J Mol Med (2005) 83:97–109
DOI 10.1007/s00109-004-0603-7

REVIEW

Julian C. Knight

Regulatory polymorphisms underlying complex disease traits

Received: 3 August 2004 / Accepted: 15 September 2004 / Published online: 9 December 2004

Abstract There is growing evidence that genetic variation plays an important role in the determination of individual susceptibility to complex disease traits. In contrast to coding sequence polymorphisms, where the consequences of non-synonymous variation may be resolved at the level of the protein phenotype, defining specific functional regulatory polymorphisms has proved problematic. This has arisen for a number of reasons, including difficulties with fine mapping due to linkage disequilibrium, together with a paucity of experimental tools to resolve the effects of non-coding sequence variation on gene expression. Recent studies have shown that variation in gene expression is heritable and can be mapped as a quantitative trait. Allele-specific effects on gene expression appear relatively common, typically of modest magnitude and context specific. The role of regulatory polymorphisms in determining susceptibility to a number of complex disease traits is discussed, including variation at the VNTR of *INS*, encoding insulin, in type 1 diabetes and polymorphism of *CTLA4*, encoding cytotoxic T lymphocyte antigen, in autoimmune disease. Examples where regulatory polymorphisms have been found to play a role in mongenic traits such as factor VII deficiency are discussed, and contrasted with those polymorphisms associated with ischaemic heart disease at the same gene locus. Molecular mechanisms operating in an allele-specific manner at the level of transcription are illustrated, with examples including the role of Duffy binding protein in malaria. The difficulty of resolving specific functional regulatory variants arising from linkage disequilibrium is demonstrated using a number of examples including polymorphism of *CCR5*, encoding CC chemokine receptor 5, and HIV-1 infection. The importance of understanding haplotypic structure to the design and interpretation of functional assays of putative regulatory variation is highlighted, together with discussion of the strategic use of experimental tools to resolve regulatory polymorphisms at a transcriptional level. A number of examples are discussed including work on the *TNF* locus which demonstrate biological and experimental context specificity. Regulatory variation may also operate at other levels of control of gene expression and the modulation of splicing at *PTPRC*, encoding protein tyrosine phosphatase receptor-type C, and of translational efficiency at *F12*, encoding factor XII, are discussed.

JULIAN C. KNIGHT graduated from the medical school at the University of Edinburgh and received his D.Phil. degree in molecular genetics from the University of Oxford. He is presently a Wellcome Senior Research Fellow in Clinical Science at the Wellcome Trust Centre for Human Genetics at the University of Oxford. His research interests include transcriptional regulation and the functional characterisation of human genetic variation.

J. C. Knight (✉)
Wellcome Trust Centre for Human Genetics,
University of Oxford,
Oxford, OX3 7BN, UK
e-mail: julian@well.ox.ac.uk
Tel.: +44-1865-287671, Fax: +44-1865-287533

Keywords Gene expression · Genetics · Gene polymorphism · Promoter · Transcription

Abbreviations *HIV:* Human immunodeficiency virus · *HNF:* Hepatic nuclear factor · *NF:* Nuclear factor · *SNP:* Single nucleotide polymorphism · *TNF:* Tumour necrosis factor · *UTR:* Untranslated region · *VNTR:* Variable number tandem repeat

Defining regulatory polymorphisms

The dissection of genetic factors defining our individual susceptibility to complex disease traits is being tackled

across the field of clinical medicine, offering the promise of novel insights into disease pathogenesis and therapeutic targets together with the tailoring of management to the individual patient. However, resolving the genetic factors underlying complex diseases has proved problematic, with typically multiple genes involved of individually modest magnitude, which together only form part of a multifactorial disease process [1, 2]. For these reasons linkage analysis has met with limited success, and while a genomic region may be localised, fine mapping typically proves problematic. Association studies have proved useful, often based on a candidate gene approach, with whole-genome association studies now becoming feasible [3]. The premise of such genetic analysis is that functionally important genetic variation results in differing clinical phenotypes. As discussed below, the challenges of mapping genetic susceptibility loci are compounded by difficulties in defining specific functional polymorphism(s) at an experimental level. At present it is typically left unresolved as to whether a disease-associated polymorphism is itself functionally important or acting only as a marker for a coinherited, perhaps as yet unidentified, genetic polymorphism.

Classification of variation

DNA sequence polymorphisms are usually defined as variation present at greater than 1% frequency in the population. The most common are single nucleotide polymorphisms (SNPs) in which one of the four possible nucleotides in the DNA sequence is substituted by another, occurring on average every 800 nucleotides across the genome. Other sequence polymorphisms include deletions and insertions of one or more nucleotides, rearrangements and repeating sequences which may be short tandemly repeated motifs of one to six nucleotides (microsatellites) or longer repeating 'minisatellites'. The majority of DNA sequence polymorphisms are of no functional importance, although some alter the structure of the resulting peptide through, for example, amino acid changes [4]. Such coding polymorphisms have classically been implicated in monogenic Mendelian disorders with their consequences typically predictable and amenable to testing. In contrast, regulatory polymorphisms occurring outside exonic regions have long been postulated to be important modulators of gene expression and evolutionary change [5], but there is only now growing evidence that this is the case [6].

Regulatory polymorphisms can be classified into two groups. The first are *cis*-acting, in other words acting on the copy of the gene present on that allele and typically present in or near the locus of the gene that it regulates. This may arise, for example, through the sequence change occurring in a regulatory DNA binding site and altering the affinity with which a regulatory protein is recruited and hence how the gene is expressed. Alternatively, the regulatory polymorphism may be *trans*-acting, a polymorphism in one gene affecting the expression of another gene. There is evidence from differences in strains of mice and yeast that *trans*-acting loci were responsible for much of the observed differences in expression [7, 8, 9], but in the context of complex disease susceptibility work has focused on potential *cis*-acting variation and forms the basis of this review.

Evidence for regulatory polymorphisms

If such regulatory polymorphisms exist, differences in expression of genes within and between populations would be expected. Variation in genome-wide gene expression has been associated with phenotypic variation in a number of different organisms, notably budding yeast. A genetic linkage analysis of global expression patterns in a cross between two strains of budding yeast demonstrated heritable differences in expression [9]. Among naturally occurring isolates of *Saccharomyes cerevisiae*, population genetic variation is correlated with phenotypic variation [10, 11]. Differences in gene expression have been associated with phenotypic variation in *Drosophila* development [12], within and between populations of teleost fish [13] and primate species [14], and in human lymphoblastoid cell lines [15]. Here variation in gene expression was observed between lines established from unrelated and related individuals, with evidence of familial aggregation [15]. Moreover, this variation in expression has recently been used as a quantitative trait to carry out linkage mapping at a genomic level in defined pedigrees of lymphoblastoid lines [16]. These data suggest that only a minority of effects are likely to be operating in *cis*, and that a number of regions show strong linkage to expression of many genes, possibly representing hotspots of transcriptional regulation. For some putative *cis*-regulators, typing additional SNP markers confirmed the association at a population and familial level [16]

There is growing evidence that allele-specific differences in gene expression (Fig. 1) occur in autosomal genes which do not show genomic imprinting (imprinted genes being the small number of genes in which allelic expression is dependent on parent of origin) [17]. The finding of allele-specific differences in transcript abundance among non-imprinted genes provides important direct supporting evidence for the existence of regulatory polymorphisms. If such differences are arising from *cis*-acting regulatory polymorphisms, we would expect allele-specific effects to be heritable, and there is evidence that this is the case [18, 19]. Allelic differences in expression appear context specific, for example, with regard to tissue-type, and to be of modest magnitude (typically 1.5- to 2.0-fold) [18, 19, 20, 21]. At present it is unclear how common allele-specific gene expression is across the genome, ranging from 6% of genes in a murine study [20] to approx. 20% in human lymphoblastoid cell lines [18, 19].

