

**Chromosome 1 Map, Sequence
and Variation**

by

Simon Gray Gregory

**Thesis submitted for the
degree of Doctor of Philosophy**

The Open University

17th February 2003

The Wellcome Trust Sanger Institute

Wellcome Trust Genome Campus

Hinxton

Cambridge, UK

*This thesis is dedicated to my wife, Deborah,
and my daughter, Olivia, for their unwavering
support, understanding, strength and love.*

*Also to my parents for their constant
encouragement and for instilling me
with a belief in my own abilities.*

Abstract

The construction of well characterised sequence-ready physical maps has been central to the generation of high quality genomic sequence by the Human Genome Project. The technological advances that made possible a clone based sequencing approach to large genomes have included the use of large insert bacterial clones and the development of high throughput fingerprinting techniques.

The first part of this thesis is devoted to development and application of these improvements in technology. The adaptation of fluorescent technologies and their application to existing fingerprinting methods described in this work has resulted in a fingerprinting technique which improves upon levels of data accuracy, increases throughput and incorporates of increased levels of safety and automation. The initial application of this and other restriction digest fingerprinting methods to the assembly of large insert P1-artificial chromosome clones (PACs) was also evaluated. PACs were used to construct a 1.4 Mb contig across a region of chromosome 13q12 that includes the breast cancer susceptibility gene *BRCA2*. These experimental and technical developments were then utilised within a hierarchical mapping strategy to construct a 13 Mb contig of human chromosome 1pcen – 1p13.

The finished sequence generated by the clone based sequencing strategy provides the basis for the elucidation of genic features and the motifs that influence their regulation within the human genome sequence. Detailed analysis of the finished genomic sequence from 1pcen – 1p13 is described. These analyses include the characterisation of base composition and

determination of repeat content within the region, as well as identification of known and novel genes by manual annotation.

The majority of differences between individuals can be attributed to allelic sequence variation. The characterisation of sequence differences and comprehension of how they may affect the expression and function of genes will be crucial for the study of molecular alterations in human disease. A subset of highly similar genes within 1pcen – 1p13, in addition to seven other genes of interest, were investigated by developing and assessing assays to determine sequence variation. The particular challenges of investigating gene families where sequences are nearly identical were explored, and enable better resolution of new and previously available data. The consequences that these sequence changes may have upon gene function is also discussed, and this provides an example of the ways in which knowledge of genomic sequence can be analysed to support new areas of structural and functional research.

Acknowledgements

I would like to thank my supervisor, David Bentley, for fulfilling a promise that he made more than 10 years ago, by seeing me through to the submission of my doctorate. I have appreciated your advice, friendship and tolerance, during not only my PhD but throughout our working relationship. I would also like to thank my second supervisor, Richard Wooster, for his guidance and many informal ‘thesis discussions’. Thanks are also extended to my long suffering office neighbour, Mark Ross, who has endured me barging into to his office with a multitude of questions and yet proffering an answer to *almost* every one.

Traditionally, a Sanger Institute PhD could not be completed without invaluable contributions from a multitude of individuals (all of whom I have aggravated at some stage). In particular, I wish to start by thanking Carol Scott for providing me with unflagging informatics support which extends beyond the work described in this thesis. I would also like to acknowledge Cordelia Langford, Nigel Carter, Laurent Baron, Andy Smith and Panos Deloukas for their help with the generation and assaying of the small insert library. Thanks to the awe-inspiring chromosome 1 sequencing team (Kirsten McLay, Becky Hall, Karen Barlow and the Sequencing Centre) for pushing my 1pcen – 1p13 clones through the sequencing pipeline and for carrying me on the C1 project! Assistance has also been gratefully received from Dave Beare, Stephen Keenan, James Gilbert, Jose Oliver and Jen Ashurst who helped me to analyse the 1pcen – 1p13 sequence. I have appreciated the help from Liz Sheridan, Jackie Bye and Liz Huckle with cDNA library screening and sequencing. Thanks also to the ExoSeq project (Andrew Bentley, Wilb Dunham, Sarah Hunt and Sarah Lindsay) for helping me to generate so much data so quickly. My thanks also go to Robert Steward, Jane Loveland and Tim Hubbard for opening my eyes to the wonderful world of proteins.

