Cell. Mol. Life Sci. 64 (2007) 961 – 978
1420-682X/07/070961-18
DOI 10.1007/s00018-007-6405-7

Cellular and Molecular Life Sciences

Human Genome & Diseases: Review

The human genome and understanding of common disease: present and future technologies

C. McKinney and T. R. Merriman*

Department of Biochemistry, University of Otago, Box 56, Dunedin, (New Zealand), Fax: +643 4797866, e-mail: tony.merriman@stonebow.otago.ac.nz

Received 11 September 2006; received after revision 17 December 2006; accepted 18 January 2007
Online First 23 February 2007

Abstract. The study of candidate genes over the past three decades has yielded notable successes in common-disease genetics. During this time, however, interpretation of genetic association studies has been hampered by the use of clinical cohorts of inadequate power and insufficient information on genetic variation in candidate genes. The unavailability of high-throughput and low-cost genotyping technologies has also limited the scope of complex-disease genetic studies. More recently, however, the sequencing and characterization of variation within the human genome has revolutionized genetic studies and enabled full genome-wide scans for genes associated with disease. The identification of disease-associated (causative) genes has illuminated disease mechanisms. The translation of this knowledge into direct clinical benefit in diagnosis, prognosis and therapy for an individual's disease still remains a challenge.

Keywords. Genome, gene, association, linkage, polymorphism, human genome project.

The sequencing of the human genome and identification of common inter-individual variation (polymorphism) is revealing the complexity of our genetic makeup [1–4]. It is clear that this variation plays a central role in the pathogenesis of common human disease. The existence of common disease-causing genetic variants is consistent with the idea that ancestral variations evolved in combination with an environment that provided a selective advantage to carriers of today's common polymorphisms. However, the presence of these variants in the modern gene pool can also result in less advantageous phenotypes. In the environment in which our ancestors lived and evolved, the positive selective advantages (e.g. resistance to infectious disease, ability to survive times of poor nutrient availability) outweighed the negative selective pressures. In contrast, largely because of the increased lifespan allowed by our modern environment and other environmental exposures (e.g. diet), previously advantageous phenotypes can now be disadvantageous.

Here we will outline contemporary approaches to identify genes conferring susceptibility to common disease. This review focuses on the genetic technologies applied to these common diseases (excluding cancer), for which the paradigm surrounding the aetiology is impact of the environment on an individual who has inherited a complement of common predisposing genetic variants. Our examples are restricted to genetic variants causing disease in the wider population and do not include those causing rarer conditions that have a more clearly identifiable familial (Mendelian) inheritance. After describing previous approaches, we will then examine the current impact of the Human Genome Project and related

* Corresponding author.

technologies on the field of the genetics of common disease. We also examine reasons why genetic association studies have been plagued by inconsistent results.

The candidate gene approach

A productive approach to understanding the genetic aetiology of human disease has been to study genes within candidate pathogenic pathways on the basis that, owing to the function they encode, they play a role in disease. The established association between the human leukocyte antigen (HLA) antigen-presenting molecules and autoimmunity [5], first discovered in the 1970s, is the first example of the success of this approach. A more recent example of the knowledge generated by this approach also comes from autoimmune disease. The key challenge of the immune system is to reliably distinguish between the proteins of a foreign invading (infecting) organism and the proteins of its own body ('self'). When the immune system makes a mistake and recognizes a normal self-peptide (antigen) as foreign, or fails to immediately down-regulate such an aberrant response, then an autoimmune disorder may ensue [e.g. type 1 diabetes (T1D), rheumatoid arthritis (RA)]. Recent attention has focused on the signalling mechanisms stimulated upon engagement of the T cell receptor (TCR) complex with self-peptides displayed by antigen-presenting cells (APCs). The protein tyrosine phosphatase 22 gene (*PTPN22*) encodes lymphoid tyrosine phosphatase (LYP), which modulates the activation of kinases involved in early signalling events following TCR engagement by APCs [6]. Stimulated by two initial reports [7, 8] the R620W variant of the *PTPN22* gene has been confirmed as a determinant of a number of autoimmune diseases by extensive replication in Caucasian patient cohorts (this variant tends to be very rare in non-Caucasian racial groups) [9]. The functional consequences of the disease-associated 620W allele seem likely to result in activated autoreactive T cells that are more potently down-regulated [10]. Perhaps this enhances the survival of autoreactive T cells that would have been activated and deleted during the self-tolerance process in the infant thymus. Combined with the confirmed association of the T cell-negative costimulatory molecule CTLA4 with autoimmunity [11], the recent genetic data have made it clear that a key checkpoint for human autoimmunity occurs in the regulation of the threshold for T cell activation towards self-antigens. Fifteen other confirmed genetic associations in common disease that provide important knowledge on disease pathogenesis are presented in Table 1.

The positional approach

Whilst the candidate gene approach can reinforce and fine-tune knowledge of disease aetiology, it is not suited to discovery of novel disease pathways and can, in fact, be hampered by the insufficiency of current knowledge on disease pathogenesis. Thus the positional approach in which disease susceptibility loci/genes are pinpointed on the basis of their position in the genome is increasingly being exploited. Knowledge of the mechanism of disease or even of the tissue in which the disease-associated variant of interest produces its function is not required. The first positional cloning experiments in human complex disease (genome-wide scans for linkage) were published in T1D in 1994 [35, 36]. More recently, Affymetrix and Illumina microarray-based technologies for whole genome association (WGA) scanning have become available and the first three WGA scans in complex disease have been published in the last year [19, 37, 38]. The positional approach often leads to a situation in which a certain variant is reproducibly associated with a specific disease, but the exact mechanism or biochemical pathway that is influenced by the gene product is unclear.

The positional approach: genome-wide linkage analysis

Identification of regions of linkage and testing of positional candidate genes within these regions for association with disease represents a hybrid approach that has led to the characterization of a number of common disease genes. In common-disease, a widely employed clinical resource used for linkage analysis is affected sibling pairs (ASPs). Linkage to a genomic region is exhibited in a cohort of ASPs when the region is coinherited ('shared') by the ASPs significantly more often than expected due to the random Mendelian segregation observed under the null hypothesis of no linkage. A genome-wide linkage analysis in common disease is typically carried out using at least several hundred ASPs (including parents if available) and 300–500 polymorphic microsatellite markers evenly spread throughout the genome, at an average density of at least one per 10 centimorgans. The magnitude of linkage of a given region to disease can be quantified in different ways, commonly based on the logarithm of the odds score. There are clear guidelines for declaration of 'suggestive' and 'significant' linkage [39]. The most important criterion, however, for general acceptance of linkage is replication between independent studies.

The first genome-wide linkage analyses in complex

Table 1. Genes reproducibly associated with common disease.

Disease	Gene	Associated variant	Accession number	Population[1]	Effect size[2]	Mechanism	Reference[3]
Age-related macular degeneration (AMD)	Complement factor H (CFM)	Y402H	rs1061170	Caucasian	Strong	The Y402H variant is within a binding site for heparin and C-reactive protein. Binding to these factors increases the affinity of CFH for complement 3b and hastens deposition of the drusen that cause AMD.	12
Alzheimer's disease	Apolipoprotein E (*APOE*)	APOE4		Caucasian Japanese	Strong Strong	The disease-associated variant increases the affinity of APOE for LDL and chylomicron receptors, and has a possible chaperone function in β-amyloid plaque formation.	13
	Angiotensin-converting enzyme (*ACE*)	Insertion/deletion variant in intron 16		Caucasian Asian	Weak Weak	The D allele is associated with increased serum ACE levels. ACE has been reported to modify β-amyloid plaque formation.	14
Autoimmune	HLA class II	Disease-associated allele(s) specific to phenotype		All populations	Strong	Exact mechanism unclear, but revolves around thymic and peripheral recognition of (self) antigen.	5
	Protein tyrosine phosphatase 22 (*PTPN22*)	R620W	rs2476601	Caucasian	Moderate	See Table 3.	6, 15
	cytotoxic T lymphocyte antigen-4 (*CTLA4*)	CT60 (3′UTR)			Weak	See Table 3.	11,16
Crohn's disease	Caspase activation recruitment domain-15 (*CARD15*)	R702W 1007InsC	rs2066844 rs2066847	Caucasian	Strong	Mutant CARD15 is unable to activate nuclear factor NF-κB in response to bacterial lipopolysaccharides.	17,18
Obesity	Insulin-induced gene 2 (*INSIG2*)	C allele	rs556605	Caucasian	Weak	Causal variant unclear, but INSIG2 inhibits synthesis of fatty acids and cholesterol.	19
Rheumatoid arthritis	Peptidylarginine deiminase 4 (*PADI4*)	A specific haplotype		Japanese Caucasian	Strong Weak	Associated haplotype affects transcript stability and levels of cyclic citrullinated peptide antibody.	20
	Fc receptor-like 3 gene (FCRL3)	−169C→T		Japanese	Weak	The −169C→T promoter variant affects NFκ-B binding and regulates *FRCL3* expression.	21,22
Schizophrenia	neuregulin 1 (NRG1)	a specific haplotype		Caucasian Asian	Weak Weak	NRG1 is a growth factor important in nervous system growth and development. Mechanism of involvement in disease unclear.	23
Systemic lupus erythematosus (SLE)	Interferon regulatory factor 5 (*IRF5*)	T allele	rs2004640	Caucasian	Moderate	The associated allele allows expression of unique isoforms of *IRF5*. It is not clear how these isoforms influence SLE pathogenesis.	24,25
Type 1 diabetes	*IFIH1*, *CD25*, *INS*, Table 3						

Table 1 *(Continued)*

Disease	Gene	Associated variant	Accession number	Population[1]	Effect size[2]	Mechanism	Reference[3]
Type 2 diabetes (T2D)	Transcription factor 7-like 2 (*TCF7L2*)	IVS3 C→T	rs7903146	Caucasian	Weak	T variant may affect transcription, splicing or stability of *TCF7L2* transcript. How this influences T2D pathogenesis is unclear.	26–32
	Calpain 10	a specific haplotype		Caucasian	Weak	Protease implicated in insulin secretion, but role in pathogenesis is unclear.	33
	Peroxysomal proliferative activated receptor-g (PPAR-γ)	P12A	rs1801282	Caucasian	Weak	Unclear, but seems to be related to insulin resistance.	34

LDL, low-density lipoprotein.
[1] Populations in which association has been demonstrated are listed.
[2] Strong is defined as odds ratio (OR) > 2.0; moderate as OR 1.5–2.0; weak as OR < 1.5.
[3] Meta-analysis data are referred to when available, otherwise multiple confirmatory studies are cited.

human disease were published for T1D in 1994 using only around 100 ASPs [35, 36], which emphasized the importance of the HLA region in T1D. Since then multiple genome-wide linkage analyses and meta-analyses, the vast majority using microsatellite markers, have been published for most major common diseases. This has been a major achievement in biomedical genetics given the enormity of the task required to recruit and clinically characterize the ASP cohorts, followed by the laboratory-based genome-wide scan and data analysis. This work has, however, been characterized by the difficulty of replicating linkages between independent studies. For example, two genome scans in T1D with a total of >1,000 sibpairs from the UK and the US yielded linkages that were, aside from the chromosome 6p HLA region, primarily non-overlapping [40–42]. This is most likely due to insufficient cohort size, with substantially larger cohorts needed to provide adequate power for replication of an initial finding of linkage [43]. Meta-analysis, whereby information from several independent studies is combined to obtain a consensus value for linkage (or association), is one method to overcome the limited statistical power inherent in genome-wide linkage analysis. There has been much international co-operation between research groups to enable this for body mass index [44], T2D [45, 46], hypertension [47], schizophrenia [48], bipolar disorder [49], inflammatory bowel disease [50], osteoporosis [51], T1D [52], systemic lupus erythematosus [53], RA [54], ankylosing spondylitis [55], psoriasis [56], multiple sclerosis [57], autism [58] and heart disease [59].