The elucidation of allele-specific gene expression using SNP markers to discriminate transcribed RNA should

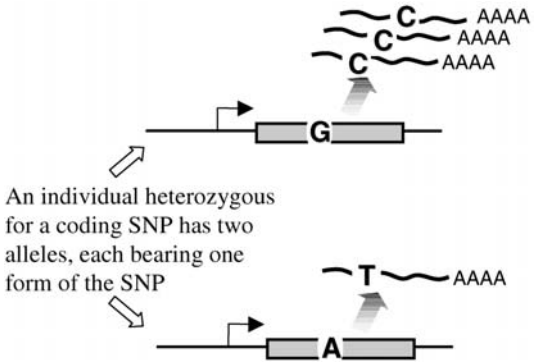

Fig. 1 Investigation of regulatory polymorphism: allele-specific analysis of mRNA. *Strategy:* This approach uses the presence of an exonic transcribed SNP (shown here as *G/A*) to resolve the allelic origin of transcribed mRNA [18]. By analysis of mRNA from cells derived from an individual heterozygous for the marker SNP, an internally controlled system is established in which relative allele-specific gene expression can be estimated. *Uses:* Allele-specific quantification of mRNA is a useful in vivo approach to resolving functionally important haplotypes using transcribed marker SNPs. It provides a direct assessment of the relative abundance of allele-specific transcript in a natural chromosomal context in which the normal regulatory machinery and chromatin environment are operating. *Limitations:* The assay requires accurate and sensitive quantification of relative transcript abundance, typically based on primer extension methods. A major limitation is that for many genes and for the majority of haplotypes no exonic marker is present to allow resolution of transcript origin. Some information may be achieved by using intronic SNPs to study relative expression of unspliced RNA; a further approach is to use a different indirect measure of gene expression, namely phosphorylated Pol II loading by haploChIP in living cells [80]. The allele-specific density of Pol II loading can be used in the same way as transcript abundance except that as the Pol II is being measured in situ by crosslinking it to DNA, any SNP marker can be used within 2 kb 5′ or 3′ to the gene including promoter and 3′UTR SNPs which considerably expands the number of haplotypes that can be interrogated

prove a powerful approach in screening for regulatory polymorphisms. It is important to note, however, that it serves to identify functionally important haplotypes (the coinherited polymorphisms on an allele) rather than specific polymorphisms, unless sufficient complexity of the underlying haplotypic structure can be resolved so as to fine map observed differences down to a specific region or individual polymorphism. The consequences of regulatory polymorphism are highly context dependent, depending, for example, on the cell type, prevailing environmental conditions and exogenous stimuli. This is compounded by the effects of covariation at other genomic sites, both within and without the locus. Such effects mean that the logical approach of investigating the effects of putative regulatory polymorphisms by correlating the levels of protein product derived from the gene between cells or individuals of differing genotype (Fig. 2) is often confounded.

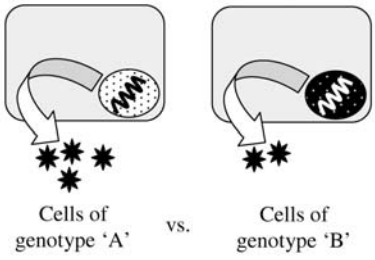

Fig. 2 Investigation of regulatory polymorphism: assays of secreted protein. *Strategy:* An intuitive approach is to compare levels of protein produced from the gene of interest in individuals of differing genotype, for example, homozygous AA or BB or heterozygous AB. *Uses:* This approach can be highly informative where sufficient numbers of individuals are assayed on multiple occasions using appropriate controls to minimise confounding by environmental or experimental factors. *Limitations:* The approach is potentially confounded at many levels including environmental factors and other variables affecting the levels of expression of a gene between individuals such as differences in receptor-ligand interaction and signal transduction as well as translation and post-translational effects. Genetic variation on the compared haplotypes may confound interpretation of the differences seen with the chosen marker SNP

Reporter gene assays

The most widely used experimental tool to interrogate the significance of specific putative regulatory polymorphisms are reporter gene assays [22] in which cells are transiently transfected with allele- or SNP-specific promoter constructs (Fig. 3). The results of reporter gene experiments are highly context dependent, in terms of both the biological system used and reporter gene construct design. A recent analysis of 144 functionally significant human genetic polymorphisms analysed by reporter gene assays in physiologically relevant cell lines revealed most were in the proximal regulatory promoter regions of genes, but over one quarter were located more than 1 kb upstream or 3′ to the transcriptional start site [23]. Overall an equal number of examples were found of effects of polymorphic activator or repressor binding and analysis of ancestral gain or loss of transcription factor binding was consistent with this. Notable among classes of transcription factor implicated as showing polymorphic modulation of allelic binding were USF, Sp1, nuclear factor (NF) κB, GATA and Oct-1, most often demonstrated in vitro by gel shift assays (Fig. 4). Overall Rockman and Wray [23] estimate that humans are heterozygous at more functional *cis*-regulatory sites than at amino acid positions, with 10,700 functional biallelic *cis*-regulatory polymorphisms in a typical human. A recent study found that 34% of promoter polymorphisms significantly modulated reporter gene expression more than 1.5-fold in at least one of three transfected human cell lines in a screening exercise of the proximal promoters of 170 opportunistically selected genes [24]. Systematic surveys of promoter haplotypes by reporter gene analysis on chromosomes 21 [25] and 22q11 [26] show functional variation among 18% and 20% of polymorphisms tested,

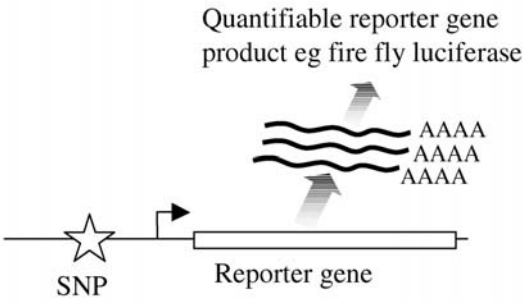

Fig. 3 Investigation of regulatory polymorphism: the reporter gene assay. *Strategy:* Reporter gene assays are a powerful approach to resolve the effects of DNA sequences on gene expression. The DNA sequence of interest, for example a promoter region spanning an SNP, is placed upstream of a reporter gene whose expression can be measured. The reporter gene construct is then inserted (transfected) into a cell and expression assayed. *Uses:* This robust approach allows the effects of polymorphisms to be assayed, for example, comparing the expression of reporter gene constructs differing only by the nucleotide(s) of interest. The assay is highly sensitive and with appropriate controls is reproducible and specific. *Limitations:* The assay is an in vitro approach as the transfected DNA sequences lack the native chromatin configuration which may be essential to accurate interpretation of the consequences of genetic polymorphism. The design of reporter constructs is critical, notably the choice of which portions of the naturally occurring regulatory regions of the gene locus within which the polymorphism is found to include in the reporter gene. Any functional effect of an SNP may also be dependent on its naturally occurring haplotype: many early studies assayed SNPs in isolation rather dissecting their naturally occurring coinherited combinations. Results of reporter gene assays are highly context specific with respect to choice of cell type to transfect; stimulus used for induction of gene expression; the mode and efficiency of transfection; and the DNA plasmid design and preparation

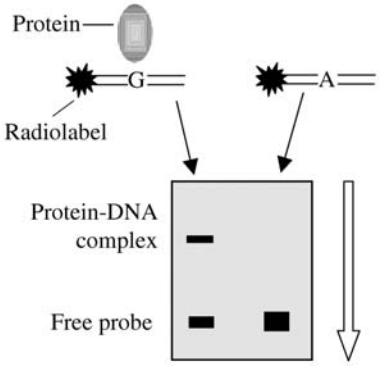

Fig. 4 Investigation of regulatory polymorphism: the gel shift assay of protein-DNA interactions. *Strategy:* The 'gel shift' or electrophoreitic mobility shift assay investigates protein-DNA binding. Short DNA probes corresponding to a genomic DNA sequence of interest are synthesised, annealed as a duplex and radiolabelled. These are then usually incubated with a crude nuclear extract or a recombinant protein of interest. On binding by protein to DNA the mobility of the probe on electrophoresis is retarded ('shifted'). The nature and specificity of these complexes can then be resolved. *Uses:* A highly specific and sensitive assay to investigate relative binding affinities of proteins to the two allelic forms of an SNP. Specificity can be resolved using unlabelled competitor probes. The nature of retarded complexes can be investigated by UV crosslinking experiments and using antibodies to abolish or 'supershift' bound complexes. *Limitations:* This is an in vitro assay, typically used as a screening tool using crude nuclear lysates for hypothesis generation which does not define whether binding actually occurs in vivo. The DNA probes used are short and lack a native chromatin structure. The absence of flanking sequences and conformational effects may lead to discrepancies in observed binding

respectively, a proportion similar to that reported on analysis of allele-specific expression for other genomic regions [18, 19].

Other strategies to resolve regulatory SNPs include using computational based approaches to predict regulatory elements, such as promoters and sites of regulatory protein binding within them, combined with interspecies comparisons to identify evolutionary conservation [27]. A number of software programmes are available which can facilitate prediction of the effect of nucleotide substitutions on likely regulatory protein binding [28]. For specific regulatory transcription factor proteins the sophistication of such predictions is rapidly increasing based on experimentally derived protein-DNA binding data, for example, for NF-κB/Rel [29].

To illustrate the approaches to defining regulatory polymorphisms that I have outlined, I will discuss the evidence for regulatory polymorphisms in the context of a number of complex disease traits. This discussion demonstrates the limitations of experimental strategies used to resolve regulatory polymorphisms and also the successes that have been achieved. Given the breadth of the field this review does not attempt to be comprehensive or to discuss in detail the evidence for disease associations at a population genetic level but rather serves to introduce the subject to a clinically based readership and to discuss the challenges that lie ahead.

