http://www.expertreviews.org/

expert reviews in molecular medicine

Regulatory SNPs in complex diseases: their identification and functional validation

Ludmila Prokunina and Marta E. Alarcón-Riquelme

Finding the genetic causes for complex diseases is a challenge. Expression studies have shown that the level of expression of many genes is altered in disease compared with normal conditions, but what lies behind these changes? Linkage studies provide hints as to where in the genome the genetic triggers – the mutations – might be located. Fine-mapping and association studies can give yet more information about which genes, and which changes in the genes, are involved in the disease. Recent examples show that single-nucleotide polymorphisms (SNPs), which are variations at the single-nucleotide level within an individual's DNA, in the regulatory regions of some genes constitute susceptibility factors in many complex diseases. This article discusses the nature of regulatory SNPs (rSNPs) and techniques for their functional validation, and looks towards what rSNPs can tell us about complex diseases.

Following the sequencing of the human genome (Ref. 1), it became clear that although the complete set of genes and DNA sequence is essentially similar between individuals, variations at single bases occur. On average, one nucleotide of every 300–570 nucleotides in human genomic DNA varies between individuals (Refs 2, 3). The term single-nucleotide polymorphism (SNP) relates to this kind of variation at a single-nucleotide level. Many of these SNPs will result in neutral changes, not affecting the functions of proteins, and will never be noticed. However, some of these variants will occur in coding parts or flanking regulatory sequences of genes and affect protein function, which could have serious consequences and possibly give rise to disease. If the changes are

Ludmila Prokunina
Postdoctoral Fellow, Department of Genetics and Pathology, Section for Medical Genetics, Rudbeck Laboratory, Uppsala University, Dag Hammarskjölds väg 20, 751 85, Uppsala, Sweden. Tel: +46 18 4714487; Fax: +46 18 471 4808; E-mail: ludmila.prokunina@genpat.uu.se

Marta E. Alarcón-Riquelme (corresponding author)
Associate Professor in Medical Genetics, Department of Genetics and Pathology, Section for Medical Genetics, Rudbeck Laboratory, Uppsala University, Dag Hammarskjölds väg 20, 751 85, Uppsala, Sweden. Tel: +46 18 4714805; Fax: +46 18 4714808; E-mail: marta.alarcon@genpat.uu.se

Institute URL: http://www.genpat.uu.se/Forskargrupper/syaudi/syaudi.html

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

not lethal or sufficiently dramatic to lead to negative selection, they might be preserved and carried to further generations.

Some SNPs change the recognition sites for restriction enzymes, and this has been utilised for restriction fragment length polymorphism (RFLP) assays. RFLPs, together with VNTRs (variable number of tandem repeats, represented by mini- and microsatellites) (Ref. 4), have been used as genetic markers for linkage studies (Refs 5, 6). Even though VNTRs are more polymorphic, they are about 100 times less frequent in the human genome than SNPs (Ref. 1) and therefore they are not widely used in association studies. The SNP Consortium (TSC) (Ref. 7) and the International Human Genome Sequencing Consortium (Refs 2, 8) aim to identify as many SNP variations as possible through the re-sequencing of a large number of individuals from different populations. Nearly six million SNPs have already been identified (Ref. 2).

SNPs and complex diseases

Monogenic diseases are usually associated with one or several alternative mutations within a single gene that frequently lead to a clear alteration of the protein function; for example, in the case of cystic fibrosis, 500 different mutations in the CFTR gene are known to cause the disease (Ref. 9). By contrast, any given complex disease might result from a combination of inherited and acquired mutations in more than one gene, as well as exposure to environmental factors, which together might modulate the susceptibility, severity and clinical manifestations of the disease. In other words, no particular mutation is absolutely necessary or sufficient as a cause in the context of a complex disease; rather, individual mutations might lead to a certain disease only on a specific genetic background (with other mutations present) or upon exposure to certain environmental factors such as infections, drugs, contaminants, UV radiation, and so on. Therefore, although complex genetic disorders demonstrate a significant underlying genetic component, they generally segregate in a nonmendelian fashion. Thus, identifying the factors that contribute to the development of complex diseases is a considerable challenge.

Many disease-associated mutations are expected to be SNPs. Some of these SNPs will be located in the coding regions and cause considerable alterations in the function of the protein as a result of substitution of one amino acid by another or by the creation of stop codons. Severe changes in the coding parts of the genes might result in early and severe forms of complex diseases. This is the case for mutations in the *BRCA1* gene for breast cancer (Ref. 10) or in the presenilin gene for Alzheimer's disease (Ref. 11).

Other SNPs associated with complex diseases might be located in the noncoding, regulatory regions of genes and might modify gene function more subtly through effects on the level of gene expression and its spatial and temporal control. These SNPs are referred to as regulatory (rSNPs). In a cis position, rSNPs might alter the regulation of transcription, mRNA stability or mRNA splicing of genes located on the same chromosome (allele-specific effects), for example by altering binding sites for regulatory proteins. rSNPs in genes encoding proteins that regulate gene expression, such as transcription factors, can also lead to effects on gene expression in trans (i.e. effects on expression of the target genes of the regulatory factor).

Cis-acting rSNPs

Transcription-factor-binding sites

rSNPs acting in cis can alter (modify, destroy or create) recognition and binding sites for transcription-regulatory proteins, resulting in a loss- or gain-of-function of the effect of the regulatory protein. As a result of this alteration, binding can be increased, decreased, abrogated or induced, leading to changes in the level, timing and localisation of gene expression. For instance, an rSNP in the promoter region of the gene encoding tumour necrosis factor α (TNF-α) creates a new binding site for the OCT-1 transcription factor and upregulates the expression of TNF-α in monocytes, which thereby increases the risk for cerebral malaria in Africans (Ref. 12). Intriguingly, the same SNP affects the expression of an upstream gene, lymphotoxin α, and this alteration is associated with myocardial infarction (Ref. 13). SNPs in intronic or intergenic regulatory sequences to which the RUNX1 transcription factor binds are associated with three autoimmune diseases – systemic lupus erythematosus (Refs 14, 15), rheumatoid arthritis (Ref. 16) and psoriasis (Ref. 17) – suggesting a common mechanism for these diseases (Ref. 18).

