Hum Genet (2002) 111 : 339–349
DOI 10.1007/s00439-002-0771-x

ORIGINAL INVESTIGATION

**Jung Ahn · Tae-Woong Won · Deborah E. Kaplan
Eric R. Londin · Petr Kuzmič · Joel Gelernter
Jeffrey R. Gruen**

A detailed physical map of the 6p reading disability locus, including new markers and confirmation of recombination suppression

Received: 2 January 2002 / Accepted: 23 May 2002 / Published online: 13 August 2002

Abstract Loci for several complex disorders have been genetically linked to markers located telomeric of the HLA class I region of the major histocompatibility complex on 6p21.3-22. However, this same region has been characterized by a large interval of recombination suppression with the potential to greatly complicate precise localization of these risk loci. Furthermore, a paucity of markers and physical mapping data has confounded precise localization of the boundaries of linkage and recombination suppression. In order to create a more detailed marker map of this region and define these boundaries we generated a minimal tiling pathway of BACs, PACs, and cosmids and a multiplexed panel of 29 short tandem repeat markers spanning 10 Mb. In addition to providing precise marker order and distances, the pathway and marker panel frame an inversion of recombination frequency that has distorted the resolution of linkage studies for 6p loci such as reading disability and others, and that should be accounted for in the design of future studies.

J. Ahn
Department of Genome Research, DNA Research Institute, Bioneer Corporation, Cheongwon-Kun, Chungbuk 363-813, Republic of Korea

T.-W. Won
Biotechnology Department, Won International Patent & Law Firm, Seoul, Korea

D.E. Kaplan · E.R. Londin · J.R. Gruen (✉)
Yale Child Health Research Center, Department of Pediatrics, Yale University School of Medicine,
464 Congress Avenue, New Haven, CT 06520-8081, USA
e-mail: jeffrey.gruen@yale.edu,
Tel.: +1-203-7372202, Fax: +1-203-7375972

P. Kuzmič
BioKin Ltd,
1652 S. Grand Avenue, Suite 337, Pullman, WA 99163, USA

J. Gelernter
Department of Psychiatry, Yale University School of Medicine, Veterans Administration Hospital,
West Haven, Connecticut, USA

Introduction

Many diseases and complex traits have been genetically linked to the major histocompatibility complex (MHC) on 6p21.3, including loci for schizophrenia (Maziade et al. 2001; Wei and Hemmings 2000; Wright et al. 2001), asthma (Immervoll et al. 2001), and hypertension (Vidan-Jeras et al. 2000). Increasingly, large insert clone contig maps and more regional markers have localized traits to broad areas even within the MHC, including telomeric to the traditional human leukocyte antigen (HLA) class I region, corresponding to 6p21.3-6p22 of the cytogenetic map. Using these resources, a gene for hemochromatosis (HFE) (Feder et al. 1996) and risk loci for Behçet's disease (Gül et al. 2001), inflammatory bowel disease (IBD3) (Cho 2000; Dechairo et al. 2001), hypotrichosis simplex (HSS) (Betz et al. 2000), insulin dependent diabetes mellitus (*IDDM*) (Lie et al. 1999a, 1999b), attention deficit hyperactivity disorder (ADHD) (Barr et al. 2001), and reading disability (DYX2) (Kaplan et al. 2002), have all been mapped to this telomeric region. Yet, precise localization of these loci, as delineated by recombination breakpoints or by peaks of linkage disequilibrium, requires an even higher density of markers, and greater accuracy in intermarker distances and order than currently available through the public domain servers.

We are particularly interested in localizing the gene for reading disability (RD) on 6p21.3-22. Also known as dyslexia, RD is a common heterogeneous syndrome with a significant genetic component. Genetic linkage studies have identified a total of five RD loci on chromosomes 1, 2, 6, 15, and 18. (Fagerheim et al. 1999; Fisher et al. 2002; Nöthen et al. 1999; Rabin et al. 1993; Smith et al. 1983) Of these, 6p is the linkage most consistently reproduced in independent samples, but frequently with different regional markers (see Fig. 1). In 1994, Cardon et al. (1994) published a peak of linkage at marker *D6S105*. Subsequent studies reported linkage regions of 10–16 cM (7.7 Mb) spanning *D6S109* (pter) through *D6S306* (Grigorenko et al. 1997), 13.4 cM (16.9 Mb) spanning *D6S422* through