Whilst the success of the positional candidate approach (that is, identification of disease-associated genes within regions of linkage) rivals that of the candidate gene approach, it is fair to say that the number of common disease genes identified by genome-wide linkage analysis has not met expectations. The best example of successful positional cloning was identification of the caspase activation recruitment domain 15 (*CARD15*) gene as a Crohn's disease (CD) susceptibility variant [60, 61]. *CARD15* is located in a region on chromosome 16 previously linked to CD [62]. Hugot et al. [60] tested single-nucleotide polymorphisms (SNPs) within a 20-Mb region of chromosome 16 for association with CD, and discovered rare *CARD15* alleles that alter the function of the protein in up to 15% of cases. Using a different approach, Ogura et al. [61] tested candidate genes within the chromosome 16 linkage peak for association with CD and also found association with *CARD15*. CD is characterized by perturbed control of inflammation in the gut and its interaction with bacteria. *CARD15* was a strong CD candidate gene because of its role in regulating NF-κB, which is a master regulator of the genes involved in inflammation and which has been implicated in gut-microbe interactions [63]. There are three variants associated with CD (R702W, G908R and a frameshift insertion L1007fsinsC) that have been widely replicated in Caucasians but, because of the absence of these variants in the gene pool, they are not a feature of disease in Asian populations [64].

The primary reason for the small number of complex-disease genes identified by genome-wide linkage analysis is the lack of power to identify and confirm linked regions – such confirmation is very important before committing to positional cloning in a region of linkage. In T1D an international consortium has been established to assemble a *de novo* cohort of 2,500 ASPs for a sufficiently powered genome-wide linkage analysis (www.t1dgc.org), followed by positional cloning of genes underlying the five most promising linkage peaks. Ideally, a second larger cohort of

ASPs would be useful for replication of these linkages, but this is clearly not practical.

A second reason for the difficulty in identifying disease genes under peaks of linkage is the likelihood that linkage signals in common disease result from a clustering of multiple disease susceptibility genes that, individually, have weak effects on disease risk but collectively exhibit enough influence to generate a linkage signal. In support of this, linkage signals in inbred animal models of human disease routinely resolve into multiple loci. In the well-characterized non-obese diabetic mouse model of T1D, the Idd5, Idd9, Idd10 and Idd21 loci all consist of >2 subloci [65–68]. In humans, there is considerable evidence that multiple loci contribute to linkage to autoimmune disease within the HLA region on chromosome 6p [69]. Even if linked regions could be reliably identified and replicated using cohorts of sufficient power, the field has until now been hampered, in addition to the reasons outlined above, by a lack of access to genotyping technology of sufficiently high throughput and low cost to enable adequate testing of more than just a few positional candidate genes within regions of linkage. Also slowing progress until very recently has been insufficient information on the haplotype structure within candidate genes.

It is clear that linkage analysis using ASPs is not a suitable strategy for a comprehensive survey of the human genome for susceptibility loci for common disease. Although linkage analysis is more powerful than association analysis for detecting rare, high-risk disease alleles, association studies are predicted to be more powerful for the genome-wide detection of common alleles that confer modest levels of disease risk [70]. A major advantage of association analysis is that recruitment of large unrelated case-control cohorts is significantly easier than recruitment of multiply affected pedigrees, especially in diseases with an older age of onset.

The positional approach: WGA

Very recently several technologies and resources have evolved to the point where the scanning of the entire genome for genes associated with common disease has become a reality: (i) information from the HapMap project to enable selection of single nucleotide polymorphisms (SNPs) that will adequately capture the differing levels of variation within individual genes [2]; (ii) microarray-based genotyping systems enabling the genotyping of hundreds of thousands of SNPs [71–73]; and (iii) large-enough case-control cohorts exist so that studies of adequate power can be designed. Here, we will discuss each of these technologies in turn, before reviewing current progress in WGA scans in common disease.

The HapMap project

Geneticists seek to understand human disease by unravelling the association between variation in the human genome and susceptibility to disease. Thus, characterization of variation in the human genome is necessary to realize the impact of the Human Genome Project on knowledge of common disease. Last year, phase II of the 'haplotype map' (HapMap) project became available [2]. Phase I (launched late in 2002 with the aim of guiding the design and analysis of medical genetic studies) genotyped >1 million SNPs in Caucasian, Nigerian, Chinese and Japanese cohorts [1]. Phase II aims to genotype an additional 4.6 million SNPs. In October 2005, dbSNP contained 9.2 million candidate SNPs, of which 3.6 million have been validated and 2.4 million genotyped. HapMap has confirmed: (i) that there is more linkage disequilibrium (LD; non-random association of alleles at closely linked markers) throughout the genome than predicted under simple population genetic models, and (ii) that LD is more segmentally structured and varied across regions than previously supposed. This can be illustrated with two very closely linked regions on chromosome 18q: the *BCL2* gene is characterized by low LD, whereas the *SERPIN* gene cluster, only 400 kb from *BCL2*, has extensive LD (Fig. 1). A significant proportion of the human genome exhibits high LD (67% in the Nigerian cohort, 87% in the Caucasian cohort), with SNP markers falling into regions ('haplotype blocks') of little or no recombination flanked by recombination hotspots that delineate the haplotype blocks (Fig. 1). The average haplotype block size is 7.3 kb (<1–94 kb) in Nigerian samples and 16.3 kb (<1–173 kb) in Caucasians [2, 75]. There are typically dozens of SNPs per haplotype block, yet owing to the high intermarker LD, the average number of common haplotypes per block is only 5.6 in the Nigerian and 4.7 in the Caucasian cohort. The Nigerian samples exhibit lower LD than the Caucasian samples; this is consistent with the ancestors of contemporary Caucasians undergoing a population bottleneck (that increases genome-wide LD) during the migration out of Africa.

The disproportionate LD in the human genome can be exploited in genetic studies. One genetic marker can be a surrogate for those with which it is in high LD. Thus only a limited number of markers have to be genotyped to capture all the major variation within a region of high LD, with more markers needed in regions of lower LD (Fig. 1). These markers are termed tag SNPs, or when applied to marking of common haplotypes within a haplotype block, haplotype tag SNPs (htSNPs) (Table 2).

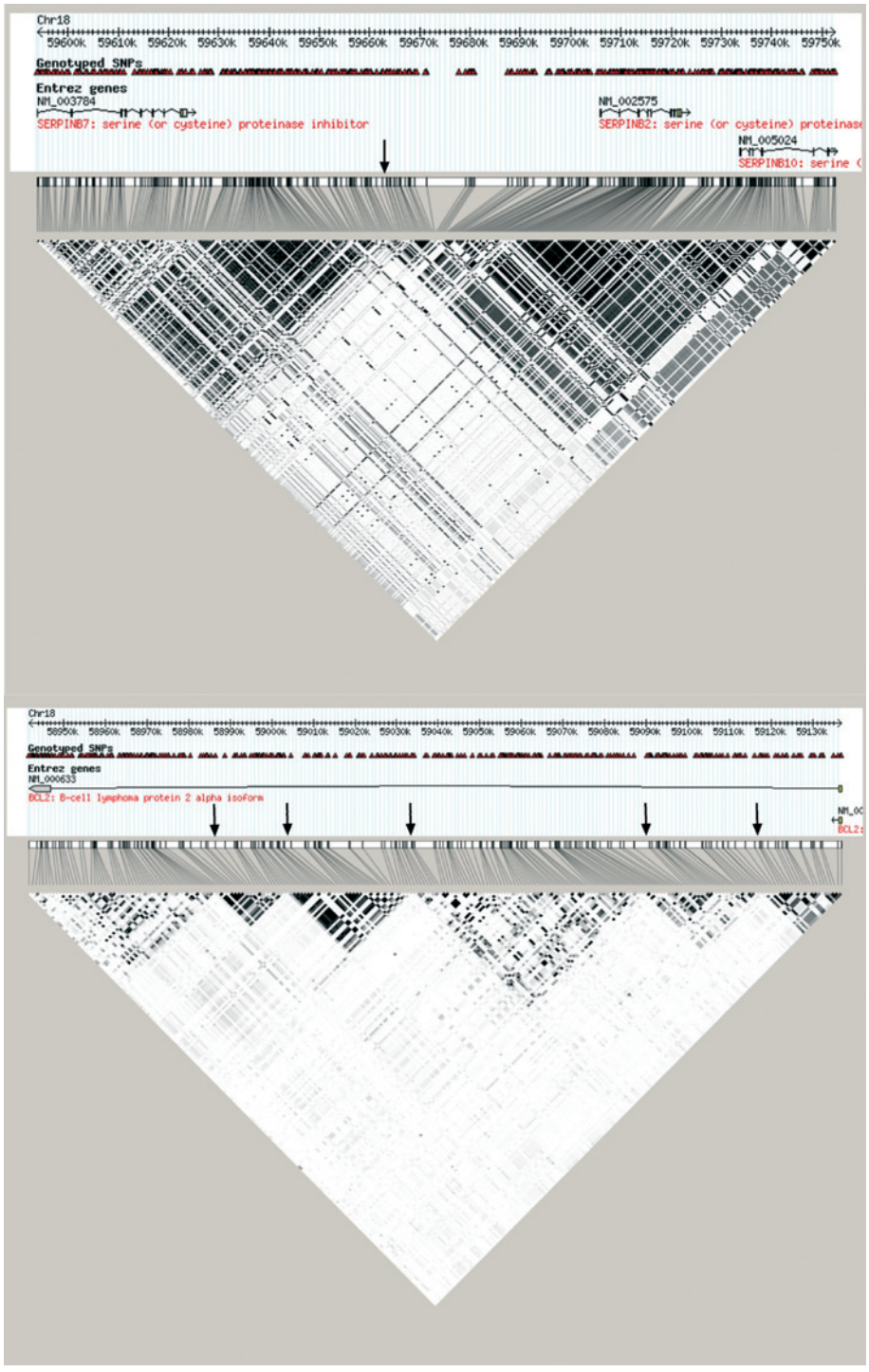

Figure 1. Haplotype structure in the chromosome 18q21.3 region. The top panel shows three of the *SERPIN* genes (*B7*,*B2*,*B10*) and the bottom panel shows the *BCL2* gene. Each image displays the haplotype block structure for 233 and 367 SNPs, respectively, spread over regions of 196 and 160 kb, respectively (the two genomic segments are separated by 400 kb). The colouring in each small square reflects the amount of linkage disequilibrium (LD) between the connected markers on the diagonal, as measured using the intermarker measure of LD r^2 [74] (r^2 is of particular importance in genetic mapping as it is inversely related to the required sample size for association mapping – for example, if only one of a pair of SNPs was genotyped and the intermarker r^2 was 0.5, then the sample size would need to be doubled at the second SNP to obtain the same statistical power). When $r^2 = 1$, one SNP completely predicts the other. Black indicates complete LD; dark colouring indicates strong LD; lighter colouring moderate LD and no colouring weak/no LD. Note that the *SERPIN* gene cluster contains two haplotype blocks of strong LD, whereas *BCL2* has considerably more haplotype diversity (six haplotype blocks); a larger number of SNPs would have to be genotyped at *BCL2* to capture the genetic diversity. Recombination hotspots are indicated by arrows. The pictures were drawn using the program HAPLOVIEW (www.hapmap.org) using CEPH genotype data from www.hapmap.org.

It is estimated that nearly 300,000 SNPs would be needed to 'capture' all common SNPs observed in phase I of the HapMap project in Caucasians, and over 470,000 in the Nigerian samples [2]. Until phase II is completed, we will not know how many tag SNPs will be needed to capture all common variation in the human genome.

Genotyping technology

Crucial to WGA scanning is the availability of inexpensive yet robust genotyping methodology suitable for the simultaneous genotyping of hundreds of thousands of SNPs. Such technologies are commercially available at a cost between 0.2 and 1.0 US cents per genotype. Two widely used systems have been developed by Affymetrix and Illumina [77]. The

Table 2. Reduction of genotyping by the use of haplotype tag SNPs: a haplotype analysis of 14 SNPs observed at the insulin gene [76] in a white European population.