Interestingly, rSNPs can provide not only genetic but also epigenetic regulation, by changing the methylation status of the binding

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

sites of regulatory proteins. Cytosines in CpG pairs are usually methylated in the mammalian genome and some regulatory proteins bind only to methylated DNA (Ref. 19). An allelic change (C to A) in an intronic SNP in the pig insulin-like growth factor 2 (IGF2) gene was recently shown to change the methylation status of the binding site and abrogate binding of an as-yet-unknown regulatory protein (possibly a transcriptional repressor) to this target (Ref. 20).

mRNA stability

rSNPs acting in cis to affect the stability of mRNA are often located in the 5' and 3' untranslated regions of the genes (UTRs) and can alter the binding of proteins that regulate the stability of newly synthesised RNA. The regulatory sequences in UTRs form regions with a high content of secondary structure (hairpins), which might serve as regulatory elements as they are recognised and bound by regulatory proteins. These proteins protect mRNA from degradation by RNases (Refs 21, 22). Specific adenylate- and uridylate-rich (AU-rich) elements (ARE) are often found in unstable mRNAs (Ref. 23). rSNPs occurring in these elements can be associated with increased stability or decay of mRNA. Silent exonic mutations can be associated with altered mRNA folding, stability and efficiency of mRNA translation (Ref. 24). Sometimes a haplotype that includes several SNPs can be associated with altered mRNA stability and it can be difficult to define which SNP is responsible for this effect (Ref. 25).

mRNA splicing

Splicing of mRNA can be regulated by 5' and 3' splice sites, branch sites, and intronic and exonic enhancers and silencers. Therefore, rSNPs occurring within these sequences can create nonfunctional forms of proteins by exclusion and inclusion of exons and/or activation of alternative splice variants (Ref. 26). Approximately 15% of mutations associated with human diseases affect mRNA splicing (Ref. 27). Mutations in exonic splicing enhancers (ESEs) are associated with human diseases such as spinal muscular atrophy (Ref. 28), Becker's muscular dystrophy (Ref. 29), breast cancer (Ref. 30) and cystic fibrosis (Refs 31, 32). Inactivation of an exonic silencer by an rSNP within the *CD45* gene and inclusion of an alternative exon was associated with susceptibility to multiple sclerosis (Ref. 33). One of the mutations in the *BRCA2* gene causing early-onset breast cancer has been associated with disruption of an intronic splicing site, causing exon skipping and creating a deleterious allele of the gene carrying this mutation (Ref. 34). An rSNP associated with susceptibility to several autoimmune diseases found in the 3' UTR of the *CTLA4* gene was a cause of alternative splicing and expression of a nonfunctional variant of the protein that lacks its ligand-binding domain (Ref. 35). More information about regulation of splicing and human diseases can be found in recent reviews (Refs 36, 37, 38, 39, 40).

Trans-acting sRNPs

rSNPs affecting genes encoding factors that control gene expression can act in trans and consequently alter regulation of the target genes of the factors, possibly leading to disease. An rSNP might lead to functional inactivation of one of the alleles of a transcription factor (e.g. by aberrant splicing or reduced stability of mRNA) and affect function through haploinsufficiency (Refs 41, 42). Recently, an intronic SNP in the *RUNX1* gene, which encodes a transcription factor that binds to the *SLC22A4* gene, has been associated with rheumatoid arthritis. As mentioned above, an rSNP that abrogates RUNX1 binding through mutation in the *SLC22A4* gene itself is also associated with this disease. Thus, it is possible that a putative rSNP in the *RUNX1* gene trans-regulates expression of the *SLC22A4* gene, and that the rSNP in *SLC22A4* alters the cis-regulated expression of this gene by altering the binding of the trans-factor (Ref. 16).

Identification of rSNPs

One of the major problems in the identification of rSNPs is that they can be located anywhere in the noncoding genome. Exons represent at most some 5–10% of the total size of genes (Ref. 1). Therefore, it takes at least ten times more sequencing to identify variation in the noncoding segments.

In the search for any kind of mutation, re-sequencing of full-size genes in DNA from patients and controls from defined populations is still the most direct way of identification. However, this is an expensive and time-consuming process, and alternative methods of SNP detection have been developed. One such method is single-strand conformation polymorphism (SSCP), in which polymerase chain reaction (PCR) fragments amplified from the DNA of different individuals

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

are separated on denaturing slab polyacrylamide gels or by using capillary sequencers (Refs 43, 44). Variants of DNA present in the amplicons result in the different conformation and hence migration of single-stranded DNA through the gel. To identify the nature of the migration difference, the PCR fragments must be sequenced. A more-recent and widely used method is denaturing high-pressure liquid chromatography (DHPLC). PCR fragments are generated and analysed by DHPLC, and fragments differing from the wild type are sequenced. This method has the advantage of allowing high-throughput and semi-automation (Refs 45, 46).

To search specifically for rSNPs, where should we look and how? Although rSNPs can be located elsewhere in the genome, promoters, which by definition contain multiple binding sites for regulatory factors, would be the first place to start the search. Promoter fragments of 0.5–2.0 kb [from the Eukaryotic Promoter Database (Ref. 47), or any promoters of interest] can be used for DHPLC analysis and re-sequencing, and then analysed by reporter assays to evaluate allele- or haplotype-specific activity of the promoters. The first studies using this approach have been successful, and promoter rSNPs modifying the activity of the reporter constructs have been found in 17 out of 59 promoters studied (Ref. 48). However, the appropriate selection of cell lines in which the activity is to be evaluated, together with the particular conditions (e.g. activation state or developmental stage of cells), are crucial for the adequate detection of promoter activity as well as for deriving conclusions as to whether the rSNPs are the underlying cause of the difference (discussed later).

Comparative studies of conserved noncoding regions using the extensive sequence information from different species can provide interesting data about putative functional elements and pathways of gene regulation (Ref. 49). Computer programs such as VISTA (Ref. 50) or NIX (Ref. 51), which allow large-scale sequence comparison, help to highlight regions of high evolutionary conservation for more-detailed searches. However, 32–40% of functional sites present in humans or rodents are not functional in other species (Ref. 52). Additionally, recent comparisons between the human and mouse genomes demonstrated that, out of 7293 amino acids in 687 human genes associated with human diseases, 2.2% of human mutations were normal (nondisease) variants in mice and 7.5% had no homologues in the mouse genome (Ref. 53). Nonconserved functional sites might reflect a dynamic system of tissue-, stage- or organism-dependent gene regulation, and could be especially important in the case of human diseases. Therefore, by looking only for conserved regulatory regions, there is a risk of overlooking human-specific mutations.

Genotyping of SNPs

Once SNPs are detected, a large selection of genotyping methods is available. RFLP is the oldest and simplest method for genotyping (Ref. 5), but it can be used only when the nucleotide change leads to a change in the recognition site of a restriction enzyme. Sequencing is also a choice in many cases, especially where automated capillary high-throughput sequencing is available. The high demand for SNP genotyping has resulted in the development of many other new and efficient methods with high-throughput capabilities (Table 1). The choice of the method will depend on the genotyping scale and the resources, as many methods require the availability of expensive equipment and reagents.