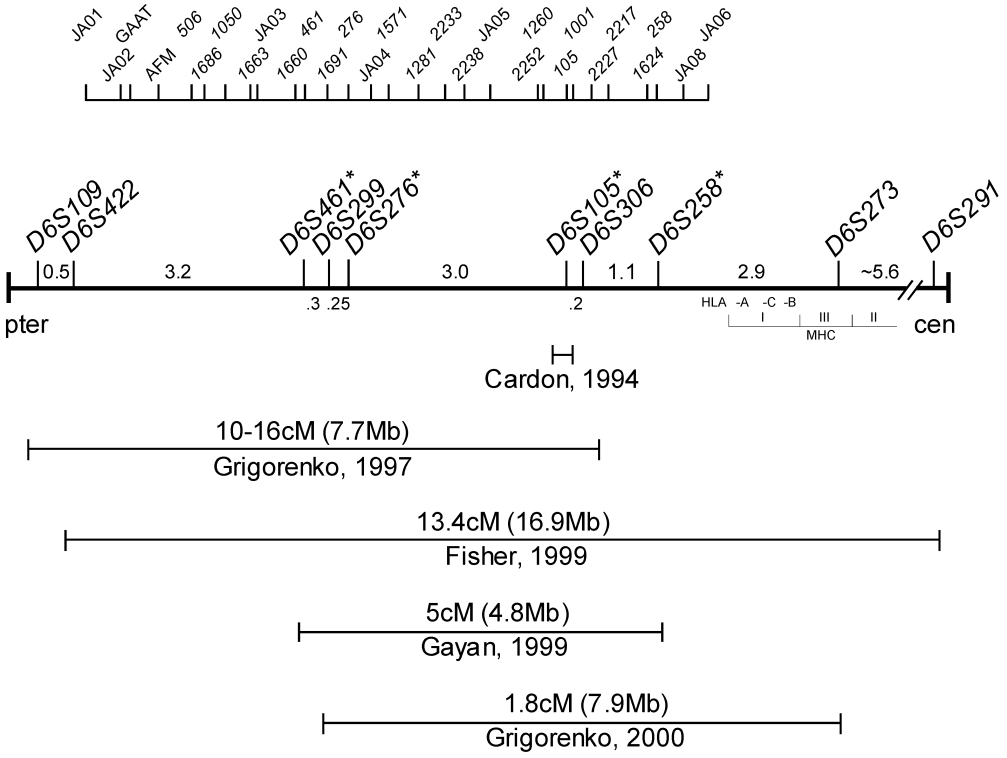

Fig. 1 Localization of the 6p21.3-22 RD linkage studies. Markers span 17.3 Mb including the MHC class I, III, and II regions at the centromeric end of the map. An *asterisk* designates a marker in the panel described in Table 2. HLA-A, HLA-C and HLA-B are represented by *-A*, *-C*, and *-B* in the MHC class I region. Physical distances (in Mb) and marker order are based on data from the contig in Fig. 2 (covering the 5.5 Mb region corresponding to markers *D6S422* through *D6S276*), the contig from Ahn and Gruen (1999) (covering 2 Mb surrounding markers *D6S464*/*D6S105*), and the most recent versions of the Sanger data (covering the markers centromeric to *D6S306*). Markers *D6S109* through *D6S273* are located on a continuous BAC/PAC contig. The most centromeric marker, *D6S291* is separated from the others by a cloning gap in the BAC/PAC contig, as denoted by the break in the median (–//–). The distance between *D6S273* and *D6S291* is estimated by a YAC contig (http://www.sanger.ac.uk/cgi-bin/humace/SCMAPS.cgi?chr=6&interval=interval18). TNFB, TNFA, and *D6S273* are located in the class III region of the MHC, and *D6S291* is located centromeric to the class II region. The intervals beneath the median indicate the markers in linkage with the RD phenotypes described in the five studies. The genetic distances (cM) above the intervals are those reported by each study. The locations of the markers comprising the multiplex genotyping panel in Table 2 are represented above the horizontal median

D6S291 (Fisher et al. 1999), 5 cM (4.8 Mb) spanning *D6S461* through *D6S258* (Gayán et al. 1999), and 1.8 cM (7.9 Mb) spanning *D6S299* through *D6S273* (Grigorenko et al. 2000) (physical distances are as described in Ahn et al. 2001). Combining these studies, linkage mapping has defined an RD locus that spans 17 Mb (14.9 cM) between markers *D6S109* and *D6S291*. The overall consistency with which evidence for a chromosome 6 locus has been replicated is remarkable and highly unusual for a complex behavioral trait attesting to the high heritability of RD in general (0.4–0.6), the quality of phenotypic assessments, and particularly to the high attributable risk for a chromosome 6p locus (Gayán et al. 1999).

Ambiguity in the precise localization of a 6p locus for RD and the other disorders listed above is partly attributable to the paucity of, imprecise order among, and distances between available markers. For example, the linkage studies described for RD reported inconsistent marker orders and intermarker distances that varied from 1.5 to 31 Mb. In addition, recombination suppression reportedly present telomeric of HLA-A would further confound precise localization (Malfroy et al. 1997). Especially in the context of the usual difficulties with localizing risk loci for complex disease traits, regional recombination suppression could create the plateau of linkage spanning large genetic and physical distances described for RD instead of a narrower peak confined to a smaller distance.

To better define the boundaries of genetic linkage and recombination suppression for RD and other loci in the same region, we set out to increase the density of markers and assign precise order and distances by constructing a minimal tiling pathway of contiguous overlapping genomic clones on 6p21.3-22. We constructed this pathway spanning markers *D6S1950* through *D6S478* with 88 overlapping BAC and PAC clones culled from our published contigs and those from the Sanger chromosome 6 genome center. From these data, we then selected and characterized a panel of 29 short tandem repeat (STR) markers, which we then optimized for high-throughput genotyping. The pathway and marker panel are presented below with a discussion of the likely ramifications of recombination suppression in this region.

Materials and methods

Construction of the minimal tiling pathway

Using the human DNA inserts amplified from a YAC clone scaffold (Bray-Ward et al. 1996; Malaspina et al. 1996) we generated a 2-Mb BAC/PAC/cosmid contig covering the histone region on 6p21.3 (Ahn and Gruen 1999), and then a contiguous 5-Mb BAC/PAC contig covering the telomeric and middle portions of the RD locus.(Ahn et al. 2001) We extended coverage through the centromeric portion with PAC and BAC clones and sequence from the Sanger Centre available through public domain and FTP servers (www.sanger.ac.uk/HGP/Chr6). Additional map elements were identified from the published literature (Bray-Ward et al. 1996; Dib et al. 1996; Feder et al. 1996; Lauer et al. 1997; Malaspina et al. 1996) and public domain servers at the Cooperative Human Linkage Center (CHLC: lpg.nci.nih.gov/CHLC), the Genome Database (GDB: gdbwww.gdb.org), the National Center for Biotechnology Information (NCBI: www.ncbi.nlm.nih.gov), the Sanger Centre (www.sanger.ac.uk/HGP/Chr6), the Stanford Human Genome Center (SHGC: www-shgc.Stanford.edu), UniSTS (www.ncbi.nlm.nih.gov/genome/sts), and the Whitehead Institute Center for Genome Research (WICGR: www-genome.wi.mit.edu).

Optimization and multiplexing

PCR conditions for individual markers were first optimized for single band clarity by agarose-gel electrophoresis, starting with the melting temperature predicted by the primer sequences and ramping 10 °C in either direction. Markers were assigned to one of four sub-panels based on overlapping molecular weight ranges (Table 2, column 5): eleven markers in the low range (90–175 nt), eleven in the middle range (176–225 nt), eight in the high range (226–283 nt), and three in the very-high range (284–309 nt). Markers were further assigned to one of three or four multiplex PCR groups based on common annealing temperature (Table 2, column 2), and one of three fluors (FAM, HEX, or TET, Life Technologies, www.invitrogen.com) to distinguish markers of similar molecular weight within each sub-panel (Table 2, column 8). After labeling one primer of each pair, primer ratios were optimized for multiplex PCR reactions within each sub-panel (Table 2, column 10). Other than annealing temperature and primer volumes, the conditions for multiplex PCR were the same: denaturation at 96 °C for 20 s, annealing for 30 s, polymerization at 72 °C for 45 s, for 35 cycles, followed by a final elongation step at 72 °C for 20 min. Reaction volumes were 15 µl with 25–100 µg of template DNA, 0.03 U/µl of HotStartTaq polymerase (Qiagen Genomics, Bothell, Wash.), and 1.5 mM $MgCl_2$ in standard 10×buffer (Qiagen). Pooling ratios for individual markers were adjusted for fluorescence (Table 2, column 9), which varied between fluors and between labeling reactions. Finally, the markers within each sub-panel were electrophoresed together along with the TAMRA-labeled molecular weight standard (Life Technologies) in a single lane (multiplex loading) on an ABI377 DNA Sequencer (Applied Biosystems, Foster City, Calif.).