1	2	**3**	**4**	5	6	7	8	9	10	**11**	**12**	**13**	**14**			
A/C	C/T	**A/T**	**C/G**	C/T	C/T	C/A	C/T	G/T	G/A	**G/A**	**C/A**	**C/T**	**G/A**	Frequences	Haplotype number	Identifying SNPs
A	C	A	C	C	C	C	C	G	G	G	C	C	G	45%	1	3, 12, 13, 14
A	C	A	C	C	C	C	C	G	G	G	A	C	G	20%	2	3, 12
C	T	T	G	C	T	A	T	T	A	G	C	C	G	13.3%	3	4
A	C	A	C	C	C	C	C	G	G	G	C	C	A	11.3%	4	14
C	C	T	C	T	C	A	C	G	G	A	A	C	G	3.8%	5	11
A	C	A	C	C	C	C	C	G	G	G	C	T	G	3.5%	6	13
C	C	A	C	C	C	C	C	G	G	G	C	C	G	1.5%	7	
C	C	T	C	T	C	A	C	G	G	G	C	C	G	0.5%	8	
A	T	T	G	C	T	A	T	T	A	G	C	C	G	0.5%	9	

Each column represents the allele data for a single SNP marker and each row can be thought of as representing the linked alleles at markers along a chromosome occurring in this population (a haplotype). Genotyping of the six 'haplotype tag' SNPs (htSNPs; SNPs 3, 4, 11, 12, 13, 14 in bold) would be sufficient to identify all common haplotypes of $> 3\%$ frequency. The right-hand column lists the htSNPs that specifically identify each haplotype. Further SNPs would have to be genotyped to identify rarer haplotypes.

Affymetrix technology represents each SNP on microarrays by 40 probes arranged in ten quartets that consist of perfect-match and mismatch pairs for each allele. Genomic DNA is restricted, amplified, labelled with a fluorescent signal and hybridized to the microarray, and genotypes are assigned according to the strength of hybridisation to the SNP-specific probes [73]. Affymetrix currently market two microarrays suitable for WGA scanning, the GeneChip Mapping 100K and 500K SNP array sets (containing 116,204 and 504,152 SNPs, respectively) with quasi-random SNPs that were selected primarily according to technical quality. The Illumina technology has two oligonucleotide probes specific for each SNP allele attached to beads arranged in microarrays. After whole-genome amplification, the DNA sample is hybridized to the microarray and SNP genotype determined by enzymatic allele-specific primer extension using fluorescently labelled dNTPs [72]. In contrast to Affymetrix, SNPs on the Illumina products were selected differently. The Illumina Human-1 100K BeadChip has 109,367 SNPs selected such that 85% of SNPs are within, or very close to, genes. The Illumina Sentrix HumanHap300 Beadchip has 317,503 tag SNPs selected from the HapMap project [71]. The differing Affymetrix and Illumina SNP selection criteria provide complementary approaches for WGA scanning, with each capturing >70% of common variation in people of Caucasian and Asian descent and ~50% in people of African descent [77]. The lower coverage in people of African ancestry is a consequence of the lower genome-wide LD in this racial group.

Study design

The optimal study design in WGA scanning in common disease is unresolved. One possibility is genome-wide scanning in case-control cohorts of sufficient size to enable stringent significance thresholds to be set (e.g. P=5x10^{-8}) [70]. A stringent threshold is necessary owing to the low prior probability of any given variant being associated with disease and the need to adjust P values for the hundreds of thousands of independent tests of association performed. It has been estimated that 6,000 cases and 6,000 controls would be needed to provide 94% power to detect a disease-associated variant having a minor allele frequency of 0.1 and OR = 1.3 [78]. This, however, has the disadvantage of high cost. To reduce cost, a staged approach could be used [74], whereby the clinical resource is divided into cohorts, one being used for the genome-wide scan and the other(s) used for replication of putative associations. For example, a reduced number of cases and controls (<1,000 of each) could be genome-wide typed with a more liberal P value threshold set in order to maintain the power to detect loci of weaker effect (e.g. the putative Wang et al. [78] OR=1.3 locus described above). To account for the fact that a large, but manageable number of false-positive associations will reach this threshold, a second cohort of larger size could be used to replicate initial associations. The larger size is necessary because the disease-risk of the initial positive associations will very likely be overestimated – the random nature of selecting cases and controls for the genome-wide scan will ensure that, by chance, some genetic associations will be 'concentrated' in the discovery case-control cohort and thus risk over-estimated in this cohort (of course, the

opposite will be true with the risk of other genes under-estimated and hence association undetected). A third cohort could be used for increased stringency. Thus, genome-wide significance is achieved by a series of SNP reduction steps.

Whilst the optimal WGA scan study design is still a matter for discussion, it will likely be determined in the empirical setting. Consistent replication in large cohorts will probably provide the most pragmatic way to identify robust genetic associations in WGA scanning. It is clear, however, that, in contrast to genome-wide linkage analysis, sufficient clinical resources are available worldwide to enable adequately powered WGA scans in common disease, if organized into research consortia.

Current WGA studies

As of July 2006, three genome-wide scans for genetic association in common disease [age-related macular degeneration (AMD), Parkinson's disease (PD) and obesity] have been published [19, 37, 38]. Although these scans had relatively limited sample sizes and covered less than 50% of the genome, the results provide much encouragement that this strategy will revolutionize complex-disease genetics.

The first WGA scan in common disease was done for AMD using the Affymetrix GeneChip Mapping 100K array [37]. The experimental approach was to genotype a small Caucasian cohort of 96 cases and 50 controls and to apply a stringent genome-wide significance cutoff of $P < 4.8 \times 10^{-7}$. Two SNPs provided genome-wide evidence for association; one of which was within an intron of the complement factor H (CFH) gene (the second SNP was not studied further). Subsequent genotyping of CFH SNPs found strongest association with the Y402H variant [37] that has been validated in other cohorts [79, 80]. It has been hypothesized that CFH is involved in immune complement deposition in the drusen found in eyes of AMD patients. The Klein et al. [37] WGA scan was successful primarily because the *CFH* susceptibility variant confers a very strong genetic risk for AMD (OR=4.6). The study, however, lacked power to detect AMD-associated variants of weaker effect at the genome-wide level of significance. Use of a replication cohort would have enabled lowering of the cutoff threshold and potential detection of other AMD genetic susceptibility variants.

In contrast, WGA scans in PD and obesity have employed a staged study design [19, 38]. Maraganore et al. [38] typed nearly 200,000 SNPs over 443 sibling pairs discordant for PD using the Perlegen service (www.perlegen.com; the Perlegen technology is very similar to the Affymetrix technology). Using a liberal cutoff of $P < 0.01$, 1,793 SNPs were then typed in 332 case-control pairs. In addition, a further 941 SNPs with $P < 0.05$, which fell in biologically plausible PD candidate genes and previously implicated regions of linkage to PD, were genotyped. The most associated SNP (OR=1.7) was within the semaphorin 5A gene (*SEMA5A*) – SEMA5A plays a role in neurogenesis and in neuronal apoptosis, consistent with current thinking in PD pathogenesis [81]. However, the *SEMA5A* genetic association was not replicated in several attempts [82–85], although none of the sample sizes were significantly larger than the cohorts used by Maraganore et al. [38]; this is necessary in order to ensure convincing replication [43].

Herbert et al. [19] used a novel analytical approach in obesity [86]. The same Caucasian family-based cohort was used for both WGA scanning (with the Affymetrix GeneChip Mapping 100K array) and replication with a reduced set of SNPs. One SNP (*rs7566605*), 10 kb upstream of the *INSIG2* (insulin-induced 2) gene, reached a genome-wide level of significance. The association of *rs7566605* with bodymass index (BMI) was replicated in two case-control cohorts (one African-American), one population-based cohort and two family-based cohorts. However, the association was not replicated in a second population-based cohort. Given the extensive replication, the *rs7566605* data most likely represent genuine association of a gene of weak effect with BMI (OR=1.22). This SNP is in a 70 kb haplotype block that includes two genes: *FLJ10996* (of unknown function) and *INSIG2*. Determination of the exact aetiological variant influencing BMI will be aided by the knowledge obtained from the HapMap project that the associated allele of *rs7566605* is present on two African haplotypes but only one Caucasian haplotype and that the haplotypes are only similar close to *rs7566605* (suggesting that the risk allele arose before migration out of Africa with the small pool of Caucasian founders taking only the one modern disease-associated haplotype). Narrowing the location of the aetiological variant within the 70 kb haplotype block can be done by more detailed analysis of these two haplotypes – the causative variant should be contained within the region shared by the two haplotypes. Finally, acknowledging that the Affymetrix 100K SNP microarray does not provide adequate coverage of the genome, the obesity WGA scan did not detect any loci of stronger effect (for example, equivalent to the OR > 1.5 effect of *PTPN22* in autoimmune disease). This is consistent with the notion that the inherited component of obesity is constituted of genes of weak population-wide effect. Alleles for complex disease are likely to be common because they have been positively selected in past

generations. They have more subtle effects on disease risk and are more likely, therefore, to include non-exonic variants with modest effects on gene expression. Of the widely replicated common-disease associations, coding or non-synonymous SNPs (nsSNPs) represent about one-third of the disease alleles (Table 1). Whether or not this will be truly representative of common disease is unknown. Nevertheless, reducing the multiple testing inherent in WGA scans by the use of nsSNPs is a justifiable experiment. Indeed, Smyth et al. [87] found convincing evidence for a genetic association of weak effect (OR=1.16) between the interferon-induced helicase (*IFIH1*) region and T1D as a result of testing >6,500 nsSNPs for association. The *IFIH1* gene is a strong functional candidate for T1D susceptibility owing to its role as a receptor for viral dsRNA and the epidemiological and experimental evidence associating viral infection with T1D [88].

One important caveat in mapping a disease-causing allele is that once a functionally relevant SNP is reliably associated with disease it is necessary to rule out the possibility that the SNP itself is not in LD with the causative second variant or whether it is not the only disease-causative variant in the immediate region. For example, the T17A variant of CTLA4 (in the leader peptide) had been reproducibly associated with autoimmune disease [89, 90] and has been posited to alter the ability of CTLA4 to regulate T cell activation [91]. However, a comprehensive survey of CTLA4 variants demonstrated that the disease-causative alleles map to the regulatory 3′ untranslated region of the gene [11], and not to the leader peptide. In the case of *IFIH1*, the genetic effect of the *IFIH1* nsSNP associated with T1D could not be distinguished from several other SNPs with which it was in LD. Functional studies are the next step; the disease-causing SNP should have an influence on IFIH1 molecular biology, at any point from control of gene expression to protein function.

Gene-gene interactions

Given that most genetic variants of weak effect are difficult to detect in isolation, attention is turning to the possibility that such variants might have a more substantive contribution to disease through non-additive (multiplicative) interactions with other disease loci (epistasis), perhaps in the same pathway [92]. It has been shown that gene-gene interactions could be one reason why replication of gene associations has been notoriously so difficult – plausible variations in allele frequencies between populations among interacting loci can markedly affect the power to detect their weak effects individually [93]. An example of epistasis is provided by the interaction of the T cell human immunodeficiency virus-1 (HIV-1) chemokine receptor 5 with its cognate ligand CCL3L1 in susceptibility to HIV-1 infection [94]. The *CCR5* gene has a common deletion variant that inactivates the protein – this variant is protective against HIV-1 infection because the virus cannot enter cells through its portal CCR5 [95]. *CCL3L1* is present in the genome in variable copies (0–7), with people having fewer copies (0, 1) more prone to HIV infection (because CCL3L1 is a potent HIV-1-suppressive chemokine that down -regulates CCR5 expression) [94]. Genetic interaction analysis demonstrated that inheritance of zero or one copy of *CCL3L1* over-rides the protective effect of the presence of the deleted form of CCR5 [94] – that is, the down-regulation of CCR5 expression by increased *CCL3L1* copy number is more important biologically than a defective HIV-1 receptor (CCR5). Turning to WGA scanning, testing for gene-gene interactions is not only computationally feasable, this approach may even be more powerful than traditional locus-by-locus analysis. A two-stage strategy whereby all pairs of loci reaching a modest level of genome-wide significance are examined in pair-wise combination could be employed [93]. The increase in power, despite the increased number of statistical tests done, arises because the use of the biologically correct model results in the most powerful genetic testing approach. Even though the evidence for association of a single gene with disease can be equivocal, demonstration of gene-gene interaction with biologically plausible partners can more convincingly implicate a pathway in pathogenesis.