Statistical considerations

SNPs, as bi-allelic markers, can have low polymorphism information content (PIC value). To increase the PIC value of the locus, haplotypes (Refs 54, 55), which are a series of alleles of markers, residing on the same chromosome and associated with each other [in linkage disequilibrium (Ref. 56)], can be derived. Several programs can help to reconstruct haplotypes from alleles of several markers using genotypes from families or from unrelated individuals; examples are MERLIN (Ref. 57), PHASE (Ref. 58) and GENEHUNTER (Ref. 59), but more than 240 other useful programs for linkage, linkage disequilibrium and association studies can be found at the Rockefeller University website (Ref. 60). Thus, several bi-allelic SNPs can be substituted by one multi-allelic haplotype marker, thereby increasing the information content of the locus.

To evaluate the association of alleles of an SNP or haplotype of a locus or gene (also treated as alleles) with a phenotype (a disease), several types of analyses can be performed. Association studies in patients and controls matched by ethnicity, age and sex are the most common type. The

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

Table 1. Methods of SNP (single-nucleotide polymorphism) genotyping

Method	Refs
Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF)	101, 102, 103
Dynamic allele-specific hybridisation (DASH)	104, 105
Pyrosequencing	106
Minisequencing	107, 108
Invader assay	109, 110
Rolling circle amplification	111, 112
Melting curve analysis (McSNP)	113
Multiplex automated primer extension analysis (MAPA)	107
MIP genotyping with molecular inversion probes	114
Survivor assay; an SNP detection method based on electrospray ionisation mass spectrometry (ESI-MS)	115
Qbead system with fluorescent Qdot semiconductor nanocrystals	116
Template-directed dye-terminator incorporation with fluorescence quenching detection (FQ-TDI)	117
SNP genotyping using single-tube fluorescent bidirectional polymerase chain reaction (PCR)	118
Multiplexed single-base extension (SBE) genotyping via end-labelled free-solution electrophoresis (ELFSE)	119
Molecular beacons and real-time PCR	120
Site-selective RNA scission	121
'ZipCodes'-oligonucleotide ligation assay (OLA) and flow cytometric analysis of fluorescent microspheres	122
Fluorescence polarisation	123
Amplifluor (allele-specific amplification and universal energy-transfer-labelled primers)	124

significance is presented as a P value of a X^2 distribution. The risk ratio (Ref. 61) indicates folds of increase in the probability of having the disease by individuals with a certain allele (genotype) compared with those with a different allele. In family-based studies, the most commonly used methods are the transmission disequilibrium test (TDT) (Ref. 62), which evaluates the transmission of alleles from heterozygous parents to their sick child or children. In the case of larger pedigrees, the pedigree disequilibrium test (PDT) is used (Ref. 63) because it can handle more than one affected individual in a family. The haplotype relative risk (HRR) test (Ref. 64) and the affected family-based controls (AFBAC) test (Ref. 65) can use information from parents both homozygous and heterozygous for the marker or SNP. The result will be a comparison of allele frequencies in chromosomes transmitted to patients from their parents versus nontransmitted (control) chromosomes. The family-based studies use a population of affected and control chromosomes obtained from the same individuals that represent the best 'matching-by-ethnicity' approximation, helping to avoid 'population stratification' (a situation when the patients and control sets are

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

not matched by ethnicity, but frequencies of alleles are different in ethnic groups). Replication of association studies in other populations and in independent (and larger) sets of patients/controls can help to support real associations and eliminate false positives.

Functional validation of rSNPs

Having found a genetic association of a disease with a gene haplotype or a few SNPs, one can attempt to define if any of the identified SNPs might 'functionally' explain the association with the disease. Since functional studies are usually time-consuming and difficult to scale up, these tend to be initiated only when a statistically significant association with a phenotype is already established and has been replicated. Table 2 outlines the main methods of functional validation of rSNPs, and these are discussed further below.

Identification of splicing variants and effects on mRNA stability

Alleles of rSNPs can affect splicing by creating or abrogating splicing sites or modifying splicing enhancers and silencers, and this can be analysed first by computational tools. Exonic splicing enhancers (ESEs) are the most common splicing regulatory elements present in majority of exons (Ref. 39) and the ESE Finder tool helps to identify these elements in the genomic DNA (Ref. 66). RESCUE-ESE was used to screen for exonic enhancer elements within 4817 human genes. ESEs of ten different types were identified through this approach (Ref. 67). NIX software also helps to predict exons and introns in the genomic sequence by means of several prediction algorithms, and the effect of a particular allelic change can also be evaluated (Ref. 51).

Of course, the predicted presence of alternative transcripts should be confirmed by experimental analysis of the RNA (cDNA) from individuals carrying the allele. This can be accomplished by northern blots or reverse transcriptase (RT)-PCR analysis, followed by sequencing of alternative fragments and correlation of the presence of the alternative transcripts in healthy and sick individuals with alleles of the rSNP. The relative ratio of alternative transcripts can be measured by quantitative RT-PCR or microarrays. The effect of rSNPs on splicing can be studied with the use of minigenes, where an exon surrounded by adjacent intronic sequences is cloned into expression vectors. The effect of particular allelic changes is evaluated by site-directed mutagenesis and transient expression of minigenes in cell cultures (Refs 68, 69). This approach is also used to study activity of predicted splicing regulatory elements (Ref. 67).

As indicated above, rSNPs can result in altered stability of mRNA, and this can be evaluated experimentally as described elsewhere (Refs 25, 70).

Analysis of transcription binding sites

The Transfac (Ref. 71) and TFSearch (Ref. 72) databases can help to analyse the genomic sequences for the presence of binding sites of known transcription factors. Both alleles of the rSNP should be evaluated as either of the alleles can abrogate or create a binding site. CONSITE (Ref. 73) is another tool that helps to compare sequences from two species and find conserved transcription-factor-binding sites present in both. Several centres provide access to many useful web-based programs, available freely or upon registration (Refs 51, 74). Computer predictions result in hundreds of potential binding sites, but many of these will be false positives and only a small proportion will be active at a certain time or in a certain tissue. Relevant predictions have to be evaluated individually and according to previous knowledge of the biology of the gene before initiating the experimental validation. Methods for the validation of the rSNPs altering transcription-factor-binding sites are described in some detail below.