Genotyping

One thousand and twenty PCR amplifications were performed on a PTC-225 Tetrad programmable temperature cycler (MJ Research, Waltham, Mass.) and genotypes resolved in four gels on an ABI377 sequencer (PE/ABI, Foster City, Calif.). Technicians were blinded by assigning random tracking numbers to clinical samples and through predetermined PCR plate, pooling plate, and gel-lane assignments. In order to avoid overflow and bleed-through errors loading was staggered between consecutive lanes by 100 scans. Each gel included two CEPH controls, 1331-1 and1331-2, and one of nine unrelated controls chosen from the Coriell sample. GeneScan and GenoTyper (ABI/Perkin-Elmer, Foster City, Calif.) tracked and converted ABI377 fluorescent chromatogram data to nucleotide assignments. Two technicians independently scored the output of every gel, compared base-pair assignments using an allele-consistency macro (Allele Comparison) in Excel (Microsoft), and resolved conflicts or flagged alleles for re-genotyping. The final nucleotide assignments were then ported to the Genetic Analysis System software package (GAS version 2.0, Alan Young, Oxford University, 1993–95) for binning into allele numbers and for identifying inconsistencies in controls.

Nonlinear least-squares regression analysis

To assess the recombination suppression previously described in this region of 6p, we compared the physical map described in terms of nucleotides to the genetic map described in terms of centimorgans by nonlinear least-square regression analysis. The independent variables for this analysis were the physical distances (D_P) between markers (Fig. 2), and the genetic distances (D_G) published by Dib et al. (1996) (Table 1). The piece-wise linear model shown in equation (1a, b) served as the fitting model.

$$D_G = s_1 \times D_P + I \quad \text{...if } D_P \le D_B \tag{1a}$$

$$D_G = s_2 \times (D_P - D_B) + s_1 \times D_B + I \quad \text{...if } D_P > D_B \tag{1b}$$

In equation (1a, b), s_1 and s_2 are the slopes before and after the breaking point, D_B is the position of the breaking point on the horizontal axis, and I is the intercept. Parameters s_1, s_2, I, and D_B were optimized in the nonlinear least-squares fit. The statistical analysis was performed by using specialized computer software written in the C++ language (Barton and Hackman 1995). The computation was based on Reich's modification of the Levenberg-Marquardt nonlinear minimization method (Marquardt 1963). Duggleby (1984) demonstrated that the Levenberg-Marquardt nonlinear minimization algorithm is very appropriate for fitting piece-wise linear data.

Results

Minimal tiling pathway

We generated an uninterrupted minimal tiling pathway (Fig. 2) spanning 10 Mb of the 6p21.3-22 RD locus, based on redundant clone contigs we published previously (Ahn and Gruen 1999; Ahn et al. 2001; Bray-Ward et al. 1996) and the published genomic contig and sequences from the Sanger Centre. The minimal tiling pathway consists of 88 genomic clones: 78 PACs, six BACs, and four cosmids. It also includes 16 sequence tagged sites (STS) and 43 STR markers (Table 1).

The overall physical distance of the region equivalent to the map presented in Figure 2 that is described on the Ensembl Human Genome Server based at Sanger (www.ensembl.org, March 8, 2002) is considerably longer, spanning 15 Mb from nucleotide 19,580,000 through 35,260,000 (nucleotide number 1 at pter and ascending towards the cen). It also contains eight gaps: one at the most telomeric end near *D6S1950* and seven surrounding *D6S276* (4 Mb in Fig. 2). The two maps are concordant at the centromeric end from *D6S1621* (5.2 Mb in Fig. 2 and 30.95 Mb in Ensembl) through *D6S478* (9.5 Mb in Fig. 2 and 35.26 Mb in Ensembl). However, telomeric to *D6S1621* the maps diverge with large genomic blocks misappropriated primarily because of the gaps in the Ensembl map; but within any single block there is good agreement in intermarker order and distance. The block spanning *D6S1950*

