HUGO J (2011) 5:13–23
DOI 10.1007/s11568-011-9155-4

REVIEW

The impact of *cis*-acting polymorphisms on the human phenotype

Bryony L. Jones · Dallas M. Swallow

Received: 9 March 2011 / Revised: 27 May 2011 / Accepted: 17 June 2011 / Published online: 20 July 2011

Abstract *Cis*-acting polymorphisms that affect gene expression are now known to be frequent, although the extent and mechanisms by which such variation affects the human phenotype are, as yet, only poorly understood. Key signatures of *cis*-acting variation are differences in gene expression that are tightly associated with regulatory SNPs or expression Quantitative Trait Loci (eQTL) and an imbalance of allelic expression (AEI) in heterozygous samples. Such *cis*-acting sequence differences appear often to have been under selection within and between populations and are also thought to be important in speciation. Here we describe the example of lactase persistence. In medical research, variants that affect regulation in *cis* have been implicated in both monogenic and polygenic disorders, and in the metabolism of drugs. In this review we suggest that by further understanding common regulatory variations and how they interact with other genetic and environmental variables it will be possible to gain insight into important mechanisms behind complex disease, with the potential to lead to new methods of diagnosis and treatments.

Keywords *Cis*-acting polymorphism · Gene expression · Regulation · Phenotypic variability · Allelic expression · Soft selective sweeps

Introduction

Understanding the mechanisms behind phenotypic variation is a key aim in human genetics, particularly in relation to disease and disease susceptibility. With approximately 1 difference per 1,000 nucleotides between any two human individuals chosen at random, deciphering which variants have function is a vital objective in modern day genetic research. The availability of a much larger range of methodologies has led to a bias towards studies focused on coding sequence variation. The functioning of each gene however, is determined not only by the protein itself but is also governed by spatiotemporal expression patterns, with differences in the timings and levels of expression of genes potentially altering phenotype. Considerable differences in gene expression have been demonstrated both between individuals and populations (Cheung et al. 2005; Stranger et al. 2005, 2007) and it has long been speculated that much of the phenotypic diversity within and between species is due to genetically determined differences in gene expression levels (King and Wilson 1975; Skelly et al. 2009).

Gene expression is controlled by genetic factors, acting both in *cis* and in *trans*, epigenetic factors and environmental influences (Stranger et al. 2007) with many complex heritable traits controlled by both *cis* and *trans* acting loci (Cheung et al. 2010; Wang et al. 2008). Mutations in *cis*-regulatory elements (including promoters, enhancers, silencers and insulators) can disrupt or enhance the binding of transcription factors and alter the state of gene expression of single genes. Transposable elements also appear to have influenced mammalian regulation with the creation of repeat associated binding sites (Bourque et al. 2008; Kunarso et al. 2010). Heterozygotes for regulatory SNPs display intermediate levels of expression and show allelic expression imbalance (AEI), one of the hallmarks of *cis*-acting variation (Fig. 1). Functional mutations in transcription factors (in *trans*), in contrast, are often pleiotropic because of modified binding to multiple transcription factor binding sites (TFBS), which may alter the expression of

B. L. Jones · D. M. Swallow (✉)
Research Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK
e-mail: d.swallow@ucl.ac.uk

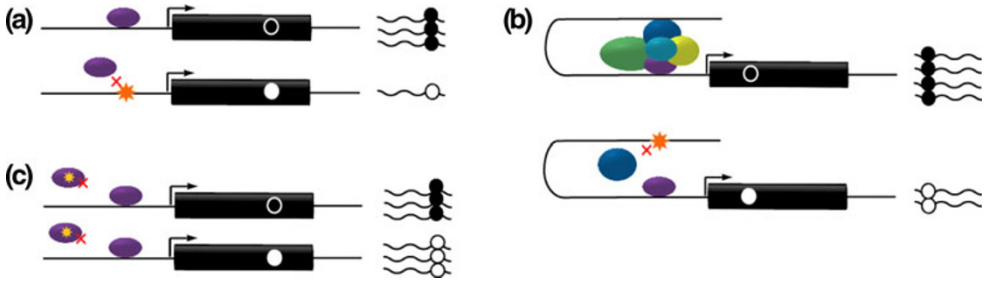

Fig. 1 Heterozygous allelic imbalance through *cis*-acting regulatory variation. **a** A mutation in a proximal promoter may prevent transcription factor binding altering expression of the allelic transcript. Marker SNPs (shown here as *black/white circles*) can be used to determine transcript ratios. **b** A mutation in a distal enhancer site may prevent combinatorial binding and affect transcription levels. **c** A mutation in *trans* will affect both alleles equally and no AEI will be seen

many genes. *Trans*-acting variants involved in gene expression operate equally on both chromosomes so no imbalance of allelic expression is seen (Fig. 1c).

TFBSs are usually 6–20 base pairs in length (Lapidot et al. 2008) and imprecise; one transcription factor may bind to a range of similar sequence TFBSs. Although these sites can differ, there are clear biases in the expectations of bases that appear at each binding site position, probabilities of which can be displayed using position weight matrices (PWMs) (Lapidot et al. 2008; Stormo 2000). Because of this flexibility in specificity, mutations in binding sites may not alter binding and thus have no phenotypic effect. In other instances, a nucleotide change within a binding site leads to an altered binding affinity or even the loss or gain of a TFBS (Doniger and Fay 2007), and it is these variations which have the potential to alter gene expression and consequently affect phenotype. Since transcription factors show different gene expression profiles in different cell types, nucleotide changes in binding sites will have different implications in different cells. This review will focus primarily on current understanding of the effects of *cis*-regulation on the human phenotype. Note that mutations which affect copy number, RNA stability or splicing also affect transcript levels in a *cis*-acting manner, but are not specifically considered in this review.

Methods of detecting *cis*-acting variation

Until high-throughput methods for detecting gene expression were developed, techniques for the evaluation of gene expression were limited to single gene analyses. Semi-quantitative reverse transcription PCR and real time PCR were commonly used to quantify levels of expression and could be adapted to measure individual allelic transcripts of a gene by making use of marker SNPs located in the exons. The allelic transcripts could be distinguished by a variety of methods including direct sequencing and RFLP primer extension assays (Loh et al. 2010; Wang et al. 1995, 1998). By comparing the relative expression levels of allelic transcripts within a heterozygous sample, *cis*-acting differences or 'allelic expression' can be measured (AE), while internally controlling for environmental and *trans*-acting factors (Pastinen and Hudson 2004; Verlaan et al. 2009). This approach also detects epigenetic effects such as imprinting (Pollard et al. 2008; Verlaan et al. 2009) although in this case, analysis of parents would show exclusively maternal or paternal inheritance.