Electrophoretic mobility shift assay (EMSA)

Predicted DNA–protein interactions can be validated with electrophoretic mobility shift assays (EMSAs; also called 'gel shift' or 'band shift' assays) (Ref. 75). In this method, oligonucleotides 20–25 bp in length representing both alleles of an SNP are labelled radioactively or nonradioactively and incubated with nuclear cell extracts prepared from cells or tissues of interest. Protein(s) from the nuclear extract binding to the oligonucleotides will cause retardation of migration in a gel owing to formation of DNA–protein complexes. Allele-specific binding will suggest the potential functional effect of the SNP studied and, if the addition of excess of unlabelled oligonucleotide containing one but not another allele decreases or eliminates the binding, and an unrelated

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

Table 2. Functional validation of regulatory SNPs (single-nucleotide polymorphisms)

Method and purpose	Input	Output
Splicing studies To evaluate whether an allele of an rSNP is creating or abrogating splicing sites	RNA/cDNA from individuals with different alleles of the rSNP; northern blotting, RT-PCR, sequencing	Allele-specific alternative transcripts of different size and sequence
EMSA To test whether alleles of an rSNP show a difference in binding to nuclear proteins	Allele-specific oligonucleotides, nuclear extracts or recombinant proteins, antibodies	Allele-specific DNA–protein interactions
DNaseI footprinting To test if there are multiple binding sites for nuclear proteins and define their positions within a DNA fragment	Allele-specific labelled DNA fragments, nuclear extracts, DNase I	Number and position of binding sites within a DNA fragment
Reporter assays To study allele-specific regulatory effects on expression of a reporter gene	Allele-specific constructs, cell lines, activation agents	Confirmation of a regulatory (enhancer or silencer) activity of allele-specific constructs
ChIP assay (chromatin immunoprecipitation) To confirm (study) in vivo DNA–protein interactions; to determine DNA targets for a particular protein	Cells, antibodies against a particular protein	Confirmation that a DNA–protein interaction occurs in vivo in particular cells and conditions; new genomic targets for a protein
Methylation studies To evaluate the effect of an SNP on the methylation status of the sequence	Bisulphite sequencing, EMSA with methylated oligonucleotides, reporter assays with allele-specific constructs methylated in vitro	Altered binding of regulatory proteins to a DNA sequence owing to its allele-specific methylation leading to aberrant expression of the gene
Expression microarrays To evaluate the trans effect caused by allele-specific variants of a regulatory protein on its targets	Allelic forms of a regulatory protein, cDNA or oligonucleotide microarrays, samples of cDNA to compare	A list of the genes differentially regulated by allelic forms of a regulatory protein
Allelic imbalance To evaluate whether both alleles of an SNP are equally expressed	cDNAs heterozygous for alleles of the SNP; any method for evaluation of allele-specific expression	Allelic imbalance when one allele of the SNP is expressed more than another (by at least 10%)

oligonucleotide does not cause this effect, it proves the specificity of the interaction. The addition of antibodies against the nuclear factor causes additional retardation of the migration of the DNA–protein complexes, and proves the involvement of the nuclear protein in the interaction. Some factors can be expressed at very low amounts in any cell type, making it difficult to detect DNA–protein interaction when the protein in question represents only a tiny proportion of all proteins present in the nuclear cell extract. In this case, a recombinant form of the transcription factor or its DNA-binding domain can be used in an EMSA (Ref. 76).

Of course, not all transcription factors are known and not all binding sites are present in the

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

databases, and EMSAs can reveal interactions with unknown proteins (Ref. 77). Methods to identify proteins present in the nuclear cell extracts interacting with the target sequences are outside the scope of this review.

DNaseI footprinting

Footprinting of DNaseI-hypersensitive sites using nuclear extracts can confirm binding of transcription factors and show the position and number of sites occupied by the protein(s) (Ref. 78). In this method, a radioactively labelled DNA fragment (usually PCR-amplified and up to 400–500 bp in length) is incubated with a nuclear extract and treated with the DNaseI enzyme. Complexes are resolved on polyacrylamide gels and analysed: sites of DNA covered by the interacting proteins are protected from DNaseI cleavage and leave 'white spots' (footprints) surrounded by a ladder of bands corresponding to digested (unprotected) DNA. Multiple binding sites of the same or different nuclear factors will result in several footprints. The exact position and identity of the nucleotides constituting the target involved in the binding can also be defined using this method.

If several proteins bind DNA at the same time to form a complex, the EMSA will not be able detect such an interaction, because not all proteins required for this complex will be able to bind to the short oligonucleotides used in EMSA. In cases like this, footprinting can give more information.

Reporter assays

Potential regulatory activities (promoter, enhancer or silencer) of a putative regulatory sequence, and the effect of a particular change caused by alleles of an rSNP, can be studied using reporter gene assays (Ref. 79). Fragments of DNA carrying alleles of the rSNP are cloned into a promoter-less vector or a vector carrying a weak promoter (Ref. 80). These vectors carry a 'reporter' gene, whose expression will depend on the inserted regulatory sequence. Such reporter genes are easy to monitor. Those commonly used are the firefly-derived luciferase gene or the CAT (chloramphenicol acetyl transferase) gene. Different allelic constructs, together with a control plasmid used for normalisation of the efficiency of transfection, are transiently transfected into appropriate cell lines. The level of reporter gene expression is evaluated for both alleles of the rSNP and compared with the expression of the 'empty' vector. An increase or decrease in reporter gene expression caused by one of the alleles will demonstrate that the allele is affecting the activity of the regulatory element in which it is found (promoter, enhancer or silencer). Some transcription factors might require partners or cofactors to perform fully their activities in some cell types; for example, binding of the RUNX1 protein to DNA is increased in the presence of its partners CBFβ (40-fold) or Ets-1 (10-fold) (Ref. 81). Plasmids carrying an interacting partner can be co-transfected with the allele-specific constructs. In this type of experiment, the choice of cell line is very important. For example, computer analysis might predict that a certain SNP disrupts the binding site for the RUNX1 transcription factor, and literature searches and the phenotype of the disease suggest that this factor could be expressed in T cells. The Jurkat cell line (T leukaemia cell line) might therefore be expected to be a good choice for analysis and has indeed been successfully used in several experiments involving the RUNX1 protein (Refs 14, 16, 17). Many genes are expressed only under certain conditions, for example upon cellular activation, making it important to select the right activating agent and to perform a pre-analysis of the kinetics of activation for each cell type used.

Chromatin immunoprecipitation

Protein–DNA interactions detected in vitro (by EMSA or footprinting) might not occur in vivo. Chromatin immunoprecipitation (ChIP) is the method of choice to study whether a particular nuclear factor is indeed binding a particular DNA target under certain conditions in vivo. Antibodies against the nuclear factor in question are required for this method. Formaldehyde crosslinking is used to make 'snapshots' of DNA–protein and protein–protein contacts at a given moment in living cells (cell cultures) or tissues. DNA bound to its interacting proteins is sonicated to produce fragments of 300–700 bp and is incubated with appropriate antibodies specific for the protein. The fragments of DNA (allele-specific fragments in the case of SNPs) bound to their interacting proteins will be immunoprecipitated with the specific antibodies and can be detected by PCR. PCR allows evaluation of the strength (degree) of the interaction by measuring the quantity of precipitated DNA and indirectly of the interacting complexes. To simulate in vitro the

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

relevant disease conditions, cells can be treated with different agents.