Regulatory polymorphisms are important in complex disease traits: lessons from analysis of a VNTR at the *IDDM 2* locus

Regulatory polymorphisms are increasingly recognised to contribute to complex disease traits although the number of examples with both a clear-cut genetic association and defined molecular mechanism remain small. A number of examples involving SNPs are discussed below, including the role of polymorphism of *CTLA4*, encoding cytotoxic T lymphocyte antigen 4, in determining susceptibility to autoimmune disease [30] and polymorphism of *DARC*, encoding Duffy antigen receptor for chemokines, in malaria [31]. Variable number tandem repeat (VNTR) sequences (minisatellites) have been associated with a number of complex disease traits, but only in a minority of cases has there been evidence of direct functional consequences arising from polymorphism within the VNTR itself. One notable example is the insulin/IGF-2 (*INS-IGF2*) VNTR found approx. 600 bp from the transcriptional start site of *INS*, encoding insulin.

The VNTR is important to *INS* transcriptional regulation [32] and is found only in primates, suggesting that

it represents a recent evolutionary change [32]. It comprises 14–15 bp tandem repeating sequences with the consensus ACAGGGGTSYGGGG [33] from which three classes can be defined based on length. Individuals homozygous for the shorter class I polymorphism were associated with higher rates of type I diabetes in case-control studies [34, 35]. This association was mapped to a 4.1 kb region spanning *INS* [36] within which the strongest disease association lies at the VNTR, but two additional SNPs in the locus may be involved [37, 38]. Overall, homozygosity for class I alleles was seen to confer increased risk while class III alleles were dominantly protective [37, 39, 40]. Genetic susceptibility to type I diabetes is a good example of a polygenic, multifactorial disease in which a number of loci other than the *IDDM 2* locus at *INS* have been implicated through association studies, linkage analysis and genome scanning [41].

There is some functional evidence that the *INS-IGF2* VNTR alters gene expression in the pancreas where class I alleles have been associated with modest levels of increased expression relative to class III alleles. Autopsy studies of human pancreas from fetal [42] and adult [37, 40] tissues show mRNA to be relatively higher for class I than class III alleles but observed differences were modest. Insulin secretion in vivo was lower in class III/II than class I/I homozygous individuals in some but not all studies [43, 44, 45, 46]. This is consistent with results obtained using reporter constructs transfected into a rodent pancreatic β cell line where class I alleles show 1.5- to 3-fold more expression than class III alleles [47]. One exception is the class I allele 698 (λHI) which showed lower expression than class III alleles when compared in fetal rat islet cells and in insulinoma cell lines using native insulin and heterologous promoter constructs [32]. Interestingly this particular allele showed no disease association with type I diabetes susceptibility, which may account for the apparent dichotomy with other class I alleles showing increased expression or alternatively be dependent on sequence rather than length differences between alleles [37]. Such sequence heterogeneity may account for differences in binding by regulatory proteins. For example, while the transcription factor Pur-1 was shown to bind the VNTR and activate transcription from a linked downstream promoter, different repeats bound Pur-1 with differing affinities [32]. These differences in affinity may relate to structural differences between repeats, in particular the presence of G-quartets. These unusual four-stranded DNA structures arise through non-Watson-Crick base-pairing as the *INS-IGF2* VNTR is guanin rich [48, 49]. Pur-1 binding was found to depend on both intra- and intermolecular G-quartet formation of the *INS-IGF2* VNTR [49].

Further clues as to how VNTR alleles may modulate disease susceptibility arose from studies of *INS* expression in the thymus [50, 51]. Given that insulin is a target autoantigen in type I diabetes it was postulated that insulin is normally expressed in the thymus to induce tolerance. Differential expression of *INS* associated with the VNTR may in turn then affect selection of insulin-specific T-lymphocytes and predispose to β-cell autoimmunity. *INS* was found to be transcribed in fetal thymi with proinsulin and insulin proteins present [50, 51]. Relative *INS* expression was significantly higher (two- to three-fold) with class III alleles than class I when mRNA was assayed in individuals heterozygous for *INS-IGF2* VNTR I/III using a linked transcribed marker polymorphism in the *INS* 3′UTR [51]. Using the same approach, the allelic ratio was found to be converse in the pancreas, suggesting tissue-specific transcriptional effects. The same differences in relative allelic expression were observed in an independent study showing protective class III VNTRs were associated with higher steady-state levels of *INS* mRNA expression [50]. The higher levels of proinsulin and insulin in the thymus associated with class III VNTRs are postulated to promote self-tolerance through negative selection of insulin-specific T-lymphocytes in development, cells which play a critical role in the pathogenesis of type I diabetes.

Whether the *INS-IGF2* VNTR also modulates expression of the adjacent imprinted gene *IGF2* is controversial and may be tissue specific. Higher levels of *IGF2* mRNA were associated with class I than class III alleles in human placenta using competitive reverse transcription PCR, consistent with reporter gene experiments in a hepatoma cell line using *INS-IGF2* constructs [52]. However, other workers have shown no difference in *IGF2* expression in thymus or pancreas associated with VNTR alleles [53]. Whether the VNTR acts as a long-range control element affecting the expression of both *INS* and *IGF2* remains to be resolved.

Defining functionality among coinherited polymorphisms: *CCR5* and AIDS progression

Complex multifactorial disease traits such as susceptibility to infectious disease have been found to typically involve several host susceptibility genes of individually modest magnitude within which a number of functionally important variants may be present [54]. Linkage disequilibrium between genetic markers may confound not only localisation on linkage mapping or association study but also the results of functional experiments to localise specific regulatory variation. Detailed definition of underlying haplotypic structure is needed in order to help resolve specific functional polymorphism(s), which may include both coding and regulatory variation as demonstrated by the work on variation at *CCR5* and AIDS progression.

A number of host genetic factors have been shown to influence transmission and disease progression of the human immunodeficiency virus (HIV) 1 [55]. Notable among these is polymorphism of *CCR5*, encoding CC chemokine receptor 5, a major chemokine coreceptor of HIV-1 necessary for viral entry into cells. A number of coding polymorphisms have been identified, including a rare 32-bp deletion of the *CCR5* open reading frame

(Δ32) which results in a failure of the protein to be transported to the cell surface. Individuals homozygous for *CCR5* Δ32 allele do not express CCR5 and are highly resistant to HIV-1 infection [56, 57]. Individuals heterozygous for *CCR5* Δ32 have a delayed progression to AIDS and have a lower HIV-1 viral load early in disease [58].

Non-coding SNPs of the *CCR5 cis*-regulatory region have also been shown to modulate disease transmission and progression [59, 60, 61, 62, 63] although localising effects to specific polymorphisms has proved problematic due to linkage disequilibrium across the locus including neighbouring genes such as *CCR2*, the product of which can also act as an HIV-1 coreceptor. In order to dissect such effects the haplotypic structure of the highly diverse *CCR5 cis*-regulatory region was defined, with seven human 'haplogroups' and an ancestral haplotype resolved through comparisons with non-human primates [64]. Both *CCR5* haplotypes and specific SNPs have been shown to modulate gene expression using reporter gene assays with evidence of differential protein-DNA binding in gel shift assays. For example, the ancestral haplotype HHA led to the lowest level of transcriptional activity in reporter gene assays using haplotype-specific constructs [64].

A G to A SNP of *CCR5* at −2459 nt has been associated with disease susceptibility [59, 60, 62, 63], with the G allele forming part of the HHA haplotype and showing lower expression than the A allele on reporter gene analysis [60]. When unstimulated $CD14^+$ monocytes from healthy donors were analysed for CCR5 density, a clear relationship was found with *CCR5*-2459 genotype: the lowest levels were observed in cells from individuals homozygous GG, intermediate with GA and highest with those homozygous −2459AA [65]. The same relationship was found when macrophage tropic HIV-1 was propagated in vitro in activated peripheral blood mononuclear cells with the G allele showing the lowest levels [65]. A study of Langerhans cells from healthy individuals heterozygous for Δ32 has also shown HIV infection levels to be associated with the *CCR5*-2459 SNP, with higher levels seen with the AA genotype [66]. Interestingly, while no difference in protein-DNA binding was observed for this SNP on gel shift assays using crude nuclear lysates, differential protein-DNA binding has been reported for a number of other *CCR5* alleles, notably at −2554 with recruitment of NF-κB/Rel proteins postulated [64, 67]. The complexity of the task ahead to fully resolve the *cis*-regulatory SNPs of *CCR5* is compounded by the multigenic host factors now shown to influence HIV-1 disease progression, including, for example, variation at chemokine/chemokine receptor loci, receptor ligands such as stromal cell derived factor 1 and RANTES ('regulated on activation normal T cell expressed and secreted'), and HLA genotype [55].