In the mid 1990's, it became possible to measure the expression levels of thousands of genes simultaneously and various approaches have since been developed which utilise hybridisation or sequence based methods with the capability of measuring complete transcriptomes (Bertone et al. 2004; Cheng et al. 2005). Microarrays exploit hybridisation methods and fluorescence, and kits using oligonucleotide sequences or cDNA probes were made available commercially that allow mapping of patterns of expression but measure both allelic transcripts of a gene simultaneously. However by demonstrating association of SNPs near to the gene with expression levels (Stranger et al. 2005), it was possible to provide evidence of, and map, *cis*-acting regulatory loci. Such 'eQTL mapping' is a way of determining the relationship between the genome and transcriptome and has led to the observation that *cis*-eQTLs are a common cause of variation in gene expression (Stranger et al. 2005, 2007). The availability of bead chips has enabled genome-wide allelic expression studies by allowing the comparison of SNP allele ratios in expressed RNA transcripts normalized against genomic DNA heterozygote ratios (Ge et al. 2009).

The manipulation of next generation sequencing (NGS) methods to measure gene expression levels (and also for the detection of alternative splicing or novel transcripts) has been one of the most striking advances in transcriptomics in recent years. RNA-sequencing (RNA-Seq) of a population of RNA (total or fractionated) involves

converting the RNA to cDNA fragments attached to adapters and direct sequencing using NGS methods. The transcripts are then aligned or assembled to produce a map of the transcriptome and expression levels for each gene (Wang et al. 2009). RNA-Seq is able to describe previously unknown or novel sequences and discover new variants in transcribed sequences. Importantly it is able to detect allele specific expression although, as with all AE methods that rely on heterozygous markers within exons, the number of SNPs within the coding regions and the number of individuals who are heterozygous for any particular gene transcript can be restrictive. A recent study that utilised both unspliced primary transcripts as well as mRNA, identified over 50% more genes showing AE differences than if exonic SNPs alone were used (Verlaan et al. 2009).

An important limitation in studying *cis*-regulation genome-wide is that analysis of the transcriptome is ultimately dependent on the cell type and developmental stage. So far the majority of genome wide studies have used a small number of cell lines giving a very restrictive representation of tissue specific gene transcripts. Because of the density and coverage of the HapMap SNP data and DNA availability from the lymphoblastoid cell lines (LCL) for HapMap individuals, the majority of regulatory studies have been conducted in LCL cells (see for example (Ge et al. 2009; Stranger et al. 2005, 2007). Concerns have been raised about the use of LCLs in determining allelic expression, because of changes that may have occurred in cell culture (Gimelbrant et al. 2007; Pastinen et al. 2004) although it has since been reported that this is not a significant source of spurious AE association and that it is possible to correct for the confounding effect (Ge et al. 2009).

Where it is possible to compare allelic expression across cell types, further information on the complexity of the regulation of gene expression can be gained. Tissue specificity of AE has been documented in mice (Campbell et al. 2008) and differences in allelic expression have been demonstrated between human brain regions (Buonocore et al. 2010). Genome wide comparison of AE in human cell lines appears to show the same effect and indicates an enrichment of gene networks associated with immunological disease in LCLs, and musculoskeletal disease in human osteoblast cells with several of these genes implicated in multifactorial disease (Verlaan et al. 2009). A study of gene expression across three cell types estimated 69–80% of regulatory variants were cell type specific (Dimas et al. 2009) and recent work by Ernst and colleagues (using chromatin profiling across nine cell types) suggests disease associated SNPs are frequently found within enhancer elements and appear to be active in specific and relevant cell types (Ernst et al. 2011). Both of these studies clearly demonstrate the importance of examining expression across multiple cell types in future studies.

Locating putative functional elements and variants

In the absence of suitable cell lines, another method of identifying potential functional regulatory variants is to look at conservation across species (Goode et al. 2010; Lomelin et al. 2010). Goode and colleagues were able to show that a very large percentage of SNPs in conserved regions, identified by evolutionary rate profiling, were in non-coding regions under evolutionary constraint (Goode et al. 2010). By making use of a transgenic mouse embryo system, Pennachio et al. further characterised some of the conserved sequences and used this information to rank and map potential enhancers across the human genome (Pennacchio et al. 2006).

However as increasing evidence suggests that many *cis*-regulatory sequence regions (and many of the non-coding alleles under purifying selection (Asthana et al. 2007)) may be poorly conserved (Burton et al. 2007; Kunarso et al. 2010; Pennacchio and Visel 2010; Yokoyama et al. 2011), an alternative method is to identify regions of the genome with a high density of TFBSs (proposed to be more likely functional than regions with single TFBSs due to the formation of transcription factor complexes) (Yu et al. 2007; Zinzen et al. 2009). By selecting relevant or interracting TFs for a set of tissues, a bioinformatics study that followed this approach detected genes that appeared to be regulated by different TF clusters/complexes in different tissues and observed that conserved regulatory modules are more likely to regulate essential genes than non-conserved modules (Yu et al. 2007). The recent development of bioinformatics tools, such as COMPASSS (COMplex Pattern of Sequence Search Software) (Maccari et al. 2010), provide a further resource in the detection of putative functional *cis*-acting elements in the genome. As the empirical datasets with binding, sequence and expression data are rapidly expanding, it is likely that the sensitivity of these methods will increase and provide further insight into the complexities of *cis*-acting regulation.

Methods to detect function of *cis*-acting regulatory regions

To identify or confirm particular functional elements in the genome, including promoters, enhancers, silencers and insulators, a number of empirical approaches can be taken. These range from methods involving transfection of reporter constructs to those examining properties of chromatin. For example, by immunoprecipitating from chromatin, ChIP-chip experiments are able to localize binding sites for any protein of interest. Approaches to identify active gene transcription include the Haplochip method

which exploits the fact that the amount of chromatin-bound active Pol II RNA polymerase enzyme is related to the transcriptional activity of the corresponding gene, so that differences in active Pol II loading between the two alleles in a heterozygous sample provides a measure of allele-specific gene expression (Knight et al. 2003). However to assess the effect of particular nucleotide substitutions it is more usual to conduct gel shift assays, often in combination with antibodies, to detect protein binding (for example Hultman et al. 2010), or transfections of constructs in which the relevant mutations have been introduced (for example Jensen et al. 2011; Troelsen et al. 2003). Both these methods have the limitation that they reflect effects in vitro which may not well replicate what happens in vivo, where for example different proportions of transcription factors may be present.