Copy number polymorphism

Chromosomal rearrangements (deletions, duplications and inversions) often occur as a result of non-allelic recombination (both intra- and inter-chromosomal) between regions of high homology [96]. Until recently, gene duplications and deletions have been regarded as mutations or abnormalities, often involving large-scale chromosomal rearrangements and associated with rare, sporadic and typically severe phenotypes [97]. During the last decade, however, a number of variable sequence-length repeats and copy number-variable (CNV) genes have been identified within healthy populations [reviewed in [98]]. Over the last 3 years, an increasing number of duplications and deletions have also been described. These range in length from tens to hundreds of kilobases, and many occur with sufficient frequency (>1 %) to be considered genuine, heritable polymorphisms. Three array-

based studies in 2004 and 2005 found some 393 individual CNVs (mainly duplications) encompassing several hundred known genes [99–101]. In a parallel approach, the potential for CNV to confound SNP-based association studies has been exploited to identify polymorphic deletions. The first such study, by Conrad et al. [102], analysed HapMap data for SNPs showing non-Mendelian inheritance patterns and estimated that a typical individual carries 30–50 hemizygous deletions of more than 5 kb. Although these regions tended to be gene poor, they included 267 known and predicted genes. The Structural Variation Database (http://paralogy.gs.washington.edu/structuralvariation) is an online database of regional CNVS which, as of October 2006, described nearly 4,000 CNV loci. Many potential or confirmed CNV genes are involved in environmentally responsive processes such as metabolism [for example various cytochromes, Lp(a), neuropeptide Y] and immune function (e.g. CCL3L1, β-defensin).

One example of a demonstrated relationship between CNV and complex disease is the chemokine gene *CCL3L1* (17q12), which arises from segmental duplication of *CCL3*. First identified in 1990 [103], the copy number ranges from 0 to 7 in Caucasian populations (median CN=2), and gene dosage correlates with both the ratio of *CCL3L1*:*CCL3* transcription and with chemokine production [104]. No disease association was identified until 15 years later, when it was found to influence susceptibility to HIV-1 [94], Kawasaki disease [105] and RA [our unpublished results].

Inconsistency in gene association studies

The vast majority of reports of genetic associations with common disease have been plagued by a lack of convincing replication between studies. Owing to publication bias towards positive associations, the typical scenario is an initial positive study that does not include replication in the study design and is not replicated in larger, well-powered studies. Some examples from T1D are the 2′5′-oligoadenylate synthetase, interleukin-4, -13 and -18, interleukin-4 receptor A, catalase, PAX4, TATA box-binding protein, NeuroD and vitamin D receptor genes – the following citations all derive from the group of John Todd (Cambridge, UK) and all describe lack of replication of previously published genetic association studies in cohorts of significantly greater power [106–113]. A combination of reasons is likely to account for this. Firstly, the initial studies reporting association of genetic variant(s) with disease may represent a false-positive association. There is little consensus on the level of significance required to report association, but this level needs to account for multiple testing which, when stringently applied, should be adjusted for the testing done by all investigators rather than in the individual laboratory. To determine the appropriate significance level requires knowledge of the prior probability of association of a trait to a specific candidate gene – these estimates are subjective and hypothesis based. Thus, the field has tended not to apply stringent significance levels to association studies in common disease. It is a characteristic of epidemiological studies that the first report of a disease-associated risk factor tends to overestimate the relative risk because of publication bias. Thus, assuming that an initial report of genetic association with autoimmunity represents a true association, most replication studies are under-powered, even if they are of greater size than the initial study [43]. These issues have been empirically investigated by meta-analysis of published genetic association studies [114, 115]. These meta-analyses demonstrate that results of an initial study reporting genetic association correlate only modestly with follow-up replication studies. In about half of the genetic studies, follow-up studies provided evidence that the initial association was likely to be real, although in most cases, the associations have not been widely replicated. A prerequisite for publishing contemporary gene association studies in higher impact journals is that study design includes replication (see e.g. association of *IRF5* and *CD25* with systemic lupus erythematosus [24] and T1D [116]).

Genotyping errors are a rarely discussed confounding variable in all association studies [117]. These can be caused by variation in DNA sequence within PCR primers, low quantity and/or quality of genomic DNA (which can result in allelic dropout and problems from contamination), biochemical artefacts and, of course, human error. In case-control studies, genotyping errors or missing data should affect both cohorts equally; however this may not be the situation when cases and controls are not genotyped in an identical manner (e.g. genotyping on different days). Even with rejection of any incomplete genotype data and duplicate genotyping of a proportion of both case and control cohorts, some spurious studies may still be published in good faith [118, 119]. Checking of Hardy-Weinberg equilibria is an important aspect of validating genotype data [107, 112, 120] (The Hardy-Weinberg equilibrium is a fundamental principle in population genetics; the genotype frequencies of a large, randomly mating population remain constant provided immigration, mutation and selection do not take place).

Other factors relate to population differences, best illustrated in identification of the *PADI4* gene in RA.

An initial report of association at SNPs within the *PADI4* gene with RA [121] was subsequently replicated in another Japanese cohort (OR=1.2–1.4) [122]. Individual studies that genotyped the same SNPs in Caucasians, however, showed no convincing evidence for association with RA in Caucasians [123–125], although meta-analysis has since demonstrated that *PADI4* does have a weak effect in this racial group (OR=1.1) [20]. There are two possible explanations: (i) *PADI4* does have a weaker effect in Caucasian and/or (ii) given that it is known that significant differences in allele frequency exist between Japanese and Caucasians at *PADI4* [126] a more appropriate, ethnically specific set of SNPs need to be typed in Caucasians before drawing any conclusions. In contrast, the *CARD15* gene is associated with CD in Caucasians [60, 61], but is not found as a susceptibility allele in the Asian population [64]. Thus, it remains imperative to study putative disease susceptibility alleles in populations of varied racial background using population-specific tag SNPs. Furthermore, examining association in different racial groups can give important clues to the identity of aetiologically important alleles [127]. This approach should prove useful in studying *INSIG2* in obesity (see above).

Heterogeneity in phenotype assessment between studies may also hamper replication of genetic associations. For example, ascertainment criteria differ considerably between cohorts reporting association of RA with the *PTPN22* R620W variant (e.g. probands of multiply affected families, patients with early RA, patients recruited from hospital-based clinics [128]). The variation in ascertainment criteria may be part of the reason for inconsistency of association with *PTPN22* R620W reported in the subgroup of RA patients defined clinically by the absence of rheumatoid factor (an anti-IgG antibody present in the majority of patients with RA that has an unclear role in disease aetiology). This source of heterogeneity can be decreased by reducing a phenotype to measurable component subphenotypes. For example, acute gout is a result of an immune response to uric acid crystals deposited in the joints, and serum urate levels are positively correlated with the risk of developing acute gout. It would be expected that the number of environmental and genetic risk factors influencing serum urate would be less than those involved in the subsequent observation of an acute gout attack. Thus, genetic factors contributing to the intermediate phenotype of serum urate levels should be easier to identify.

Genetic heterogeneity (the production of identical or similar phenotypes by different genetic mechanisms) can be reduced by stratifying cases according to the presence or absence of established susceptibility alleles. For example, it is routine to examine HLA-DRB1 'positive' and 'negative' cases in autoimmunity and *CARD15* 'positive' and 'negative' cases in CD.

Concluding remarks

Testing of candidate and positional candidate genes has resulted in a limited set of genes confirmed as playing a role in common disease (Table 1). Most putative genetic associations are plagued by lack of replication. One by one, the reasons for this are being widely acknowledged, understood and applied by the biomedical genetics community. At present, the best gene association studies are addressing these issues using tag SNPs that adequately assess variation in candidate genes, using large cohorts, reducing heterogeneity and incorporating replication into study design. The number of common disease genes being identified is increasing exponentially. We expect the next 5 years to bring dozens of validated genetic associations, largely as an outcome of WGA scanning. Results of full-genome association analysis are likely to produce a bewildering array of data, some of which will provide vital new biological insights into autoimmunity that may hold the key to novel therapies. Of previously identified genes, aside from a handful of examples (Table 1), the individual effect of each variant is weak. However, the insight these associations give into molecular pathways of disease is incalculable and the major motivation for biomedical genetics. Not only do genetic associations uncover molecular pathways, they also highlight disease checkpoints, progression through which requires inheritance of particular genetic variants and which can be blocked by others. For example, inheritance of the *HLA-DRB1*02* allele is sufficient to block T1D, perhaps by ensuring that autoreactive T cells are negatively selected in the thymus [129]. The six confirmed genetic associations in T1D have significantly added to the picture we have of disease pathogenesis (Table 3).

Knowledge of both the disease pathways *per se* and, of most importance, possible checkpoints (as defined by genetic association) should identify good targets for intervention – 'if this pathway is blocked at a certain point by the presence of a specific variant allele, can we functionally mimic this action by intervention?' For example: the checkpoint revealed by the PPARγ association in T2D is able to be targeted by agonists able to treat insulin resistance and dyslypidaemia [133]; clinical trials of a CTLA4Ig (abatacept) that modulates the CD28-mediated T cell co-stimulatory signal have provided preliminary evidence of its

Table 3. Information on T1D pathogenesis generated by genetic association studies.

Disease compartment	'Normal' function	Genetic data	Postulated mechanism in T1D
Thymic tolerance			
HLA-DRB1	Highly variant gene that encodes antigen-presenting molecules expressed on the surface of antigen-presenting cells such as macrophages and dendritic cells. Bind and present foreign antigen to antigen-specific T lymphocytes, thereby initiating an immune response.	The *HLA-DRB1*1501* and *HLA-DQB1*0602* alleles are strongly protective of T1D.	The *HLA-DRB1*1501* allele binds insulin peptides with much greater affinity than other alleles [129]. The disease protective allele ('class III') results in higher expression in the thymus and lower expression in the pancreas [130]. It is possible that infants with the *HLA-DRB1-DR2* allele and the protective *INS* allele are better able to delete auto-reactive T cells in the thymus (i.e. more insulin binds better to DR2, leading to strong recognition by auto-reactive T cells in the thymus and their consequent deletion). In the pancreas, people with the susceptibility *INS* allele produce higher levels of insulin, thereby increasing the amount of self-antigen available for autoimmune recognition.
INS	Hormone required for uptake of glucose from the bloodstream into cells.	A variable number of repeats (VNTR) variant in the insulin regulatory gene is associated with T1D [131].	
PTPN22	Associates with the Csk tyrosine kinase to mediate negative regulation of TCR signalling.	The 620W allele predisposes to T1D [8].	The *620W* allele is a gain-of-function variant, suppressing TCR signalling better than 620R [10]. This may aid the survival of autoreactive T cells that would have been activated and deleted in the thymus.
T cell regulation			
CTLA4	Regulates CD28-mediated TCR activation by inhibiting the TCR-signalling complex.	Disease-associated variants in the 3′UTR lead to reduced amounts of a soluble CTLA4 isoform that lacks exon 3 [11].	Unclear. Less-soluble CTLA4 could result in increased CD28-mediated TCR signalling.
CD25	Expression on regulatory T cells is essential for their function of suppressing T cell immune responses.	Haplotype tag SNPs were used to demonstrate association of *CD25* with T1D [117].	Unclear. A disease-associated *CD25* variant could result in T regulatory cells less able to suppress autoreactive T cells.
Pancreas			
HLA-DRB1	See above.	The *DR4* allele is strongly associated with T1D susceptibility.	The majority of DR4 subjects show autoreactivity to a proinsulin epitope, mediated by a specific DR4-restricted T cell clone [132]. The disease-associated VNTR allele at *INS* ('class I') results in higher insulin expression in the pancreas [130]. This allied to the presence of DR4-restricted insulin-specific T cells (and aberration in T cell regulation) increases the chance of β-islet autoimmunity.
INS	See above.	See above.	
Environment			
IFIH1	Receptor for double-stranded RNA from viral infections.	An A946T substitution is the most likely disease-causative variant in *IFIH1* [87].	Residue 946 is not in an obvious functional domain. However, the susceptibility Ala allele is fully conserved in mammals, suggesting it could be the functional variant [87].

efficacy in RA [134], and a range of agents that prevent the CCR5-mediated entry of HIV into T cells have been developed [135].