On a personal note, I would like to thank my many friends and colleagues at the Sanger whose help in large or small ways has made my molecular biological rites of passage more bearable. To Alison ‘Pod’ Coffey, Ian ‘Roo’ Barrett and Tamsin ‘Mate’ van de Ven thanks for all the proof reading, years of encouragement (Pod) and endless clichés to help me realise that I was *actually* making progress. To all the chromosome 1 mapping group members, past and present, many thanks for being a great bunch to work with and for allowing me to be a very ‘hands off’ group leader, particularly in the past 18 months. Thanks to Jane Rogers for your support, especially during the final stages of the thesis, and for accepting the same tired old excuses as to why the C1 project wasn’t progressing as quickly you’d hoped. Thanks Sean and Lisa for being good friends over the years. To Gareth, thanks for treading the long grass before me; without your calming influence, good humour and friendship I could have finished the thesis a lot earlier, but been the worse for it!

“He was calm; however, he had to be supported during the journey through the long corridors, since he planted his feet unsteadily, like a child who has just learned to walk, or as if he were about to fall through like a man who has dreamt that he is walking on water only to have a sudden doubt: but is this possible?” (Vladimir Nabokov; 1899–1977)

Thanks to my mother and father for equipping me with the tools to make this arduous journey and to my wife, Deborah, and daughter, Olivia, without whose confidence and support I would ever have arrived.

Table of Contents	page
Abstract	iii
Acknowledgements	v
Table of Contents	vii
List of Figures	xiii
List of Table	xvi
Glossary of Abbreviations	xviii
Publications	xxii
Chapter 1: Introduction	1
1.1 Mapping and sequencing model genomes	4
1.2 Mapping and sequencing the human genome	9
1.2.1 Cytogenetic mapping	9
1.2.2 Genetic mapping	11
1.2.3 Radiation hybrid mapping	12
1.2.4 Physical mapping	14
1.2.4.1 <i>YAC Maps</i>	14
1.2.4.2 <i>Bacterial Clone Maps</i>	15
1.3 Generating human genomic sequence	16
1.4 Interpreting the human genome landscape	18
1.4.1 Sequence composition	18
1.4.2 CpG island identification	19
1.4.3 Repeat content	20
1.5 Gene identification	21
1.6 Computational Genomics	25
1.6.1 <i>In silico</i> gene prediction	25
1.6.2 Sequence Analysis	27
1.7 Allelic variation	31
1.7.1 SNP discovery	32
1.7.2 Utilising SNPs	34
1.8 Chromosome 1	36
1.9 Aims of this thesis	41

Chapter 2: Materials and Methods	43
<u>Materials</u>	44
2.1 Chemical reagents	45
2.2 Enzymes and commercially prepared kits	45
2.3 Nucleotides	46
2.4 Solutions	46
2.4.1 Buffers	46
2.4.2 Electrophoresis and Filter preparation solutions	47
2.4.3 Media	48
2.4.4 DNA labelling and hybridisation solutions	49
2.4.5 General DNA preparation solutions	49
2.5 Size markers	50
2.6 Hybridisation membranes and X-ray and photographic film	50
2.7 Sources of genomic DNA	50
2.8 Bacterial clone libraries	51
2.8.1 Cosmid libraries	51
2.8.2 PAC and BAC libraries	51
2.8.3 cDNA libraries	51
2.9 Primer sequences	52
2.10 World Wide Web addresses	56
<u>Methods</u>	57
2.11 Isolation of bacterial clone DNA	57
2.11.1 Miniprep of cosmid	57
2.11.2 Microprep of cosmid, PAC and BAC DNA for restriction digest fingerprinting	58
2.11.3 Filterprep of PAC and BAC DNA for restriction digest fingerprinting	59
2.12 Bacterial clone fingerprinting	60
2.12.1 Radioactive fingerprinting	60
2.12.2 Fluorescent fingerprinting	61
2.12.3 <i>Hind</i> III fingerprinting	62