Cis-acting variation and evolution

The existence of conserved non-coding elements (CNE) has been helpful in identifying tissue specific enhancers (Lee et al. 2011; Pennacchio et al. 2006) as described above. However, regulatory elements may be lineage specific and also show evidence of rapid evolution in some groups, as described, for example, in Teleost fishes (Lee et al. 2011) where a high level of loss is associated with a whole genome duplication. It is now evident that *cis*-regulatory mutations have been important in the adaptive evolution of populations both in terms of speciation and environmental adaptations or adaptive divergence of particular species (Fay and Wittkopp 2008; Bourque et al. 2008; Kunarso et al. 2010; He et al. 2011; Lee et al. 2011; Yokoyama et al. 2011). It has been proposed that certain adaptations or traits such as those involving immune responses, behaviour, reproduction and development and also gain of function mutations are more likely to occur through *cis*-regulatory mutations where transcription can be 'fine-tuned' to meet demands (Wray 2007). While a non-synonymous coding mutation usually affects the protein regardless of where or when it is expressed, a mutation in a *cis*-regulatory element has the potential to affect gene expression during a particular stage of development or in a specific cell type. Positive or adaptive selection is potentially able operate more efficiently on *cis*-regulatory regions than coding regions because single nucleotide changes are less likely to have an all or nothing effect, while at the same time they are usually co-dominant (i.e heterozygotes show intermediate phenotype) and therefore directly available to natural selection in the presence of the ancestral allele (Wray 2007). Repeat associated binding sites often appear to be lineage specific leading to the possibility that many binding sites have arisen fairly recently and rewired regulatory pathways.

Common *cis*-variation and selection in human evolution

It was first suggested some time ago that the differences between the proteins of the chimpanzee and human were insufficient to explain the differences in phenotypic characteristics and that differences in regulation could play an important role (King and Wilson 1975). One example of sequence changes in *cis*-regulatory regions in the human/chimpanzee divergence involves a 68 bp tandem repeat element 1,250 bp upstream from the start of transcription of *PDYN* (which encodes a neuropeptide with roles in cognition) for which there is evidence of function. Non-human primates have only one copy while human haplotypes contain between 1 and 4 copies (Rockman et al. 2005). Multiple other upstream polymorphisms are also thought to influence the expression of *PDYN*, some in a cell-specific or sex-specific manner (Babbitt et al. 2010). A selection of examples of polymorphic *cis*-regulatory variants with fairly certain phenotypic effect in humans are shown in Table 1. Several of these variations are involved in the inflammatory response, while others are involved in disease resistance or susceptibility and dietary adaptation.

Lactase persistence; an example of dietary adaptation through *cis*-acting regulatory polymorphisms

The ongoing expression of lactase into adulthood, in some humans but not others, is one of the classic examples of an environmental adaptation. The expression of lactase enzyme in the intestine is necessary for the breakdown of lactose, the main carbohydrate in milk. As milk is the primary source of nutrition for all newborn mammals, functional lactase enzyme is critical for survival (excluding the Pinnepedia where milk has a high fat content and low lactose content) until other food sources can be consumed. Usually lactase is down-regulated after weaning (termed 'lactase non-persistence') although approximately 35% of the human population continue to produce lactase into adulthood ('lactase persistence') (Ingram et al. 2009a). The ability to use milk as a source of nutrition without digestive complications is thought to have put some people at a selective advantage, with the expansion of this genetic trait reflecting the onset of animal domestication and milking over the last 10,000 years (Ingram et al. 2009a). It was first directly demonstrated that the inter-individual differences in expression of lactase were *cis*-acting using allelic expression techniques (Wang et al. 1995, 1998). Extended sequencing identified a C/T SNP 13,910 base pairs upstream of the lactase gene (see Fig. 2) for which the T allele was 100% associated with lactase persistence in Finns (Enattah et al. 2002), and also directed differential levels of promoter construct expression (Lewinsky et al.

Table 1 Examples of *cis*-acting functional variants thought to affect human phenotypic characteristics

Gene	Variant	Effect	Phenotype/disease risk	Methods to show function	Reference(s)
ADH1B	C>A rs1229982 in proximal promoter	Increases promoter activity 1.4 fold	Associated with alcoholism	Transfections	Pochareddy and Edenberg (2011)
DARC	T-46C	Impairs promoter activity	Resistance to malaria	Transfections	Tournamille et al. (1995)
FMO3	Promoter haplotypes (1 kb)	'2' 8 fold increase in activity, '8' & '15' loss of activity	'8' & '15' may contribute to trimethylaminuria	Transfections	Koukouritaki et al. (2005)
IL6	Promoter haplotype variation	Altered *IL6* expression	Various disease associations	Transfections	Terry et al. (2000)
IL10	Promoter haplotype variation	TA2 haplotype has higher levels of *IL10* expression	High risk TA2 haplotype associated with trachoma susceptibility	AE	Natividad et al. (2008)
IL13	C-1055T	T increases *IL13* expression	Association with asthma & allergy	Gel shift assay	van der Pouw Kraan et al. (1999)
LCT	Several in enhancer ~14 kb upstream within *MCM6*	Higher expression of lactase in adulthood	Lactase persistence into adult life	Transfections & gel shift assay	See Ingram et al. (2009a)
LTC_4S	A-444C	Altered expression	Linked to aspirin associated asthma & chronic hyperplastic eosinophilic sinusitis	AE	de Alarcon et al. (2006)
MUC5B	Several promoter haplotypes	Altered *MUC5B* expression	Underrepresentation of low expression haplotypes in patients with diffuse panbronchiolitis	AE & transfections	Kamio et al. (2005), Loh et al. (2010)
OCA2	Rs12913832 (21,152 bp upstream within *HERC2*)	Derived allele lowers promoter activity	Associated with blue eye colour in homozygotes	Transfections & gel shift assays	Eiberg et al. (2008)
SHH	Several mutations 1mb upstream in enhancer within *LMBR1* (& in other species)	Altered TF binding	Preaxial polydactyly	Gel shift assay & transgenic assay in mice	Farooq et al. (2010), Lettice et al. (2008)
UGT1A1	(TA)n repeat in TATA box	(TA)7 (low expression)	Gilberts syndrome, adverse drug reactions	Transfections	Bosma et al. (1995)
UGT2B15	Promoter haplotypes (rs34010522 & rs35513228)	Altered expression	Possible links to oxazepam glucuronidation, breast and prostate cancer risk	Transfections	Sun et al. (2010)