A combination of microarray chips and ChIP technology has produced a new method called 'ChIP-on-chips' (Ref. 82). Genomic microarrays representing sequences of the whole genome, separate chromosomes or genomic fragments (for example, promoters or CpG islands) can be used for this type of analysis. The first examples of this hybrid technology using arrays containing spotted PCR fragments of promoters have been successful (Ref. 83). Another example is the use of CpG island microarrays, often representing 5' regions of active genes (Ref. 84). One possible application of this approach is for identification of novel gene targets (as potential candidate genes for complex diseases) for transcription factors.

Methylation studies

The addition of a methyl group to a cytosine (C) residue in CpG dinucleotides in the mammalian genome creates 5-methylcytosine (5-MeC). When deaminated, the two forms of cytosine produce different bases: cytosine is converted to uracil (which is found in RNA but not DNA and therefore is eliminated); and 5-MeC is converted to tyrosine (T). During evolution, the deamination process has contributed to a gradual decrease in the frequency of CpG dinucleotides in the genome (CpG pairs make up 41% of the human genome, and most are methylated) (Refs 1, 85). The only regions where CpGs are usually unmethylated, and therefore protected from mutations, are important regulatory regions called CpG islands (Ref. 85). If an SNP occurs in a CpG dinucleotide where cytosine is methylated (5-MeC), the allelic nucleotide change of either the C or G base will cause demethylation of this site and alteration of the binding of methylase-sensitive regulatory proteins (Ref. 20). Somatic hypermethylation of genes is associated with repression of gene transcription (Ref. 86), whereas hypomethylation is associated with activation of gene transcription and is found in several cancers, autoimmunity and ageing (Refs 87, 88).

Bisulphite treatment allows C and 5-MeC to be distinguished, by converting nonmethylated Cs to Ts but leaving the methylated C residues unchanged. DNA treated with bisulphite is then amplified by PCR, the fragments are cloned and 15–20 colonies for each fragment are sequenced to determine the proportion of methylation for each CpG site. The functional importance of the methylated allele can be established with EMSA using methylated oligonucleotides and/or by investigating the correlation between the level of methylation of a regulatory region containing the SNP with the level of expression of the corresponding gene (Ref. 20). The effect of methylation can also be evaluated in reporter assays where allele-specific constructs are methylated in vitro with SssI methylase and used for transfection. If the DNA–protein interaction depends on methylation, the level of expression of the constructs will be different between the methylated and nonmethylated forms (Ref. 89).

Expression microarrays

Large-scale expression studies became possible with the use of microarrays. The two main types of microarrays are oligonucleotide arrays [Affymerix (Ref. 90)] and custom cDNA arrays (Ref. 91). Affymetrix arrays contain a set of oligonucleotides representing several thousand genes from the human genome. cDNA arrays usually contain smaller and customised sets of genes (such as specific genes relevant for pathways or from certain chromosomal regions). The latter arrays are cheaper than oligonucleotide arrays, as they are spotted PCR fragments of sequence-verified cDNA clones. The effect of rSNPs in genes encoding regulatory proteins (transcription factors) on the expression of their target genes is often impossible to predict, but is possible to detect by analysing expression of thousands of genes at the same time using microarrays. Some of these genes will be directly regulated by the transcription factor through binding to its target sites, whereas other genes will be regulated indirectly through a chain of remote responses. Analysis software for microarrays can help to cluster genes according to their gene expression, and suggest their involvement in certain pathways. For more information, excellent reviews on microarrays (Refs 92, 93, 94) or useful websites (Refs 90, 91, 95, 96) can be consulted.

Allelic imbalance

The functional effect of rSNPs in the expression of genes located on the same DNA molecule (cis position) can be evaluated in assays measuring allelic imbalance. Whatever the nature of the alteration associated with variants of SNP – loss or gain of binding sites for regulatory factors, or altered efficiency of this binding, or epigenetic change – this alteration can be detected by

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

measuring the proportion of expression of both allelic variants of the gene. If there is no preferential expression of one allele over another, they will both be equally expressed, in a proportion of 50:50. A 10% deviation from equimolar expression of alleles is an arbitrary threshold for identification of allelic imbalance (where the proportion of the transcripts is at least 40:60). An extreme case of allelic imbalance is the monoallelic expression of imprinted genes, where one allele (maternal or paternal) is always repressed and the proportion of transcripts is therefore 100:0.

Because rSNPs are usually located in noncoding regions, their relationship or correlation with expression levels is difficult to measure in transcripts directly. Often another heterozygous SNP located in the coding region and tightly linked to the rSNP is used for expression analysis (in linkage disequilibrium and therefore in the same haplotype as the rSNP). This implies that some genotyping work has to be done before allelic imbalance studies can be started. In fact, expression of intronic rSNPs can also be measured, because unspliced introns are often present in transcripts and can be detected when random hexamers are used for cDNA synthesis (Ref. 97). These measurements should be validated by analysis of allelic gene expression with additional coding SNPs. Several recent studies of SNPs have shown deviation from equimolar allelic expression in at least one individual in 18–54% of the analysed genes (this wide range probably depends on the study, number of individuals and thresholds used) (Refs 97, 98, 99). Various methods can be used to detect allele-specific mRNA expression, such as RT-PCR, microarrays, and padlock probes (Ref. 100). Identification of allelic imbalance might be the most direct method for functional implication of a given rSNP in the regulation of a given gene, and provides a first step towards an understanding of the role of that rSNP in disease susceptibility.

Animal models

Eventually, the final proof for the role of a mutated gene in a disease is through the experimental evidence that can be provided by in vivo models. Methods to produce so called knockin models (where a specific mutation can be inserted), or conditional or knockout models, have been extensively discussed and fall outside the scope of this review.

Conclusions

Millions of SNPs have already been identified. The time has now come to evaluate which of these alter gene function and gene expression. rSNPs are a rapidly emerging class of functional SNPs that affect the level and tissue specificity of gene expression. In a complex combination with other factors, rSNPs can increase the susceptibility to different diseases. Future directions in the genetics of complex diseases will be increasingly orientated towards the biological and functional understanding of mutations, perhaps leading to the demonstration of new biological pathways and networks. This will help in understanding the complex interplay between different mutations contributing to different phenotypes, and in applying this knowledge to the development of new therapies and diagnostics for complex diseases. It is possible that in the future we will be able to modify these rSNPs so that new therapeutic alternatives can be developed, as, for example, in the use of artificial transcription factors.

