



## Post-GWAS methodologies for localisation of functional non-coding variants: *ANGPTL3*



Federico Oldoni <sup>a, b, 1</sup>, Jutta Palmen <sup>a, 1</sup>, Claudia Giambartolomei <sup>c</sup>, Philip Howard <sup>a</sup>, Fotios Drenos <sup>a</sup>, Vincent Plagnol <sup>c</sup>, Steve E. Humphries <sup>a</sup>, Philippa J. Talmud <sup>a</sup>, Andrew J.P. Smith <sup>a, \*</sup>

<sup>a</sup> Department of Cardiovascular Genetics, Institute of Cardiovascular Sciences, University College London, London, UK

<sup>b</sup> Department of Molecular Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

<sup>c</sup> UCL Genetics Institute, University College London, London, UK

### ARTICLE INFO

#### Article history:

Received 21 October 2015

Received in revised form

29 November 2015

Accepted 4 December 2015

Available online 12 December 2015

#### Keywords:

*ANGPTL3*

Polymorphism

Regulation

Chromatin

Genome-wide

Functional polymorphism

FAIRE

LDL-C

### ABSTRACT

Genome-wide association studies have confirmed the involvement of non-coding angiotensin-like 3 (*ANGPTL3*) gene variants with coronary artery disease, levels of low-density lipoprotein cholesterol (LDL-C), triglycerides and *ANGPTL3* mRNA transcript. Extensive linkage disequilibrium at the locus, however, has hindered efforts to identify the potential functional variants. Using regulatory annotations from ENCODE, combined with functional *in vivo* assays such as allele-specific formaldehyde-assisted isolation of regulatory elements, statistical approaches including eQTL/lipid colocalisation, and traditional *in vitro* methodologies including electrophoretic mobility shift assay and luciferase reporter assays, variants affecting the *ANGPTL3* regulome were examined. From 253 variants associated with *ANGPTL3* mRNA expression, and/or lipid traits, 46 were located within liver regulatory elements and potentially functional. One variant, rs10889352, demonstrated allele-specific effects on DNA-protein interactions, reporter gene expression and chromatin accessibility, in line with effects on LDL-C levels and expression of *ANGPTL3* mRNA. The *ANGPTL3* gene lies within *DOCK7*, although the variant is within non-coding regions outside of *ANGPTL3*, within *DOCK7*, suggesting complex long-range regulatory effects on gene expression. This study illustrates the power of combining multiple genome-wide datasets with laboratory data to localise functional non-coding variation and provides a model for analysis of regulatory variants from GWAS.

© 2015 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

### 1. Introduction

Genome-wide association studies (GWAS) have revealed novel gene loci and confirmed many candidate genes for numerous traits including lipid levels and coronary artery disease (CAD) risk [1,2]. Despite the increasing number of loci associated with such traits, identification of functional or causal variants at each locus has not been met with similar advances. This lack of progress is largely due

to two important characteristics: the presence of many variants in strong linkage disequilibrium (LD) with the leading GWAS signal, and the fact that the majority of potentially causal SNPs are located in non-coding regions of the genome, for which *in silico* tools to reliably predict SNP functionality are lacking.

A typical example of this occurs at the angiotensin-like 3 (*ANGPTL3*) gene locus, associated with triglyceride (TG) levels, low-density lipoprotein cholesterol (LDL-C) levels, and CAD risk in several independent GWAS [1,3,4], where the functional variant(s) associated with these traits have not been identified. *ANGPTL3* is a member of the angiotensin-like proteins, a family of secreted glycoproteins, expressed in the liver, containing a signal peptide, N-terminal coiled coil domain and a C-terminal fibrinogen-like domain. *ANGPTL3* can inhibit lipoprotein lipase [5] and endothelial lipase [6], and through the fibrinogen-like domain, induce endothelial cell adhesion and migration [7].

**Abbreviations:** GWAS, genome-wide association study; LDL-C, low density lipoprotein cholesterol; HDL-C, high density lipoprotein cholesterol; TG, triglycerides; EMSA, electrophoretic mobility shift assay; SNP, single nucleotide polymorphism.

\* Corresponding author. Centre for Cardiovascular Genetics, Department of Medicine, University College London, 5 University Street, London WC1E 6JF, UK.

E-mail address: [Andrew.J.P.Smith@ucl.ac.uk](mailto:Andrew.J.P.Smith@ucl.ac.uk) (A.J.P. Smith).

<sup>1</sup> These authors contributed equally.

In a large meta-analysis comprising >100,000 individuals, rs2131925, the lead SNP at the *ANGPTL3* locus for TG association (4.9 mg/dL decrease) was also associated with LDL-C levels (1.6 mg/dL decrease) [1]. In addition to these associations with common polymorphism, rare missense mutations in *ANGPTL3* have also been identified through exome-sequencing to be associated with familial combined hypolipidemia, confirming this locus as an important mediator in lipid metabolism [8].

*ANGPTL3* lies on the forward strand of chromosome 1p31 within an intronic region of *DOCK7* (located on the reverse strand). Examining the *ANGPTL3* locus, it can be seen that there are over 250 SNPs in very strong LD with the lead TG GWAS SNP that span ~250 kb and include 3 genes [1,9], *ANGPTL3*, *DOCK7* and *USP1* (Fig. 1A), leading to the strong likelihood that rs2131925 is simply acting as a marker for the true functional variant(s). Interestingly, SNPs at this locus show a lack of significant association with HDL-C (Fig. 1B), but association with LDL-C shows a similar pattern as for TG, suggesting a common functional polymorphism (Fig. 1C and D).

One commonly employed method to identify potential causal non-coding variants is through *in silico* tools, including databases used to predict the effect of a SNP on transcription factor binding motifs. Such methods, unfortunately, result in many false-positive findings that are not borne out when tested in the laboratory [10]. Functional examination of all SNPs showing association to a trait can be undertaken in the laboratory where the number of SNPs is small [11,12], although when strong LD within the association signal results in hundreds of potentially causal variants, such analysis is not feasible. The ENCODE and Roadmap Epigenomics