Translating genetic information into direct clinical benefit is a challenge. The combined positive predictive value of an array of modest genetic effects may well be specific, but not sufficiently sensitive for screening in the general population. However, understanding the genetic basis *per se* to common disease and the application of this knowledge to appropriate clinical trials may provide clinical benefits in several ways. First, individuals at the highest genetic risk will be able to be followed in prospective trials aimed at defining the role of tractable environmental influen-

ces on disease; second, susceptible individuals can be enrolled in trials to test preventative therapies; third, genetic profiles can be aligned with prognosis or response to specific treatments – for example, severity in RA – so that more appropriate or aggressive treatments can be targeted at an earlier stage, if beneficial; fourth, negative predictive value may be useful clinically – if a patient does not have known risk alleles for a disease, this may be useful in diagnosis; finally, analysis of combinations of susceptibility alleles may help to subclassify heterogenous diseases (e.g. diabetes, RA) into subgroups with distinguishing clinical features that could be utilized in clinical trials.

1 The International HapMap Consortium. (2003) The International HapMap Project. Nature 426, 789 – 796.

2 The International HapMap Consortium. (2005) A haplotype map of the human genome. Nature 437, 1299 – 1320.

3 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S. and Chen, Y. J. (2001) Initial sequencing and analysis of the human genome. Nature 409, 860 – 921.

4 Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. and Zhu, X. (2001) The sequence of the human genome. Science 291, 1304 – 1351.

5 Jones, E. Y., Fugger, L., Strominger, J. L. and Siebold, C. (2006) MHC class II proteins and disease: a structural perspective. Nat. Rev. Immunol. 6, 271 – 282.

6 Gregersen, P. K. and Batliwalla, F. (2005) PTPN22 and rheumatoid arthritis: gratifying replication. Arthritis Rheum. 52, 1952 – 1955.

7 Begovich, A. B., Carlton, V. E., Honigberg, L. A., Schrodi, S. J., Chokkalingam, A. P., Alexander, H. C., Ardlie, K. G., Huang, Q., Smith, A. M., Spoerke, J. M., Conn, M. T., Chang, M., Chang, S. Y., Saiki, R. K., Catanese, J. J., Leong, D. U., Garcia, V. E., McAllister, L. B., Jeffery, D. A., Lee, A. T., Batliwalla, F., Remmers, E., Criswell, L. A., Seldin, M. F., Kastner, D. L., Amos, C. I., Sninsky, J. J. and Gregersen, P. K. (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am. J. Hum. Genet. 75, 330 – 337.

8 Bottini, N., Musumeci, L., Alonso, A., Rahmouni, S., Nika, K., Rostamkhani, M., MacMurray, J., Meloni, G. F., Lucarelli, P., Pellecchia, M., Eisenbarth, G. S., Comings, D. and Mustelin, T. (2004) A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat. Genet. 36, 337 – 338.

9 Gregersen, P. K., Lee, H. S., Batliwalla, F. and Begovich, A. B. (2006) PTPN22: setting thresholds for autoimmunity. Semin. Immunol. 18, 214 – 223.

10 Vang, T., Congia, M., Macis, M. D., Musumeci, L., Orru, V., Zavattari, P., Nika, K., Tautz, L., Tasken, K., Cucca, F., Mustelin, T. and Bottini, N. (2005) Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. Nat. Genet. 37, 1317 – 1319.

11 Ueda, H., Howson, J. M., Esposito, L., Heward, J., Snook, H., Chamberlain, G., Rainbow, D. B., Hunter, K. M., Smith, A. N., Di Genova, G., Herr, M. H., Dahlman, I., Payne, F., Smyth, D., Lowe, C., Twells, R. C., Howlett, S., Healy, B., Nutland, S., Rance, H. E., Everett, V., Smink, L. J., Lam, A. C., Cordell, H. J., Walker, N. M., Bordin, C., Hulme, J., Motzo, C., Cucca, F., Hess, J. F., Metzker, M. L., Rogers, J., Gregory, S., Allahabadia, A., Nithiyananthan, R., Tuomilehto-Wolf, E., Tuomilehto, J., Bingley, P., Gillespie, K. M., Undlien, D. E., Ronningen, K. S., Guja, C., Ionescu-Tirgoviste, C., Savage, D. A., Maxwell, A. P., Carson, D. J., Patterson, C. C., Franklyn, J. A., Clayton, D. G., Peterson, L. B., Wicker, L. S., Todd, J. A. and Gough, S. C. (2003) Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 423, 506 – 511.

12 Thakkinstian, A., Han, P., McEvoy, M., Smith, W., Hoh, J., Magnusson, K., Zhang, K. and Attia, J. (2006) Systematic review and meta-analysis of the association between complementary factor H Y402H polymorphisms and age-related macular degeneration. Hum. Mol. Genet. 15, 2784 – 2790.

13 Farrer, L. A., Cupples, L. A., Haines, J. L., Hyman, B., Kukull, W. A., Mayeux, R., Myers, R. H., Pericak-Vance, M. A., Risch, N. and van Duijn, C. M. (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. JAMA 278, 1349 – 1356.

14 Lehmann, D. J., Cortina-Borja, M., Warden, D. R., Smith, A. D., Sleegers, K., Prince, J. A., van Duijn, C. M. and Kehoe, P. G. (2005) Large meta-analysis establishes the ACE insertion-deletion polymorphism as a marker of Alzheimer's disease. Am. J. Epidemiol. 162, 305 – 317.

15 Lee, Y. H., Rho, Y. H., Choi, S. J., Ji, J. D., Song, G. G., Nath, S. K. and Harley, J. B. (2007) The PTPN22 C1858T functional polymorphism and autoimmune diseases – a meta-analysis. Rheumatology 46, 49 – 56.

16 Plenge, R. M., Padyukov, L., Remmers, E. F., Purcell, S., Lee, A. T., Karlson, E. W., Wolfe, F., Kastner, D. L., Alfredsson, L., Altshuler, D., Gregersen, P. K., Klareskog, L. and Rioux, J. D. (2005) Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. Am. J. Hum. Genet. 77, 1044 – 1060.

17 Economou, M., Trikalinos, T. A., Loizou, K. T., Tsianos, E. V. and Ioannidis, J. P. (2004) Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. Am. J. Gastroenterol. 99, 2393 – 2404.

18 Oostenbrug, L. E., Nolte, I. M., Oosterom, E., van der Steege, G., Te Meerman, G. J., van Dullemen, H. M., Drenth, J. P., de Jong, D. J., van der Linde, K., Jansen, P. L. and Kleibeuker, J. H. (2006) CARD15 in inflammatory bowel disease and Crohn's disease phenotypes: an association study and pooled analysis. Dig. Liver Dis. 38, 834 – 845.

19 Herbert, A., Gerry, N. P., McQueen, M. B., Heid, I. M., Pfeufer, A., Illig, T., Wichmann, H. E., Meitinger, T., Hunter, D., Hu, F. B., Colditz, G., Hinney, A., Hebebrand, J., Koberwitz, K., Zhu, X., Cooper, R., Ardlie, K., Lyon, H., Hirschhorn, J. N., Laird, N. M., Lenburg, M. E., Lange, C. and Christman, M. F. (2006) A common genetic variant is associated with adult and childhood obesity. Science 312, 279 – 283.

20 Iwamoto, T., Ikari, K., Nakamura, T., Kuwahara, M., Toyama, Y., Tomatsu, T., Momohara, S. and Kamatani, N. (2006) Association between PADI4 and rheumatoid arthritis: a meta-analysis. Rheumatology 45, 804 – 807.

21 Ikari, K., Momohara, S., Nakamura, T., Hara, M., Yamanaka, H., Tomatsu, T. and Kamatani, N. (2006) Supportive evidence for a genetic association of the FCRL3 promoter polymorphism with rheumatoid arthritis. Ann. Rheum. Dis. 65, 671 – 673.

22 Kochi, Y., Yamada, R., Suzuki, A., Harley, J. B., Shirasawa, S., Sawada, T., Bae, S. C., Tokuhiro, S., Chang, X., Sekine, A., Takahashi, A., Tsunoda, T., Ohnishi, Y., Kaufman, K. M., Kang, C. P., Kang, C., Otsubo, S., Yumura, W., Mimori, A., Koike, T., Nakamura, Y., Sasazuki, T. and Yamamoto, K. (2005) A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. Nat. Genet. 37, 478 – 485.

23 Li, D., Collier, D. A. and He, L. (2006) Meta-analysis shows strong positive association of the neuregulin 1 (NRG1) gene with schizophrenia. Hum. Mol. Genet. 15, 1995 – 2002.

24 Graham, R. R., Kozyrev, S. V., Baechler, E. C., Reddy, M. V., Plenge, R. M., Bauer, J. W., Ortmann, W. A., Koeuth, T., Gonzalez Escribano, M. F., Pons-Estel, B., Petri, M., Daly, M., Gregersen, P. K., Martin, J., Altshuler, D., Behrens, T. W. and Alarcon-Riquelme, M. E. (2006) A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. Nat. Genet. 38, 550 – 555.

25 Sigurdsson, S., Nordmark, G., Goring, H. H., Lindroos, K., Wiman, A. C., Sturfelt, G., Jonsen, A., Rantapaa-Dahlqvist, S., Moller, B., Kere, J., Koskenmies, S., Widen, E., Eloranta, M. L., Julkunen, H., Kristjansdottir, H., Steinsson, K., Alm, G., Ronnblom, L. and Syvanen, A. C. (2005) Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. Am. J. Hum. Genet. 76, 528 – 537.

26 Cauchi, S., Meyre, D., Dina, C., Choquet, H., Samson, C., Gallina, S., Balkau, B., Charpentier, G., Pattou, F., Stetsyuk, V., Scharfmann, R., Staels, B., Fruhbeck, G. and Froguel, P. (2006) Transcription factor TCF7L2 genetic study in the french population: expression in human {beta}-cells and adipose tissue and strong association with type 2 diabetes. Diabetes 55, 2903 – 2908.

27 Damcott, C. M., Pollin, T. I., Reinhart, L. J., Ott, S. H., Shen, H., Silver, K. D., Mitchell, B. D. and Shuldiner, A. R. (2006) Polymorphisms in the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in the Amish: replication and evidence for a role in both insulin secretion and insulin resistance. Diabetes 55, 2654 – 2659.

28 Florez, J. C., Jablonski, K. A., Bayley, N., Pollin, T. I., de Bakker, P. I., Shuldiner, A. R., Knowler, W. C., Nathan, D. M. and Altshuler, D. (2006) TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. N. Engl. J. Med. 355, 241 – 250.

29 Grant, S. F., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J., Helgason, A., Stefansson, H.,

Emilsson, V., Helgadottir, A., Styrkarsdottir, U., Magnusson, K. P., Walters, G. B., Palsdottir, E., Jonsdottir, T., Gudmundsdottir, T., Gylfason, A., Saemundsdottir, J., Wilensky, R. L., Reilly, M. P., Rader, D. J., Bagger, Y., Christiansen, C., Gudnason, V., Sigurdsson, G., Thorsteinsdottir, U., Gulcher, J. R., Kong, A. and Stefansson, K. (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat. Genet. 38, 320 – 323.