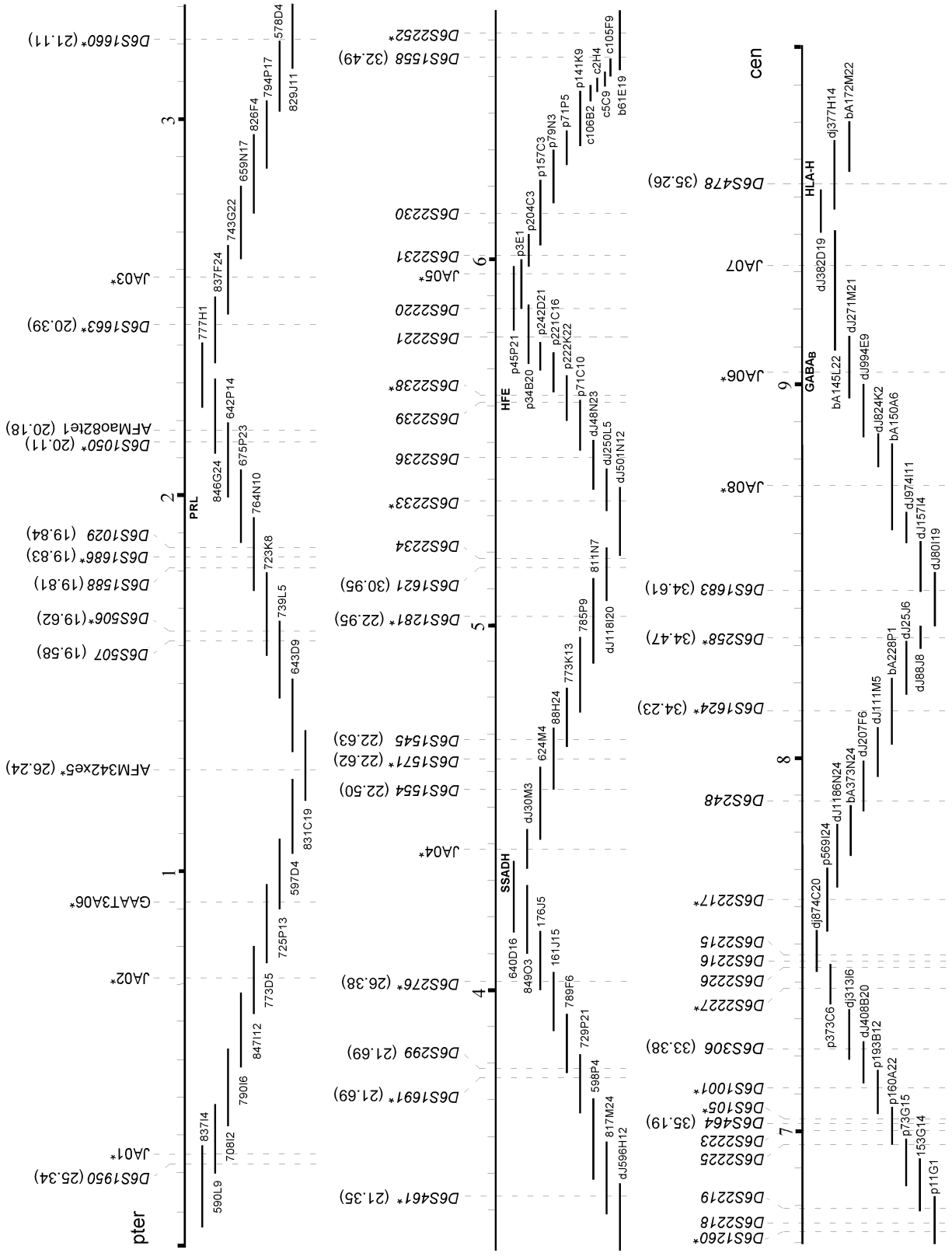

Fig. 2 Minimal tiling pathway of BACs/PACs/cosmids spanning the 10 Mb of the 6p21.3-22 RD locus. *Scale* is in megabases (Mb), with pter at the top left hand corner and cen at the bottom right hand corner. Beneath the scale are five genes (PRL, SSADH, HFE, $GABA_B$, and HLA-H) included as regional landmarks. Above the scale are the 59 map elements (16 STS, and 43 STR) described in the text. An *asterisk* associated with a map element designates a STR included in the Table 2 marker panel. *Numbers in parentheses* specify the corresponding Ensembl location in Mb from pter (from March 8, 2002). Discontinuities between consecutive Ensembl localizations are discussed in the text. Beneath the scale are the 88 genomic clones: *c_* designates a cosmid (four clones), *b_* designates a BAC (six clones), the others are PACs (78 clones)

Table 1 Map elements from Fig. 2

Map element[a]	Origin	Type	GenBank accession no.	RH^W distance from pter (cR)[b]	RH^N distance from pter (cR)[c]	Genetic distance from pter (cM)	Physical distance in Fig. 1 (Mb)[f]
D6S1950	Généthon	STS	G05373	118.48			0.225
JA01	This paper	STR	G72378				0.25
JA02	This paper	STR	G72379				0.7
GAAT3A06	CHLC	STS	G12405			36.37[e]	0.9
AFM342xe5	Généthon	STR	Z76085	118.38			1.25
D6S507	Orr	STR	GDB:588550				1.625
D6S506	Orr	STR	GDB:229129				1.65
D6S1588	Généthon	STS	Z52855		287	38.1[d]	1.825
D6S1686	Généthon	STR	Z51587		285.5	39.3[d]	1.85
D6S1029	CHLC	STS	G08502			39.2[e]	1.875
D6S1050	CHLC	STS	G08554			42.3[e]	2.125
AFMa082te1	Généthon	STR	Z67232				2.175
D6S1663	Généthon	STR	Z53981			40.2[d]	2.45
JA03	This paper	STR	G72380				2.575
D6S1660	Généthon	STR	Z53853		289.3	40.2[d]	3.2
D6S461	Généthon	STR	Z24329		303	41.3[d]	3.45
D6S1691	Généthon	STR	Z51673		296	42.7[d]	3.75
D6S299	Généthon	STR	Z16986			41.3[d]	3.775
D6S276	Généthon	STR	Z16711			44.9[d]	4.025
JA04	This paper	STR	G72384				4.395
D6S1554	Généthon	STS	Z52373	130.36	308	42[d]	4.55
D6S1571	Généthon	STR	Z52578			42.7[d]	4.65
D6S1545	Généthon	STR	Z52117			42.7[d]	4.695
D6S1281	CHLC	STS	G08569		317.2	44.1[e]	5.05
D6S1621	Généthon	STR	Z51159			44.3[d]	5.15
D6S2234	Feder	STR	GDB:1297611				5.175
D6S2233	Feder	STR	GDB:1297594				5.35
D6S2236	Feder	STR	GDB:1297641				5.45
D6S2239	Mercator	STR	GDB:1297778				5.6
D6S2238	Mercator	STR	GDB:1297759				5.625
D6S2221	Mercator	STR	GDB:1296777				5.8
D6S2220	Mercator	STR	GDB:1296760				5.85
JA05	This paper	STR	G72381				5.975
D6S2231	Mercator	STR	GDB:1297567				6.025
D6S2230	Mercator	STR	GDB:1297551			44.4[e]	6.125
D6S1558	Généthon	STS	Z52438		382	44.3[d]	6.55
D6S2252	Shearman	STR	GDB:1336893				6.625
D6S1260	Stone	STR	GDB:454966				6.725
D6S2218	Mercator	STR	GDB:1278269				6.75
D6S2219	Mercator	STR	GDB:1296748				6.775
D6S2225	Mercator	STR	GDB:1297462				6.975
D6S2223	Mercator	STR	GDB:1296788				7
D6S464	Généthon	STS	Z24380		353.4	44.3[d]	7.015
D6S105	Généthon	STS	X59425			44.4[e]	7.025
D6S1001	Stone	STS	GDB:363825				7.115
D6S306	Généthon	STS	Z17120			44.3[d]	7.215
D6S2227	Mercator	STR	GDB:1297514				7.375
D6S2226	Mercator	STR	GDB:1297480				7.425
D6S2216	Mercator	STR	GDB:1297480				7.45
D6S2215	Mercator	STR	GDB:1278219				7.475
D6S2217	Mercator	STR	GDB:1278219				7.625
D6S248	Boyle	STR	GDB:187112				7.8775
D6S1624	Généthon	STS	Z53323			44.9[d]	8.125
D6S258	Généthon	STS	Z53323	164.64		44.3[d]	8.325
D6S1683	Généthon	STS	Z51515			44.3[d]	8.45
JA08	This paper	STR	G72385				8.725
JA06	This paper	STR	G72382				9.025
JA07	This paper	STR	G72383				9.3
D6S478	CHLC	STS	G09109	167.86	495.3		9.525

[a]The 59 map elements from Fig. 2, with accession numbers are presented in order from pter at the top

[b]Radiation hybrid (*RH*) distance in centirays (cR) from the WICGR radiation hybrid map (Hudson et al. 1995; Walter et al. 1994)

[c]Radiation hybrid (*RH*) distance in centirays (cR) from the NCBI radiation hybrid map (Agarwala et al. 2000)

[d]Genetic distance from pter in centimorgans (cM) from the Généthon genetic linkage map (Dib et al. 1996)

[e]Genetic distance from pter in centimorgans (cM) from the Marshfield genetic linkage map (Broman et al. 1998)

[f]Physical distance in megabases (Mb) from the most telomeric clone in Fig. 2

Table 2 Multiplex PCR and loading schemes. The 33 markers for high-throughput genotyping are divided into four sub-panels. Horizontal line spaces within each sub-panel divide the multiplex PCRs by common annealing temperature. The number of alleles, size range in nucleotides, and heterozygosity are as observed in the 30 controls described in the text. The primer sequences for the markers were as described in public domain servers, except where indicated