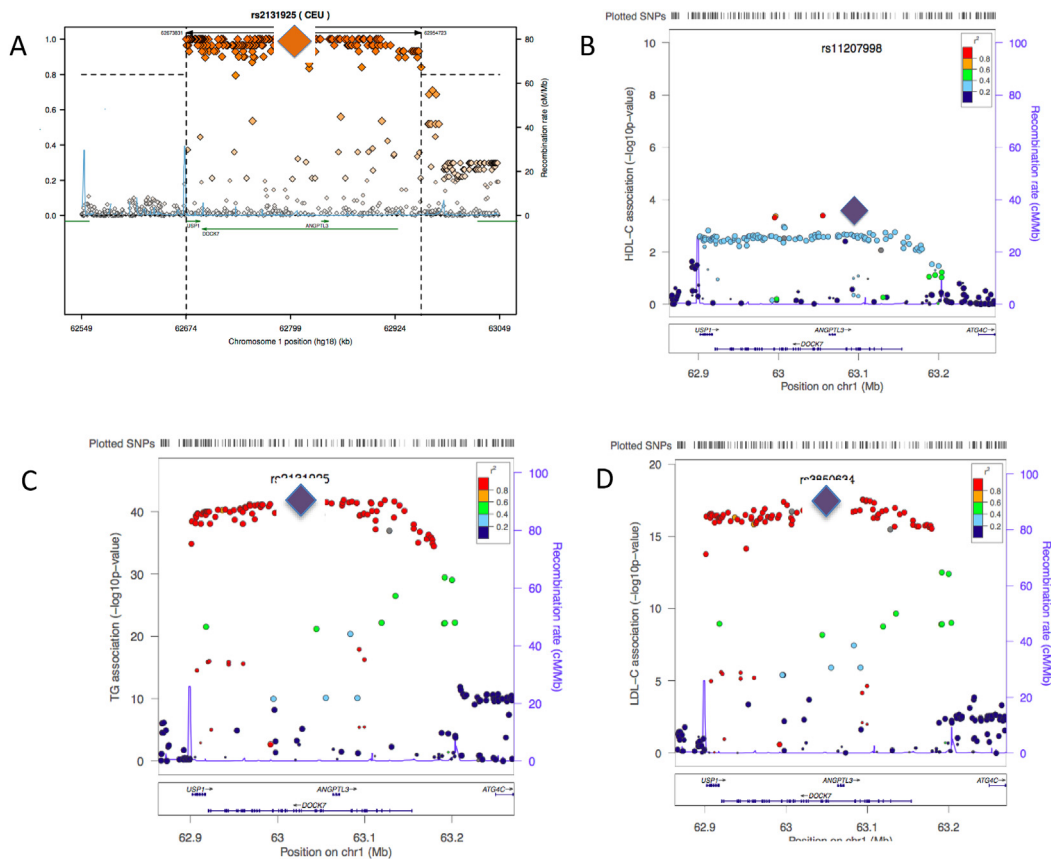
projects, however, provide additional tools that can be used to identify non-coding regions of the genome that are likely to harbour important regulatory features [13,14]. Where such features also contain a genetic variant associated with an epidemiologic trait, this may suggest potentially causality of the variant, rather than simply a marker for the trait.

The lead SNP, rs2131925 is unlikely to represent the functional SNP at the locus, with 92 SNPs in perfect LD, and 202 SNPs in very strong LD ( $r^2 > 0.8$ ) [9]. We used data from the ENCODE Project and Roadmap Epigenomics, in addition to epidemiological, statistical and functional assays, to assist in the identification of potentially causal variants at the *ANGPTL3* locus, and examined these variants for functionality in the laboratory.

## 2. Materials and methods

### 2.1. Identification of SNPs associated with lipid biomarkers

Variants in strong LD ( $r^2 \geq 0.8$ ) with the GWAS lead SNP (rs2131925) for TG and LDL-C [14] were identified using 1000 Genomes Project CEU Caucasian population Phase 1 data (1KGP) [9]. Additionally, all variants within 500 kb from *ANGPTL3* reaching genome-wide significance ( $p < 5 \times 10^{-8}$ ) for LDL-C or TG levels using a large-scale meta-analysis of >100,000 individuals were identified [1] and selected for further analysis.



**Fig. 1.** A) Linkage disequilibrium with the lead TG GWAS SNP, rs2131925 (indicated by large diamond), at the *ANGPTL3* locus. Over 200 SNPs are in strong LD at this locus ( $r^2 > 0.8$ ), which span 3 genes, *USP1*, *DOCK7* and *ANGPTL3* (<http://www.broadinstitute.org/mpg/snap/ldplot.php>). 1B) Association plot of HDL-C at the *ANGPTL3* locus, revealing no significant association. 1C) Association plot of TG at *ANGPTL3* locus, showing rs2131925 as lead SNP. 1D) Association plot of LDL-C at *ANGPTL3* locus, showing lead SNP as rs3850634 [1] (<http://csg.sph.umich.edu/locuszoom>).

## 2.2. Identification of variants associated with *ANGPTL3* mRNA transcript levels

Gene expression and genotype data were obtained for 966 post-mortem and surgical resection human liver samples from unrelated European-American individuals [15]. Genotyping was carried out using the Illumina 650Y BeadChip array, and Agilent human gene expression arrays used to profile 39,000 expression probes. Imputation of genotypes not covered on the array was carried out as previously described, and the full methods for this, and the procedures for colocalisation have been detailed previously [16].

## 2.3. Identification of candidate functional SNPs using ENCODE

Published ENCODE annotations for the HepG2 liver cell line for DNaseI-seq, FAIRE-seq, ChIP-seq (transcription factors and histone methylation signatures H3K4me1 and H3K27Ac) were examined in relation to variants associated with lipid biomarkers and *ANGPTL3* transcript levels. Roadmap Epigenomics annotations for liver tissue were examined for DNaseI-seq and histone modifications. SNPs were identified which occurred within narrow peaks for DNaseI hypersensitivity or within 500 bp of SPP narrow peaks for ChIP-seq as previously described [17]. SNPs which were located close to regulatory signatures (<500 bp) were also included.

## 2.4. Tissue culture

HepG2 cells were cultured in high-glucose Dulbecco's modified Eagle's medium (PAA, Austria) supplemented with 2 mM L-glutamine and 10% foetal bovine serum (FBS; PAA). EBV-transformed lymphoblast cell lines from the Centre d'Etude du Polymorphisme Humain (CEPH) panel (Coriell Cell Repositories), were cultured in RPMI 1640 (PAA) with 2 mM L-glutamine and 15% foetal bovine serum (PAA) at 37 °C, 5% CO<sub>2</sub>. Cell viability was verified using the ADAM-MC cell counter (Digital Bio), and minimum cell viability for experiments was ≥99%. Cytokine stimulation was performed by overnight incubation in serum-free media, followed by addition of 5 ng/ml IL-1β, two hours prior to cell fixing.

## 2.5. In vitro analysis of DNA-protein interactions

Nuclear extracts were obtained from HepG2 cells using the NE-PER Nuclear and Cytoplasmic Extraction Reagents Kit (Pierce, USA) as described in the manual, with the addition of Complete Protease Inhibitor (Roche, UK) to buffers CER I and NER I. EMSA probes were designed using 15 bp of genomic sequence on each side of the candidate SNP, and reverse complement for each forward strand was synthesised and annealed (sequences for all probes available on request). Forward sequence for rs6690733: 5'-ATCAAAGGGGGTAA(A/C)CACTACTACGTATAC and rs10889352: 5'-GCAAACAGAAAAA(T/C)ATAGTACTAGTAAAA-3'.