2.13	Marker preparation	63
	2.13.1 Radioactive fingerprinting	63
	2.13.2 Fluorescent fingerprinting	63
	2.13.3 <i>Hind</i> III fingerprinting	63
2.14	Gel preparation and electrophoresis	64
	2.14.1 Agarose gel preparation and electrophoresis	64
	2.14.2 Gel preparation and electrophoresis for radioactive fingerprinting	64
2.15	Construction of small insert library	65
	2.15.1 Library preparation	65
	2.15.2 Electroporation and library plating	65
2.16	Applications using the polymerase chain reaction	66
	2.16.1 Primer design	66
	2.16.2 Oligonucleotide preparation	66
	2.16.3 Amplification of genomic DNA by PCR	67
2.17	Radiolabelling of DNA probes	67
	2.17.1 Radiolabelling of PCR products	67
	2.17.2 Pre-reassociation of radiolabelled probes	68
2.18	Hybridisation of radiolabelled DNA probes	68
	2.18.1 Hybridisation of DNA probes derived from STSs	68
	2.18.2 Stripping radiolabelled probes from hybridisation filters	68
2.19	Restriction endonuclease digestion of cosmid DNA	69
2.20	Clone library screening	69
	2.20.1 cDNA library screening by PCR	69
	2.20.2 Vectorette PCR on cDNA	69
2.21	Exon Amplification	71
2.22	Mapping and sequence analysis software and databases	71
	2.22.1 IMAGE	71
	2.22.2 FPC	72
	2.22.3 lace	73
	2.22.4 BLIXEM	74
	2.22.5 RepeatMasker	74

Chapter 3: Using large insert clones to construct contigs:**The development of fluorescent fingerprinting**

3.1	Introduction	76
3.2	Large insert clones	78
	3.2.1 Application of large insert clones to restriction enzyme fingerprinting	78
	3.2.2 Validation of PAC inserts	80
3.3	Fluorescent fingerprinting	84
	3.3.1 Fluorescent labelling of cosmid and lambda DNA	85
	3.3.2 Residual dye removal	87
	3.3.3 First position labelling	90
	3.3.4 One step reaction	92
	3.3.5 DNA prep modifications	94
	3.3.6 New size standard	96
	3.3.7 Data collection and processing	99
	3.3.8 Reproducibility	99
	3.3.9 Validation of fluorescent fingerprinting	101
3.4	Discussion	105

Chapter 4: Construction of a sequence-ready bacterial clone contig of 1pcen – 1p13 108**clone contig of 1pcen – 1p13**

4.1	Introduction	109
4.2	Construction of sequence-ready map of 1pcen – 1p13	110
	4.2.1 Small insert library construction	112
	4.2.2 Hybrid mapping of SIL markers	115
	4.2.3 Bacterial clone contig construction	116
4.3	Evaluation of SIL marker distribution in chromosome 1	123
4.4	Comparisons of Published Maps	126
	4.4.1 Physical maps	126

4.4.2	Genetic map	127
4.4.3	Radiation hybrid map	128
4.4.4	A comparison of three maps	129
4.5	Discussion	131
Chapter 5: Sequence analysis of 1pcen – 1p13.2		134
5.1	Introduction	136
5.2	Sequence Composition Analysis	137
5.2.1	G-Banding	137
5.2.2	Isochores	140
5.2.3	Repeats	142
5.2.4	Low copy repeats	144
5.2.5	CpG Islands	148
5.2.6	Eponine	149
5.3	Gene Identification	149
5.3.1	Known genes	151
5.3.2	Novel genes	154
	5.3.2.1 <i>Splicing ESTs support the structure of a gene</i>	159
	5.3.2.2 <i>mRNA support of novel coding features</i>	162
5.3.3	Novel transcripts	165
5.3.4	Pseudogenes	166
5.4	Gene assessment	168
5.4.1	Alternative splicing	171
5.4.2	Genic features	173
	5.4.2.1 <i>Putative bidirectional promoters</i>	174
	5.4.2.2 <i>Overlapping genes</i>	175
5.5	Inferring function by protein homology	176
5.5.1	Identifying function through sequence homology	177
5.5.2	Identifying function by structural homology	179
5.6	Discussion	182
5.7	Appendix	187

Chapter 6: The identification and analysis of single nucleotide polymorphisms	192
6.1 Introduction	193
6.2 Gene Annotation	195
6.3 Identifying SNPs within Gene Families	195
6.4 Primer Design	199
6.5 DNA screening	201
6.6 Sequence Generation and Assembly	209
6.7 Exon coverage of sequence contigs	210
6.7.1 Validation and localisation of known SNPs	213
6.7.2 Identification of novel SNPs	215
6.8 SNP Analysis	216
6.8.1 Validating SNPs within highly homologous genes	217
6.8.2 Validating SNPs	219
6.8.2.1 <i>Known</i>	219
6.8.2.2 <i>Novel</i>	220
6.8.2.3 <i>Suspect candidate SNPs</i>	221
6.8.2.4 <i>Rejected candidate SNPs</i>	224
6.8.3 Effect of Sequence variation upon gene structure	225
6.9 Discussion	229
Chapter 7: Discussion	233
7.1 Genome Mapping and Sequencing	234
7.2 The determination of coding features	240
7.3 Assigning gene function	241
7.4 Sequence variation	245
7.5 Conclusion and future work	246
Chapter 8: References	250