30 Saxena, R., Gianniny, L., Burtt, N. P., Lyssenko, V., Giuducci, C., Sjogren, M., Florez, J. C., Almgren, P., Isomaa, B., Orho-Melander, M., Lindblad, U., Daly, M. J., Tuomi, T., Hirschhorn, J. N., Ardlie, K. G., Groop, L. C. and Altshuler, D. (2006) Common single nucleotide polymorphisms in TCF7L2 are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. Diabetes 55, 2890 – 2895.

31 Scott, L. J., Bonnycastle, L. L., Willer, C. J., Sprau, A. G., Jackson, A. U., Narisu, N., Duren, W. L., Chines, P. S., Stringham, H. M., Erdos, M. R., Valle, T. T., Tuomilehto, J., Bergman, R. N., Mohlke, K. L., Collins, F. S. and Boehnke, M. (2006) Association of transcription factor 7-like 2 (TCF7L2) variants with type 2 diabetes in a Finnish sample. Diabetes 55, 2649 – 2653.

32 Zhang, C., Qi, L., Hunter, D. J., Meigs, J. B., Manson, J. E., van Dam, R. M. and Hu, F. B. (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene and the risk of type 2 diabetes in large cohorts of, U.S. women and men. Diabetes 55, 2645 – 2648.

33 Tsuchiya, T., Schwarz, P. E., Bosque-Plata, L. D., Geoffrey Hayes, M., Dina, C., Froguel, P., Wayne Towers, G., Fischer, S., Temelkova-Kurktschiev, T., Rietzsch, H., Graessler, J., Vcelak, J., Palyzova, D., Selisko, T., Bendlova, B., Schulze, J., Julius, U., Hanefeld, M., Weedon, M. N., Evans, J. C., Frayling, T. M., Hattersley, A. T., Orho-Melander, M., Groop, L., Malecki, M. T., Hansen, T., Pedersen, O., Fingerlin, T. E., Boehnke, M., Hanis, C. L., Cox, N. J. and Bell, G. I. (2006) Association of the calpain-10 gene with type 2 diabetes in Europeans: results of pooled and meta-analyses. Mol. Genet. Metab. 89, 174 – 184.

34 Altshuler, D., Hirschhorn, J. N., Klannemark, M., Lindgren, C. M., Vohl, M. C., Nemesh, J., Lane, C. R., Schaffner, S. F., Bolk, S., Brewer, C., Tuomi, T., Gaudet, D., Hudson, T. J., Daly, M., Groop, L. and Lander, E. S. (2000) The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat. Genet. 26, 76 – 80.

35 Davies, J. L., Kawaguchi, Y., Bennett, S. T., Copeman, J. B., Cordell, H. J., Pritchard, L. E., Reed, P. W., Gough, S. C., Jenkins, S. C., Palmer, S. M., Balfour, K. M., Rowe, B. R., Farrall, M., Barnett, A. H., Bain, S. C. and Todd, J. (1994) A genome-wide search for human type 1 diabetes susceptibility genes. Nature 371, 130 – 136.

36 Hashimoto, L., Habita, C., Beressi, J. P., Delepine, M., Besse, C., Cambon-Thomsen, A., Deschamps, I., Rotter, J. I., Djoulah, S. and James, M. R. (1994) Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. Nature 371, 161 – 164.

37 Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C. and Hoh, J. (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308, 385 – 389.

38 Maraganore, D. M., de Andrade, M., Lesnick, T. G., Strain, K. J., Farrer, M. J., Rocca, W. A., Pant, P. V., Frazer, K. A., Cox, D. R. and Ballinger, D. G. (2005) High-resolution whole-genome association study of Parkinson disease. Am. J. Hum. Genet. 77, 685 – 693.

39 Lander, E. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat. Genet. 11, 241 – 247.

40 Concannon, P., Gogolin-Ewens, K. J., Hinds, D. A., Wapelhorst, B., Morrison, V. A., Stirling, B., Mitra, M., Farmer, J., Williams, S. R., Cox, N. J., Bell, G. I., Risch, N. and Spielman, R. S. (1998) A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. Nat. Genet. 19, 292 – 296.

41 Lernmark, A. and Ott, J. (1998) Sometimes it's hot, sometimes it's not. Nat. Genet. 19, 213 – 214.

42 Mein, C. A., Esposito, L., Dunn, M. G., Johnson, G. C., Timms, A. E., Goy, J. V., Smith, A. N., Sebag-Montefiore, L., Merriman, M. E., Wilson, A. J., Pritchard, L. E., Cucca, F., Barnett, A. H., Bain, S. C. and Todd, J. A. (1998) A search for type 1 diabetes susceptibility genes in families from the United Kingdom. Nat. Genet. 19, 297 – 300.

43 Suarez, B., Hampe, C. and Eerdewegh, P. V. (1994) Problems of replicating linkage chlaims in psychiatry. in: Genetic Approaches to Mental Disorders (Gershon, E. and Cloninger, C., Eds); pp 23 – 46, American Psychiatric Press, Washington DC.

44 Wu, X., Cooper, R. S., Borecki, I., Hanis, C., Bray, M., Lewis, C. E., Zhu, X., Kan, D., Luke, A. and Curb, D. (2002) A combined analysis of genomewide linkage scans for body mass index from the National Heart, Lung, and Blood Institute Family Blood Pressure Program. Am. J. Hum. Genet. 70, 1247 – 1256.

45 An, P., Freedman, B. I., Hanis, C. L., Chen, Y. D., Weder, A. B., Schork, N. J., Boerwinkle, E., Province, M. A., Hsiung, C. A., Wu, X., Quertermous, T. and Rao, D. C. (2005) Genome-wide linkage scans for fasting glucose, insulin, and insulin resistance in the National Heart, Lung, and Blood Institute Family Blood Pressure Program: evidence of linkages to chromosome 7q36 and 19q13 from meta-analysis. Diabetes 54, 909 – 914.

46 Demenais, F., Kanninen, T., Lindgren, C. M., Wiltshire, S., Gaget, S., Dandrieux, C., Almgren, P., Sjogren, M., Hattersley, A., Dina, C., Tuomi, T., McCarthy, M. I., Froguel, P. and Groop, L. C. (2003) A meta-analysis of four European genome screens (GIFT Consortium) shows evidence for a novel region on chromosome 17p11.2-q22 linked to type 2 Diabetes. Hum. Mol. Genet. 12, 1865 – 1873.

47 Wu, X., Kan, D., Province, M., Quertermous, T., Rao, D. C., Chang, C., Mosley, T. H., Curb, D., Boerwinkle, E. and Cooper, R. S. (2006) An updated meta-analysis of genome scans for hypertension and blood pressure in the NHLBI Family Blood Pressure Program (FBPP). Am. J. Hypertens. 19, 122 – 127.

48 Lewis, C. M., Levinson, D. F., Wise, L. H., DeLisi, L. E., Straub, R. E., Hovatta, I., Williams, N. M., Schwab, S. G., Pulver, A. E., Faraone, S. V., Brzustowicz, L. M., Kaufmann, C. A., Garver, D. L., Gurling, H. M., Lindholm, E., Coon, H., Moises, H. W., Byerley, W., Shaw, S. H., Mesen, A., Sherrington, R., O'Neill, F. A., Walsh, D., Kendler, K. S., Ekelund, J., Paunio, T., Lonnqvist, J., Peltonen, L., O'Donovan, M. C., Owen, M. J., Wildenauer, D. B., Maier, W., Nestadt, G., Blouin, J. L., Antonarakis, S. E., Mowry, B. J., Silverman, J. M., Crowe, R. R., Cloninger, C. R., Tsuang, M. T., Malaspina, D., Harkavy-Friedman, J. M., Svrakic, D. M., Bassett, A. S., Holcomb, J., Kalsi, G., McQuillin, A., Brynjolfson, J., Sigmundsson, T., Petursson, H., Jazin, E., Zoega, T. and Helgason, T. (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder. II. Schizophrenia. Am. J. Hum. Genet. 73, 34 – 48.

49 Segurado, R., Detera-Wadleigh, S. D., Levinson, D. F., Lewis, C. M., Gill, M., Nurnberger, J. I., Jr., Craddock, N., DePaulo, J. R., Baron, M., Gershon, E. S., Ekholm, J., Cichon, S., Turecki, G., Claes, S., Kelsoe, J. R., Schofield, P. R., Badenhop, R. F., Morissette, J., Coon, H., Blackwood, D., McInnes, L. A., Foroud, T., Edenberg, H. J., Reich, T., Rice, J. P., Goate, A., McInnis, M. G., McMahon, F. J., Badner, J. A., Goldin, L. R., Bennett, P., Willour, V. L., Zandi, P. P., Liu, J., Gilliam, C., Juo, S. H., Berrettini, W. H., Yoshikawa, T., Peltonen, L., Lonnqvist, J., Nothen, M. M., Schumacher, J., Windemuth, C., Rietschel, M., Propping, P., Maier, W., Alda, M., Grof, P., Rouleau, G. A., Del-Favero, J., Van Broeckhoven, C., Mendlewicz, J., Adolfsson, R., Spence, M. A.,

Luebbert, H., Adams, L. J., Donald, J. A., Mitchell, P. B., Barden, N., Shink, E., Byerley, W., Muir, W., Visscher, P. M., Macgregor, S., Gurling, H., Kalsi, G., McQuillin, A., Escamilla, M. A., Reus, V. I., Leon, P., Freimer, N. B., Ewald, H., Kruse, T. A., Mors, O., Radhakrishna, U., Blouin, J. L., Antonarakis, S. E. and Akarsu, N. (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder. III. Bipolar disorder. Am. J. Hum. Genet. 73, 49 – 62.

50 van Heel, D. A., Fisher, S. A., Kirby, A., Daly, M. J., Rioux, J. D. and Lewis, C. M. (2004) Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. Hum. Mol. Genet. 13, 763 – 770.

51 Lee, Y. H., Rho, Y. H., Choi, S. J., Ji, J. D. and Song, G. G. (2006) Meta-analysis of genome-wide linkage studies for bone mineral density. J. Hum. Genet. 51, 480 – 486.

52 Concannon, P., Erlich, H. A., Julier, C., Morahan, G., Nerup, J., Pociot, F., Todd, J. A. and Rich, S. S. (2005) Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. Diabetes 54, 2995 – 3001.

53 Lee, Y. H. and Nath, S. K. (2005) Systemic lupus erythematosus susceptibility loci defined by genome scan meta-analysis. Hum. Genet. 118, 434 – 443.

54 Choi, S. J., Rho, Y. H., Ji, J. D., Song, G. G. and Lee, Y. H. (2006) Genome scan meta-analysis of rheumatoid arthritis. Rheumatology 45, 166 – 170.

55 Lee, Y. H., Rho, Y. H., Choi, S. J., Ji, J. D. and Song, G. G. (2005) Ankylosing spondylitis susceptibility loci defined by genome-search meta-analysis. J. Hum. Genet. 50, 453 – 459.

56 Sagoo, G. S., Tazi-Ahnini, R., Barker, J. W., Elder, J. T., Nair, R. P., Samuelsson, L., Traupe, H., Trembath, R. C., Robinson, D. A. and Iles, M. M. (2004) Meta-analysis of genome-wide studies of psoriasis susceptibility reveals linkage to chromosomes 6p21 and 4q28-q31 in Caucasian and Chinese Hans population. J. Invest. Dermatol. 122, 1401 – 1405.

57 GAMES. Transatlantic Multiple Sclerosis Genetics Cooperative. (2003) A meta-analysis of whole genome linkage screens in multiple sclerosis. J. Neuroimmunol. 143, 39 – 46.