Marker	Annealing temp (°C)	Repeat unit	No. of alleles	Allele size (nt)	Hetero-zygosity value	Primer name	Fluor	Primer vol. (μl)[b]	Pooling ratio[c]	Primer sequence
Sub-panel 1										
JA02	56	(GGAT)	4	100–121	0.6094	JA02TTF[a]	FAM	0.6	4	TGCCTGAGGAACACTGAAG
						JA02TTR		0.6		TCAGTCAATCAATTCTGTCCC
D6S276	56	(CA)	10	196–225	0.6816	AFM158ye9a[a]	FAM	0.75		TCAATCAAATCATCCCCAGAAG
						AFM158ye9b		0.75		GGGTGCAACTTGTTCCTCCT
DYS389	56	(GATA)	3	247–255	0.4609	30F10F	FAM	1.2		CCAACTCTCATCTGTATTATCTATG
						30F10R[a]		1.2		TCTTATCTCCACCCACCAGA
D6S1571	55	(CA)	7	159–174	0.7088	D6S1571R[a]	HEX	0.9	6	TGGCTCTAATGGTTACTTTTTACA
						D6S1571F		0.9		GGACCTACGCATCTGGTG
D6S1660	55	(CA)	6	203–215	0.745	AFMb355wg5a[a]	HEX	0.6		GAGTCTTGAGTAACTCCCACG
						AFMb355wg5b		0.6		GACAATGAGTATCCCCCAC
JA06	57	(CA)	10	90–126	0.7822	JA06DNF[a]	TET	1.2	3	CACTGTTTCCTCTCCCAGC
						JA06DNR		1.2		GGCTATACCTTGAAGTGGAGG
D6S1050	57	(GATA)	7	180–212	0.7522	GATA52D05F[a]	TET	0.75		CACACTTCTGCAAAGCAATG[d]
						D6S1050R3		0.75		TGCTGCCAAGAACATATATGTG
JA08	57	(GATA)	8	252–283	0.6561	JA08F[a]	TET	0.6		TGGCTGTAAAGAGAAGGGAC
						JA08R		0.6		AAATCAACACAAACTCCATGTC
Sub-panel 2										
D6S2252	52	(CA)	10	144–163	0.8155	Aggie3GT[a]	FAM	1.2	2	GATTTAGAAATGTAGGCCAG
						Aggie3CA		1.2		CTAATCTCCAAATGCCTAAG
D6S1686	52	(CA)	5	257–267	0.5761	D6S1686R[a]	FAM	0.6		CACTGGGTGCTAGGAATG
						D6S1686F		0.6		TGGCAAGCCACTTTCA
D6S1624	52	(CA)	7	195–210	0.6472	D6S1624F[a]	FAM	0.75	5	ATAACCCCAGGTGTTTGTG
						D6S1624R		0.75		TGGAAGTCTTCAGTGGAGAG
D6S1663	55	(CA)	5	115–139	0.7305	D6S1663R[a]	HEX	0.4	6	GGCTTCCATGCAGAGGT
						D6S1663F		0.4		AATGGCCTTGTGACACATAG
D6S1001	55	(CA)	11	198–219	0.6888	D6S1001B[a]	HEX	1.2		AGATCTCTGGGATTCCTGTC[d]
						D6S1001A		1.2		CATTTATTGGTGTATCCCTGAC
AFM342xe5	52	(CA)	4	252–258	0.4294	AFM342xe5a[a]	HEX	0.75	5	CCTAAATGTGCCGCTG
						AFM342xe5b		0.75		TGGGTCATGCTGACAAC
JA05	57	(CA)	8	149–170	0.8138	JA05DNF[a]	TET	0.6	2	CTACTTGCCTCTTCTCCATGTG
						JA05DNR		0.6		ATCGCTGAGAGTGGTGAGTG
JA04	57	(CA)	13	273–309	0.8672	JA04DNR[a]	TET	0.9		AGATACCGTACCTGTGGTCAC
						JA04DNF2		0.9		AGCAGTAGCTTCAGCCTGTC
Sub-panel 3										
D6S1260	57	(CA)	8	136–152	0.6655	CS-5A[a]	FAM	0.9	4	TTGGAGTTGATTCCCAGTG
						CS-5B		0.9		TGGTGGTACATGCCTTTG

D6S258	57	(CA)	8	189–210	0.7366	D6S258R[a]	FAM	0.9		CTTCCAATCCATAAGCATGG
						D6S258F		0.9		GCAAATCAAGAATGTAATTCCC
DXS9908	57	(GATA)	9	220–237	0.7194	182E04F[a]	FAM	0.5		TAGGGTGCAGCAAATCACCA
						182E04R		0.5		GCATTTCTATCCCTCTGCAA
CHLC.GAAT3A06	55	(GAAT)	5	176–192	0.6861	3A06F[a]	HEX	0.4	6	CCCATGAATGCTGAGACTTT
						3A06R		0.4		TTGCAGTCCTTTTCAGTAAGG
D6S1691	55	(CA)	12	213–240	0.8872	AFMa052vh1a[a]	HEX	1		AGGACAGAATTTTGCCTC
						AFMa052vh1b		1		GCTGCTCCTGTATAAGTAATAAAC
D6S2227C/D	56	(CA)	8	151–167	0.7122	D6S2227C[a]	TET	1.2	3	CTCTCATGCAGCCTCTTCTC[d]
						D6S2227D		1.2		CAACCCAGAATCACATCTAGTG
D6S2238	56	(CA)	6	188–203	0.8122	D6S2238R[a]	TET	0.4		GTCCCAGCTCTGTTTCAGAC
						D6S2238F		0.4		GCAATGACACACCCTCCCATCACC
D6S461	56	(CA)	9	252–270	0.7622	D6S461F[a]	TET	0.9	3	ACAACCCATCAGCCCACT
						D6S461R		0.9		TATGACTTCTGGACAGTTAGGGG
DXS7132	56	GATA	4	280–292	0.7116	72E05F	TET	0.6	3	AGCCCATTTTCATAATAAATCC
						72E05R[a]		0.6		AATCAGTGCTTTCTGTACTATTGG
Sub-panel 4										
D6S105	55	(CA)	13	113–139	0.825	Mfd61GT[a]	FAM	0.75	2	GAAGGAGAATTGTAATTCCG
						Mfd61CA		0.75		GCCCTATAAAATCCTAATTAAC
JA01	55	(CA)	7	185–207	0.7494	JA01TRF[a]	FAM	0.4		GCTGCAAACTTCAGGCTAG
						JA01TRR		0.4		TGTGTACATTGCTACTCATTGC
D6S506	65	(GT)	6	130–142	0.6983	LRgt1Ba[b]	HEX	0.75	4	AGATAACGCCACCACACTCCA
						LRgt1Bb		0.75		ATGATTTGGGCAGAGAACTTG
D6S2217A/B	56	(CA)	7	191–204	0.6694	D6S2217B[a]	HEX	0.9	8	ACCATCCTGTACTACCCAAGC[d]
						D6S2217A		0.9		TCAGAGAGTTGGAGAGATAGGTG
JA03	56	(CA)	13	229–256	0.8317	JA03DNF[a]	HEX	1.2		TGTGGCTAGTGAGAGAGTGAC
						JA03DNR		1.2		CACGTTTATTTCTCCTTACAGC
D6S2233	52	(CA)	6	140–150	0.7777	D6S2233B[a]	TET	0.75	8	TTAGTCTTTCTTGTAGCTCAGAC[d]
						D6S2233A		0.75		ACTGAGATCATTTACTGTTACTAGAC
D6S1281	52	(CTAT)	8	174–210	0.7527	GATA89B07R[a]	TET	0.75		AGAAGCAGCTGTGCTTTGTT
						GATA89B07F		0.75		GATGCCACGTTTTAAAATGC
DYS391	56	GATA	3	284–292	0.539	32C10F	TET	0.4	8	CTATTCATTCAATCATACACCCA
						32C10R[a]		0.4		GATTCTTTGTGGTGGGTCTG