Probes were labelled using the Biotin 3'-End DNA Labelling Kit (Pierce, USA) as described in the manual, and complementary labelled oligonucleotides annealed by heating to 95 °C for 10 min, followed by slow cooling (−1 °C/min to 30 °C). Each binding reaction consisted of 2 µl 10× binding buffer (100 mM Tris, 500 mM KCl; pH 7.5), 1 µg p[dl-dC], 2 nmol biotin-labelled DNA, made to a total of 20 µl with H<sub>2</sub>O, and incubated at 25 °C for 30 min, followed by the addition of 5× loading buffer. Samples were electrophoresed on a 6% polyacrylamide gel for 150 min at 120 V. Transfer to nylon membrane was achieved through Southern transfer and detected using the Chemiluminescent Nucleic Acid Detection Module (Pierce, USA).

## 2.6. In silico analysis of transcription factor binding

To determine which proteins may be interacting with EMSA oligonucleotide probes, the MatInspector (Genomatix, Germany) [18] database was examined using the Transcription Factor Binding Sites (Weight matrices) library for general core promoter elements and vertebrates matrix group. Core similarity (the proportion of bases exactly matching the highest conserved positions of the matrix) was set to 0.75. HaploReg v3 was used as an addition source for predicted SNP effects on transcription factor binding [17].

## 2.7. In vitro analysis of gene expression

For the putative *ANGPTL3* regulatory elements surrounding the SNPs of interest, oligonucleotides were designed with the addition of BamHI and Sall restriction sites at the 5'- and 3'-ends and PCR amplification of homozygous wild-type individuals performed. Primer sequences for cloning: rs6690733 - 5'-GGGGTACTCTTCAAATCCCT-3' forward, 5'-GGCAGCTTCTCTCCCTAT-3' reverse; rs10889352 - 5'-TGACACAGGCTAGGTGACAAA-3' forward, 5'-AGAAGCAGATGCCACCAAGG-3' reverse. PCR products were digested with BamHI and Sall, and ligated into the BamHI/Sall-linearized pGL3 Promoter vector (Promega), downstream of the SV40 late poly(A) signal. Allelic variants were created using the Site-Directed-Mutagenesis kit (Invitrogen), the variant verified by sequencing and insert digested and re-inserted into a fresh pGL3 vector. The size of insert for rs6690733 was 249 bp and 192 bp for rs10889352.

HepG2 cells were seeded at  $2.0 \times 10^4$ /well in a 96 well plate and grown to confluence in high-glucose Dulbecco's modified Eagle's medium (PAA, Austria) supplemented with 2 mM L-glutamine and 10% fetal bovine serum (FBS). Cells were transfected with 250 ng pGL3 reporter construct with 10 ng pRLTK as a transfection control. Transfection was carried out using Opti-Mem serum-free media (Sigma, Dorset, UK) and Lipofectamine 2000 (Invitrogen, Paisley, UK) as described in the manual. Media was replaced 24 h following transfection, with serum-containing media, and the cells left for 2 days before harvesting. Cells were lysed using Passive Lysis Buffer (Promega, Southampton, UK) and luciferase expression was determined using the Dual Luciferase Reporter Assay System (Promega, Southampton, UK), and measured in the Tropic TR717 Microplate Luminometer (PE Applied Biosystems, UK). Luciferase activity was determined as the mean of 12 transfections with the assay performed in triplicate.

## 2.8. Allele-specific FAIRE

The methodology for allele-specific FAIRE used for this study is described elsewhere [19]. Briefly, 20 unrelated lymphoblast cell lines of European origin (Coriell Repositories) were fixed in the log-phase of growth, and the chromatin extracted. Following sonication, one portion of chromatin was reverse-crosslinked and purified, the other underwent FAIRE enrichment using phenol:chloroform:isoamyl alcohol. Following reversal of crosslinks and DNA purification, all samples were genotyped using the Illumina Metachip array. Identification of heterozygous individuals at each allele was determined using the B allele frequency (BAF) from the samples not enriched by FAIRE, and the BAF for these individuals compared to that from the FAIRE-treated samples. For rs6690733, which was not present on the Metachip array, a custom KASPar (KBioscience) genotyping assay was created to determine the allelic imbalance in heterozygotes from extrapolation from a standard curve, and this data was analysed in the same manner as with the Metachip-derived genotype data (KASPar probes available on request).

## 2.9. Statistical analysis

Luciferase expression was normalised to the pRLTK control vector, and a 2-sided t-test performed between each run of 12 samples. Assays were performed in triplicate. Analysis of FAIRE-gen was performed using a paired t-test between the BAF from SNPs of all heterozygous individuals, comparing non-FAIRE vs FAIRE-treated chromatin. Imputation of genotypes not covered on the array was carried out as described [16] using EUR reference haplotypes available from the 1000 Genomes Project. eQTL analysis was run by fitting a linear regression between expression of the gene and all SNPs between 200 kilobases upstream and downstream from the probe. Colocalisation analysis between eQTL and lipid biomarkers uses summary statistics downloaded from a publicly available meta-analysis (<http://www.sph.umich.edu/csg/abecasis/public/lipids2010/>) and the statistical methodology is described in full in [16]. A posterior probability >75% was considered as strong evidence for colocalisation.

## 3. Results

### 3.1. Selection of SNPs associated with lipid biomarkers

To identify the pool of variants most likely to contain the functional variant/s responsible for TG and LDL-C plasma levels at the *ANGPTL3* locus, all variants from the 1000 Genomes Project [9] in strong LD ( $r^2 > 0.8$ ) with the GWAS lead SNPs (rs2131925 and rs3850634, respectively) were identified, providing 202 SNPs (Supplementary Table 1). Additionally, meta-analysis data from >100,000 individuals with lipid measurements was examined [1], and association at the *ANGPTL3* locus with TG and LDL-C was determined. From this dataset, 151 SNPs reaching genome-wide significance ( $p < 5 \times 10^{-8}$ ) were selected. All 106 LDL-C associated SNPs were present within the subset associated with TG (Supplementary Fig. 1). Since the associated polymorphisms are all non-coding variants and therefore likely to be affecting gene expression levels, liver eQTLs for *ANGPTL3* were examined from 966 liver samples, including imputed SNPs from 1KGP not directly genotyped. This data corroborated the association of epidemiological data for lipid levels with *ANGPTL3* transcript levels (Supplementary Figs. 2 and 3). In total, 253 SNPs were identified to show either genome-wide association to a lipid trait or a liver eQTL for *ANGPTL3* within the 0.8  $r^2$  cut-off with the TG lead SNP from GWAS, rs2131925 (Supplementary Table 1). Interestingly, it was observed that *DOCK7* transcript levels are also affected by the same variants that affect *ANGPTL3* expression.

### 3.2. Colocalisation of lipid associations and liver eQTLs at the *ANGPTL3* locus

To determine if the association with TG and LDL-C levels at the *ANGPTL3* locus shared a causal variant/s with liver *ANGPTL3* mRNA expression, a Bayesian test to examine colocalisation between pairs of traits was performed (procedure detailed in [16]). Summary statistics from associations with TG and LDL-C were derived from the lipid meta-analysis data used to select *ANGPTL3* variants [1], and the 966-sample liver expression dataset described previously was used for eQTL analysis, with imputation used to detect variants not directly genotyped. The data demonstrated that both TG (posterior probability 92.4%) and LDL-C (posterior probability 90.6%) colocalise with *ANGPTL3* expression, indicating a shared causal variant or variants (Supplementary Fig. 2). Due to the very strong LD at the region, however, colocalisation did not reduce the pool of potential causal variants, and all 253 SNPs were included for further examination to assess functional potential.