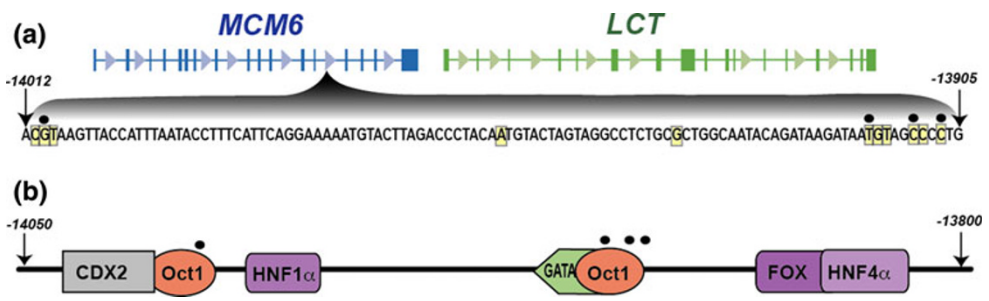

Fig. 2 The lactase enhancer region. **a** The *LCT* enhancer is located upstream in intron 13 of *MCM6*. Multiple derived alleles are found clustered in this small region, indicated by *yellow boxes*. *Dots* above the variant sites show the 4 SNPs that have been associated with LP, or shown to have function so far. The position of this region upstream of *LCT* is shown above the sequence. **b** Known transcription factor binding sites (see Lewinsky et al. 2005; Jensen et al. 2011) in the enhancer region are shown in relation to the 4 LP associated SNPs (*black dots*) with approximate positions relative to *LCT* indicated

2005; Troelsen et al. 2003) in CaCo2 cells, a colon carcinoma cell line that expresses lactase. This allele is found at high frequency (in many European populations) on an exceptionally long haplotype, indicative of directional selection (Bersaglieri et al. 2004; Poulter et al. 2003).

It was subsequently found however that this particular mutation was not responsible for the ongoing expression of lactase in all humans (Mulcare et al. 2004) and at least three further functional alleles (and likely more (Itan et al. 2010)) appear to have been selected independently (Enattah et al. 2008; Imtiaz et al. 2007; Ingram et al. 2007; Tishkoff et al. 2007). These SNPs are clustered in a region that acts as an enhancer for the lactase gene (Lewinsky et al. 2005) with several known transcription factor binding sites (Jensen et al. 2011; Lewinsky et al. 2005), outlined in Fig. 2. Interestingly, several of these SNPs can occur in a single ethnic group providing evidence in humans of a 'soft selective sweep' of the kind described below (Ingram et al. 2009b). Since Caco-2 is the only human cell line known to express significant levels of lactase, the genome wide expression studies so far have missed the *LCT* quantitative trait loci, which highlights the fact that there must be many more similar *cis*-acting variants to be discovered once more cell lines/tissues are used in genome wide expression studies.

Cis-regulation and selection for disease resistance, disease susceptibility and differing drug responses

Domestication also exposed populations to new pathogens and the resulting villages and towns facilitated the spread of diseases (Diamond 2002; Wolfe et al. 2007). Malaria is the disease for which there is the most evidence for human adaptation against infectious agents, in that a very large number of associations have been described with gene variants that confer resistance. Although many of these variants affect protein sequences, one of these mutations is a *cis*-acting regulatory mutation that disrupts the binding of the GATA1 transcription factor, preventing the expression of the Duffy blood group chemokine receptor (DARC) in erythrocytes (Tournamille et al. 1995). As the DARC protein is the usual point of entry for the malarial parasite *Plasmodium vivax*, individuals homozygous for the mutation are resistant to the disease and in fact selection has been so strong that most individuals in the endemic areas are homozygous for the mutation. This particular example is illustrative of the ability of a *cis*-acting mutation to cause a change in gene expression in one cell type only without affecting expression levels elsewhere, as individuals with the mutation still express the DARC protein in other cell types (Chaudhuri et al. 1995) and show no adverse health affects.

Variation in the number of TA repeats in the TATA box for the enzyme UDP-glucuronosyltransferase 1A1 (UGT1A1) which catalyses the conjugation of glucuronic acid to bilirubin, and also a number of pharmaceutical drugs, has been shown directly by transfection studies to affect expression of *UGT1A1,* and is associated with bilirubin level as well as with drug sensitivity (Borlak and Klutcka 2004; Bosma et al. 1995). Interestingly, the geographic distribution of the low activity alleles (which are proposed to be protective against various infectious blood pathogens) in Africa, suggest that this is another polymorphism that might have been under selection by malaria (Horsfall et al. 2011b) and is an example of polymorphism in which there are both costs and benefits, since while bilirubin is potentially toxic, moderately high levels seem to be beneficial to respiratory and cardiac health (Horsfall et al. 2011a). There are several genes involved in the metabolism of drugs, for which there is evidence for the effect of *cis*-variation in their expression (for two excellent reviews see (Hines et al. 2008; Johnson et al. 2005)). There are also many other examples of TATA box polymorphisms, some of which are involved in disease (see Savinkova et al. 2009).

Indeed there are now many examples in which promoter and other regulatory variants cause or modify Mendelian disease or disorders. Several examples come from the globin loci and have been reviewed by others (see Sankaran et al. 2010). One example of particular interest involves enhancer mutations in a *cis*-regulatory element upstream of the *SHH* gene that lead to aberrant developmental expression and incorrect limb bud formation resulting in extra digits in humans and certain other species. Several different causal mutations are located close together in this long-range (1 Mb upstream) enhancer, which is, as in the case of lactase, in an intron of another gene (Gurnett et al. 2007; Lettice et al. 2003).

Genome-wide association studies and *cis*-acting regulation

Genome-wide association studies to find the genetic components of multifactorial disease are able to locate regions along the genome that contain variants that underlie diseases or phenotype and it is now evident that many of these will be regulatory. Recent developments are providing more precision in mapping of the causal loci (Andrew et al. 2008), but at present bioinformatics methods are lacking power to identify the likely functional variants or mechanism behind the differences in gene expression levels. With the complication of different transcription factor complexes initiating transcription in different cell types, it is currently not possible to predict whether a non-coding

variant will be functional without experimental evidence in an appropriate environment. However identification of variants in binding sites that are likely to alter gene expression and cause disease is becoming more attainable on a genome-wide scale, with the use of computational prediction methods which depend on accumulation of published experimental data (Lapidot et al. 2008). For example, in *S. cerevisiae*, it appears that not all nucleotide substitutions are equal; a substitution involving guanine appears to have more of an effect than a substitution of adenine (Lapidot et al. 2008) and although this effect has not yet been shown in the human genome, it may be that similar computational methods will eventually be able to predict the variants in regulatory elements that are more likely to alter transcription in humans and thus cause disease or disease susceptibility. It may also be the case that multiple independent mutations, (perhaps closely located as in the case of the enhancer polymorphisms of *SHH* and *LCT*), influence disease susceptibility and resistance and are potentially good candidates for multifactorial disease. These are characteristically harder to identify than single mutations because they may not be efficiently tagged by the SNPs tested, and this could account for some of the 'missing heritability' in genome wide association studies.