Acknowledgements and funding

The authors acknowledge the financial support of the Swedish Research Council, the Swedish Association against Rheumatism, the Beijer Foundation, the Clas Groschinsky Memorial Foundation and the Swedish Medical Association. Participants and instructors of the Eukaryotic Gene Expression Course 2003, Cold Spring Harbor, USA, are acknowledged for fruitful discussions reflected in this paper. This article was refereed anonymously and the authors are grateful for the constructive comments of the referees.

References

1 Venter, J.C. et al. (2001) The sequence of the human genome. Science 291, 1304-1351, PubMed: 11181995
2 http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi
3 http://hgvbase.cgb.ki.se/cgi-bin/main.pl?page=snp_info.htm
4 Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314, 67-73, PubMed: 3856104
5 Botstein, D. et al. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32, 314-331, PubMed: 6247908

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

6 Telenius, H. et al. (1990) Minisatellite DNA profiles: rapid sample identification in linkage analysis. Hum Hered 40, 77-80, PubMed: 1970804
7 http://snp.cshl.org/
8 http://www.ensembl.org/genome/central/
9 (1994) Population variation of common cystic fibrosis mutations. The Cystic Fibrosis Genetic Analysis Consortium. Hum Mutat 4, 167-177, PubMed: 7530552
10 Hall, J.M. et al. (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250, 1684-1689, PubMed: 2270482
11 Sherrington, R. et al. (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 375, 754-760, PubMed: 7596406
12 Knight, J.C. et al. (1999) A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. Nat Genet 22, 145-150, PubMed: 10369255
13 Ozaki, K. et al. (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 32, 650-654, PubMed: 12426569
14 Prokunina, L. et al. (2002) A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. Nat Genet 32, 666-669, PubMed: 12402038
15 Prokunina, L. et al. (2004) The systemic lupus erythematosus-associated PDCD1 polymorphism PD1.3A in lupus nephritis. Arthritis Rheum 50, 327-328, PubMed: 14730631
16 Tokuhiro, S. et al. (2003) An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. Nat Genet 35, 341-348, PubMed: 14608356
17 Helms, C. et al. (2003) A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis. Nat Genet 35, 349-356, PubMed: 14608357
18 Alarcon-Riquelme, M.E. (2003) A RUNX trio with a taste for autoimmunity. Nat Genet 35, 299-300, PubMed: 14647282
19 Novik, K.L. et al. (2002) Epigenomics: genome-wide study of methylation phenomena. Curr Issues Mol Biol 4, 111-128, PubMed: 12432963
20 Van Laere, A.S. et al. (2003) A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 425, 832-836, PubMed: 14574411
21 Miller, G.M. and Madras, B.K. (2002) Polymorphisms in the 3'-untranslated region of human and monkey dopamine transporter genes affect reporter gene expression. Mol Psychiatry 7, 44-55, PubMed: 11803445
22 Di Paola, R. et al. (2002) A variation in 3' UTR of hPTP1B increases specific gene expression and associates with insulin resistance. Am J Hum Genet 70, 806-812, PubMed: 11833006
23 Bevilacqua, A. et al. (2003) Post-transcriptional regulation of gene expression by degradation of messenger RNAs. J Cell Physiol 195, 356-372, PubMed: 12704645
24 Duan, J. et al. (2003) Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. Hum Mol Genet 12, 205-216, PubMed: 12554675
25 Suzuki, A. et al. (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat Genet 34, 395-402, PubMed: 12833157
26 Faustino, N.A. and Cooper, T.A. (2003) Pre-mRNA splicing and human disease. Genes Dev 17, 419-437, PubMed: 12600935
27 Krawczak, M., Reiss, J. and Cooper, D.N. (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 90, 41-54, PubMed: 1427786
28 Lorson, C.L. et al. (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc Natl Acad Sci U S A 96, 6307-6311, PubMed: 10339583
29 Shiga, N. et al. (1997) Disruption of the splicing enhancer sequence within exon 27 of the dystrophin gene by a nonsense mutation induces partial skipping of the exon and is responsible for Becker muscular dystrophy. J Clin Invest 100, 2204-2210, PubMed: 9410897
30 Liu, H.X. et al. (2001) A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. Nat Genet 27, 55-58, PubMed: 11137998
31 Pagani, F. et al. (2003) Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. J Biol Chem 278, 26580-26588, PubMed: 12732620
32 Pagani, F. et al. (2003) New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

exon 12. Hum Mol Genet 12, 1111-1120, PubMed: 12719375
33 Lynch, K.W. and Weiss, A. (2001) A CD45 polymorphism associated with multiple sclerosis disrupts an exonic splicing silencer. J Biol Chem 276, 24341-24347, PubMed: 11306584
34 Pyne, M.T. et al. (2000) The BRCA2 genetic variant IVS7 + 2T—>G is a mutation. J Hum Genet 45, 351-357, PubMed: 11185744
35 Ueda, H. et al. (2003) Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 423, 506-511, PubMed: 12724780
36 Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci 25, 106-110, PubMed: 10694877
37 Caceres, J.F. and Kornblihtt, A.R. (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet 18, 186-193, PubMed: 11932019
38 Hastings, M.L. and Krainer, A.R. (2001) Pre-mRNA splicing in the new millennium. Curr Opin Cell Biol 13, 302-309, PubMed: 11343900
39 Cartegni, L., Chew, S.L. and Krainer, A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3, 285-298, PubMed: 11967553
40 Kuersten, S. and Goodwin, E.B. (2003) The power of the 3' UTR: translational control and development. Nat Rev Genet 4, 626-637, PubMed: 12897774
41 Song, W.J. et al. (1999) Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. Nat Genet 23, 166-175, PubMed: 10508512
42 Fisher, E. and Scambler, P. (1994) Human haploinsufficiency—one for sorrow, two for joy. Nat Genet 7, 5-7, PubMed: 8075640
43 Kozlowski, P. and Krzyzosiak, W.J. (2001) Combined SSCP/duplex analysis by capillary electrophoresis for more efficient mutation detection. Nucleic Acids Res 29, E71, PubMed: 11452040
44 Kaczanowski, R., Trzeciak, L. and Kucharczyk, K. (2001) Multitemperature single-strand conformation polymorphism. Electrophoresis 22, 3539-3545, PubMed: 11669539
45 Ribas, G., Neville, M.J. and Campbell, R.D. (2001) Single-nucleotide polymorphism detection by denaturing high-performance liquid chromatography and direct sequencing in genes in the MHC class III region encoding novel cell surface molecules. Immunogenetics 53, 369-381, PubMed: 11486274
46 Xiao, W. et al. (2001) Multiplex capillary denaturing high-performance liquid chromatography with laser-induced fluorescence detection. Biotechniques 30, 1332-1338, PubMed: 11414227
47 http://www.epd.isb-sib.ch
48 Hoogendoorn, B. et al. (2003) Functional analysis of human promoter polymorphisms. Hum Mol Genet 12, 2249-2254, PubMed: 12915441
49 Hardison, R.C., Oeltjen, J. and Miller, W. (1997) Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. Genome Res 7, 959-966, PubMed: 9331366
50 Mayor, C. et al. (2000) VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16, 1046-1047, PubMed: 11159318
51 http://www.hgmp.mrc.ac.uk/Bioinformatics/
52 Dermitzakis, E.T. and Clark, A.G. (2002) Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. Mol Biol Evol 19, 1114-1121, PubMed: 12082130
53 Waterston, R.H. et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520-562, PubMed: 12466850
54 Hoehe, M.R. (2003) Haplotypes and the systematic analysis of genetic variation in genes and genomes. Pharmacogenomics 4, 547-570, PubMed: 12943464
55 Cardon, L.R. and Abecasis, G.R. (2003) Using haplotype blocks to map human complex trait loci. Trends Genet 19, 135-140, PubMed: 12615007
56 Clark, A.G. (2003) Finding genes underlying risk of complex disease by linkage disequilibrium mapping. Curr Opin Genet Dev 13, 296-302, PubMed: 12787793
57 Abecasis, G.R. et al. (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30, 97-101, PubMed: 11731797
58 http://www.stat.washington.edu/stephens/phase.html
59 Kruglyak, L. et al. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58, 1347-1363, PubMed: 8651312
60 http://linkage.rockefeller.edu/soft/
61 Lathrop, G.M. (1983) Estimating genotype relative risks. Tissue Antigens 22, 160-166,