Interpreting functional analysis of putative regulatory variation: *TNF* polymorphisms in malaria and rheumatoid arthritis

The functional characterisation of regulatory polymorphisms has been made more difficult by the many potential confounders of commonly used functional assays. Apparently conflicting results often relate to the context specificity within which data should be seen. As noted previously, understanding haplotypic structure often reveals that while a given SNP marker may have been used to distinguish between cells from different individuals, there has been confounding by coinherited functional variants within a subset of the population. Context specificity also extends to the precise experimental conditions used, with results, for example, specific to particular stimuli or cell type. Moreover, the design of experiments can be critical. Underpowered studies may be highly misleading while the particular genetic construct used can be highly specific in the results obtained for a given SNP in a reporter gene assay. These points are well illustrated by work on the *TNF* locus, which has been the subject of several studies dissecting the functional basis of observed genetic associations.

There has been considerable interest in genetic variation in *TNF*, encoding tumour necrosis factor (TNF) and susceptibility to a number of infectious and autoimmune conditions, often based on strong evidence of a role for TNF dysregulation in disease pathogenesis. Stable 'high' and 'low' TNF producers can be identified in a population with a significant genetic component to circulating TNF levels. There is evidence from linkage analysis implicating the *TNF* locus in susceptibility to mild malaria in African populations [68, 69] while possession of a G to A promoter SNP at −308 nt has been associated with susceptibility to severe malaria in Gambia [70] and Sri Lanka [71]. This SNP does not appear to modulate susceptibility to rheumatoid arthritis, but there is some evidence that possession of *TNF*-308 may be a useful marker of a patient's response to anti-TNF treatment [72, 73].

The functional significance of *TNF*-308 remains controversial (reviewed in [74, 75]). There is some evidence to support an increased level of expression with the *TNF*-308A allele from reporter gene studies using a range of cell lines and range of stimuli, approx. one-half show a modest increase in levels of transcription with the A allele (predominantly in human cell types) while the remainder show no effect. Effects are highly context specific depending on cell type and stimulus [76] and features of reporter gene construct design. For example, detecting a difference between alleles may be dependent on inclusion of the *TNF* 3′ UTR in reporter constructs [77]. A number of studies correlating circulating TNF levels in individuals, or induced levels of TNF from isolated cells using a range of stimuli, have shown either a small increase with the *TNF*-308A allele or no effect. The small reported study sizes have limited power to detect a modest difference and may well be confounded by haplotypic heterogeneity when comparing groups using a single SNP

marker. There is evidence in human B cell lines but not monocyte cell lines that this SNP lies in a site of protein-DNA occupancy using in vitro DNA footprinting [78, 79], and contrasting evidence of whether differential allelic recruitment of protein binding is present [77, 78]; the nature of any binding complex remains unresolved. Allele-specific analysis of gene expression in vivo using human lymphoblastoid cell lines [80] suggested no difference in allelic expression in the specific context analysed.

The *TNF*-308 SNP forms an extended haplotype with polymorphisms in neighbouring gene *LTA*, encoding lymphotoxin α, including *LTA*+252. Intriguingly, allele-specific differences in gene expression were observed at *LTA* with haplotypes bearing this SNP [80]. There is some evidence that *LTA*+252 modulates reporter gene expression and possibly protein-DNA binding [81] but, strikingly, it shows a very strong association with susceptibility to myocardial infarction on a genome-wide association study [81]. There is also evidence of association between the *TNF*-308/*LTA*+252 haplotype and rheumatoid arthritis in a study of individuals possessing HLA DRB1*04, the major determinant of disease susceptibility [82]. The picture is potentially further complicated by other *LTA* haplotypes showing functional effects. For example, the *LTA*+80 SNP shows evidence of allele-specific binding by a transcriptional repressor, ABF-1 [83]. This highlights the importance of a full understanding of haplotypic structure if both the functional effects of regulatory SNPs are to be resolved and disease associations finely mapped.

A number of other *TNF* SNPs found on different haplotypes have also been implicated in susceptibility to disease. For example, a *TNF* SNP at −238 nt has been associated with severe malarial anaemia [84] and protection from severe malaria [79, 85] while a linked SNP at −376 nt is associated with susceptibility to severe malaria [79]. There is evidence that *TNF*-376 lies in a complex area of protein-DNA binding and acts to modulate binding of the transcription factor Oct-1 in human monocytes in vitro, increasing constitutive gene expression in a reporter gene model [79]. There is also some evidence that *TNF*-238 lies in the binding site of an as yet unidentified transcriptional repressor [86] and is associated with disease severity in rheumatoid arthritis [87, 88].

Regulatory polymorphisms in monogenic disease and complex traits: lessons from analysis of variation at *F7*

There has been considerable interest in genetic variation of *F7*, the gene encoding coagulation factor VII, both to determine the molecular basis of the autosomal recessive condition factor VII deficiency and in complex disease susceptibility, notably ischaemic heart disease. Factor VII is a vitamin K dependent factor essential for haemostasis which initiates the extrinsic pathway of coagulation and leads to localised generation of thrombin. It is striking, but not perhaps surprising, that the effects of the regulatory polymorphisms incriminated in factor VII deficiency and heart disease should be found to differ so markedly in their magnitude of effect on gene expression. Here, as in other examples of regulatory polymorphisms associated with complex multifactorial traits, the effects on gene expression appear modest. This highlights the need for experimental designs to be sufficiently sensitive to detect such differences.

The clinical phenotype of factor VII deficiency is highly variable and has been associated almost exclusively with coding changes in the gene, predominantly missense mutations but also deletions, splice site abnormalities and nonsense mutations [89]. Analysis of the promoter regions of individuals with severe factor VII deficiency has identified rare examples of patients possessing promoter SNPs with drastic effects on factor VII expression through modulation of transcription factor binding.

A naturally occurring T to G transversion at −61 nt was found to reduce reporter gene expression to 6.7% vs. wild type on transfection into a human hepatocyte cell line [90]. This SNP was found to prevent binding of the transcription factor hepatic nuclear factor (HNF) 4 on gel shift assay using human liver nuclear extracts. HNF4 had been previously shown to bind to this region of the promoter by DNA footprinting analysis and gel shift assay, and to be functionally important to promoter activity [91]. On cotransfection of an HNF4 expression construct in a non-hepatocyte cell line that does not constitutively support factor VII activity, transactivation was seen with the wild-type factor VII promoter construct but not with the rarer allele bearing the polymorphism [90]. A further C to T SNP at −55 nt also modulated binding by HNF4 in vitro on electrophoretic mobility shift assay, reducing reporter gene expression to 9.7% vs. wild type on transient transfection into HepG2 cells [92]. A C to G SNP at −94 nt prevents recruitment of Sp1 on electrophoretic mobility shift assay, reducing expression to 5.8% of wild type on reporter gene expression in HepG2 cells [93]. This area had been shown to be important to factor VII gene expression on promoter analysis [94].

These findings contrast with functional analysis of putative regulatory polymorphisms associated with susceptibility to ischaemic heart disease. Atherosclerotic plaque disruption which results in binding of tissue factor to circulating factor VII is a major cause of thrombosis in myocardial infarction. High levels of plasma factor VII are a significant predictor of death due to ischaemic heart disease [95], and genetic factors are reported to account for approx. one-third of plasma factor VII levels between individuals [96]. Resolving the functionally important variant(s) has, however, proved problematic. A coding region G to A SNP which results in a change in amino acid 353 from arginine to glutamine was associated with a significant reduction in plasma factor VII coagulant levels [97], and with susceptibility to ischaemic heart disease [98, 99] although this remains controversial [100, 101, 102]. Strong linkage disequilibrium was found between

this SNP and a 10-bp insertion polymorphism in the 5′ untranslated region (position −323) with evidence that the disease association [99, 103] and functional effect may lie rather at this site. Low levels of plasma factor VII [104] and a reduction in reporter gene expression by one-third in a human hepatic cell line [91] were reported. However, evidence from a population in which the two polymorphisms were found in some individuals on different haplotypes suggest that the amino acid change does affect hepatic secretion independently of the insertion polymorphism [105].

To add to this complexity, the 10-bp insertion is in linkage with SNPs at −122 (T/C) and −401 (G/T); reporter gene experiments in a human hepatic cell line suggests that the maximal repression of expression was seen only when the naturally occurring haplotype containing all three polymorphisms was used rather than the polymorphisms in isolation [106]. Moreover, a further disease associated haplotype of *F7* has been identified (−670C/−402A) [102] and associated with increased reporter gene expression in hepatoma cells [102]. The *F7*-402 A allele has been independently associated with increased transcriptional activity on transient transfection and with increased factor VII plasma levels [107]. Other non-coding polymorphisms of *F7* have also been associated with levels of factor VII, notably of intron 7, where higher numbers of copies of a 37-bp repeat sequence were associated with mRNA expression levels through differential efficiency in mRNA splicing [108]. There is a report that this intronic repeat is associated with risk of myocardial infarction [98].