List of figures:

Chapter 1	Figure 1.1 A plot of the increase in complexity of genomic sequencing.	8
	Figure 1.2 The alignment of syntenic region between human and mouse chromosomes 1.	40
Chapter 2		
Figure 2.1:	Strategy for vectorette PCR screening of cDNA libraries.	71
Chapter 3		
Figure 3.1:	A representation of the mapping of BRCA2 region.	79
Figure 3.2:	An agarose gel fingerprint of large insert bacterial clones in IMAGE.	82
Figure 3.3:	A comparison of <i>Hind</i> III fingerprint fragments and genomic sequence for the BRCA2 contig.	83
Figure 3.4:	The result of the first fluorescent fingerprinting experiment.	86
Figure 3.5:	Fluorescent fingerprint data collection using an extended run time.	87
Figure 3.6:	A comparison of labelled fragments when investigating removal.	89
Figure 3.7:	Labelling with spectrally distinct fluorophores.	91
Figure 3.8:	Testing one-step labelling.	93
Figure 3.9:	Testing one-step labelling and DNA prep protocols.	95
Figure 3.10:	A comparison of labelled lambda digest fragments using <i>Sau</i> 3A I, <i>Bsa</i> J I and <i>Taq</i> α 1 restriction enzymes.	98
Figure 3.11:	An FPC display of band labelling uniformity.	100
Figure 3.12:	A comparison between fingerprinting methods of using 14 clones comprising a minimum tiling path.	103
Chapter 4		
Figure 4.1:	A representation of the two strategies used to construct a sequence-ready bacterial clone map of 1pc – 1p13.	111
Figure 4.2:	The construction of a chromosome 1 specific small insert library.	113
Figure 4.3:	Sequence length and frequency of SILs passing STS design stage.	115

Figure 4.4:	Generation of sequence ready bacterial coverage using the hierarchal strategy.	117
Figure 4.5:	The assimilation of PAC contigs into whole genome and chromosome specific fingerprint databases.	120
Figure 4.6:	A representation of PAC and BAC contig coverage of 1pcen – 1p13.	122
Figure 4.7:	Chromosomal distribution of flow sorted markers and comparison to of radiation hybrid and physical maps.	125
Figure 4.8:	The distribution of genetic mapped markers positioned within the 1pc – 1p13 contig by hybridisation.	128
Figure 4.9:	The distribution of radiation hybrid mapped markers positioned within the 1pc – 1p13 contig by hybridisation.	129
Figure 4.10:	A comparison of marker distribution between genetic, physical and radiation hybrid maps of 1pcen – 1p13.	130

Chapter 5

Figure 5.1	The genomic characterisation of human chromosome 1pc – 1p13.	141
Figure 5.2:	Low copy repeat detected within 1pc – 1p13.	146
Figure 5.3:	An ACeDB display of two annotated genes, including coding sequences, on opposite strands of DNA.	150
Figure 5.4:	Primer combinations used to validate putative gene structures.	155
Figure 5.5:	The annotation of a novel gene from <i>de novo</i> prediction and splicing EST alignment.	161
Figure 5.6:	The annotation of a novel gene from <i>de novo</i> prediction, splicing EST and homologous mRNA alignment.	164
Figure 5.7:	The characterisation of a processed pseudogene to 1p12, the original of which localised to 1p35.	167
Figure 5.8:	Incomplete polyA primed mRNA.	170
Figure 5.9:	Splice variants of adenosine monophosphate deaminase 2 (AMPD2).	173
Figure 5.10:	Genes in genomic context.	176
Figure 5.11:	Putative assignment of structure and function of a novel gene.	178
Figure 5.12:	Identification of putative functional domain of a novel protein.	181
Figure 5.13:	Generic structure of a gene.	185

