 4.6 shows that the DNA from chromosome 22 clones spotted onto the array does not always show a full copy number change when an extra copy of chromosome 22 is added into the hybridisation mix. Cross hybridisation with other regions of the genome would mask the copy number change that occurs when an extra copy of chromosome 22 is added. If this is the case, then there should be a correlation between the slopes obtained from the add-in experiments and the ratios reported in the DiGeorge experiment. To test if this was so, the DiGeorge ratios reported were plotted against the slope obtained in the chromosome 22 add-in experiments, for clones in the DiGeorge region (Figure 7.21)

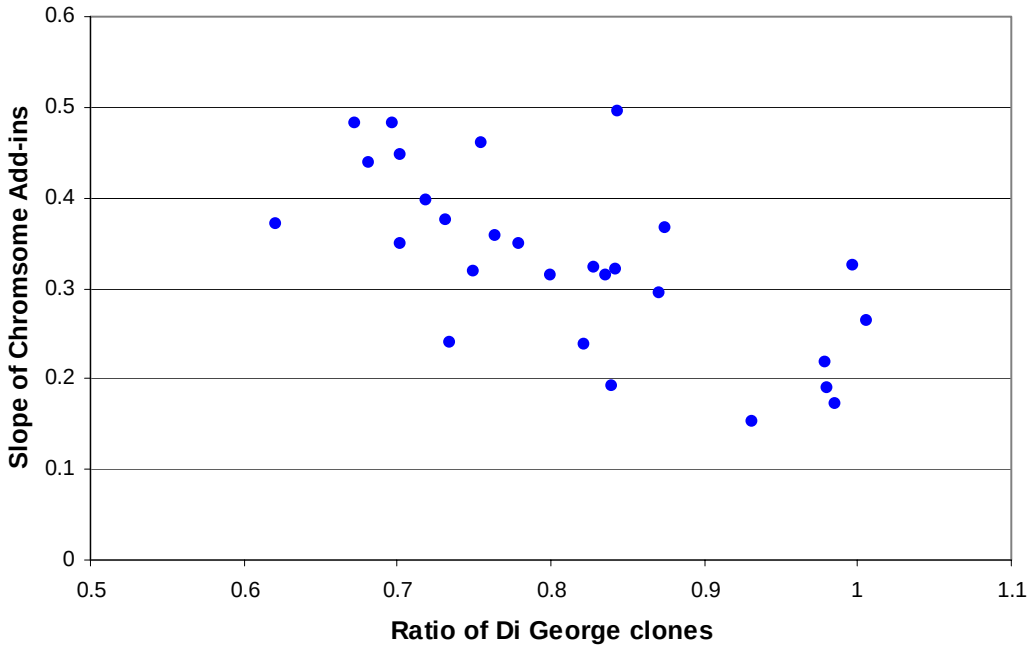


Figure 7.21: Correlation between DiGeorge ratios reported and the slope obtained from the chromosome 22 add-in experiments, for the clones in the DiGeorge region.

There is a negative correlation (regression coefficient 0.66) between the slope and the DiGeorge ratio. Clones that reported the largest slope (i.e. responded best to the add-in experiment) also reported the lowest DiGeorge ratio (those most consistent with a single copy deletion). Conversely, those clones that show a suppressed response to the chromosome 22 add-in experiments are those that show an incomplete single copy deletion ratio when hybridised with DiGeorge DNA. However the chromosome 22add-in experiment will not report the effect of intrachromosomal deletions. This analysis shows that the reason for the suppressed DiGeorge ratio is due to a

characteristic of the clone, and not due to differences in copy number within the deleted DiGeorge region.

It might be expected that the FISH experiments utilising clones with segmental duplications would show the regions with homology as secondary FISH signals. However, this was not seen, but it is unclear how the hybridisation kinetics of the arrays relate to the hybridisation kinetics of FISH. The DNA involved in the hybridisation is of different complexities, there are different relative amounts of Cot 1 present and different washing stringencies are used. Duplications would also have to be at least 2Mb apart to be resolved by metaphase FISH. As seen in Figure 7.20, most of the intrachromosomal duplications at the DiGeorge locus are not more than 2Mb apart. It is therefore unsurprising that segmental duplications have different consequences for arrays and metaphase FISH experiments.