Resolving a clear molecular mechanism whereby a regulatory polymorphism modulates transcriptional regulation: Duffy binding protein and susceptibility to malaria

An elegant example of how a regulatory polymorphism may act to alter gene expression through a highly context specific effect on transcriptional regulation is shown by work on the Duffy binding protein. The presence of the Duffy blood group antigen on the surface of erythrocytes is essential for invasion by the malarial parasite *Plasmodium vivax*. Those individuals who lack the Duffy protein are completely protected from malaria due to this parasite [109, 110]. The molecular basis for this was found to be a promoter SNP in the *FY* gene (alias *DARC*, 'Duffy antigen receptor for chemokines') [31]. In the presence of the T to C SNP at -46 nt, a binding site for the transcription factor GATA-1 was found to be disrupted which in a reporter gene system reduced gene expression 20-fold. It is striking that in Duffy negative individuals, possessing the FY*0 allele bearing this SNP, the Duffy protein is not expressed on erythrocytes but is expressed on other cell types. This arises because GATA-1 is an erythroid-specific transcription factor, and hence the functional modulation of gene expression by the *FY*-46 SNP is cell-specific. The FY*0 allele found in Duffy negative individuals is at or near fixation in most sub-Saharan African populations but is very rare outside Africa, and this has been postulated as strong evidence of the action of natural selection. Given that vivax malaria is present at significant levels in non-African populations, this suggests that in evolutionary terms the mutation occurred after the proposed major human migrations out of Africa. It has been postulated that the strong subsequent selection pressure resulted in non-African populations showing different and greater numbers of SNPs at the Duffy locus than African populations, the opposite of the situation characteristically seen in other genomic regions where African populations are typically more diverse [111].

Looking beyond transcription: regulatory polymorphism and gene regulation

In order to advance further our understanding of regulatory polymorphisms in complex disease traits, it will be necessary to consider more broadly the potential role of DNA sequence polymorphisms in terms of the different facets of gene regulation which may be modulated (for reviews of gene expression see [112, 113]). In the examples described above, we have been concerned largely with *cis*-acting regulatory polymorphisms putatively modulating transcriptional initiation. This is the multifactorial process whereby gene expression is initiated through formation of a preinitiation complex including the enzyme RNA polymerase II and associated factors. Transcription factors and coregulator proteins facilitate and define this process, with modulation through DNA binding site sequence polymorphism altering the affinity of protein-DNA interactions and the control of transcriptional initiation. Following phosphorylation of the carboxyterminal domain of Pol II, nascent preRNA is transcribed through elongation with subsequent processing through capping, cleavage and polyadenylation with splicing of the transcript. The stability and localisation of RNA, and its translation into protein, are also all highly regulated and orchestrated. There is evidence that many of these steps occur concurrently, both physically and functionally. For example, splicing may be occurring while transcript is being synthesised. The relevance of this to our discussion of regulatory polymorphisms in complex disease traits is that the DNA, and in turn RNA, sequence remains pivotal to regulation throughout gene expression: sequence specificity for transcription factor binding is equally applicable to the rapidly growing list of other regulatory factors such as elongation and splicing factors. Many of the data to date have derived from allele-specific transcription factor binding and promoter analysis using reporter gene assays, but in many instances no definitive mechanism has been demonstrated beyond relative allelic differences in levels of transcription. A broader appreciation is therefore needed of the process of gene expression and how it may be potentially modulated.

Control of splicing

Genetic variation in both coding and non-coding DNA has been shown to affect the complex process of splicing whereby coding RNA sequences are identified and joined together. Polymorphism in exonic or intronic splicing regulatory elements can lead to exon skipping, activate cryptic splice sites or alter the balance of alternatively spliced isoforms [114]. Modulation of alternative splicing has been proposed as a mechanism underlying the strong observed disease association between susceptibility to autoimmune disease and variation at *CTLA4*, encoding cytotoxic T lymphocyte antigen 4 [30]. Disease susceptibility mapped to a non-coding region 6.1 kb 3′ to *CTLA4* and mRNA isoforms were found to be correlated with genotype. In particular, a disease-susceptibility haplotype bearing the marker CT60G produced less of the soluble isoform of CTLA-4 which lacks exon 3 [30]. The molecular mechanism for this remains to be resolved. This contrasts with a C to G nucleotide transition in exon 4 of *PTPRC*, encoding a transmembrane protein tyrosine phosphatase, receptor-type C (also known as *CD45*) which was found to modulate splicing [115]. Possession of this SNP, 77 nucleotides downstream of the splice acceptor junction, was associated with susceptibility to multiple sclerosis although this remains controversial [116, 117]. *PTPRC* is expressed on all nucleated haemopoietic cells and is essential for T cell receptor signal transduction. In the presence of the SNP, an exonic splicing silencer element (designated ESS1) was found to be disrupted which normally functions to repress the weak 5′ splice site of CD45 exon 4 [115]. This results in expression of high levels of high molecular weight isoforms in affected individuals by activated lymphocytes [118, 119] and has been demonstrated on transfection of DNA constructs [120].

Translational efficiency

DNA sequence variation may potentially modulate translational initiation and efficiency. An example of the latter in the context of complex disease susceptibility is a polymorphism of *F12*, encoding the serine protease factor XII which is the first coagulation factor in the intrinsic pathway of the coagulation cascade. This SNP, denoted 46C/T, has been associated with high risk of coronary heart disease, possibly as a result of reduced fibrinolysis due to low plasma factor XII levels [121], and with risk of venous thromboembolism [122]. The SNP lies in the 5′UTR of *F12*, 4 bp upstream of ATG translation initiation codon [123]. The T allele, and in particular the TT genotype, is significantly associated with lower levels of factor XII [123, 124, 125] and creates a novel methionine initiating codon that reduces the translation efficiency of factor XII. Both alleles are equally transcribed; however, transcription/translational analysis showed less factor XII was produced from cDNA containing 46T than 46C [123].

Conclusions

Genetic variation plays a significant role in determining susceptibility to complex, multifactorial disease traits ranging from autoimmune to infectious disease. Typically multiple genes have been implicated, each of modest magnitude of effect and fine mapping has proved problematic. There is increasing evidence at a population and experimental level that regulatory polymorphisms are important, and given the strong recent data that genetic variation underlies differences in gene expression this is only likely to increase. Data from in vitro reporter gene assays and allele-specific quantification of gene expression suggest the proportion of non-coding SNPs which may be functional could be much higher than expected. This may also reflect the need for greater rigor in the technical interpretation of data from such approaches. A number of examples where functional regulatory polymorphisms have been found to contribute to complex traits have been discussed, including disease associated genes where both coding and regulatory SNPs have been implicated such as *CCR5* and HIV-1 disease. Regulatory polymorphisms have also been found to underlie rare monogenic traits. In the case of *F7* the difference in magnitude of effect was striking for SNPs underlying factor VII deficiency and those implicated in ischaemic heart disease. In general for complex traits, a modest magnitude of effect has typically been found which has implications for the sensitivity of experimental approaches which should be applied for their detection. Differences in expression associated with regulatory polymorphisms appear highly context specific, with effects dependent on the specific stimulus, tissue or cell type analysed. This is highlighted by the Duffy binding protein and *INS* VNTR which showed altered expression in erythrocytes and the thymus, respectively, context specificity which appears critical to the resulting biological phenotypes of protection from malarial parasite invasion and type I diabetes. To date most work has focused on modulation of the process of transcriptional initiation, with data derived largely from reporter gene analysis and gel shift assays of protein-DNA binding. The value and limitations of such approaches have been discussed, with often apparently contradictory results relating to differences, for example, in the biological systems studied, and the design of DNA constructs.

The precise combination of coinherited polymorphisms on a given allele or haplotype are critical to both mapping polygenic disease and in designing and interpreting assays investigating putative regulatory polymorphisms. In many cases it is not possible to demonstrate that a specific SNP is functional, but rather that the haplotype shows allele-specific differences in expression with further resolution needed. Improved approaches for in vivo analysis of regulatory polymorphisms are required which allow the naturally occurring regulatory machinery to operate in a native chromatin environment. There is a need for a broader appreciation of the complexities of gene regulation, and the multiple points at which se-

quence polymorphism may modulate control of gene expression so that alteration of processes as diverse as transcriptional elongation, splicing, RNA stability and translation can also be considered.