Chapter 6

Figure 6.1:	An ACeDB display of GSTM 1 – 5 and a generic GSTM gene structure.	197
Figure 6.2	A genomic sequence alignment of GSTM 1 – 5:	232
Figure 6.3:	A CEPH pedigree.	202
Figure 6.4:	Screening of CEPH DNAs with exon specific primer pairs designed to GSTM4.	203
Figure 6.5	PCR products from exon primers using CEPH DNA as template.	206
Figure 6.6:	Assembly of <i>de novo</i> exon specific sequences shown in Gap4.	210
Figure 6.7:	A summary representation of the <i>de novo</i> sequence coverage of exons from the target genes.	211
Figure 6.8:	Identification of a known SNP which is within intron 3 of GSTM1.	214
Figure 6.9:	The identification of a known T/C SNP, dbSNP: 737497.	220
Figure 6.10:	A novel A/C SNP identified within intron 3 of GSTM2.	221
Figure 6.11	The alignment of GSTM 1 – 5 coding sequence.	223
Figure 6.12:	A 3-D representation within ICMLite of the homodimeric GST model, 3LJR.	227
Figure 6.13:	A vertical cross-sectional view of glutathione conjugating amino acid residues of 3LJR.	228

List of tables:**Chapter 1**

Table 1.1:	Comparison of G-bands and R-bands.	10
Table 1.2:	A comparison of marker content within genetic maps.	12
Table 1.3:	Genes in the human genome.	27
Table 1.4:	Prediction programs used to identify gene features.	27
Table 1.5:	Sequence queries available using BLAST alignment.	28
Table 1.6:	A list of the large scale comparative organisms sequencing projects.	30
Table 1.7:	SNP totals contained within or adjacent to coding features.	35
Table 1.8:	Diseases elucidated as a result of the Sanger Institute chromosome 1 mapping and sequencing project.	37
Table 1.9:	Disease loci mapping to 1pcen -1p13.	41

Chapter 2

Table 2.1:	Clones and appropriate antibiotics.	48
Table 2.2:	cDNA libraries used.	51
Table 2.3:	Vector-specific primer used in vectorette PCR.	52
Table 2.4:	STSS designed for cDNA screening and Link PCR product synthesis.	52
Table 2.5:	Exon specific primer pairs designed to pharmacogenomic gene targets.	54

Chapter 5

Table 5.1:	Fluorescence <i>in situ</i> hybridisation data of selected bacterial clones from 1pcen – p13.	138
Table 5.2:	The breakdown of repeat content within 1pcen – 1p13.2.	143
Table 5.3:	Known genes localising to 1pcen – 1p13.	152
Table 5.4:	cDNA primary pool and link PCR screening results.	156
Table 5.5:	Minimum tile path clones and accessions from the 1pcen – 1p13.2 contig (October 2002).	187
Table 5.6:	Primer pairs designed for the validation of predicted gene structures by cDNA library screening.	189

Chapter 6

Table 6.1:	A summary of the primers designed to the exons of 12 genes for the detection of coding polymorphisms.	200
Table 6.2:	A summary of the CEPH/Utah DNA used for the generation of exon sequence.	201
Table 6.3:	A summary of exon specific PCR reaction results using CEPH DNAs 1 – 8 as a template.	207
Table 6.4:	A summary of the known and novel SNPs associated with 12 target genes.	215
Table 6.5:	Expected occurrence of transitions and transversions in genomic sequence.	216
Table 6.6:	The observed number of transitions and transversions of known and novel SNPs within the 12 target genes.	216
Table 6.7:	Exonic SNP analysis.	217
Table 6.8:	Summary of categorised GSTM 1 – 5 SNPs.	218

Chapter 7

Table 7.1:	Organisms for which genome-wide fingerprint databases have or are being constructed.	240
------------	--	-----

Glossary of Abbreviations

1ace	1 chromosome version of ACeDB
ACeDB	<i>A. C. elegans</i> database
AMPD2	adenosone monophosphate deaminase 2
<i>Alu</i> -PCR	<i>Alu</i> -element-mediated polymerase chain reaction
ATP (dATP, ddATP)	adenosine 5'-triphosphate (deoxy-, dideoxy-)
BAC	bacterial artificial chromosome
BLAST	basic local alignment search tool
BLIXEM	BLAST In an X-windows Embedded Multiple Alignment
β -ME	β -mercaptoethanol
bp	base pair
BSA	bovine serum albumin
$^{\circ}\text{C}$	degrees Celsius
CaM	calmodulin
cDNA	complementary deoxyribonucleic acid
chr	chromosome
CEPH	Centre d'Etude du Polymorphisme Humain
(c)M	(centi)Morgan
cm	centimetre
CDD	CONSERVED domain database
CpG	cytidyl phosphoguanosine dinucleotide
cR	centiRays
CTP (dCTP, ddCTP)	cytidine 5'-triphosphate (deoxy-, dideoxy-)
dbEST	database of expressed sequence tags
DNA	deoxyribonucleic acid
dNTP	2'-deoxyribonucleoside 5'-triphosphate
EDTA	ethylenediamine tetra-acetic acid
EMBL	European Molecular Biology Laboratory
EST	expressed sequence tag
FISH	fluorescence <i>in situ</i> hybridisation
FP	forward primer