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

PubMed: 6623452
62 Spielman, R.S., McGinnis, R.E. and Ewens, W.J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52, 506-516, PubMed: 8447318
63 Martin, E.R. et al. (2000) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 67, 146-154, PubMed: 10825280
64 Ott, J. (1989) Statistical properties of the haplotype relative risk. Genet Epidemiol 6, 127-130, PubMed: 2731704
65 Thomson, G. (1995) Mapping disease genes: family-based association studies. Am J Hum Genet 57, 487-498, PubMed: 7668275
66 http://exon.cshl.org/ESE/
67 Fairbrother, W.G. et al. (2002) Predictive identification of exonic splicing enhancers in human genes. Science 297, 1007-1013, PubMed: 12114529
68 Graham, I.R., Hamshere, M. and Eperon, I.C. (1992) Alternative splicing of a human alpha-tropomyosin muscle-specific exon: identification of determining sequences. Mol Cell Biol 12, 3872-3882, PubMed: 1508190
69 Tacke, R. and Goridis, C. (1991) Alternative splicing in the neural cell adhesion molecule pre-mRNA: regulation of exon 18 skipping depends on the 5'-splice site. Genes Dev 5, 1416-1429, PubMed: 1869048
70 Cuadrado, A. et al. (2002) HuD binds to three AU-rich sequences in the 3'-UTR of neuroserpin mRNA and promotes the accumulation of neuroserpin mRNA and protein. Nucleic Acids Res 30, 2202-2211, PubMed: 12000840
71 http://transfac.gbf.de/TRANSFAC/
72 http://molsun1.cbrc.aist.go.jp/research/db/TFSEARCH.html
73 http://mordor.cgb.ki.se/cgi-bin/CONSITE/consite
74 http://www.123genomics.com/files/home.html
75 Fried, M. and Crothers, D.M. (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res 9, 6505-6525, PubMed: 6275366
76 Kelly, S. et al. (2003) Recombinant expression, purification and characterisation of the HMG domain of human SRY. Protein Pept Lett 10, 281-286, PubMed: 12871148
77 Horikawa, Y. et al. (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet 26, 163-175, PubMed: 11017071
78 Galas, D.J. and Schmitz, A. (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res 5, 3157-3170, PubMed: 212715
79 Alam, J. and Cook, J.L. (1990) Reporter genes: application to the study of mammalian gene transcription. Anal Biochem 188, 245-254, PubMed: 2121064
80 Nordeen, S.K. (1988) Luciferase reporter gene vectors for analysis of promoters and enhancers. Biotechniques 6, 454-458, PubMed: 2908509
81 Gu, T.L. et al. (2000) Auto-inhibition and partner proteins, core-binding factor beta (CBFbeta) and Ets-1, modulate DNA binding by CBFalpha2 (AML1). Mol Cell Biol 20, 91-103, PubMed: 10594012
82 Kurdistani, S.K. and Grunstein, M. (2003) In vivo protein-protein and protein-DNA crosslinking for genomewide binding microarray. Methods 31, 90-95, PubMed: 12893178
83 Ren, B. et al. (2002) E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev 16, 245-256, PubMed: 11799067
84 Weinmann, A.S. and Farnham, P.J. (2002) Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. Methods 26, 37-47, PubMed: 12054903
85 Bird, A.P. (1980) DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res 8, 1499-1504, PubMed: 6253938
86 Meehan, R. et al. (1992) Transcriptional repression by methylation of CpG. J Cell Sci Suppl 16, 9-14, PubMed: 1297654
87 Richardson, B.C. (2002) Role of DNA methylation in the regulation of cell function: autoimmunity, aging and cancer. J Nutr 132, 2401S-2405S, PubMed: 12163700
88 Deng, G. et al. (2001) Methylation in hMLH1 promoter interferes with its binding to transcription factor CBF and inhibits gene expression. Oncogene 20, 7120-7127, PubMed: 11704838
89 Murumagi, A., Vahamurto, P. and Peterson, P. (2003) Characterization of regulatory elements and methylation pattern of the autoimmune regulator (AIRE) promoter. J Biol Chem 278, 19784-19790, PubMed: 12651856
90 http://www.affymetrix.com/index.affx
91 http://cmgm.stanford.edu/pbrown/
92 Duggan, D.J. et al. (1999) Expression profiling