58 Ylisaukko-oja, T., Alarcon, M., Cantor, R. M., Auranen, M., Vanhala, R., Kempas, E., von Wendt, L., Jarvela, I., Geschwind, D. H. and Peltonen, L. (2006) Search for autism loci by combined analysis of Autism Genetic Resource Exchange and Finnish families. Ann. Neurol. 59, 145 – 155.

59 Chiodini, B. D. and Lewis, C. M. (2003) Meta-analysis of 4 coronary heart disease genome-wide linkage studies confirms a susceptibility locus on chromosome 3q. Arterioscler. Thromb. Vasc. Biol. 23, 1863 – 1868.

60 Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C. A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J. F., Sahbatou, M. and Thomas, G. (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411, 599 – 603.

61 Ogura, Y., Bonen, D. K., Inohara, N., Nicolae, D. L., Chen, F. F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R. H., Achkar, J. P., Brant, S. R., Bayless, T. M., Kirschner, B. S., Hanauer, S. B., Nunez, G. and Cho, J. H. (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 411, 603 – 606.

62 Hugot, J. P., Laurent-Puig, P., Gower-Rousseau, C., Olson, J. M., Lee, J. C., Beaugerie, L., Naom, I., Dupas, J. L., Van Gossum, A., Orholm, M., Bonaiti-Pellie, C., Weissenbach, J., Mathew, C. G., Lennard-Jones, J. E., Cortot, A., Colombel, J. F. and Thomas, G. (1996) Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature 379, 821 – 823.

63 Mahida, Y. R. and Johal, S. (2001) NF-kappa B may determine whether epithelial cell-microbial interactions in the intestine are hostile or friendly. Clin. Exp. Immunol. 123, 347 – 349.

64 Gaya, D. R., Russell, R. K., Nimmo, E. R. and Satsangi, J. (2006) New genes in inflammatory bowel disease: lessons for complex diseases? Lancet 367, 1271 – 1284.

65 Hill, N. J., Lyons, P. A., Armitage, N., Todd, J. A., Wicker, L. S. and Peterson, L. B. (2000) NOD Idd5 locus controls insulitis and diabetes and overlaps the orthologous CTLA4/IDDM12 and NRAMP1 loci in humans. Diabetes 49, 1744 – 1747.

66 Hollis-Moffatt, J. E., Hook, S. M. and Merriman, T. R. (2005) Colocalization of mouse autoimmune diabetes loci Idd21.1 and Idd21.2 with IDDM6 (human) and Iddm3 (rat). Diabetes 54, 2820 – 2825.

67 Lyons, P. A., Hancock, W. W., Denny, P., Lord, C. J., Hill, N. J., Armitage, N., Siegmund, T., Todd, J. A., Phillips, M. S., Hess, J. F., Chen, S. L., Fischer, P. A., Peterson, L. B. and Wicker, L. S. (2000) The NOD Idd9 genetic interval influences the pathogenicity of insulitis and contains molecular variants of Cd30, Tnfr2, and Cd137. Immunity 13, 107 – 115.

68 Podolin, P. L., Denny, P., Lord, C. J., Hill, N. J., Todd, J. A., Peterson, L. B., Wicker, L. S. and Lyons, P. A. (1997) Congenic mapping of the insulin-dependent diabetes (Idd) gene, Idd10, localizes two genes mediating the Idd10 effect and eliminates the candidate Fcgr1. J. Immunol. 159, 1835 – 1843.

69 Steer, S., Abkevich, V., Gutin, A., Cordell, H. J., Gendall, K. L., Merriman, M. E., Rodger, R. A., Rowley, K. A., Chapman, P., Gow, P., Harrison, A. a., Highton, J., Jones, P. B. B., O'Donnell, J., stamp, L., Fitzgerald, L., Iliev, D., Kouzmine, A., Tran, T., Skolnick, M. H., Timms, K. M., Lanchbury, J. S. and Merriman, T. R. (2007) Genomic DNA pooling for whole genome association scans in complex disease: empirical demonstration of efficacy in rheumatoid arthritis. Genes Immun. 8, 57 – 68.

70 Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. Science 273, 1516 – 1517.

71 Gunderson, K. L., Kuhn, K. M., Steemers, F. J., Ng, P., Murray, S. S. and Shen, R. (2006) Whole-genome genotyping of haplotype tag single nucleotide polymorphisms. Pharmacogenomics 7, 641 – 648.

72 Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G. and Chee, M. S. (2005) A genome-wide scalable SNP genotyping assay using microarray technology. Nat. Genet. 37, 549 – 554.

73 Matsuzaki, H., Dong, S., Loi, H., Di, X., Liu, G., Hubbell, E., Law, J., Berntsen, T., Chadha, M., Hui, H., Yang, G., Kennedy, G. C., Webster, T. A., Cawley, S., Walsh, P. S., Jones, K. W., Fodor, S. P. and Mei, R. (2004) Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. Nat. Methods 1, 109 – 111.

74 Hirschhorn, J. N. and Daly, M. J. (2005) Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6, 95 – 108.

75 Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J. and Altshuler, D. (2002) The structure of haplotype blocks in the human genome. Science 296, 2225 – 2229.

76 Johnson, G. C., Esposito, L., Barratt, B. J., Smith, A. N., Heward, J., Di Genova, G., Ueda, H., Cordell, H. J., Eaves, I. A., Dudbridge, F., Twells, R. C., Payne, F., Hughes, W., Nutland, S., Stevens, H., Carr, P., Tuomilehto-Wolf, E., Tuomilehto, J., Gough, S. C., Clayton, D. G. and Todd, J. A. (2001) Haplotype tagging for the identification of common disease genes. Nat. Genet. 29, 233 – 237.

77 Pe'er, I., de Bakker, P. I., Maller, J., Yelensky, R., Altshuler, D. and Daly, M. J. (2006) Evaluating and improving power in whole-genome association studies using fixed marker sets. Nat. Genet. 38, 663 – 667.

78 Wang, W. Y., Barratt, B. J., Clayton, D. G. and Todd, J. A. (2005) Genome-wide association studies: theoretical and practical concerns. Nat. Rev. Genet. 6, 109 – 118.

79 Hageman, G. S., Anderson, D. H., Johnson, L. V., Hancox, L. S., Taiber, A. J., Hardisty, L. I., Hageman, J. L., Stockman, H. A., Borchardt, J. D., Gehrs, K. M., Smith, R. J., Silvestri, G., Russell, S. R., Klaver, C. C., Barbazetto, I., Chang, S., Yannuzzi, L. A., Barile, G. R., Merriam, J. C., Smith, R. T., Olsh, A. K., Bergeron, J., Zernant, J., Merriam, J. E., Gold,

B., Dean, M. and Allikmets, R. (2005) A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. Proc. Natl. Acad. Sci. USA 102, 7227 – 7232.
80 Souied, E. H., Leveziel, N., Richard, F., Dragon-Durey, M. A., Coscas, G., Soubrane, G., Benlian, P. and Fremeaux-Bacchi, V. (2005) Y402H complement factor H polymorphism associated with exudative age-related macular degeneration in the French population. Mol. Vis. 11, 1135 – 1140.
81 Bonuccelli, U. and Del Dotto, P. (2006) New pharmacologic horizons in the treatment of Parkinson disease. Neurology 67, S30 – S38.
82 Clarimon, J., Scholz, S., Fung, H. C., Hardy, J., Eerola, J., Hellstrom, O., Chen, C. M., Wu, Y. R., Tienari, P. J. and Singleton, A. (2006) Conflicting results regarding the semaphorin gene (SEMA5A) and the risk for Parkinson disease. Am. J. Hum. Genet. 78, 1082 – 1084.
83 Farrer, M. J., Haugarvoll, K., Ross, O. A., Stone, J. T., Milkovic, N. M., Cobb, S. A., Whittle, A. J., Lincoln, S. J., Hulihan, M. M., Heckman, M. G., White, L. R., Aasly, J. O., Gibson, J. M., Gosal, D., Lynch, T., Wszolek, Z. K., Uitti, R. J. and Toft, M. (2006) Genomewide association, Parkinson disease, and PARK10. Am. J. Hum. Genet. 78, 1084 – 1088.
84 Goris, A., Williams-Gray, C. H., Foltynie, T., Compston, D. A., Barker, R. A. and Sawcer, S. J. (2006) No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. Am. J. Hum. Genet. 78, 1088 – 1090.
85 Li, Y., Rowland, C., Schrodi, S., Laird, W., Tacey, K., Ross, D., Leong, D., Catanese, J., Sninsky, J. and Grupe, A. (2006) A case-control association study of the 12 single-nucleotide polymorphisms implicated in Parkinson disease by a recent genome scan. Am. J. Hum. Genet. 78, 1090 – 1092.
86 Van Steen, K., McQueen, M. B., Herbert, A., Raby, B., Lyon, H., Demeo, D. L., Murphy, A., Su, J., Datta, S., Rosenow, C., Christman, M., Silverman, E. K., Laird, N. M., Weiss, S. T. and Lange, C. (2005) Genomic screening and replication using the same data set in family-based association testing. Nat. Genet. 37, 683 – 691.
87 Smyth, D. J., Cooper, J. D., Bailey, R., Field, S., Burren, O., Smink, L. J., Guja, C., Ionescu-Tirgoviste, C., Widmer, B., Dunger, D. B., Savage, D. A., Walker, N. M., Clayton, D. G. and Todd, J. A. (2006) A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. Nat. Genet. 38, 617 – 619.
88 Knip, M., Veijola, R., Virtanen, S. M., Hyoty, H., Vaarala, O. and Akerblom, H. K. (2005) Environmental triggers and determinants of type 1 diabetes. Diabetes 54 (Suppl. 2), S125 – S136.
89 Marron, M. P., Raffel, L. J., Garchon, H. J., Jacob, C. O., Serrano-Rios, M., Martinez Larrad, M. T., Teng, W. P., Park, Y., Zhang, Z. X., Goldstein, D. R., Tao, Y. W., Beaurain, G., Bach, J. F., Huang, H. S., Luo, D. F., Zeidler, A., Rotter, J. I., Yang, M. C., Modilevsky, T., Maclaren, N. K. and She, J. X. (1997) Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA4 polymorphisms in multiple ethnic groups. Hum. Mol. Genet. 6, 1275 – 1282.
90 Nistico, L., Buzzetti, R., Pritchard, L. E., Van der Auwera, B., Giovannini, C., Bosi, E., Larrad, M. T., Rios, M. S., Chow, C. C., Cockram, C. S., Jacobs, K., Mijovic, C., Bain, S. C., Barnett, A. H., Vandewalle, C. L., Schuit, F., Gorus, F. K., Tosi, R., Pozzilli, P. and Todd, J. A. (1996) The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. Hum. Mol. Genet. 5, 1075 – 1080.
91 Maurer, M., Loserth, S., Kolb-Maurer, A., Ponath, A., Wiese, S., Kruse, N. and Rieckmann, P. (2002) A polymorphism in the human cytotoxic T-lymphocyte antigen 4 (CTLA4) gene (exon 1 +49) alters T-cell activation. Immunogenetics 54, 1 – 8.
92 Daly, M. J. and Altshuler, D. (2005) Partners in crime. Nat. Genet. 37, 337 – 338.
93 Marchini, J., Donnelly, P. and Cardon, L. R. (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat. Genet. 37, 413 – 417.
94 Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R. J., Freedman, B. I., Quinones, M. P., Bamshad, M. J., Murthy, K. K., Rovin, B. H., Bradley, W., Clark, R. A., Anderson, S. A., O'Connell R J., Agan, B. K., Ahuja, S. S., Bologna, R., Sen, L., Dolan, M. J. and Ahuja, S. K. (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307, 1434 – 1440.
95 Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., Goedert, J. J., Buchbinder, S. P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R. and O'Brien, S. J. (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science 273, 1856 – 1862.
96 Stankiewicz, P. and Lupski, J. R. (2002) Molecular-evolutionary mechanisms for genomic disorders. Curr. Opin. Genet. Dev. 12, 312 – 319.
97 Lupski, J. R. (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends. Genet. 14, 417 – 422.
98 Buckland, P. R. (2003) Polymorphically duplicated genes: their relevance to phenotypic variation in humans. Ann. Med. 35, 308 – 315.
99 Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., Scherer, S. W. and Lee, C. (2004) Detection of large-scale variation in the human genome. Nat. Genet. 36, 949 – 951.
100 Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T. C., Trask, B., Patterson, N., Zetterberg, A. and Wigler, M. (2004) Large-scale copy number polymorphism in the human genome. Science 305, 525 – 528.
101 Sharp, A. J., Locke, D. P., McGrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U., Pertz, L. M., Clark, R. A., Schwartz, S., Segraves, R., Oseroff, V. V., Albertson, D. G., Pinkel, D. and Eichler, E. E. (2005) Segmental duplications and copy-number variation in the human genome. Am. J. Hum. Genet. 77, 78 – 88.
102 Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E. and Pritchard, J. K. (2006) A high-resolution survey of deletion polymorphism in the human genome. Nat. Genet. 38, 75 – 81.
103 Irving, S. G., Zipfel, P. F., Balke, J., McBride, O. W., Morton, C. C., Burd, P. R., Siebenlist, U. and Kelly, K. (1990) Two inflammatory mediator cytokine genes are closely linked and variably amplified on chromosome 17q. Nucleic Acids Res. 18, 3261 – 3270.
104 Townson, J. R., Barcellos, L. F. and Nibbs, R. J. (2002) Gene copy number regulates the production of the human chemokine CCL3-L1. Eur. J. Immunol. 32, 3016 – 3026.
105 Burns, J. C., Shimizu, C., Gonzalez, E., Kulkarni, H., Patel, S., Shike, H., Sundel, R. S., Newburger, J. W. and Ahuja, S. K. (2005) Genetic variations in the receptor-ligand pair CCR5 and CCL3L1 are important determinants of susceptibility to Kawasaki disease. J. Infect. Dis. 192, 344 – 349.
106 Maier, L. M., Chapman, J., Howson, J. M., Clayton, D. G., Pask, R., Strachan, D. P., McArdle, W. L., Twells, R. C. and Todd, J. A. (2005) No evidence of association or interaction between the IL4RA, IL4, and IL13 genes in type 1 Diabetes. Am. J. Hum. Genet. 76, 517 – 521.
107 Maier, L. M., Cooper, J. D., Walker, N., Smyth, D. J. and Todd, J. A. (2005) Comment to: Biason-Lauber A, Boehm B, Lang-Muritano M et al. (2005) association of childhood type 1