References

1. Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. Science 298:2345–2349
2. Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517
3. Carlson CS, Eberle MA, Kruglyak L, Nickerson DA (2004) Mapping complex disease loci in whole-genome association studies. Nature 429:446–452
4. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22:231–238
5. King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188:107–116
6. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003) The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol 20:1377–1419
7. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–302
8. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L (2003) Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet 35:57–64
9. Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296:752–755
10. Townsend JP, Cavalieri D, Hartl DL (2003) Population genetic variation in genome-wide gene expression. Mol Biol Evol 20:955–963
11. Fay JC, McCullough HL, Sniegowski PD, Eisen MB (2004) Population genetic variation in gene expression is associated with phenotypic variation in Saccharomyces cerevisiae. Genome Biol 5:R26
12. Rifkin SA, Kim J, White KP (2003) Evolution of gene expression in the Drosophila melanogaster subgroup. Nat Genet 33:138–144
13. Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. Nat Genet 32:261–266
14. Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrop RE, Paabo S (2002) Intra- and interspecific variation in primate gene expression patterns. Science 296:340–343
15. Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, Spielman RS (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. Nat Genet 33:422–425
16. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG (2004) Genetic analysis of genome-wide variation in human gene expression. Nature 430:743–747
17. Knight JC (2004) Allele-specific gene expression uncovered. Trends Genet 20:113–116
18. Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW (2002) Allelic variation in human gene expression. Science 297:1143
19. Pastinen T, Sladek R, Gurd S, Sammak A, Ge B, Lepage P, Lavergne K, Villeneuve A, Gaudin T, Brandstrom H, Beck A, Verner A, Kingsley J, Harmsen E, Labuda D, Morgan K, Vohl MC, Naumova AK, Sinnett D, Hudson TJ (2004) A survey of genetic and epigenetic variation affecting human gene expression. Physiol Genomics 16:184–193
20. Cowles CR, Hirschhorn JN, Altshuler D, Lander ES (2002) Detection of regulatory variation in mouse genes. Nat Genet 32:432–437
21. Bray NJ, Buckland PR, Owen MJ, O'Donovan MC (2003) Cis-acting variation in the expression of a high proportion of genes in human brain. Hum Genet 113:149–153
22. Alam J, Cook JL (1990) Reporter genes: application to the study of mammalian gene transcription. Anal Biochem 188:245–254
23. Rockman MV, Wray GA (2002) Abundant raw material for cis-regulatory evolution in humans. Mol Biol Evol 19:1991–2004
24. Hoogendoorn B, Coleman SL, Guy CA, Smith K, Bowen T, Buckland PR, O'Donovan MC (2003) Functional analysis of human promoter polymorphisms. Hum Mol Genet 12:2249–2254
25. Buckland PR, Coleman SL, Hoogendoorn B, Guy C, Smith SK, O'Donovan MC (2004) A high proportion of chromosome 21 promoter polymorphisms influence transcriptional activity. Gene Expr 11:233–239
26. Hoogendoorn B, Coleman SL, Guy CA, Smith SK, O'Donovan MC, Buckland PR (2004) Functional analysis of polymorphisms in the promoter regions of genes on 22q11. Hum Mutat 24:35–42
27. Banerjee P, Bahlo M, Schwartz JR, Loots GG, Houston KA, Dubchak I, Speed TP, Rubin EM (2002) SNPs in putative regulatory regions identified by human mouse comparative sequencing and transcription factor binding site data. Mamm Genome 13:554–557
28. Bulyk ML (2003) Computational prediction of transcription-factor binding site locations. Genome Biol 5:201
29. Linnell J, Mott R, Field S, Kwiatkowski DP, Ragoussis J, Udalova IA (2004) Quantitative high-throughput analysis of transcription factor binding specificities. Nucleic Acids Res 32:e44
30. Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, Rainbow DB, Hunter KM, Smith AN, Di Genova G, Herr MH, Dahlman I, Payne F, Smyth D, Lowe C, Twells RC, Howlett S, Healy B, Nutland S, Rance HE, Everett V, Smink LJ, Lam AC, Cordell HJ, Walker NM, Bordin C, Hulme J, Motzo C, Cucca F, Hess JF, Metzker ML, Rogers J, Gregory S, Allahabadia A, Nithiyananthan R, Tuomilehto-Wolf E, Tuomilehto J, Bingley P, Gillespie KM, Undlien DE, Ronningen KS, Guja C, Ionescu-Tirgoviste C, Savage DA, Maxwell AP, Carson DJ, Patterson CC, Franklyn JA, Clayton DG, Peterson LB, Wicker LS, Todd JA, Gough SC (2003) Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 423:506–511
31. Tournamille C, Colin Y, Cartron JP, Le Van Kim C (1995) Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat Genet 10:224–228
32. Kennedy GC, German MS, Rutter WJ (1995) The minisatellite in the diabetes susceptibility locus IDDM 2 regulates insulin transcription. Nat Genet 9:293–298
33. Bell GI, Selby MJ, Rutter WJ (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature 295:31–35
34. Bell GI, Horita S, Karam JH (1984) A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes 33:176–183
35. Hitman GA, Tarn AC, Winter RM, Drummond V, Williams LG, Jowett NI, Bottazzo GF, Galton DJ (1985) Type 1 (insulin-dependent) diabetes and a highly variable locus close to the insulin gene on chromosome 11. Diabetologia 28:218–222