Many of the chromosomal micro-deletion syndromes occur within regions of segmental duplication. It is therefore likely that underestimation of a full single copy loss on arrays will not be a characteristic unique to the DiGeorge region. By understanding how the duplications affect the arrays, their effects could be subtracted from the ratios obtained. This can be achieved in one of two ways;

Firstly, the utilisation of degenerate oligonucleotide primers to amplify the clone DNA ensures that full coverage of the clone DNA is represented on the array, including any segmental duplication contained within the clone DNA. To avoid the effect of segmental duplication on the ratios obtained on the array, the whole segmental duplication can be removed from the array by using an alternative strategy to amplify the clone DNA. Specifically designed PCR primers can be used to amplify all unique sequences within a clone, without amplifying segmental duplications or common repeat elements (Buckley, Mantripragada et al. 2002). In this way, segmental duplications would not be present on the array, and therefore would have no influence on the ratios reported. In addition, the removal of repeat elements would also improve the quantitation of DNA ratios reported by arrays. A secondary advantage may be a reduction in the amount of Cot 1 needed in the hybridisation mix, therefore reducing hybridisation costs. The drawback of this approach is that the design and production of individual PCR primers used for the amplification of the clone DNA is much more

expensive than using a universal primer to amplify all clone DNA. The removal of segmental duplications from the array also means the whole genome will not be covered. As many microdeletions and chromosomal breakpoints occur within segmental duplication regions, the removal of duplicate regions from arrays may limit their use in investigating these phenomena.

A second way of removing the effect of segmental duplications on the array ratio is to address the problem using a bioinformatic approach. Since the publication of the draft human genome sequence, work has been underway to map segmentally duplicated regions in the human genome. This has been achieved by identifying sequences that are over-represented in the Celera shotgun sequence and mapping them back against the draft sequence, or by repeat-masking regions of the sequence (to remove common repeat elements) and performing a global BLAST comparison with the rest of the genome (Bailey, Yavor et al. 2001; Bailey, Gu et al. 2002). By correctly identifying the amount of segmental duplication present in each clone, the number of duplications present elsewhere in the genome, and the degree of homology required for cross hybridisation, it should be possible to predict what ratio a clone containing a segmental duplication will produce on the array if a deletion or amplification is present. However the arrays will not be as sensitive when detecting single copy number changes. Currently, a problem with this analysis is that much of the genome sequence is still present in a draft form. Misalignment of the genome at duplicated regions in the draft sequence (as reviewed by Eichler (Eichler 2001)) will underestimate the amount of segmental duplications in the genome, and so make any correlation between reported array ratio and duplication inaccurate. Once all chromosomes have been sequenced to a ‘finished’ status these misalignments will be minimal and correlations between segmental duplication and the ratio reported by arrays should be possible. Work is currently underway in our group to correlate repeat content with ratios reported by the arrays. In summary the chromosome 22q tile path array can be used to detect DiGeorge deletions. However the presence of segmental duplications can make interpretation difficult and their presence should be taken into account when analysing these arrays.

7.4.2: Analysis of Patients showing the DiGeorge phenotype with no 22q11 deletion.

DNA from patients showing a DiGeorge phenotype but with no deletion in the DiGeorge critical region were analysed on the chromosome 22q tile path and the 1Mb arrays. Results from the 22q tile path arrays show that there was no deletion seen in any of these patients at the DiGeorge critical region. The DNA from all six patients was applied to the 1Mb array. One patient (patient 4) showed a 3-4Mb single copy deletion at 5q11.2. This patient's karyotype had previously been examined cytogenetically and no deletion had been detected on chromosome 5. This demonstrates that the arrays are more sensitive at detection of deletion than conventional cytogenetic methods.

Further studies were performed at the Institute of Child Health to confirm the results obtained by the array. Seven microsatalite markers analysed across the region were found to be homozygous, supporting the observation that this region has a single copy deletion. FISH analysis using selected clones confirmed the deletion (Table 7.9).

Table 7.9: Results from FISH experiments performed on metaphase chromosomes from patient 4.