[a]Indicates which of the primer pair was labeled with the flour in the adjacent column
[b]The volume of a 10 μM/μl solution of each primer that was added to the multiplex PCR (total volume 15 μl)
[c]The products of all multiplexed PCR from each sub-panel were pooled by the volumetric ratios in this column and electrophoresed in a single lane
[d]The primer sequence was modified to fit the multiplex loading scheme

through AFM342xe5 (0–1.25 Mb in Fig. 2) is positioned telomeric to *D6S276* (4 Mb in Fig. 2), with an intervening gap. Figure 2 shows that *D6S276* is positioned at 4 Mb between *D6S299* and *D6S1554*. Ensembl places *D6S276* just centromeric of AFM342xe5. And between *D6S276* and *D6S1621* (5.2 Mb in Fig. 2), Ensembl has an additional 3 Mb of genomic sequence containing six gaps and no markers in common with Fig. 2.

6p21.3-22 RD locus marker panel

In order to increase the density of genetic markers, we searched for STR sequences within the end-sequences from 19 YACs, 70 BACs and eight PACs generated by TAIL PCR (Liu and Whittier 1995) and inverse PCR methods (Silver 1991), The Institute for Genomic Research (TIGR) BAC end sequence consortium (www.tigr.org/tdb/humgen/bac_end_search/bac_end_anno.html), and available genomic sequence from Sanger (www.sanger.ac.uk/HGP/Chr6) as described in Ahn et al. (2001). The analysis identified 12 potentially polymorphic STRs available for generating novel markers, if demonstrated to be sufficiently variable. Of these, eight had adequate flanking sequence for generating unique primers for PCR. From these eight, seven STRs proved suitably robust for high-throughput genetic mapping (*JA0_* markers, Tables 1 and 2). The new markers were evaluated for heterozygosity in 21 whites from the Coriell Apparently Normal Collection (http://locus.umdnj.edu/ccr/), two CEPH individuals (1331-1 and 1331-2, Coriell), and an additional seven subjects of mixed ethnic backgrounds. For the seven markers the average heterozygosity was 0.758 (range: 0.609–0.867), with an average of nine alleles (range: 4–13) per marker (Table 2).

We aimed for a relative evenly spaced set of markers that would be suitable for genetic linkage and linkage disequilibrium studies of the RD locus. From the available candidates we chose a panel of 29 STRs (Fig. 2). The panel is comprised of 24 dinucleotide and five tetranucleotide repeats with an average heterozygosity of 0.73 (range 0.5–0.8) determined in 21 whites from the Coriell Cell Repository. The average intermarker distance is 300 kb, with a range of 80–680 kb. The panel also contains two X-chromosome (*DXS9908*, *DXS7132*) and two Y-chromosome markers (*DYS389*, *DYS391*) for sex assignment and sample registration controls. Through multiplexing, the entire panel can be amplified in a total of 17 PCR reactions; multiplex loading allows the entire 33-marker panel to be resolved in four gel lanes (Table 2).

Discussion

The primary goal of these studies was to develop an ordered marker panel from 6p21.3-22 for high resolution genetic association mapping and for resolving differences in order among published linkage studies. Secondary goals were: (1) to determine the precise physical length between markers, (2) to provide a resource for generating additional markers and for mapping transcriptional elements and genes, and (3) to precisely map the limits of recombination suppression, which could affect genetic linkage studies of RD or other traits in this region.

The minimal tiling pathway and markers with the corresponding Ensembl localizations presented in Fig. 2 illustrate several interesting points. First, the order and distances described by Ensembl create different localizations for the linkage peaks of 6p *IDDM* at *D6S2223* (5.3 Mb), for Behçet's disease at *D6S285* (not in Fig. 2 but near 2 Mb), and for HHS at *D6S276* (4 Mb), and would significantly affect the multipoint linkage equilibrium studies of RD described by Kaplan et al. (2002), and Grigorenko et al. (1997, 2000). Second, the overlapping peaks of linkage centrally located between markers *D6S276* (4 Mb) and *D6S105* (7 Mb) that were independently described by Gayán et al. (1999) and Fisher et al. (1999) span a physical distance of only 3 Mb; these same peaks also overlap with the region of linkage reported by Grigorenko et al. (1997, 2000) and Cardon et al. (1994). Third, where only three markers were previously available to define the linkage peaks described by Gayán and Fisher, there are now 26, of which, ten are in the multiplex panel in Table 2. Fourth, while there is significant central overlap among the studies, there is more of a plateau of linkage rather than a peak, which significantly diminishes the resolution and hinders precise localization of potential candidate-genes.

For complex inherited disorders in general, a high degree of linkage resolution is not expected to be absent a huge sample. Certain characteristics of the published RD

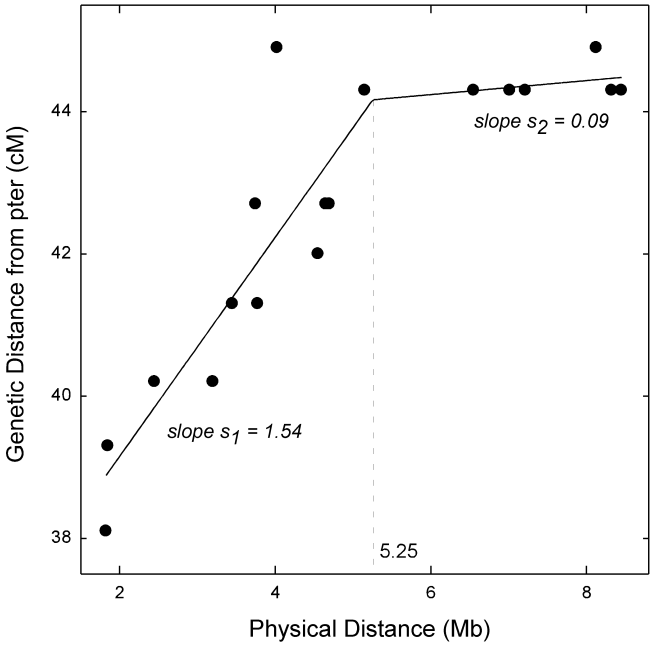

Fig. 3 Comparison of genetic distance from pter (cM) to physical distance (Mb) from the telomeric end of the contig in Fig. 2. Only markers that have both genetic and physical data from Dib et al. (1996) in Table 2 are graphed. Trend line is a non-linear regression best-fit using a piece-wise linear model. Standard error and significance for slopes, intercept, and break point are listed in Table 3