### 3.3. Functional assessment of potential regulatory *ANGPTL3* SNPs

Since variants at the *ANGPTL3* locus that determine TG and LDL-C levels colocalise with eQTL markers for *ANGPTL3* transcript levels, a common variant (or variants) affecting gene regulatory processes is likely to be responsible for both associations. The potential for this selection of 253 variants to affect gene regulation was assessed by examining data from the ENCODE Project and Roadmap Epigenomics [14,13]. Since none of these variants occur within *ANGPTL3* coding regions, 5' or 3' UTR, or predicted splice sites, allele-specific regulation is likely to be through effects on promoter, enhancer or silencer elements, rather than effects on mRNA stability. Since the variants are also associated with *DOCK7* transcript levels, a common enhancer/silencer element is more likely to be affected than an *ANGPTL3* promoter element. To identify variants that may affect such regulatory elements, ENCODE and Roadmap Epigenomics datasets were searched for universal signatures of regulatory potential, including DNaseI hypersensitivity (DHS) peaks, histone modifications and FAIRE-seq peaks, and also for transcription factor binding sites. Variants present within, or close (500 bp) to these signatures were also selected for further analysis. 33 SNPs were present directly within an annotated regulatory region in liver cells, and 13 were present close to a regulatory annotation and were included for further analysis due to possible longer-range regulatory roles (Table 1).

### 3.4. Allele-specific protein-binding effects of SNPs within regulatory regions

For a non-coding functional variant to affect gene expression, it would be expected to alter binding affinities of transcription factor/s. To assess the potential for each of the 46 lipid-associated variants from ENCODE/Roadmap Epigenomics-derived regulatory regions to affect the interaction of DNA-binding proteins, electrophoretic mobility shift assays (EMSAs) were performed for each of the alleles, using biotin-labelled double-stranded oligonucleotide probes comprising 30bp genomic sequence surrounding each variant, with the variant located directly at the centre of each probe. Each oligonucleotide probe was incubated with a HepG2 liver nuclear extract, and the resulting complexes run on a polyacrylamide gel. From a pool of 46 SNPs, 44 either did not result in transcription factor binding from either allele, or both alleles resulted in identical binding of equal intensity, with the binding site likely to be located outside of the polymorphic site. Two variants demonstrated replicated allele-specific binding to the liver nuclear extract, rs6690733 and rs10889352, indicating changes in transcription factor binding potential, and a possible regulatory role (Fig. 2). Interestingly both variants appear to show multiple proteins bound to the oligonucleotide probes, suggesting a complex means of allele-specific regulation. Replication with a Huh7 nuclear extract demonstrated identical binding patterns to HepG2 nuclear extract (data not shown) indicating a factor common to hepatic cells.

Examination of predicted changes to transcription factor binding based on each allele was examined using two related tools: MatInspector [18] and HaploReg v2 [17], which use extensive libraries of matrix descriptions for transcription factor binding sites to identify matches in DNA sequences. Both variants are predicted to alter binding sites for a number of transcription factors (Table 2). Although the two algorithms produced differing results, certain similarities were evident, such as a forkhead box protein with rs10889352 C, and promyelocytic leukaemia zinc finger binding with rs10889352 T. Examining MatInspector or HaploReg v2 results in isolation, there was a similarity in the number of predicted binding factors for each allele as seen in EMSA analysis. To determine if the *in silico* predicted binding was borne out *in vivo*,

**Table 1**  
Variants located within ENCODE-derived gene regulatory sites.

SNP	Promoter histone signature	Enhancer histone signature	DNaseI hypersensitivity	Proteins bound
rs10889356	X		X	X
rs9436661	X			
rs1168046		X		
rs10789117		X		
rs10889344		X		
rs11207979		X		
rs1168047		X		
rs12029068	X			
rs638714	X			
rs6690733		X		
rs9436221	X			
rs9436222	X			
rs4587594			X	
rs6678483		X		
rs11207997		X		
rs12090886		X		
rs1168045		X		
rs1627591		X		
rs10889337		X		
rs10493322	X		X	
rs1168114		X	X	X
rs1168113		X	X	X
rs12136083		X		
rs6679002		X		
rs12130333		X		
rs10889377		X		
rs17123817		X		
rs10889378		X		
rs13375691			X	X
rs646179			X	X
rs9436223	X			
rs631106	X			X
rs626787	X			
rs10889347				(X)
rs6587980				(X)
rs1184547				(X)
rs10157265				(X)
rs1168015			(X)	(X)
rs12062275				(X)
rs10889336			(X)	(X)
rs1168042			(X)	(X)
rs12037659				(X)
rs11207976				(X)
rs1184865				(X)
rs11207992				(X)
rs10889352		(X)		

X = SNP located within regulatory site.

(X) = SNP located close to regulatory site.

competitor assays were performed using a panel of unlabelled consensus sequences for 80 transcription factors, including 70 well-characterised transcription factors, and those predicted to be altered *in silico* [20]. This analysis, however, failed to verify the binding of predicted transcription factors or identify additional factors (data not shown), indicating a lack of reliability from prediction algorithms, or tissue-specific effects.

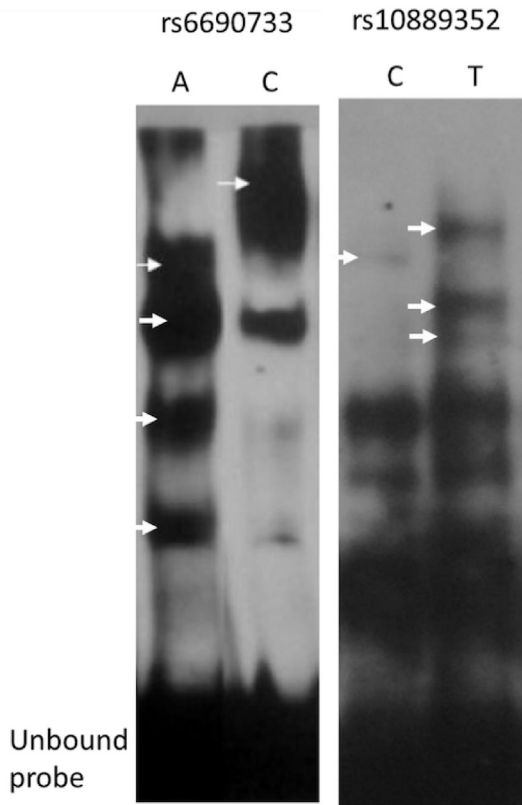
### 3.5. *In vitro* analysis of gene expression

The two SNPs associated with TG levels, *ANGPTL3* transcription and allele-specific DNA-protein interactions are strong candidates for functionality. To examine the effects of these two polymorphisms on gene expression, luciferase reporter assays were performed by designing reporter constructs containing the genomic region surrounding each SNP (200–250 bp) downstream of *luc* reporter gene (acting as a putative enhancer), driving the SV40 promoter in the pGL3 promoter vector. Constructs were transiently transfected into the HepG2 cell line and luciferase activity measured.