36. Lucassen AM, Julier C, Beressi JP, Boitard C, Froguel P, Lathrop M, Bell JI (1993) Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. Nat Genet 4:305–310
37. Bennett ST, Lucassen AM, Gough SC, Powell EE, Undlien DE, Pritchard LE, Merriman ME, Kawaguchi Y, Dronsfield MJ, Pociot F et al (1995) Susceptibility to human type 1 diabetes at IDDM 2 is determined by tandem repeat variation at the insulin gene minisatellite locus. Nat Genet 9:284–292
38. Barratt BJ, Payne F, Lowe CE, Hermann R, Healy BC, Harold D, Concannon P, Gharani N, McCarthy MI, Olavesen MG, McCormack R, Guja C, Ionescu-Tirgoviste C, Undlien DE, Ronningen KS, Gillespie KM, Tuomilehto-Wolf E, Tuomilehto J, Bennett ST, Clayton DG, Cordell HJ, Todd JA (2004) Remapping the insulin gene/IDDM 2 locus in type 1 diabetes. Diabetes 53:1884–1889
39. Pugliese A, Awdeh ZL, Alper CA, Jackson RA, Eisenbarth GS (1994) The paternally inherited insulin gene B allele (1,428 FokI site) confers protection from insulin-dependent diabetes in families. J Autoimmunol 7:687–694
40. Bennett ST, Wilson AJ, Cucca F, Nerup J, Pociot F, McKinney PA, Barnett AH, Bain SC, Todd JA (1996) IDDM 2-VNTR-encoded susceptibility to type 1 diabetes: dominant protection and parental transmission of alleles of the insulin gene-linked minisatellite locus. J Autoimmunol 9:415–421
41. Florez JC, Hirschhorn J, Altshuler D (2003) The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. Annu Rev Genomics Hum Genet 4:257–291
42. Vafiadis P, Bennett ST, Colle E, Grabs R, Goodyer CG, Polychronakos C (1996) Imprinted and genotype-specific expression of genes at the IDDM 2 locus in pancreas and leucocytes. J Autoimmunol 9:397–403
43. Weaver JU, Kopelman PG, Hitman GA (1992) Central obesity and hyperinsulinaemia in women are associated with polymorphism in the 5′ flanking region of the human insulin gene. Eur J Clin Invest 22:265–270
44. Cocozza S, Riccardi G, Monticelli A, Capaldo B, Genovese S, Krogh V, Celentano E, Farinaro E, Varrone S, Avvedimento VE (1988) Polymorphism at the 5′ end flanking region of the insulin gene is associated with reduced insulin secretion in healthy individuals. Eur J Clin Invest 18:582–586
45. Permutt MA, Rotwein P, Andreone T, Ward WK, Porte D Jr (1985) Islet beta-cell function and polymorphism in the 5′-flanking region of the human insulin gene. Diabetes 34:311–314
46. Owerbach D, Poulsen S, Billesbolle P, Nerup J (1982) DNA insertion sequences near the insulin gene affect glucose regulation. Lancet I:880–883
47. Lucassen AM, Screaton GR, Julier C, Elliott TJ, Lathrop M, Bell JI (1995) Regulation of insulin gene expression by the IDDM associated, insulin locus haplotype. Hum Mol Genet 4:501–506
48. Catasti P, Chen X, Moyzis RK, Bradbury EM, Gupta G (1996) Structure-function correlations of the insulin-linked polymorphic region. J Mol Biol 264:534–545
49. Lew A, Rutter WJ, Kennedy GC (2000) Unusual DNA structure of the diabetes susceptibility locus IDDM 2 and its effect on transcription by the insulin promoter factor Pur-1/MAZ. Proc Natl Acad Sci U S A 97:12508–12512
50. Pugliese A, Zeller M, Fernandez A Jr, Zalcberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD (1997) The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM 2 susceptibility locus for type 1 diabetes. Nat Genet 15:293–297
51. Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C (1997) Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM 2 locus. Nat Genet 15:289–292
52. Paquette J, Giannoukakis N, Polychronakos C, Vafiadis P, Deal C (1998) The INS 5′ variable number of tandem repeats is associated with IGF2 expression in humans. J Biol Chem 273:14158–14164
53. Vafiadis P, Grabs R, Goodyer CG, Colle E, Polychronakos C (1998) A functional analysis of the role of IGF2 in IDDM 2-encoded susceptibility to type 1 diabetes. Diabetes 47:831–836
54. Cooke GS, Hill AV (2001) Genetics of susceptibility to human infectious disease. Nat Rev Genet 2:967–977
55. Theodorou I, Capoulade C, Combadiere C, Debre P (2003) Genetic control of HIV disease. Trends Microbiol 11:392–397
56. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, Saragosti S, Lapoumeroulie C, Cognaux J, Forceille C, Muyldermans G, Verhofstede C, Burtonboy G, Georges M, Imai T, Rana S, Yi Y, Smyth RJ, Collman RG, Doms RW, Vassart G, Parmentier M (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382:722–725
57. Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T, Kang S, Ceradini D, Jin Z, Yazdanbakhsh K, Kunstman K, Erickson D, Dragon E, Landau NR, Phair J, Ho DD, Koup RA (1996) The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat Med 2:1240–1243
58. Anastassopoulou CG, Kostrikis LG (2003) The impact of human allelic variation on HIV-1 disease. Curr HIV Res 1:185–203
59. Martin MP, Dean M, Smith MW, Winkler C, Gerrard B, Michael NL, Lee B, Doms RW, Margolick J, Buchbinder S, Goedert JJ, O'Brien TR, Hilgartner MW, Vlahov D, O'Brien SJ, Carrington M (1998) Genetic acceleration of AIDS progression by a promoter variant of CCR5. Science 282:1907–1911
60. McDermott DH, Zimmerman PA, Guignard F, Kleeberger CA, Leitman SF, Murphy PM (1998) CCR5 promoter polymorphism and HIV-1 disease progression. Multicenter AIDS Cohort Study (MACS). Lancet 352:866–870
61. Mummidi S, Ahuja SS, Gonzalez E, Anderson SA, Santiago EN, Stephan KT, Craig FE, O'Connell P, Tryon V, Clark RA, Dolan MJ, Ahuja SK (1998) Genealogy of the CCR5 locus and chemokine system gene variants associated with altered rates of HIV-1 disease progression. Nat Med 4:786–793
62. Clegg AO, Ashton LJ, Biti RA, Badhwar P, Williamson P, Kaldor JM, Stewart GJ (2000) CCR5 promoter polymorphisms, CCR5 59029A and CCR5 59353C, are under represented in HIV-1-infected long-term non-progressors. The Australian Long-Term Non-Progressor Study Group. AIDS 14:103–108
63. Knudsen TB, Kristiansen TB, Katzenstein TL, Eugen-Olsen J (2001) Adverse effect of the CCR5 promoter −2459A allele on HIV-1 disease progression. J Med Virol 65:441–444
64. Mummidi S, Bamshad M, Ahuja SS, Gonzalez E, Feuillet PM, Begum K, Galvis MC, Kostecki V, Valente AJ, Murthy KK, Haro L, Dolan MJ, Allan JS, Ahuja SK (2000) Evolution of human and non-human primate CC chemokine receptor 5 gene and mRNA. Potential roles for haplotype and mRNA diversity, differential haplotype-specific transcriptional activity, and altered transcription factor binding to polymorphic nucleotides in the pathogenesis of HIV-1 and simian immunodeficiency virus. J Biol Chem 275:18946–18961
65. Salkowitz JR, Bruse SE, Meyerson H, Valdez H, Mosier DE, Harding CV, Zimmerman PA, Lederman MM (2003) CCR5 promoter polymorphism determines macrophage CCR5 density and magnitude of HIV-1 propagation in vitro. Clin Immunol 108:234–240
66. Kawamura T, Gulden FO, Sugaya M, McNamara DT, Borris DL, Lederman MM, Orenstein JM, Zimmerman PA, Blauvelt A (2003) R5 HIV productively infects Langerhans cells, and infection levels are regulated by compound CCR5 polymorphisms. Proc Natl Acad Sci U S A 100:8401–8406
67. Bream JH, Young HA, Rice N, Martin MP, Smith MW, Carrington M, O'Brien SJ (1999) CCR5 promoter alleles and specific DNA binding factors. Science 284:223

68. Jepson A, Sisay-Joof F, Banya W, Hassan-King M, Frodsham A, Bennett S, Hill AVS, Whittle H (1997) Genetic linkage of mild malaria to the major histocompatibility complex in Gambian children: study of affected sibling pairs. BMJ 315:96–97
69. Flori L, Sawadogo S, Esnault C, Delahaye NF, Fumoux F, Rihet P (2003) Linkage of mild malaria to the major histocompatibility complex in families living in Burkina Faso. Hum Mol Genet 12:375–378
70. McGuire W, Hill AVS, Allsopp CEM, Greenwood BM, Kwiatkowski D (1994) Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. Nature 371:508–511
71. Wattavidanage J, Carter R, Perera KL, Munasingha A, Bandara S, McGuinness D, Wickramasinghe AR, Alles HK, Mendis KN, Premawansa S (1999) TNFalpha*2 marks high risk of severe disease during Plasmodium falciparum malaria and other infections in Sri Lankans. Clin Exp Immunol 115:350–355
72. Padyukov L, Lampa J, Heimburger M, Ernestam S, Cederholm T, Lundkvist I, Andersson P, Hermansson Y, Harju A, Klareskog L, Bratt J (2003) Genetic markers for the efficacy of tumour necrosis factor blocking therapy in rheumatoid arthritis. Ann Rheum Dis 62:526–529
73. Mugnier B, Balandraud N, Darque A, Roudier C, Roudier J, Reviron D (2003) Polymorphism at position −308 of the tumor necrosis factor alpha gene influences outcome of infliximab therapy in rheumatoid arthritis. Arthritis Rheum 48:1849–1852
74. Allen RD (1999) Polymorphism of the human TNF-alpha promoter-random variation or functional diversity? Mol Immunol 36:1017–1027
75. Bayley JP, Ottenhoff TH, Verweij CL (2004) Is there a future for TNF promoter polymorphisms? Genes Immunol 5:315–329
76. Kroeger KM, Steer JH, Joyce DA, Abraham LJ (2000) Effects of stimulus and cell type on the expression of the −308 tumour necrosis factor promoter polymorphism. Cytokine 12:110–119
77. Kroeger KM, Carville KS, Abraham LJ (1997) The −308 tumor necrosis factor-alpha promoter polymorphism effects transcription. Mol Immunol 34:391–399
78. Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW (1997) Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation. Proc Natl Acad Sci USA 94:3195–3199
79. Knight JC, Udalova I, Hill AV, Greenwood BM, Peshu N, Marsh K, Kwiatkowski D (1999) A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. Nat Genet 22:145–150
80. Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP (2003) In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. Nat Genet 33:469–475
81. Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Hori M, Nakamura Y, Tanaka T (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 32:650–654
82. Newton J, Brown MA, Milicic A, Ackerman H, Darke C, Wilson JN, Wordsworth BP, Kwiatkowski D (2003) The effect of HLA-DR on susceptibility to rheumatoid arthritis is influenced by the associated lymphotoxin alpha-tumor necrosis factor haplotype. Arthritis Rheum 48:90–96
83. Knight JC, Keating BJ, Kwiatkowski DP (2004) Allele-specific repression of lymphotoxin-alpha by activated B cell factor-1. Nat Genet 36:394–399
84. McGuire W, Knight JC, Hill AV, Allsopp CE, Greenwood BM, Kwiatkowski D (1999) Severe malarial anemia and cerebral malaria are associated with different tumor necrosis factor promoter alleles. J Infect Dis 179:287–290
85. Mombo LE, Ntoumi F, Bisseye C, Ossari S, Lu CY, Nagel RL, Krishnamoorthy R (2003) Human genetic polymorphisms and asymptomatic Plasmodium falciparum malaria in Gabonese schoolchildren. Am J Trop Med Hyg 68:186–190
86. Fong CL, Siddiqui AH, Mark DF (1994) Identification and characterization of a novel repressor site in the human tumor necrosis factor alpha gene. Nucleic Acids Res 22:1108–1114
87. Fabris M, Di PE, D'Elia A, Damante G, Sinigaglia L, Ferraccioli G (2002) Tumor necrosis factor-alpha gene polymorphism in severe and mild-moderate rheumatoid arthritis. J Rheumatol 29:29–33
88. Brinkman BM, Huizinga TW, Kurban SS, van der Velde EA, Schreuder GM, Hazes JM, Breedveld FC, Verweij CL (1997) Tumour necrosis factor alpha gene polymorphisms in rheumatoid arthritis: association with susceptibility to, or severity of, disease? Br J Rheumatol 36:516–521
89. Perry DJ (2002) Factor VII Deficiency. Br J Haematol 118:689–700
90. Arbini AA, Pollak ES, Bayleran JK, High KA, Bauer KA (1997) Severe factor VII deficiency due to a mutation disrupting a hepatocyte nuclear factor 4 binding site in the factor VII promoter. Blood 89:176–182
91. Pollak ES, Hung HL, Godin W, Overton GC, High KA (1996) Functional characterization of the human factor VII 5′-flanking region. J Biol Chem 271:1738–1747
92. Carew JA, Pollak ES, Lopaciuk S, Bauer KA (2000) A new mutation in the HNF4 binding region of the factor VII promoter in a patient with severe factor VII deficiency. Blood 96:4370–4372
93. Carew JA, Pollak ES, High KA, Bauer KA (1998) Severe factor VII deficiency due to a mutation disrupting an Sp1 binding site in the factor VII promoter. Blood 92:1639–1645
94. Greenberg D, Miao CH, Ho WT, Chung DW, Davie EW (1995) Liver-specific expression of the human factor VII gene. Proc Natl Acad Sci USA 92:12347–12351
95. Meade TW, Mellows S, Brozovic M, Miller GJ, Chakrabarti RR, North WR, Haines AP, Stirling Y, Imeson JD, Thompson SG (1986) Haemostatic function and ischaemic heart disease: principal results of the Northwick Park Heart Study. Lancet II:533–537
96. Bernardi F, Marchetti G, Pinotti M, Arcieri P, Baroncini C, Papacchini M, Zepponi E, Ursicino N, Chiarotti F, Mariani G (1996) Factor VII gene polymorphisms contribute about one third of the factor VII level variation in plasma. Arterioscler Thromb Vasc Biol 16:72–76
97. Green F, Kelleher C, Wilkes H, Temple A, Meade T, Humphries S (1991) A common genetic polymorphism associated with lower coagulation factor VII levels in healthy individuals. Arterioscler Thromb 11:540–546
98. Iacoviello L, Di Castelnuovo A, De Knijff P, D'Orazio A, Amore C, Arboretti R, Kluft C, Benedetta Donati M (1998) Polymorphisms in the coagulation factor VII gene and the risk of myocardial infarction. N Engl J Med 338:79–85
99. Girelli D, Russo C, Ferraresi P, Olivieri O, Pinotti M, Friso S, Manzato F, Mazzucco A, Bernardi F, Corrocher R (2000) Polymorphisms in the factor VII gene and the risk of myocardial infarction in patients with coronary artery disease. N Engl J Med 343:774–780
100. Doggen CJ, Manger Cats V, Bertina RM, Reitsma PH, Vandenbroucke JP, Rosendaal FR (1998) A genetic propensity to high factor VII is not associated with the risk of myocardial infarction in men. Thromb Haemost 80:281–285
101. Batalla A, Alvarez R, Reguero JR, Gonzalez P, Alvarez V, Cubero GI, Cortina A, Coto E (2001) Lack of association between polymorphisms of the coagulation factor VII and myocardial infarction in middle-aged Spanish men. Int J Cardiol 80:209–212
102. Carew JA, Basso F, Miller GJ, Hawe E, Jackson AA, Humphries SE, Bauer KA (2003) A functional haplotype in the 5′ flanking region of the factor VII gene is associated with an increased risk of coronary heart disease. J Thromb Haemost 1:2179–2185
103. Di Castelnuovo A, D'Orazio A, Amore C, Falanga A, Donati MB, Iacoviello L (2000) The decanucleotide insertion/deletion