Table 3 Results of fit from non-linear regression using the piecewise linear model

	Fit	SE	*t*	*P*
slope s_1	1.54	0.25	6.21	<0.0001
slope s_2	0.10	0.52[a]	0.19	0.85
intercept *I*	36.1	0.94	38.6	<0.0001
break point D_B	5.25	0.95	5.55	<0.0001

[a]The parameter s_2 has a fairly large formal standard error (SE) and high *P* value due to its very small best-fit value

studies, however, probably contribute to a loss of resolution. Certainly the paucity and inconsistent use of different markers and order used by the different studies diminish the resolution. Also, while there is significant overlap among the groups there are also differences in the way the reading phenotypes were defined. But beyond these factors we found that the minimal tiling pathway also frames a remarkable inversion of recombination frequency that likely has significant impact for studies of this important chromosomal region. In Fig. 3, the Généthon genetic distances, as reported by Dib et al. (1996), are compared to the physical distances derived from the pathway (Table 1). Between 0 and 5.25 Mb the best-fit curve shares a nearly 1:1.5 relationship between genetic (measured in cM) and physical (measured in Mb) distances with an interval slope (s_1) equal to 1.54 (Table 3). However, at the breakpoint, located at 5.25 Mb and extending to 9 Mb, the curve acutely flattens with an interval slope (s_2) equal to 0.098 (Table 3). Here the relationship between genetic and physical distance reflects marked recombination suppression. This suppression, located in the centromeric region of the pathway and coinciding with all of the published RD linkage studies, was first described by Malfroy et al. (1997) beginning at a location near the HFE gene (5.6 Mb in Fig. 2) and extending 6 Mb to the HLA-A gene in the MHC. The breakpoint in Fig. 3 at 5.25 Mb ($P<0.0001$, Table 3) maps 300 kb centromeric of HFE and close to the marker, *D6S2233* (Fig. 2). This location places the telomeric boundary of the recombination suppression interval squarely within the RD linkage peaks described by Gayán et al. (1999) and Fisher et al. (1999), and at the linkage peaks for *IDDM*, Behçet's disease, and HHS. Suppression of meiotic recombination non-selectively applied to probands and unaffected siblings, and confined to a specific chromosomal segment, will distort any positive linkage peaks that fall into that segment by broadening them. This distortion could well explain the linkage plateau described for RD, and, in particular, the broad expanse of linkage extending into the MHC class I and III regions described by Grigorenko et al. (2000), and even beyond the class II region as described by Fisher et al. (1999) (Fig. 1).

What affect would localized recombination suppression have on other studies of this region? Studying hemochromatosis and complex traits on 6p, Collins et al. (2001) and Herr et al. (2000) showed that regional differences in recombination frequency and fluctuations in intermarker linkage disequilibrium can have profound affects on genetic association studies of simple Mendelian and complex traits. Analogous to the affect on genetic linkage, recombination suppression distorts the resolution of association studies by sustaining regional disequilibrium that would have typically decayed through historical recombination events (Yu et al. 2001). However, they further showed that, by adjusting marker density relative to regional intermarker linkage disequilibrium, the distortion caused by inherent suppression can be minimized by factoring regional differences into the analysis.

It is, finally, valuable to consider this finding, and past linkage findings for RD, in the context of the usual difficulties expected in the fine mapping of complex traits. We have demonstrated a region of recombination suppression that is very likely to have an effect on peak morphology for linkage studies of this region, particularly if a dense marker map is based on the physical map rather than the genetic map. With this consideration, the consistency of RD linkage results suggesting linkage of trait to this chromosome region becomes even more remarkable. We conclude that genetic association studies in this region anticipated in the future for RD and other loci should account for these differences in the initial study design, marker density and selection, and subsequent analyses (Collins et al. 2001; Herr et al. 2000). Furthermore, such studies for RD are likely to be rewarded by identification of a risk locus of rather large effect.

Acknowledgments We wish to thank Elizabeth Pugh (The Center for Inherited Disease Research, Baltimore) for guidance with our high-throughput genotyping, and Mrs. Marcel Nabors for editing this manuscript. Support for J.R.G. is through The Charles H. Hood Foundation, and National Institutes of Health (NIH) grants 2P01 HD21887-11, R01 DA12849, R01 DA12690, and U01 HD98002. Support for J.A. is through the March of Dimes Birth Defects Foundation and NIH grant 2P01 HD21887-11. Support for T.-W.W. is from NIH grant 2P01 HD21887-11. Support for D.E.K. is from NIH grant 2P01 HD21887-11. Support for E.L. is through The Robert Leet and Clara Guthrie Patterson Trust. This work was supported in part by funds from the U.S. Department of Veterans Affairs [the VA Medical Research Program, and the VA Connecticut-Massachusetts Mental Illness Research, Education and Clinical Center (MIRECC)].

References

Agarwala R, Applegate DL, Maglott D, Schuler GD, Schäffer AA (2000) A fast and scalable radiation hybrid map construction and integration strategy. Genome Res 10:350–364

Ahn J, Gruen JR (1999) The genomic organization of the histone clusters on human 6p21.3. Mamm Genome 10:768–770

Ahn J, Won T-W, Zia A, Reutter H, Kaplan DE, Sparks R, Gruen JR (2001) Peaks of linkage are localized by a BAC/PAC contig of the 6p reading disability locus. Genomics 78:19–29

Barr CL, Shulman R, Wigg K, Schachar R, Tannock R, Roberts W, Malone M, Kennedy JL (2001) Linkage study of polymorphisms in the gene for myelin oligodendrocyte glycoprotein located on chromosome 6p and attention deficit hyperactivity disorder. Am J Med Genet 105:250–254

Barton J, Hackman L (1995) Scientific and Engineering C++. Addison-Wesley, New York, pp 603–611

Betz RC, Lee Y-A, Bygum A, Brandrup F, Bernal AI, Toribio J, Alvarez JI, Kukuk GM, Ibsen HH, Rasmussen HB, Wienker TF, Reis A, Propping P, Kruse R, Cichon S, Nöthen MM (2000) A gene for hypotrichosis simplex of the scalp maps to chromosome 6p21.3. Am J Hum Genet 66:1979–1983

Bray-Ward P, Bowlus C, Choi J, Le Paslier DL, Weissenbach J, Gruen JR (1996) FISH-Mapped CEPH YACs spanning 0 to 46 cM on human chromosome 6. Genomics 36:104–111

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63: 861–869

Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC (1994) Quantitative trait locus for reading disability on chromosome 6. Science 266:276–279

Cho J (2000) Linkage of inflammatory bowel disease to human chromosome 6p. Inflamm Bowel Dis 6:259–261