Both rs6690733 and rs10889352 showed allele-specific effects on reporter gene expression, with each demonstrating modest reductions (~10% reduction from each variant;  $p < 1 \times 10^{10}$ ) in gene expression for the minor alleles (Fig. 3). Interestingly, both polymorphisms appear to act as silencers in this gene expression model.

### 3.6. *In vivo* analysis of allele-specific effects on chromatin structure

Variants that affect gene transcription which are not located either within a transcription factor protein-coding sequence or regulatory RNA-coding sequence (e.g. miRNA), as is the case at the *ANGPTL3* locus, are likely to be functioning through alteration of DNA-protein interactions at promoter, enhancer or silencer elements. These interactions can either result in, or be altered by, disruption to the local chromatin structure [19]. To assess whether the two potentially functional variants at the *ANGPTL3* locus may have an effect on chromatin structure and to substantiate *in vitro* functionality *in vivo*, allele-specific formaldehyde-assisted-isolation-of-regulatory-elements (FAIRE) was performed. The allelic imbalance was assessed from heterozygotes following FAIRE-

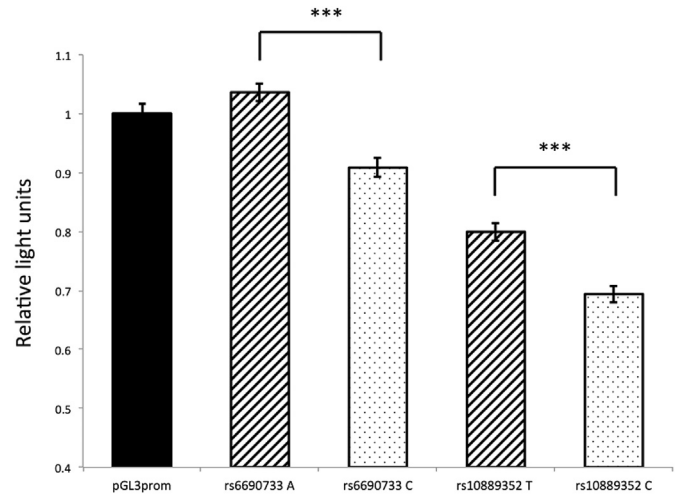


**Fig. 2.** Allele-specific effects on DNA-protein interactions. A HepG2 liver nuclear extract was mixed with allele-specific sequences for all potential regulatory variants from ENCODE analysis and allele-specific FAIRE signals. Binding differences were observed from two variants: rs6690733 and from rs10889352. Arrows indicate where a binding difference was identified between alleles, suggesting a potential effect on gene regulation.

enrichment for open chromatin, compared to non-enriched chromatin, using lymphoblast chromatin samples from 20 unrelated individuals.

Interestingly, rs10889352 T > C did indeed demonstrate a robust allele-specific effect on chromatin accessibility (20.4% enrichment of T allele from 10 FAIRE-treated heterozygotes compared to T allele-enrichment from non-treated cells,  $p = 0.002$ ), suggesting that the T allele allows for an increased regulatory potential than the C allele. This data fits with the increased binding of nuclear proteins to the T allele, and higher expression levels in reporter assay, along with the T allele associated with higher *ANGPTL3* transcript levels. The other potentially functional variant at the locus, rs6690733, however did not demonstrate effects on chromatin accessibility ( $p = 0.53$ ), and is therefore a less plausible causal candidate SNP.

Since expression of the *ANGPTL*-family respond to inflammatory stimuli [21], allele-specific FAIRE was repeated with



**Fig. 3.** Representative luciferase reporter assays for the two SNPs showing relative allele-specific effects on DNA-protein interactions compared to pGL3 promoter vector. Diagonal hatched columns represent the alleles associated with increased gene expression. Error bars represent standard error, \*\*\* =  $p < 0.001$ . Assay was repeated three times with 12 wells containing each vector per assay.

lymphoblast cells treated with the inflammatory cytokine IL-1 $\beta$  prior to the fixing of cells for FAIRE, with the hypothesis that transcription factors from inflammatory pathways may exacerbate allele-specific effects on chromatin accessibility. Upon stimulation, rs10889352 showed a similar effect (25.5% enrichment of T allele following FAIRE) and significance level ( $p = 0.0019$ ), replicating the original effects, suggesting in this case that inflammation does not interact with the functional variant.

#### 4. Discussion

Identification of functional variants from GWAS can be a laborious process with large blocks of LD surrounding the lead SNP often resulting in unfeasible numbers of variants to examine in the laboratory. The sequential methodology described here outlines a series of *in silico* and laboratory-based assays to localise functional loci at non-coding regions of the genome, using the *ANGPTL3* locus and its association with TG and LDL-C as an example (Supplementary Fig. 4). Having defined the set of potentially functional SNPs using both epidemiological data and eQTL data from the relevant tissue (liver), data from the ENCODE Project and Roadmap Epigenomics were examined in liver cells to identify regions of regulatory potential surrounding each variant, reducing the number of potentially functional variants to 46. Three complementary laboratory methods were employed to examine the effect of this reduced set of variants on regulatory potential: EMSA, a relatively inexpensive and quick method to identify the effect of variants on DNA-protein interactions; luciferase reporter assay to examine the role of these variants in gene expression; and FAIRE-

**Table 2**

Comparison of predicted transcription factor binding to putative functional *ANGPTL3* variants using MatInspector and HaploReg v2 databases.

SNP	Allele	MatInspector predicted transcription factor binding	HaploReg v2 predicted transcription factor binding
rs10889352	C	Forkhead family	Forkhead Box Protein D3
rs10889352	T	NKX-homeodomain factors, promyelocytic leukaemia zinc-finger	Ecotropic viral integration site 1/4, Glucocorticoid receptor, promyelocytic leukaemia zinc-finger, p300
rs6690733	A	Barbiturate-inducible element box from pro + eukaryotic genes, forkhead family, GLI zinc finger family, pleomorphic adenoma gene	
rs6690733	C	Barbiturate-inducible element box from pro + eukaryotic genes, hepatic nuclear factor 1, pleomorphic adenoma gene	forkhead box D3, forkhead box O4

gen [19], to confirm an *in vivo* role of these variants on chromatin accessibility and therefore regulatory potential.