polymorphism in the promoter region of the coagulation factor VII gene and the risk of familial myocardial infarction. Thromb Res 98:9–17
104. Humphries S, Temple A, Lane A, Green F, Cooper J, Miller G (1996) Low plasma levels of factor VIIc and antigen are more strongly associated with the 10 base pair promoter (-323) insertion than the glutamine 353 variant. Thromb Haemost 75:567–572
105. Hunault M, Arbini AA, Lopaciuk S, Carew JA, Bauer KA (1997) The Arg353Gln polymorphism reduces the level of coagulation factor VII. In vivo and in vitro studies. Arterioscler Thromb Vasc Biol 17:2825–2829
106. Kudaravalli R, Tidd T, Pinotti M, Ratti A, Santacroce R, Margaglione M, Dallapiccola B, Bernardi F, Fortina P, Devoto M, Pollak ES (2002) Polymorphic changes in the 5′ flanking region of factor VII have a combined effect on promoter strength. Thromb Haemost 88:763–767
107. Hooft FM van 't, Silveira A, Tornvall P, Iliadou A, Ehrenborg E, Eriksson P, Hamsten A (1999) Two common functional polymorphisms in the promoter region of the coagulation factor VII gene determining plasma factor VII activity and mass concentration. Blood 93:3432–3441
108. Pinotti M, Toso R, Girelli D, Bindini D, Ferraresi P, Papa ML, Corrocher R, Marchetti G, Bernardi F (2000) Modulation of factor VII levels by intron 7 polymorphisms: population and in vitro studies. Blood 95:3423–3428
109. Miller LH, Mason SJ, Clyde DF, McGinniss MH (1976) The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. N Engl J Med 295:302–304
110. Horuk R, Chitnis CE, Darbonne WC, Colby TJ, Rybicki A, Hadley TJ, Miller LH (1993) A receptor for the malarial parasite Plasmodium vivax: the erythrocyte chemokine receptor. Science 261:1182–1184
111. Hamblin MT, Di Rienzo A (2000) Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am J Hum Genet 66:1669–1679
112. Maniatis T, Reed R (2002) An extensive network of coupling among gene expression machines. Nature 416:499–506
113. Orphanides G, Reinberg D (2002) A unified theory of gene expression. Cell 108:439–451
114. Pagani F, Baralle FE (2004) Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet 5:389–396
115. Lynch KW, Weiss A (2001) A CD45 polymorphism associated with multiple sclerosis disrupts an exonic splicing silencer. J Biol Chem 276:24341–24347
116. Jacobsen M, Schweer D, Ziegler A, Gaber R, Schock S, Schwinzer R, Wonigeit K, Lindert RB, Kantarci O, Schaefer-Klein J, Schipper HI, Oertel WH, Heidenreich F, Weinshenker BG, Sommer N, Hemmer B (2000) A point mutation in PTPRC is associated with the development of multiple sclerosis. Nat Genet 26:495–499
117. Barcellos LF, Caillier S, Dragone L, Elder M, Vittinghoff E, Bucher P, Lincoln RR, Pericak-Vance M, Haines JL, Weiss A, Hauser SL, Oksenberg JR (2001) PTPRC (CD45) is not associated with the development of multiple sclerosis in U.S. patients. Nat Genet 29:23–24
118. Schwinzer R, Wonigeit K (1990) Genetically determined lack of CD45R-T cells in healthy individuals. Evidence for a regulatory polymorphism of CD45R antigen expression. J Exp Med 171:1803–1808
119. Thude H, Hundrieser J, Wonigeit K, Schwinzer R (1995) A point mutation in the human CD45 gene associated with defective splicing of exon A. Eur J Immunol 25:2101–2106
120. Zilch CF, Walker AM, Timon M, Goff LK, Wallace DL, Beverley PC (1998) A point mutation within CD45 exon A is the cause of variant CD45RA splicing in humans. Eur J Immunol 28:22–29
121. Zito F, Lowe GD, Rumley A, McMahon AD, Humphries SE (2002) Association of the factor XII 46C>T polymorphism with risk of coronary heart disease (CHD) in the WOSCOPS study. Atherosclerosis 165:153–158
122. Tirado I, Manuel Soria J, Mateo J, Oliver A, Carlos Souto J, Santamaria A, Felices R, Borrell M, Fontcuberta J (2004) Association after linkage analysis indicates that homozygosity for the 46C->T polymorphism in the F12 gene is a genetic risk factor for venous thrombosis. Thromb Haemost 91:899–904
123. Kanaji T, Okamura T, Osaki K, Kuroiwa M, Shimoda K, Hamasaki N, Niho Y (1998) A common genetic polymorphism (46 C to T substitution) in the 5′-untranslated region of the coagulation factor XII gene is associated with low translation efficiency and decrease in plasma factor XII level. Blood 91:2010–2014
124. Kohler HP, Futers TS, Grant PJ (1999) FXII (46C->T) polymorphism and in vivo generation of FXII activity-gene frequencies and relationship in patients with coronary artery disease. Thromb Haemost 81:745–747
125. Zito F, Drummond F, Bujac SR, Esnouf MP, Morrissey JH, Humphries SE, Miller GJ (2000) Epidemiological and genetic associations of activated factor XII concentration with factor VII activity, fibrinopeptide A concentration, and risk of coronary heart disease in men. Circulation 102:2058–2062