Since regulatory elements interact with *trans*-acting factors they are often the target of environmental response and can by definition be regulated, there is potential for manipulating gene expression as a method of disease prevention or therapy. A good example of this comes from the haemoglobins, in which attempts have been made to up-regulate fetal haemoglobin (HBF) to compensate for the defective adult haemoglobin in haemoglobinopathies, as reviewed in (Sankaran et al. 2010). The switch from fetal to adult globin expression is under relatively tight developmental control, for which many of the *cis*- and *trans*-acting interactions are now known, but show inter-individual differences, as well as non-genetic alterations (e.g stress response). With the understanding of the key players involved in this regulation it is now possible to work on and refine methods of manipulating the expression of HBF.

Since changes in gene expression are an important source of evolutionary adaptation, Kudaravalli and colleagues combined eQTL data from HapMap lymphoblastoid cell lines with a haplotype based method for detecting signals of selection (Kudaravalli et al. 2009). They found a strong overlap between signals, particularly for genes known to be associated with diseases with immunological involvement, such as susceptibility to HIV infection, as might be expected for LCLs. Several recent genome wide disease association studies have attempted to overlap GWA signals with eQTLs or variants that may be involved in the regulation of genes (either through imprinting or allelic variability) (Heid et al. 2010; Kong et al. 2009; Nica et al. 2010; Voight et al. 2010) and statistical methods are being developed, that take into account LD, to make these matches more robust. However further studies to determine function are necessary and analysis of data from comparable studies in multiple cell lines may provide further evidence on the function of *cis*-regulatory variants under selection or those associated with disease.

Parallel evolution due to *cis*-acting variation: soft selective sweeps

Care must be taken not to over-simplify the characteristics of selection (Przeworski et al. 2005), and it has recently been proposed that hard sweeps, have been too infrequent in the human population over the last 250,000 years to have been responsible for much human genetic adaptation (Hernandez et al. 2011). Populations can adapt to new environmental pressures via multiple mutations, with similar phenotypic effect that arise independently. If several mutations with similar effect are selected in parallel, by what has been termed a 'soft selective sweep' (Hermisson and Pennings 2005; Pennings and Hermisson 2006a, b), they may, for example, all get to intermediate frequencies in one population or go to fixation in different populations. Soft selective sweeps can occur from standing, new or migration of mutations and as human population densities have increased over time, so has the likelihood of parallel adaptation occurring (Ralph and Coop 2010). It has also been noted that a number of more recent human adaptations have involved repeated mutations at small mutational target sites and often occur in relatively small geographic areas, while older adaptations from single changes are more widely spread (Ralph and Coop 2010).

Changes to the pigmentation in fruit flies (Gompel et al. 2005; Jeong et al. 2008), pelvic reduction in freshwater sticklebacks (Chan et al. 2010) and lactase persistence in humans (Enattah et al. 2008; Ingram et al. 2007, 2009b; Tishkoff et al. 2007) are all well described cases of parallel evolution where multiple independent *cis*-regulatory changes are responsible for a convergent phenotype. Currently, most standard tests for selection in humans are based on the 'hard sweep model' (Smith and Haigh 1974) where one allele under selection rises quickly in frequency and is consequently found at high frequency on an extended haplotypic background (Sabeti et al. 2002, 2007; Voight et al. 2006; Zhang et al. 2006). These tests may miss the diversity characteristic of soft sweeps in populations where there are multiple alleles with the same functional effect which are likely to occur on distinct haplotype backgrounds. Furthermore recent selection may have increased the allele frequency of older 'standing' variation which will also have greater haplotype diversity (Ralph and Coop 2010).

The frequency and significance of *cis*-variation in humans

Many recent developments have enabled a rapid accumulation of regulatory data in this fast advancing field. We now know that human genomes contain thousands of *cis*-regulatory variants (Rockman and Wray 2002) and considerable differences in gene expression have been demonstrated between individuals and populations (Cheung et al. 2005; Stranger et al. 2007). For example the work of Stranger et al. (2007) revealed 831 genes that displayed a significant *cis* association in lymphoblastoid cells alone, and the work of Ge and colleagues showed the phenomenon to be extremely frequent and found several thousand 'windows' of imbalance of allelic expression (Ge et al. 2009). By considering genome wide SNP frequency, distributions in non-coding DNA, and patterns of conservation it is now predicted that as much as 90% of the important functional variation in the human genome may be regulatory (Goode et al. 2010). *Cis*-regulatory variants, which have limited pleiotropic effects, are less likely to be deleterious than those acting in *trans* and are thus more likely to be favoured in evolutionary adaptation. It is tempting to speculate that because of the combinatorial nature of *cis*-acting regulatory elements there will be many more examples of soft selective sweeps, and it will be important to develop techniques for their detection, since they may otherwise be missed in genome wide association studies. Some of these regulatory variants may have reached high frequencies as a result of selection in our past, but may now be responsible for present day disease susceptibilities.

References

Andrew T, Maniatis N, Carbonaro F, Liew SH, Lau W, Spector TD, Hammond CJ (2008) Identification and replication of three novel myopia common susceptibility gene loci on chromosome 3q26 using linkage and linkage disequilibrium mapping. PLoS Genet 4:e1000220

Asthana S, Noble WS, Kryukov G, Grant CE, Sunyaev S, Stamatoyannopoulos JA (2007) Widely distributed noncoding purifying selection in the human genome. Proc Natl Acad Sci USA 104:12410–12415

Babbitt CC, Silverman JS, Haygood R, Reininga JM, Rockman MV, Wray GA (2010) Multiple functional variants in cis modulate PDYN expression. Mol Biol Evol 27:465–479

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN (2004) Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111–1120

Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M (2004) Global identification of human transcribed sequences with genome tiling arrays. Science 306:2242–2246

Borlak J, Klutcka T (2004) Expression of basolateral and canalicular transporters in rat liver and cultures of primary hepatocytes. Xenobiotica 34:935–947

Bosma PJ, Chowdhury JR, Bakker C, Gantla S, de Boer A, Oostra BA, Lindhout D, Tytgat GN, Jansen PL, Oude Elferink RP et al (1995) The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. N Engl J Med 333:1171–1175