Allele-specific effects on DNA-protein interactions were observed only for two SNPs (rs6690733 and rs10889352) from of the 46 variants selected from ENCODE, and effects of this manageable number of variants on gene expression were examined by luciferase reporter assays. Both variants also showed allele-specific effects on gene expression, with each allele that showed increased expression levels also associated with increases in *ANGPTL3* transcript expression. Examining *in vivo* effects on chromatin accessibility, only one of the two variants, rs10889352, demonstrated allele-specific effects with the T allele associated with a 20% increase in chromatin accessibility. The rs10889352 T allele may be associated with chromatin accessibility by altering the stability of the nucleosome complex directly from alterations in the DNA sequence, or, more likely from EMSA evidence, by altering the affinity of transcription factors which would in turn, alter chromatin dynamics.

Despite >200 variants in very strong LD with the lead SNP, only one SNP demonstrated likely functionality based on all bioinformatics, *in vivo* and *in vitro* analyses. The variant rs10889352 is located outside of the coding region for *ANGPTL3*, within an intron of *DOCK7* (Supplementary Fig. 5), and close to accessible chromatin determined by FAIRE-seq in HepG2 cells (Supplementary Fig. 6). Several transcription factor binding motifs are predicted to be altered by this variant: EVI1, FOXD3, GR, PLZF and p300 (Table 2). Higher affinity for these factors are predicted for the T allele, which supports the increased protein-DNA interactions for this allele observed in EMSA, and increased reporter gene expression. We could not, however, confirm the identity of these DNA-binding factors, despite assaying for those predicted to bind. This does not necessarily limit the ability to assign functionality to this variant, but does highlight the limitations of using *in silico* tools to correctly predict transcription factor binding. It does, however, restrict the use of chromatin immunoprecipitation to demonstrate *in vivo* interactions, another technique that would help characterise functional variants and determine the precise mechanism of action.

The study described here utilises multiple lines of evidence to demonstrate SNP functionality. Where it is necessary to isolate individual SNPs for analysis, methods such as EMSA and reporter assays are important, but take no account of chromatin accessibility, requiring such information from sources such as ENCODE and Roadmap Epigenomics. FAIRE-gen largely overcomes this weakness, with long-range LD no longer a limiting factor. A potential limitation with the assay described here is the use of allele-specific FAIRE data from lymphoblasts for examining traits seen in the liver. However, although absolute levels of accessible chromatin are tissue-specific, and largely colocalise with tissue-specific transcription factor binding sites [22], evidence from multiple cell lines indicates a correlation of allele-specific effects on accessible chromatin between cell lines [23], providing validity to such analysis. The use of cell lines rather than primary hepatocytes adds one limitation to *in vitro* analysis, and although we replicated DNA-protein interactions with an additional liver cell line, primary hepatocytes may provide a different result.

This study demonstrates the complexity of identification of SNP functionality in regions where strong LD is present, and where several variants in near-complete LD may be having additive/opposite effects. Such information will be important when examining loci in populations with different LD structures, where knowledge of the functional variant will be able to improve the risk prediction for individuals in these populations compared using a lead tag SNP derived from another population. The methodology described here provides an outline to localise potentially causal regulatory variants based on the use of complementary tools.

Reducing the number of functional variants at GWAS loci will allow for the emerging genome-editing techniques such as transcription activator-like effector nucleases [24] and CRISPR/Cas9 [25] to confirm causality, with the ability to examine effects of single variants in cell-based models on gene expression. As the regulome continues to be examined using increasing numbers of cell lines/tissues, and annotations refined with more sophisticated tools, identification of functional elements will significantly improve our ability to predict likely effects of non-coding genetic variants. We have demonstrated here how such annotations can currently help identify potentially functional variants associated with GWAS traits by examining data from primary hepatocytes and HepG2 cells to define regulatory elements in liver cells. The data from these two cell types are likely to cover the majority of regulatory elements, but with further genome-wide analyses, such as ChIP-seq for transcription factors in primary hepatocytes, additional regulatory regions may be discovered at the *ANGPTL3* locus.

### Funding sources

This work was supported by the British Heart Foundation [PG2008/008 to JP and FD, FS/13/6/29977 to AS, SH is a British Heart Foundation chair] and the European Commission [GA No 278397 to PH].

### Disclosures

None.

### Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.atherosclerosis.2015.12.009>.

### References

- [1] T.M. Teslovich, K. Musunuru, A.V. Smith, A.C. Edmondson, I.M. Stylianou, M. Koseki, J.P. Pirruccello, S. Ripatti, D.I. Chasman, C.J. Willer, C.T. Johansen, S.W. Fouchier, A. Isaacs, G.M. Peloso, M. Barbalic, S.L. Ricketts, J.C. Bis, Y.S. Aulchenko, G. Thorleifsson, M.F. Feitosa, J. Chambers, M. Orho-Melander, O. Melander, T. Johnson, X. Li, X. Guo, M. Li, Y. Shin Cho, M. Jin Go, Y. Jin Kim, J.Y. Lee, T. Park, K. Kim, X. Sim, R. Twee-Hee Ong, D.C. Croteau-Chonka, L.A. Lange, J.D. Smith, K. Song, J. Hua Zhao, X. Yuan, J. Luan, C. Lamina, A. Ziegler, W. Zhang, R.Y. Zee, A.F. Wright, J.C. Witteman, J.F. Wilson, G. Willemssen, H.E. Wichmann, J.B. Whitfield, D.M. Waterworth, N.J. Wareham, G. Waeber, P. Vollenweider, B.F. Voight, V. Vitart, A.G. Uitterlinden, M. Uda, J. Tuomilehto, J.R. Thompson, T. Tanaka, I. Surakka, H.M. Stringham, T.D. Spector, N. Soranzo, J.H. Smit, J. Sinisalo, K. Silander, E.J. Sijbrands, A. Scuteri, J. Scott, D. Schlessinger, S. Sanna, V. Salomaa, J. Saharinen, C. Sabatti, A. Ruokonen, I. Rudan, L.M. Rose, R. Roberts, M. Rieder, B.M. Psaty, P.P. Pramstaller, I. Pichler, M. Perola, B.W. Penninx, N.L. Pedersen, C. Pattaro, A.N. Parker, G. Pare, B.A. Oostra, C.J. O'Donnell, M.S. Nieminen, D.A. Nickerson, G.W. Montgomery, T. Meitinger, R. McPherson, M.I. McCarthy, W. McArdle, D. Masson, N.G. Martin, F. Marroni, M. Mangino, P.K. Magnusson, G. Lucas, R. Luben, R.J. Loos, M.L. Lokki, G. Lettre, C. Langenberg, L.J. Launer, E.G. Lakatta, R. Laaksonen, K.O. Kyvik, F. Kronenberg, I.R. Konig, K.T. Khaw, J. Kaprio, L.M. Kaplan, A. Johansson, M.R. Jarvelin, A.C. Janssens, E. Ingelsson, W. Igl, G. Kees Hovingh, J.J. Hottenga, A. Hofman, A.A. Hicks, C. Hengstenberg, I.M. Heid, C. Hayward, A.S. Havulinna, N.D. Hastie, T.B. Harris, T. Haritunians, A.S. Hall, U. Gyllenstein, C. Guiducci, L.C. Groop, E. Gonzalez, C. Gieger, N.B. Freimer, L. Ferrucci, J. Erdmann, P. Elliott, K.G. Ejebe, A. Doring, A.F. Dominiczak, S. Demissie, P. Deloukas, E.J. de Geus, U. de Faire, G. Crawford, F.S. Collins, Y.D. Chen, M.J. Caulfield, H. Campbell, N.P. Burt, L.L. Bonnycastle, D.I. Boomsma, S.M. Boekholdt, R.N. Bergman, I. Barroso, S. Bandinelli, C.M. Ballantyne, T.L. Assimes, T. Quertermous, D. Altshuler, M. Seielstad, T.Y. Wong, E.S. Tai, A.B. Feranil, C.W. Kuzawa, L.S. Adair, H.A. Taylor Jr., I.B. Borecki, S.B. Gabriel, J.G. Wilson, H. Holm, U. Thorsteinsdottir, V. Gudnason, R.M. Krauss, K.L. Mohlke, J.M. Ordovas, P.B. Munroe, J.S. Kooner, A.R. Tall, R.A. Hegele, J.J. Kastelein, E.E. Schadt, J.I. Rotter, E. Boerwinkle, D.P. Strachan, V. Mooser, K. Stefansson, M.P. Reilly, N.J. Samani, H. Schunkert, L.A. Cupples, M.S. Sandhu, P.M. Ridker, D.J. Rader, C.M. van Duijn, L. Peltonen, G.R. Abecasis, M. Boehnke, S. Kathiresan, Biological, clinical and population relevance of 95 loci for blood lipids, *Nature* 466 (2010) 707–713.
- [2] S. Kathiresan, D. Srivastava, Genetics of human cardiovascular disease, *Cell*