Collins A, Ennis S, Taillon-Miller P, Kwok P-Y, Morton NE (2001) Allelic association with SNPs: metrics, populations, and the linkage disequilibrium map. Hum Mutat 17:255–262

Dechairo B, Dimon C, van Heel D, Mackay I, Edwards M, Scambler P, Jewell D, Cardon L, Lench N, Carey A (2001) Replication and extension studies of inflammatory bowel disease susceptibility regions confirm linkage to chromosome 6p (IBD3). Eur J Hum Genet 9:627–633

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Duggleby RG (1984) Regression analysis of nonlinear Arrhenius plots: an empirical model and a computer program. Comput Biol Med 14:447–455

Fagerheim T, Raeymaekers P, Tønnessen FE, Pedersen M, Tranebjærg L, Lubs HA (1999) A new gene (*DYX3*) for dyslexia is located on chromosome 2. J Med Genet 36:664–669

Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, Dormishian F, Domingo R Jr, Ellis MC, Fullan A, Hinton LM, Jones NL, Kimmel BE, Kronmal GS, Lauer P, Lee VK, Loeb DB, Mapa FA, McClelland E, Meyer NC, Mintier GA, Moeller N, Moore T, Morikang E, Wolff RK, et al. (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 13:399–408

Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, Richardson AJ, Weeks DE, Stein JF, Monaco AP (1999) A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. Am J Hum Genet 64:146–156

Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR, Ishikawa-Brush Y, Richardson AJ, Talcott JB, Gayán J, Olson RK, Pennington BF, Smith SD, DeFries JC, Stein JF, Monaco AP (2002) Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. Nat Genet 30:86–91

Gayán J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, Olson RK, Pennington BF, DeFries JC (1999) Quantitative-trait locus for specific language and reading deficits on chromosome 6p. A J Hum Genet 64:157–164

Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, Pauls DL (1997) Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. Am J Hum Genet 60:27–39

Grigorenko EL, Wood FB, Meyer MS, Pauls DL (2000) Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. Am J Hum Genet 66:715-723

Gül A, Hajeer AH, Worthington J, Ollier WE, Silman AJ (2001) Linkage mapping of a novel susceptibility locus for Behçet's disease to chromosome 6p22–23. Arthritis Rheum 44:2693–2696

Herr M, Dudbridge F, Zavattari P, Cucca F, Guja C, March R, Campbell RD, Barnett AH, Bain SC, Todd JA, Koeleman BP (2000) Evaluation of fine mapping strategies for a multifactorial disease locus: systematic linkage and association analysis of *IDDM1* in the HLA region on chromosome 6p21. Hum Mol Genet 9:1291–1301

Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu S-H, et al. (1995) An STS-based map of the human genome. Science 270:1945–1954

Immervoll T, Loesgen S, Dütsch G, Gohlke H, Herbon N, Klugbauer S, Dempfle A, Bickeböller H, Becker-Follmann J, Rüschendorf F, Saar K, Reis A, Wichmann H-E, Wjst M (2001) Fine mapping and single nucleotide polymorphism association results of candidate genes for asthma and related phenotypes. Hum Mutat 18:327–336

Kaplan D, Gayán J, Ahn J, Won T, Pauls D, Olson R, DeFries J, Wood F, Pennington B, Page G, Smith S, Gruen J (2002) Linkage and Association Studies of Reading Disability on 6p21.3–22. Am J Hum Genet 70:1287–1298

Lauer P, Meyer NC, Prass CE, Starnes SM, Wolff RK, Gnirke A (1997) Clone-contig and STS maps of the hereditary hemochromatosis region on human chromosome 6p21.3-p22. Genome Res 7:457–470

Lie BA, Sollid LM, Ascher H, Ek J, Akselsen HE, Rønningen KS, Thorsby E, Undlien DE (1999a) A gene telomeric of the HLA class I region is involved in predisposition to both type 1 diabetes and coeliac disease. Tissue Antigens 54:162–168

Lie BA, Todd JA, Pociot F, Nerup J, Akselsen HE, Joner G, Dahl-Jørgensen K, Rønningen KS, Thorsby E, Undlien DE (1999b) The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene. Am J Hum Genet 64:793–800

Liu Y-G, Whittier RF (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25:674–681

Malaspina P, Roetto A, Trettel F, Jodice C, Blasi P, Frontali M, Carella M, Franco B, Camaschella C, Novelletto A (1996) Construction of a YAC contig covering human chromosome 6p22. Genomics 36:399–407

Malfroy L, Roth MP, Carrington M, Borot N, Volz A, Ziegler A, Coppin H (1997) Heterogeneity in rates of recombination in the 6-Mb region telomeric to the human major histocompatibility complex. Genomics 43:226–231

Marquardt DW (1963) An algorithm for the least-squares estimation of nonlinear parameters. J Soc Indust Appl Math 11:431–441

Maziade M, Roy M-A, Rouillard É, Bissonnette L, Fournier J-P, Roy A, Garneau Y, Montgrain N, Potvin A, Cliche D, Dion C, Wallot H, Fournier A, Nicole L, Lavallée J-C, Mérette C (2001) A search for specific and common susceptibility loci for schizophrenia and bipolar disorder: a linkage study in 13 target chromosomes. Mol Psychiatry 6:684–693

Nöthen MM, Schulte-Korne G, Grimm T, Cichon S, Vogt IR, Muller-Myhsok B, Propping P, Remschmidt H (1999) Genetic linkage analysis with dyslexia: evidence for linkage of spelling disability to chromosome 15. Eur Child Adolesc Psychiatry 8:56–59

Rabin M, Wen XL, Hepburn M, Lubs HA, Feldman E, Duara R (1993) Suggestive linkage of developmental dyslexia to chromosome 1p34-p36. Lancet 342:178

Silver J (1991) Inverse polymerase chain reaction. In: McPherson M, Quirke P, Taylor G (eds) PCR: a practical approach. IRL Press, New York, pp 136–146

Smith SD, Kimberling WJ, Pennington BF, Lubs HA (1983) Specific reading disability: identification of an inherited form through linkage analysis. Science 219:1345–1347

Vidan-Jeras B, Gregoric A, Jurca B, Jeras M, Bohinjec M (2000) Possible influence of genes located on chromosome 6 within or near to the major histocompatibility complex on development of essential hypertension. Pflugers Arch 439:R60–62

Walter MA, Spillett DJ, Thomas P, Weissenbach J, Goodfellow PN (1994) A method for constructing radiation hybrid maps of whole genomes. Nat Genet 7:22–28

Wei J, Hemmings GP (2000) The *NOTCH4* locus is associated with susceptibility to schizophrenia. Nat Genet 25:376–377

Wright P, Nimgaonkar VL, Donaldson PT, Murray RM (2001) Schizophrenia and HLA: a review. Schizophr Res 47:1–12

Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW, Weber JL (2001) Comparison of human genetic and sequence-based physical maps. Nature 409:951–953