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, Liu ET (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 18:1752–1762

Buonocore F, Hill MJ, Campbell CD, Oladimeji PB, Jeffries AR, Troakes C, Hortobagyi T, Williams BP, Cooper JD, Bray NJ (2010) Effects of cis-regulatory variation differ across regions of the adult human brain. Hum Mol Genet 19:4490–4496

Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ, Todd JA, Donnelly P, Barrett JC, Davison D, Easton D, Evans DM, Leung HT, Marchini JL, Morris AP, Spencer CC, Tobin MD, Attwood AP, Boorman JP, Cant B, Everson U, Hussey JM, Jolley JD, Knight AS, Koch K, Meech E, Nutland S, Prowse CV, Stevens HE, Taylor NC, Walters GR, Walker NM, Watkins NA, Winzer T, Jones RW, McArdle WL, Ring SM, Strachan DP, Pembrey M, Breen G, St Clair D, Caesar S, Gordon-Smith K, Jones L, Fraser C, Green EK, Grozeva D, Hamshere ML, Holmans PA, Jones IR, Kirov G, Moskivina V, Nikolov I, O'Donovan MC, Owen MJ, Collier DA, Elkin A, Farmer A, Williamson R, McGuffin P, Young AH, Ferrier IN, Ball SG, Balmforth AJ, Barrett JH, Bishop TD, Iles MM, Maqbool A, Yuldasheva N, Hall AS, Braund PS, Dixon RJ, Mangino M, Stevens S, Thompson JR, Bredin F, Tremelling M, Parkes M, Drummond H, Lees CW, Nimmo ER, Satsangi J, Fisher SA, Forbes A, Lewis CM, Onnie CM, Prescott NJ, Sanderson J, Matthew CG, Barbour J, Mohiuddin MK, Todhunter CE, Mansfield JC, Ahmad T, Cummings FR, Jewell DP et al (2007) Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 39:1329–1337

Campbell CD, Kirby A, Nemesh J, Daly MJ, Hirschhorn JN (2008) A survey of allelic imbalance in F1 mice. Genome Res 18:555–563

Chan YF, Marks ME, Jones FC, Villarreal G Jr, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, Myers RM, Petrov D, Jonsson B, Schluter D, Bell MA, Kingsley DM (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327:302–305

Chaudhuri A, Polyakova J, Zbrzezna V, Pogo AO (1995) The coding sequence of Duffy blood group gene in humans and simians: restriction fragment length polymorphism, antibody and malarial parasite specificities, and expression in nonerythroid tissues in Duffy-negative individuals. Blood 85:615–621

Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, Sementchenko V, Piccolboni A, Bekiranov S, Bailey DK, Ganesh M, Ghosh S, Bell I, Gerhard DS, Gingeras TR (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 308:1149–1154

Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT (2005) Mapping determinants of human gene expression by regional and genome-wide association. Nature 437:1365–1369

Cheung VG, Nayak RR, Wang IX, Elwyn S, Cousins SM, Morley M, Spielman RS (2010) Polymorphic cis- and trans-regulation of human gene expression. PLoS Biol 8(9):e1000480

de Alarcon A, Steinke JW, Caughey R, Barekzi E, Hise K, Gross CW, Han JK, Borish L (2006) Expression of leukotriene C4 synthase and plasminogen activator inhibitor 1 gene promoter polymorphisms in sinusitis. Am J Rhinol 20:545–549

Diamond J (2002) Evolution, consequences and future of plant and animal domestication. Nature 418:700–707

Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, Gagnebin M, Nisbett J, Deloukas P, Dermitzakis ET, Antonarakis SE (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 325:1246–1250

Doniger SW, Fay JC (2007) Frequent gain and loss of functional transcription factor binding sites. PLoS Comput Biol 3:e99

Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, Hansen L (2008) Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Hum Genet 123:177–187

Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I (2002) Identification of a variant associated with adult-type hypolactasia. Nat Genet 30:233–237

Enattah NS, Jensen TG, Nielsen M, Lewinski R, Kuokkanen M, Rasinpera H, El-Shanti H, Seo JK, Alifrangis M, Khalil IF, Natah A, Ali A, Natah S, Comas D, Mehdi SQ, Groop L, Vestergaard EM, Imtiaz F, Rashed MS, Meyer B, Troelsen J, Peltonen L (2008) Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. Am J Hum Genet 82:57–72

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473:43–49

Farooq M, Troelsen JT, Boyd M, Eiberg H, Hansen L, Hussain MS, Rehman S, Azhar A, Ali A, Bakhtiar SM, Tommerup N, Baig SM, Kjaer KW (2010) Preaxial polydactyly/triphalangeal thumb is associated with changed transcription factor-binding affinity in a family with a novel point mutation in the long-range cis-regulatory element ZRS. Eur J Hum Genet 18:733–736

Fay JC, Wittkopp PJ (2008) Evaluating the role of natural selection in the evolution of gene regulation. Heredity 100:191–199

Ge B, Pokholok DK, Kwan T, Grundberg E, Morcos L, Verlaan DJ, Le J, Koka V, Lam KC, Gagne V, Dias J, Hoberman R, Montpetit A, Joly MM, Harvey EJ, Sinnett D, Beaulieu P, Hamon R, Graziani A, Dewar K, Harmsen E, Majewski J, Goring HH, Naumova AK, Blanchette M, Gunderson KL, Pastinen T (2009) Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. Nat Genet 41:1216–1222

Gimelbrant A, Hutchinson JN, Thompson BR, Chess A (2007) Widespread monoallelic expression on human autosomes. Science 318:1136–1140

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB (2005) Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature 433:481–487

Goode DL, Cooper GM, Schmutz J, Dickson M, Gonzales E, Tsai M, Karra K, Davydov E, Batzoglou S, Myers RM, Sidow A (2010) Evolutionary constraint facilitates interpretation of genetic variation in resequenced human genomes. Genome Res 20:301–310

Gurnett CA, Bowcock AM, Dietz FR, Morcuende JA, Murray JC, Dobbs MB (2007) Two novel point mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial polydactyly. Am J Med Genet A 143:27–32

He BZ, Holloway AK, Maerkl SJ, Kreitman M (2011) Does positive selection drive transcription factor binding site turnover? A test with Drosophila cis-regulatory modules. PLoS Genet 7:e1002053

Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, Thorleifsson G, Zillikens MC, Speliotes EK, Magi R, Workalemahu T, White CC, Bouatia-Naji N, Harris TB, Berndt SI, Ingelsson E, Willer CJ, Weedon MN, Luan J, Vedantam S, Esko T, Kilpelainen TO, Kutalik Z, Li S, Monda KL, Dixon AL, Holmes CC, Kaplan LM, Liang L, Min JL, Moffatt MF, Molony C, Nicholson G, Schadt EE, Zondervan KT, Feitosa MF, Ferreira T, Allen HL, Weyant RJ, Wheeler E, Wood AR, Estrada K, Goddard ME, Lettre G, Mangino M, Nyholt DR, Purcell S, Smith AV, Visscher PM, Yang J, McCarroll SA, Nemesh J, Voight BF, Absher D, Amin N, Aspelund T, Coin L, Glazer NL, Hayward C, Heard-Costa NL, Hottenga JJ, Johansson A, Johnson T, Kaakinen M, Kapur K, Ketkar S, Knowles JW, Kraft P, Kraja AT, Lamina C, Leitzmann MF, McKnight B, Morris AP, Ong KK, Perry JR, Peters MJ, Polasek O, Prokopenko I, Rayner NW, Ripatti S, Rivadeneira F, Robertson NR, Sanna S, Sovio U, Surakka I, Teumer A, van Wingerden S, Vitart V, Zhao JH, Cavalcanti-Proenca C, Chines PS, Fisher E, Kulzer JR, Lecoeur C, Narisu N, Sandholt C, Scott LJ, Silander K, Stark K, Tammesoo ML et al (2010) Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet 42:949–960

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335–2352

Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, Sella G, Przeworski M (2011) Classic selective sweeps were rare in recent human evolution. Science 331:920–924

Hines RN, Koukouritaki SB, Poch MT, Stephens MC (2008) Regulatory polymorphisms and their contribution to interindividual differences in the expression of enzymes influencing drug and toxicant disposition. Drug Metab Rev 40:263–301

Horsfall LJ, Rait G, Walters K, Swallow DM, Pereira SP, Nazareth I, Petersen I (2011a) Serum bilirubin and risk of respiratory disease and death. JAMA 305:691–697

Horsfall LJ, Zeitlyn D, Tarekegn A, Bekele E, Thomas MG, Bradman N, Swallow DM (2011b) Prevalence of clinically relevant UGT1A alleles and haplotypes in African populations. Ann Hum Genet 75:236–246

Hultman K, Tjarnlund-Wolf A, Odeberg J, Eriksson P, Jern C (2010) Allele-specific transcription of the PAI-1 gene in human astrocytes. Thromb Haemost 104:998–1008

Imtiaz F, Savilahti E, Sarnesto A, Trabzuni D, Al-Kahtani K, Kagevi I, Rashed MS, Meyer BF, Jarvela I (2007) The T/G 13915 variant upstream of the lactase gene (LCT) is the founder allele of lactase persistence in an urban Saudi population. J Med Genet 44:e89

Ingram CJ, Elamin MF, Mulcare CA, Weale ME, Tarekegn A, Raga TO, Bekele E, Elamin FM, Thomas MG, Bradman N, Swallow DM (2007) A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? Hum Genet 120:779–788

Ingram CJ, Mulcare CA, Itan Y, Thomas MG, Swallow DM (2009a) Lactose digestion and the evolutionary genetics of lactase persistence. Hum Genet 124:579–591

Ingram CJ, Raga TO, Tarekegn A, Browning SL, Elamin MF, Bekele E, Thomas MG, Weale ME, Bradman N, Swallow DM (2009b) Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. J Mol Evol 69:579–588

Itan Y, Jones BL, Ingram CJ, Swallow DM, Thomas MG (2010) A worldwide correlation of lactase persistence phenotype and genotypes. BMC Evol Biol 10:36

Jensen TG, Liebert A, Lewinsky R, Swallow DM, Olsen J, Troelsen JT (2011) The -14010*C variant associated with lactase persistence is located between an Oct-1 and HNF1α binding site and increases lactase promoter activity. Hum Genet 14(24):3945–3953

Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB (2008) The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. Cell 132:783–793

Johnson AD, Wang D, Sadee W (2005) Polymorphisms affecting gene regulation and mRNA processing: broad implications for pharmacogenetics. Pharmacol Ther 106:19–38

Kamio K, Matsushita I, Hijikata M, Kobashi Y, Tanaka G, Nakata K, Ishida T, Tokunaga K, Taguchi Y, Homma S, Azuma A, Kudoh S, Keicho N (2005) Promoter analysis and aberrant expression of the MUC5B gene in diffuse panbronchiolitis. Am J Respir Crit Care Med 171:949–957

King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188:107–116

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

Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, Jonasdottir A, Sigurdsson A, Kristinsson KT, Frigge ML, Gylfason A, Olason PI, Gudjonsson SA, Sverrisson S, Stacey SN, Sigurgeirsson B, Benediktsdottir KR, Sigurdsson H, Jonsson T, Benediktsson R, Olafsson JH, Johannsson OT, Hreidarsson AB, Sigurdsson G, Ferguson-Smith AC, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K (2009) Parental origin of sequence variants associated with complex diseases. Nature 462:868–874

Koukouritaki SB, Poch MT, Cabacungan ET, McCarver DG, Hines RN (2005) Discovery of novel flavin-containing monooxygenase 3 (FMO3) single nucleotide polymorphisms and functional analysis of upstream haplotype variants. Mol Pharmacol 68:383–392

Kudaravalli S, Veyrieras JB, Stranger BE, Dermitzakis ET, Pritchard JK (2009) Gene expression levels are a target of recent natural selection in the human genome. Mol Biol Evol 26:649–658

Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet 42:631–634

Lapidot M, Mizrahi-Man O, Pilpel Y (2008) Functional characterization of variations on regulatory motifs. PLoS Genet 4:e1000018

Lee AP, Kerk SY, Tan YY, Brenner S, Venkatesh B (2011) Ancient vertebrate conserved noncoding elements have been evolving rapidly in teleost fishes. Mol Biol Evol 28:1205–1215

Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12:1725–1735

Lettice LA, Hill AE, Devenney PS, Hill RE (2008) Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. Hum Mol Genet 17:978–985

Lewinsky RH, Jensen TG, Moller J, Stensballe A, Olsen J, Troelsen JT (2005) T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. Hum Mol Genet 14:3945–3953

Loh AX, Johnson L, Ng W, Swallow DM (2010) Cis-acting allelic variation in MUC5B mRNA expression is associated with different promoter haplotypes. Ann Hum Genet 74:498–505

Lomelin D, Jorgenson E, Risch N (2010) Human genetic variation recognizes functional elements in noncoding sequence. Genome Res 20:311–319