- 148 (2012) 1242–1257.
- [3] S. Kathiresan, O. Melander, C. Guiducci, A. Surti, N.P. Burtt, M.J. Rieder, G.M. Cooper, C. Roos, B.F. Voight, A.S. Havulinna, B. Wahlstrand, T. Hedner, D. Corella, E.S. Tai, J.M. Ordovas, G. Berglund, E. Vartiainen, P. Joussilahti, B. Hedblad, M.R. Taskinen, C. Newton-Cheh, V. Salomaa, L. Peltonen, L. Groop, D.M. Altshuler, M. Orho-Melander, Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans, *Nat. Genet.* 40 (2008) 189–197.
- [4] C.J. Willer, S. Sanna, A.U. Jackson, A. Scuteri, L.L. Bonnycastle, R. Clarke, S.C. Heath, N.J. Timpson, S.S. Najjar, H.M. Stringham, J. Strait, W.L. Duren, A. Maschio, F. Busonero, A. Mulas, G. Albai, A.J. Swift, M.A. Morken, N. Narisu, D. Bennett, S. Parish, H. Shen, P. Galan, P. Meneton, S. Herberg, D. Zelenika, W.M. Chen, Y. Li, L.J. Scott, P.A. Scheet, J. Sundvall, R.M. Watanabe, R. Nagaraja, S. Ebrahim, D.A. Lawlor, Y. Ben-Shlomo, G. Davey-Smith, A.R. Shuldiner, R. Collins, R.N. Bergman, M. Uda, J. Tuomilehto, A. Cao, F.S. Collins, E. Lakatta, G.M. Lathrop, M. Boehnke, D. Schlessinger, K.L. Mohlke, G.R. Abecasis, Newly identified loci that influence lipid concentrations and risk of coronary artery disease, *Nat. Genet.* 40 (2008) 161–169.
- [5] E.C. Lee, U. Desai, G. Gololobov, S. Hong, X. Feng, X.C. Yu, J. Gay, N. Wilganowski, C. Gao, L.L. Du, J. Chen, Y. Hu, S. Zhao, L. Kirkpatrick, M. Schneider, B.P. Zambrowicz, G. Landes, D.R. Powell, W.K. Sonnenburg, Identification of a new functional domain in angiotensin-like 3 (angptl3) and angiotensin-like 4 (angptl4) involved in binding and inhibition of lipoprotein lipase (Lpl), *J. Biol. Chem.* 284 (2009) 13735–13745.
- [6] M. Shimamura, M. Matsuda, H. Yasuno, M. Okazaki, K. Fujimoto, K. Kono, T. Shimizugawa, Y. Ando, R. Koishi, T. Kohama, N. Sakai, K. Kotani, R. Komuro, T. Ishida, K. Hirata, S. Yamashita, H. Furukawa, I. Shimomura, Angiotensin-like protein3 regulates plasma hdl cholesterol through suppression of endothelial lipase, *Arterioscler. Thromb. Vasc. Biol.* 27 (2007) 366–372.
- [7] G. Camenisch, M.T. Pisabarro, D. Sherman, J. Kowalski, M. Nagel, P. Hass, M.H. Xie, A. Gurney, S. Bodary, X.H. Liang, K. Clark, M. Beresini, N. Ferrara, H.P. Gerber, Angptl3 stimulates endothelial cell adhesion and migration via integrin alpha vbeta 3 and induces blood vessel formation in vivo, *J. Biol. Chem.* 277 (2002) 17281–17290.
- [8] K. Musunuru, J.P. Pirruccello, R. Do, G.M. Peloso, C. Guiducci, C. Sougnez, K.V. Garimella, S. Fisher, J. Abreu, A.J. Barry, T. Fennell, E. Banks, L. Ambrogio, K. Cibulskis, A. Kerynitsky, E. Gonzalez, N. Rudzicz, J.C. Engert, M.A. DePristo, M.J. Daly, J.C. Cohen, H.H. Hobbs, D. Altshuler, G. Schonfeld, S.B. Gabriel, P. Yue, S. Kathiresan, Exome sequencing, angptl3 mutations, and familial combined hypolipidemia, *N. Engl. J. Med.* 363 (2010) 2220–2227.
- [9] G.R. Abecasis, D. Altshuler, A. Auton, L.D. Brooks, R.M. Durbin, R.A. Gibbs, M.E. Hurles, G.A. McVean, A map of human genome variation from population-scale sequencing, *Nature* 467 (2010) 1061–1073.
- [10] R.C. Hardison, J. Taylor, Genomic approaches towards finding cis-regulatory modules in animals, *Nat. Rev. Genet.* 13 (2012) 469–483.
- [11] D.X. Pang, A.J. Smith, S.E. Humphries, Functional analysis of tcf7l2 genetic variants associated with type 2 diabetes, *Nutr. Metab. Cardiovasc. Dis.* 23 (6) (2012) 550–556.
- [12] A.J. Smith, J. Palmen, W. Putt, P.J. Talmud, S.E. Humphries, F. Drenos, Application of statistical and functional methodologies for the investigation of genetic determinants of coronary heart disease biomarkers: lipoprotein lipase genotype and plasma triglycerides as an exemplar, *Hum. Mol. Genet.* 19 (2010) 3936–3947.
- [13] C. Roadmap Epigenomics, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M.J. Ziller, V. Amin, J.W. Whitaker, M.C. Schultz, L.D. Ward, A. Sarkar, G. Quon, R.S. Sandstrom, M.L. Eaton, Y.C. Wu, A.R. Pfening, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R.A. Harris, N. Shores, C.B. Epstein, E. Gjoneska, D. Leung, W. Xie, R.D. Hawkins, R. Lister, C. Hong, P. Gascard, A.J. Mungall, R. Moore, E. Chuah, A. Tam, T.K. Canfield, R.S. Hansen, R. Kaul, P.J. Sabo, M.S. Bansal, A. Carles, J.R. Dixon, K.H. Farh, S. Feizi, R. Karlic, A.R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T.R. Mercer, S.J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R.C. Sallari, K.T. Siebenthal, A.N. Sinnott-Armstrong, M. Stevens, R.E. Thurman, J. Wu, B. Zhang, X. Zhou, A.E. Beaudet, L.A. Boyer, P.L. De Jager, P.J. Farnham, S.J. Fisher, D. Haussler, S.J. Jones, W. Li, M.A. Marra, M.T. McManus, S. Sunyaev, J.A. Thomson, T.D. Tlsty, L.H. Tsai, W. Wang, R.A. Waterland, M.Q. Zhang, L.H. Chadwick, B.E. Bernstein, J.F. Costello, J.R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J.A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes, *Nature* 518 (2015) 317–330.