Clone	**Chromosome**	**Position**	**Array ratio**	**FISH Results**	**FISH Comments**
RP11-497H16	5	21747447	0.82	present	Cross hybridises elsewhere on 5p and 5q
RP11-269M20	5	51407665	0.99	present	Secondary on chr 1
RP11-92M7	5	54387267	0.87	deleted	-
RP11-506H20	5	56074899	0.71	deleted	-
RP4-572A3	5	58515147	1.02	present	-

7.4.3: Analysis of the Immunoglobulin light chain λ recombination region.

During B cell development the loci encoding the immunoglobulin light chain (IgL) undergo rearrangement to produce antibody diversity. The IgL has 2 different classes;

κ and λ. The κ chain is encoded at 2p11.2 and the λ chain is encoded at 22q11. The rearrangement can occur on either allele at the κ of λ loci. Production of a functional protein initiates a feedback mechanism once a successful rearrangement has occurred; the other alleles are epigenetically silenced and not rearranged (Gorman and Alt 1998).

In Figure 7.11, a single copy deletion at 22q11 can be seen in two of the five cell lines analysed (HRC 575 and HRC 160) reflecting that the IgL λ has undergone rearrangement in these cell lines. The two rearranged cell lines studied show deletions of different sizes. HRC 575 has a large deletion of approx 560Kb. HRC 159 has a smaller 76Kb deletion. The clones involved in the deletion contain segmental duplications and so show incomplete reduction in ratios from those expected for a single copy loss. This makes defining the boundaries of the deletion difficult. However it can be seen that the two deleted cell lines do not share proximal or distal breakpoints and therefore it is likely that different V and J segments have been fused during recombination in these cell lines. Comparison of the deleted region with the sequence map (Kawasaki, Minoshima et al. 1997) indicates which constant and variable regions are involved in IgL λ rearrangements.

The other three cell lines do not show any rearrangement in the immunoglobulin light chain λ locus. This could be due to several different reasons. Firstly, the rearrangement and associated deletion may be too small to detect on the tile path array. Secondly, during B cell development it is the immunoglobulin heavy chain (IgH) on chromosome 14 that undergoes rearrangement first. Only one in three of these rearrangements are successful (Bassing, Swat et al. 2002). If IgH rearrangement is unsuccessful on both alleles the IgL will be prevented from rearrangement and no rearrangement will be seen at either the IgL κ or λ loci. Lastly, in humans IgL κ rearrangement occurs before IgL λ rearrangement (Nemazee and Weigert 2000). If either rearrangement at the IgL κ loci is successful, negative feedback by the transcribed IgL will prevent rearrangement at the IgL λ locus.

FISH was performed on two of the cell lines studied; one (HRC 575) showed a large deletion due to recombination of disparate V and J segments at one allele, the other

(HRC 160) showed no deletion. The FISH results, in the main, confirm the results obtained by the array analysis of the immunoglobulin light chain λ locus. The only clone that showed disparate results was cN92H4, which was absent in both cell lines when analysed by FISH. The ratio reported for this clone in the hybridisation using HRC 160 DNA (0.92) is not outside that expected due to experimental variation, and therefore it is not classed as deleted, however it is lower then the other clones in the immediate vicinity. This clone does include regions of segmental duplication. The deletion that would be reported on the array at this location may have been masked by the cross hybridisation of other regions of the genome, as described in 7.4.1.

These experiments show how the arrays can detect physiological rearrangements of the genome. Higher resolution arrays may allow the exact constant and variable genes rearranged in the B cells to be determined. Physiological rearrangements can be specific to one cell type. For example, IgL rearrangement is specific to B cells, whilst rearrangement of the T cell receptor family is specific to T cells (Turner 2001) and therefore should be taken into account when the type of DNA being used as control is decided. Knowledge of these regions means false deletions and amplifications are not detected as described in section 7.2.1.

7.4.4. Summary

This Chapter has shown how DNA microarrays can assess microdeletions at a much higher resolution than conventional cytogenetic techniques. They also give much more information about the size of the deletion than assessment of specific syndromes by commercial probes. The analysis of patient samples on a genome wide array enables detection of copy number gains and losses that might be missed if just one region of the genome was being screened.