Maccari G, Gemignani F, Landi S (2010) COMPASSS (COMplex PAttern of Sequence Search Software), a simple and effective tool for mining complex motifs in whole genomes. Bioinformatics 26:1777–1778

Mulcare CA, Weale ME, Jones AL, Connell B, Zeitlyn D, Tarekegn A, Swallow DM, Bradman N, Thomas MG (2004) The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (LCT) (C-13.9 kbT) does not predict or cause the lactase-persistence phenotype in Africans. Am J Hum Genet 74:1102–1110

Natividad A, Holland MJ, Rockett KA, Forton J, Faal N, Joof HM, Mabey DC, Bailey RL, Kwiatkowski DP (2008) Susceptibility to sequelae of human ocular chlamydial infection associated with allelic variation in IL10 cis-regulation. Hum Mol Genet 17:323–329

Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET (2010) Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet 6:e1000895

Pastinen T, Hudson TJ (2004) Cis-acting regulatory variation in the human genome. Science 306:647–650

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

Pennacchio LA, Visel A (2010) Limits of sequence and functional conservation. Nat Genet 42:557–558

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, Rubin EM (2006) In vivo enhancer analysis of human conserved non-coding sequences. Nature 444:499–502

Pennings PS, Hermisson J (2006a) Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol 23:1076–1084

Pennings PS, Hermisson J (2006b) Soft sweeps III: the signature of positive selection from recurrent mutation. PLoS Genet 2:e186

Pochareddy S, Edenberg HJ (2011) Variation in the ADH1B proximal promoter affects expression. Chem Biol Interact 191:38–41

Pollard KS, Serre D, Wang X, Tao H, Grundberg E, Hudson TJ, Clark AG, Frazer K (2008) A genome-wide approach to identifying novel-imprinted genes. Hum Genet 122:625–634

Poulter M, Hollox E, Harvey CB, Mulcare C, Peuhkuri K, Kajander K, Sarner M, Korpela R, Swallow DM (2003) The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. Ann Hum Genet 67:298–311

Przeworski M, Coop G, Wall JD (2005) The signature of positive selection on standing genetic variation. Evolution 59:2312–2323

Ralph P, Coop G (2010) Parallel adaptation: one or many waves of advance of an advantageous allele? Genetics 186:647–668

Rockman MV, Wray GA (2002) Abundant raw material for cis-regulatory evolution in humans. Mol Biol Evol 19:1991–2004

Rockman MV, Hahn MW, Soranzo N, Zimprich F, Goldstein DB, Wray GA (2005) Ancient and recent positive selection transformed opioid cis-regulation in humans. PLoS Biol 3:e387

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF,

Lander ES, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC et al (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449:913–918

Sankaran VG, Xu J, Orkin SH (2010) Advances in the understanding of haemoglobin switching. Br J Haematol 149:181–194

Savinkova LK, Ponomarenko MP, Ponomarenko PM, Drachkova IA, Lysova MV, Arshinova TV, Kolchanov NA (2009) TATA box polymorphisms in human gene promoters and associated hereditary pathologies. Biochemistry (Mosc) 74:117–129

Skelly DA, Ronald J, Akey JM (2009) Inherited variation in gene expression. Annu Rev Genomics Hum Genet 10:313–332

Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23:23–35

Stormo GD (2000) DNA binding sites: representation and discovery. Bioinformatics 16:16–23

Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavare S, Deloukas P, Dermitzakis ET (2005) Genome-wide associations of gene expression variation in humans. PLoS Genet 1:e78

Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavare S, Deloukas P, Dermitzakis ET (2007) Population genomics of human gene expression. Nat Genet 39:1217–1224

Sun C, Southard C, Witonsky DB, Olopade OI, Di Rienzo A (2010) Allelic imbalance (AI) identifies novel tissue-specific cis-regulatory variation for human UGT2B15. Hum Mutat 31:99–107

Terry CF, Loukaci V, Green FR (2000) Cooperative influence of genetic polymorphisms on interleukin 6 transcriptional regulation. J Biol Chem 275:18138–18144

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P (2007) Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 39:31–40

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

Troelsen JT, Olsen J, Moller J, Sjostrom H (2003) An upstream polymorphism associated with lactase persistence has increased enhancer activity. Gastroenterology 125:1686–1694

Van der Pouw Kraan TC, van Veen A, Boeije LC, van Tuyl SA, de Groot ER, Stapel SO, Bakker A, Verweij CL, Aarden LA, van der Zee JS (1999) An IL-13 promoter polymorphism associated with increased risk of allergic asthma. Genes Immun 1:61–65

Verlaan DJ, Ge B, Grundberg E, Hoberman R, Lam KC, Koka V, Dias J, Gurd S, Martin NW, Mallmin H, Nilsson O, Harmsen E, Dewar K, Kwan T, Pastinen T (2009) Targeted screening of cis-regulatory variation in human haplotypes. Genome Res 19:118–127

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4:e72

Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson Bostrom K, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B et al (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42:579–589

Wang Y, Harvey CB, Pratt WS, Sams VR, Sarner M, Rossi M, Auricchio S, Swallow DM (1995) The lactase persistence/non-persistence polymorphism is controlled by a cis-acting element. Hum Mol Genet 4:657–662

Wang Y, Harvey CB, Hollox EJ, Phillips AD, Poulter M, Clay P, Walker-Smith JA, Swallow DM (1998) The genetically programmed down-regulation of lactase in children. Gastroenterology 114:1230–1236

Wang HY, Fu Y, McPeek MS, Lu X, Nuzhdin S, Xu A, Lu J, Wu ML, Wu CI (2008) Complex genetic interactions underlying expression differences between Drosophila races: analysis of chromosome substitutions. Proc Natl Acad Sci USA 105:6362–6367

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Wolfe ND, Dunavan CP, Diamond J (2007) Origins of major human infectious diseases. Nature 447:279–283

Wray GA (2007) The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8:206–216

Yokoyama KD, Thorne JL, Wray GA (2011) Coordinated genome-wide modifications within proximal promoter cis-regulatory elements during vertebrate evolution. Genome Biol Evol 3:66–74

Yu X, Lin J, Zack DJ, Qian J (2007) Identification of tissue-specific cis-regulatory modules based on interactions between transcription factors. BMC Bioinformatics 8:437

Zhang C, Bailey DK, Awad T, Liu G, Xing G, Cao M, Valmeekam V, Retief J, Matsuzaki H, Taub M, Seielstad M, Kennedy GC (2006) A whole genome long-range haplotype (WGLRH) test for detecting imprints of positive selection in human populations. Bioinformatics 22:2122–2128

Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE (2009) Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature 462:65–70