FPC	FingerPrinted Contigs
g	gram
GDAP2	ganglioside-induced differentiation-associated protein 2
G banding	Geimsa banding
GDB	Genome Database
GSC	Genome Sequencing Centre, St Louis
GST (M)(T)(P)	glutathione S-transferase (mu) (theta) (pi)
GTP (dGTP, ddGTP)	guanine 5'-triphosphate (deoxy-, dideoxy-)
HapMap	haplotype block map
HGMP	Human Genome Mapping Resource Centre
HGNC	Human Genome Nomenclature Committee
HGP	Human Genome Project
H-W	Hardy-Weinberg
IHGSC	International Human Genome Sequencing Consortium
INSNPMWG	International SNP Map Working Group
kb	kilobase pairs
l	litre
LD	linkage disequilibrium
LINE	long interspersed nuclear element
LOH	loss of heterozygosity
M	molar
Mb	megabase pairs
MDS	myelodysplastic syndromes
μg	microgram
μl	microlitre
μM	micromolar
min(s)	minute(s)
mg	milligram
ml	millilitre
mm	millimetre
mM	millimolar
NCBI	National Centre for Biotechnology Information

NGFB	nerve growth factor – beta
NFE2L2	nuclear factor erythroid 2-like 2
NRAS	neuroblastoma RAS viral oncogene homolog
ng	nanogram
nm	nanometre
O/N	overnight
OD	optical density
OMIM	On-line Mendelian Inheritance in Man
ORF	open reading frame
PAC	P1-derived artificial chromosome
(e)PCR	(electronic) polymerase chain reaction
PDB	Protein Data Bank
PFAM	Protein Family
PFGE	pulsed-field gel electrophoresis
PNRC2	proline-rich nuclear receptor co-regulatory protein 2
poly(dT)	poly-deoxyribothymidyl oligonucleotide
R banding	Reverse Geimsa banding
RH	radiation hybrid
RFLP	restriction fragment length polymorphism
RNA (mRNA, rRNA, tRNA)	ribonucleic acid (messenger-, ribosomal-, transfer-)
RP	reverse primer
Rnase A	ribonuclease A
rpm	revolutions per minute
RT	room temperature
RT-PCR	reverse transcription polymerase chain reaction
SCL	stem cell leukaemia
SDS	sodium dodecyl sulphate
sec(s)	second(s)
seq	sequence
SIL	small insert library
SINE	short interspersed nuclear element
snoRNA	small nucleolar RNA

SNP	single nucleotide polymorphism
STS	sequence tagged site
TEMED	N,N,N',N'-tetramethylethylenediamine
TrEMBL	Translated EMBL
TSS	transcription start site
TSC	The SNP Consortium
TIGR	The Institute of Genome Research
Tris	tris(hydroxymethyl)aminomethane
U	unit
UCSC	University of California Santa Cruz
UNR	upstream of NRAS, gene
UTR	untranslated region
uv	ultraviolet
V	volt
v/v	volume/volume
VNTR	variable number of tandem repeats
W	watt
w/v	weight/volume
Wash U.	Washington University
WG(S)	whole genome (shotgun)
XLA	X-linked agammaglobulinaemia
YAC	yeast artificial chromosome

Publications:

Parts of the work presented in this thesis have appeared previously in the following publications which are bound at the back of this thesis:

Gregory, S.G., Howell, G.R. and Bentley, D.R. Genome Mapping by Fluorescent Fingerprinting. *Genome Research* (1997) 7: 1162-1168.

The International Human Genome Mapping Consortium. A physical map of the human genome *Nature*. 2001 Feb 15;409 (6822):934-41.

Bentley D.R., Deloukas P., Dunham A., French L., **Gregory S.G.**, Humphray S.J., Mungall A.J., Ross M.T., Carter N.P., Dunham I., *et al.* The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Nature*. 2001 Feb 15;409(6822):942-3.

Parts of the work presented in this thesis have appeared previously in the following publications which are not bound at the back of this thesis:

The International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001 Feb 15;409, 860-921.