diabetes mellitus with a variant of PAX4: possible link to beta cell regenerative capacity. Diabetologia 48:900 – 905, 2180 – 2182.
108 Nejentsev, S., Cooper, J. D., Godfrey, L., Howson, J. M., Rance, H., Nutland, S., Walker, N. M., Guja, C., Ionescu-Tirgoviste, C., Savage, D. A., Undlien, D. E., Ronningen, K. S., Tuomilehto-Wolf, E., Tuomilehto, J., Gillespie, K. M., Ring, S. M., Strachan, D. P., Widmer, B., Dunger, D. and Todd, J. A. (2004) Analysis of the vitamin D receptor gene sequence variants in type 1 diabetes. Diabetes 53, 2709 – 2712.
109 Pask, R., Cooper, J. D., Walker, N. M., Nutland, S., Hutchings, J., Dunger, D. B., Nejentsev, S. and Todd, J. A. (2006) No evidence for a major effect of two common polymorphisms of the catalase gene in type 1 diabetes susceptibility. Diabetes Metab. Res. Rev. 22, 356 – 360.
110 Payne, F., Smyth, D. J., Pask, R., Cooper, J. D., Masters, J., Wang, W. Y., Godfrey, L. M., Bowden, G., Szeszko, J., Smink, L. J., Lam, A. C., Burren, O., Walker, N. M., Nutland, S., Rance, H., Undlien, D. E., Ronningen, K. S., Guja, C., Ionescu-Tirgoviste, C., Todd, J. A. and Twells, R. C. (2005) No evidence for association of the TATA-box binding protein glutamine repeat sequence or the flanking chromosome 6q27 region with type 1 diabetes. Biochem. Biophys. Res. Commun. 331, 435 – 441.
111 Smyth, D. J., Cooper, J. D., Lowe, C. E., Nutland, S., Walker, N. M., Clayton, D. G. and Todd, J. A. (2006) No evidence for association of OAS1 with type 1 diabetes in unaffected siblings or type 1 diabetic cases. Diabetes 55, 1525 – 1528.
112 Szeszko, J. S., Howson, J. M., Cooper, J. D., Walker, N. M., Twells, R. C., Stevens, H. E., Nutland, S. L. and Todd, J. A. (2006) Analysis of polymorphisms of the interleukin-18 gene in type 1 diabetes and Hardy-Weinberg equilibrium testing. Diabetes 55, 559 – 562.
113 Vella, A., Howson, J. M., Barratt, B. J., Twells, R. C., Rance, H. E., Nutland, S., Tuomilehto-Wolf, E., Tuomilehto, J., Undlien, D. E., Ronningen, K. S., Guja, C., Ionescu-Tirgoviste, C., Savage, D. A. and Todd, J. A. (2004) Lack of association of the Ala(45)Thr polymorphism and other common variants of the NeuroD gene with type 1 diabetes. Diabetes 53, 1158 – 1161.
114 Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A. and Contopoulos-Ioannidis, D. G. (2001) Replication validity of genetic association studies. Nat. Genet. 29, 306 – 309.
115 Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S. and Hirschhorn, J. N. (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat. Genet. 33, 177 – 182.
116 Vella, A., Cooper, J. D., Lowe, C. E., Walker, N., Nutland, S., Widmer, B., Jones, R., Ring, S. M., McArdle, W., Pembrey, M. E., Strachan, D. P., Dunger, D. B., Twells, R. C., Clayton, D. G. and Todd, J. A. (2005) Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. Am. J. Hum. Genet. 76, 773 – 779.
117 Pompanon, F., Bonin, A., Bellemain, E. and Taberlet, P. (2005) Genotyping errors: causes, consequences and solutions. Nat. Rev. Genet. 6, 847 – 859.
118 Biason-Lauber, A., Boehm, B., Lang-Muritano, M., Gauthier, B. R., Brun, T., Wollheim, C. B. and Schoenle, E. J. (2005) Association of childhood type 1 diabetes mellitus with a variant of PAX4: possible link to beta cell regenerative capacity. Diabetologia 48, 900 – 905.
119 Potter, C., Worthington, J., Silman, A. and Barton, A. (2005) TNFR2 is not associated with rheumatoid arthritis susceptibility in a Caucasian population. Arthritis Rheum. 52, 2579 – 2581.
120 Salanti, G., Amountza, G., Ntzani, E. E. and Ioannidis, J. P. (2005) Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. Eur. J. Hum. Genet. 13, 840 – 848.
121 Suzuki, A., Yamada, R., Chang, X., Tokuhiro, S., Sawada, T., Suzuki, M., Nagasaki, M., Nakayama-Hamada, M., Kawaida, R., Ono, M., Ohtsuki, M., Furukawa, H., Yoshino, S., Yukioka, M., Tohma, S., Matsubara, T., Wakitani, S., Teshima, R., Nishioka, Y., Sekine, A., Iida, A., Takahashi, A., Tsunoda, T., Nakamura, Y. and Yamamoto, K. (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat. Genet. 34, 395 – 402.
122 Ikari, K., Kuwahara, M., Nakamura, T., Momohara, S., Hara, M., Yamanaka, H., Tomatsu, T. and Kamatani, N. (2005) Association between PADI4 and rheumatoid arthritis: a replication study. Arthritis Rheum. 52, 3054 – 3057.
123 Barton, A., Bowes, J., Eyre, S., Spreckley, K., Hinks, A., John, S. and Worthington, J. (2004) A functional haplotype of the PADI4 gene associated with rheumatoid arthritis in a Japanese population is not associated in a United Kingdom population. Arthritis Rheum. 50, 1117 – 1121.
124 Harney, S. M., Meisel, C., Sims, A. M., Woon, P. Y., Wordsworth, B. P. and Brown, M. A. (2005) Genetic and genomic studies of PADI4 in rheumatoid arthritis. Rheumatology 44, 869 – 872.
125 Martinez, A., Valdivia, A., Pascual-Salcedo, D., Lamas, J. R., Fernandez-Arquero, M., Balsa, A., Fernandez-Gutierrez, B., de la Concha, E. G. and Urcelay, E. (2005) PADI4 polymorphisms are not associated with rheumatoid arthritis in the Spanish population. Rheumatology 44, 1263 – 1266.
126 Mori, M., Yamada, R., Kobayashi, K., Kawaida, R. and Yamamoto, K. (2005) Ethnic differences in allele frequency of autoimmune-disease-associated SNPs. J. Hum. Genet. 50, 264 – 266.
127 Todd, J. A., Mijovic, C., Fletcher, J., Jenkins, D., Bradwell, A. R. and Barnett, A. H. (1989) Identification of susceptibility loci for insulin-dependent diabetes mellitus by trans-racial gene mapping. Nature 338, 587 – 589.
128 Simkins, H. M., Merriman, M. E., Highton, J., Chapman, P. T., O'Donnell, J. L., Jones, P. B., Gow, P. J., McLean, L., Pokorny, V., Harrison, A. A. and Merriman, T. R. (2005) Association of the PTPN22 locus with rheumatoid arthritis in a New Zealand Caucasian cohort. Arthritis Rheum. 52, 2222 – 2225.
129 Geluk, A., van Meijgaarden, K. E., Schloot, N. C., Drijfhout, J. W., Ottenhoff, T. H. and Roep, B. O. (1998) HLA-DR binding analysis of peptides from islet antigens in IDDM. Diabetes 47, 1594 – 1601.
130 Vafiadis, P., Bennett, S. T., Todd, J. A., Nadeau, J., Grabs, R., Goodyer, C. G., Wickramasinghe, S., Colle, E. and Polychronakos, C. (1997) Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. Nat. Genet. 15, 289 – 292.
131 Barratt, B. J., Payne, F., Lowe, C. E., Hermann, R., Healy, B. C., Harold, D., Concannon, P., Gharani, N., McCarthy, M. I., Olavesen, M. G., McCormack, R., Guja, C., Ionescu-Tirgoviste, C., Undlien, D. E., Ronningen, K. S., Gillespie, K. M., Tuomilehto-Wolf, E., Tuomilehto, J., Bennett, S. T., Clayton, D. G., Cordell, H. J. and Todd, J. A. (2004) Remapping the insulin gene/IDDM2 locus in type 1 diabetes. Diabetes 53, 1884 – 1889.
132 Durinovic-Bello, I., Rosinger, S., Olson, J. A., Congia, M., Ahmad, R. C., Rickert, M., Hampl, J., Kalbacher, H., Drijfhout, J. W., Mellins, E. D., Al Dahouk, S., Kamradt, T., Maeurer, M. J., Nhan, C., Roep, B. O., Boehm, B. O., Polychronakos, C., Nepom, G. T., Karges, W., McDevitt, H. O. and Sonderstrup, G. (2006) DRB1*0401-restricted human T cell clone specific for the major proinsulin73 – 90 epitope expresses a down-regulatory T helper 2 phenotype. Proc. Natl. Acad. Sci. USA 103, 11683 – 11688.
133 Cox, S. L. (2006) Tesaglitazar: a promising approach in type 2 diabetes Drugs Today (Barc.). 42, 139 – 146.
134 Ruderman, E. M. and Pope, R. M. (2005) The evolving clinical profile of abatacept (CTLA4-Ig): a novel co-stimulatory modulator for the treatment of rheumatoid arthritis. Arthritis Res. Ther. 7 (Suppl. 2), S21 – S25.
135 Ryser, H. J. and Fluckiger, R. (2005) Progress in targeting HIV-1 entry. Drug Discov. Today 10, 1085 – 1094.