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

using cDNA microarrays. Nat Genet 21, 10-14, PubMed: 9915494
93 Churchill, G.A. (2002) Fundamentals of experimental design for cDNA microarrays. Nat Genet 32 Suppl, 490-495, PubMed: 12454643
94 Quackenbush, J. (2002) Microarray data normalization and transformation. Nat Genet 32 Suppl, 496-501, PubMed: 12454644
95 http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html
96 http://www.ebi.ac.uk/arrayexpress/
97 Pastinen, T. et al. (2004) A survey of genetic and epigenetic variation affecting human gene expression. Physiol Genomics 16, 184-193, PubMed: 14583597
98 Yan, H. et al. (2002) Allelic variation in human gene expression. Science 297, 1143, PubMed: 12183620
99 Lo, H.S. et al. (2003) Allelic variation in gene expression is common in the human genome. Genome Res 13, 1855-1862, PubMed: 12902379
100 Nilsson, M. et al. (2000) Enhanced detection and distinction of RNA by enzymatic probe ligation. Nat Biotechnol 18, 791-793, PubMed: 10888852
101 Stoerker, J. et al. (2000) Rapid genotyping by MALDI-monitored nuclease selection from probe libraries. Nat Biotechnol 18, 1213-1216, PubMed: 11062445
102 Ye, S. et al. (2002) Detection of single nucleotide polymorphisms by the combination of nuclease S1 and PNA. Nucleic Acids Res Suppl 235-236, PubMed: 12903192
103 Bray, M.S., Boerwinkle, E. and Doris, P.A. (2001) High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. Hum Mutat 17, 296-304, PubMed: 11295828
104 Prince, J.A. et al. (2001) Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. Genome Res 11, 152-162, PubMed: 11156624
105 Jobs, M. et al. (2003) DASH-2: flexible, low-cost, and high-throughput SNP genotyping by dynamic allele-specific hybridization on membrane arrays. Genome Res 13, 916-924, PubMed: 12727908
106 Fakhrai-Rad, H., Pourmand, N. and Ronaghi, M. (2002) Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. Hum Mutat 19, 479-485, PubMed: 11968080
107 Makridakis, N.M. and Reichardt, J.K. (2001) Multiplex automated primer extension analysis: simultaneous genotyping of several polymorphisms. Biotechniques 31, 1374-1380, PubMed: 11768667
108 Shapero, M.H. et al. (2001) SNP genotyping by multiplexed solid-phase amplification and fluorescent minisequencing. Genome Res 11, 1926-1934, PubMed: 11691857
109 Mein, C.A. et al. (2000) Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. Genome Res 10, 330-343, PubMed: 10720574
110 Hsu, T.M. et al. (2001) Genotyping single-nucleotide polymorphisms by the invader assay with dual-color fluorescence polarization detection. Clin Chem 47, 1373-1377, PubMed: 11468224
111 Qi, X. et al. (2001) L-RCA (ligation-rolling circle amplification): a general method for genotyping of single nucleotide polymorphisms (SNPs). Nucleic Acids Res 29, E116, PubMed: 11713336
112 Faruqi, A.F. et al. (2001) High-throughput genotyping of single nucleotide polymorphisms with rolling circle amplification. BMC Genomics 2, 4, PubMed: 11511324
113 Akey, J.M. et al. (2001) Melting curve analysis of SNPs (McSNP): a gel-free and inexpensive approach for SNP genotyping. Biotechniques 30, 358-362, 364, 366-357, PubMed: 11233605
114 Hardenbol, P. et al. (2003) Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat Biotechnol 21, 673-678, PubMed: 12730666
115 Zhang, S. et al. (2002) Detection of single nucleotide polymorphisms using electrospray ionization mass spectrometry: validation of a one-well assay and quantitative pooling studies. J Mass Spectrom 37, 1039-1050, PubMed: 12375278
116 Xu, H. et al. (2003) Multiplexed SNP genotyping using the Qbead system: a quantum dot-encoded microsphere-based assay. Nucleic Acids Res 31, e43, PubMed: 12682378
117 Xiao, M. and Kwok, P.Y. (2003) DNA analysis by fluorescence quenching detection. Genome Res 13, 932-939, PubMed: 12727909
118 Waterfall, C.M. and Cobb, B.D. (2002) SNP genotyping using single-tube fluorescent bidirectional PCR. Biotechniques 33, 80, 82-84, 86 passim, PubMed: 12139261
119 Vreeland, W.N., Meagher, R.J. and Barron, A.E. (2002) Multiplexed, high-throughput genotyping by single-base extension and end-labeled free-

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press

solution electrophoresis. Anal Chem 74, 4328-4333, PubMed: 12236339
120 Mhlanga, M.M. and Malmberg, L. (2001) Using molecular beacons to detect single-nucleotide polymorphisms with real-time PCR. Methods 25, 463-471, PubMed: 11846616
121 Kuzuya, A. et al. (2002) Novel approach for SNP genotyping based on site-selective RNA scission. Nucleic Acids Res Suppl 129-130, PubMed: 12903139
122 Iannone, M.A. et al. (2000) Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. Cytometry 39, 131-140, PubMed: 10679731
123 Chen, X., Levine, L. and Kwok, P.Y. (1999) Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res 9, 492-498, PubMed: 10330129
124 Myakishev, M.V. et al. (2001) High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers. Genome Res 11, 163-169, PubMed: 11156625

Further reading, resources and contacts

Books

Strachan, T. and Read, A.P. (1999) Human Molecular Genetics (2nd edn), BIOS Scientific Publishers, Abingdon, Oxfordshire, UK

Lodish, H. et al. (2003) Molecular Cell Biology (4th edn), W.H. Freeman, New York

Websites

The National Center for Biotechnology Information provides data on the human genome sequence, species homologies, gene and protein variation, protein structure and more:

http://www.ncbi.nlm.nih.gov/

The European Molecular Biology Laboratory's STRING database collates known and predicted protein–protein interactions:

http://string.embl.de/

TFSEARCH (Computational Biology Research Center, AIST, Japan) allows searching for transcription factor binding sites:

http://www.cbrc.jp/research/db/TFSEARCH.html

A discussion of MALDI-TOF can be found at the website of the Department of Microbiology, University of Nijmegen, The Netherlands:

http://www.micrbiol.sci.kun.nl/tech/malditof.html

Features associated with this article

Tables

Table 1. Methods of SNP (single-nucleotide polymorphism) genotyping.
Table 2. Functional validation of regulatory SNPs (single-nucleotide polymorphisms).

Citation details for this article

Ludmila Prokunina and Marta E. Alarcón-Riquelme (2004) Regulatory SNPs in complex diseases: their identification and functional validation. Expert Rev. Mol. Med. Vol. 6, Issue 10, 27 April, DOI: 10.1017/S1462399404007690

Accession information: DOI: 10.1017/S1462399404007690; Vol. 6; Issue 10; 27 April 2004
©2004 Cambridge University Press

https://doi.org/10.1017/S1462399404007690 Published online by Cambridge University Press