- [14] I. Dunham, A. Kundaje, S.F. Aldred, P.J. Collins, C.A. Davis, F. Doyle, C.B. Epstein, S. Fretz, J. Harrow, R. Kaul, J. Khatun, B.R. Lajoie, S.G. Landt, B.K. Lee, F. Pauli, K.R. Rosenbloom, P. Sabo, A. Safi, A. Sanyal, N. Shores, J.M. Simon, L. Song, N.D. Trinklein, R.C. Altshuler, E. Birney, J.B. Brown, C. Cheng, S. Djebali, X. Dong, J. Ernst, T.S. Furey, M. Gerstein, B. Giardine, M. Greven, R.C. Hardison, R.S. Harris, J. Herrero, M.M. Hoffman, S. Iyer, M. Kellis, P. Kheradpour, T. Lassmann, Q. Li, X. Lin, G.K. Marinov, A. Merkel, A. Mortazavi, S.C. Parker, T.E. Reddy, J. Rozowsky, F. Schlesinger, R.E. Thurman, J. Wang, L.D. Ward, T.W. Whitfield, S.P. Wilder, W. Wu, H.S. Xi, K.Y. Yip, J. Zhuang, B.E. Bernstein, E.D. Green, C. Gunter, M. Snyder, M.J. Pazin, R.F. Lowdon, L.A. Dillon, L.B. Adams, C.J. Kelly, J. Zhang, J.R. Wexler, P.J. Good, E.A. Feingold, G.E. Crawford, J. Dekker, L. Elinitski, P.J. Farnham, M.C. Giddings, T.R. Gingeras, R. Guigo, T.J. Hubbard, M. Kellis, W.J. Kent, J.D. Lieb, E.H. Margulies, R.M. Myers, J.A. Stamatoyannopoulos, S.A. Tennebaum, Z. Weng, K.P. White, B. Wold, Y. Yu, J. Wrabel, B.A. Risk, H.P. Gunawardena, H.C. Kuiper, C.W. Maier, L. Xie, X. Chen, T.S. Mikkelsen, S. Gillespie, A. Goren, O. Ram, X. Zhang, L. Wang, R. Issner, M.J. Coyne, T. Durham, M. Ku, T. Truong, M.L. Eaton, A. Dobin, A. Tanzer, J. Lagarde, W. Lin, C. Xue, B.A. Williams, C. Zaleski, M. Roder, F. Kokocinski, R.F. Abdelhamid, T. Alioto, I. Antoshechkin, M.T. Baer, P. Batut, I. Bell, K. Bell, S. Chakraborty, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttgupta, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M.J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, C. Howard, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, G. Li, O.J. Luo, E. Park, J.B. Preall, K. Presaud, P. Ribeca, D. Robyr, X. Ruan, M. Sammeth, K.S. Sandu, L. Schaeffer, L.H. See, A. Shahab, J. Skancke, A.M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, Y. Hayashizaki, A. Raymond, S.E. Antonarakis, G.J. Hannon, Y. Ruan, P. Carninci, C.A. Sloan, K. Learned, V.S. Malladi, M.C. Wong, G.P. Barber, M.S. Cline, N.R. Dreszer, S.G. Heitner, D. Karolchik, V.M. Kirkup, L.R. Meyer, J.C. Long, M. Maddren, B.J. Raney, L.L. Grassef, P.G. Giresi, A. Battenhouse, N.C. Sheffield, K.A. Showers, D. London, A.A. Bhinge, C. Shestak, M.R. Schaner, S.K. Kim, Z.Z. Zhang, P.A. Mieczkowski, J.O. Mieczkowska, Z. Liu, R.M. McDaniell, Y. Ni, N.U. Rashid, M.J. Kim, S. Adar, Z. Zhang, T. Wang, D. Winter, D. Keefe, V.R. Iyer, K.S. Sandhu, M. Zheng, P. Wang, J. Gertz, J. Vielmetter, E.C. Partridge, K.E. Varley, C. Gasper, A. Bansal, S. Pepke, P. Jain, H. Amrhein, K.M. Bowling, M. Anaya, M.K. Cross, M.A. Muratet, K.M. Newberry, K. McCue, A.S. Nesmith, K.I. Fisher-Aylor, B. Pusey, G. DeSalvo, S.L. Parker, S. Balasubramanian, N.S. Davis, S.K. Meadows, T. Eggleston, J.S. Newberry, S.E. Levy, D.M. Absher, W.H. Wong, M.J. Blow, A. Visel, L.A. Pennachio, L. Elnitski, H.M. Petrykowska, A. Abyzov, B. Aken, D. Barrell, G. Barson, A. Berry, A. Bignelli, V. Boychenko, G. Bussotti, C. Davidson, G. Despacio-Reyes, M. Diekhans, I. Ezkurdia, A. Frankish, J. Gilbert, J.M. Gonzalez, E. Griffiths, R. Harte, D.A. Hendrix, T. Hunt, I. Jungreis, M. Kay, E. Khurana, J. Leng, M.F. Lin, J. Loveland, Z. Lu, D. Mantharavadi, M. Mariotti, J. Mudge, G. Mukherjee, C. Notredame, B. Pei, J.M. Rodriguez, G. Saunders, A. Sboner, S. Searle, C. Sisu, C. Snow, C. Steward, E. Tapanari, M.L. Tress, M.J. van Baren, S. Washietl, L. Wilming, A. Zedissa, Z. Zhengdong, M. Brent, D. Haussler, A. Valencia, A. Raymond, N. Addelman, R.P. Alexander, R.K. Auerbach, K. Bettinger, N. Bhardwaj, A.P. Boyle, A.R. Cao, P. Cayting, A. Charos, Y. Cheng, C. Eastman, G. Euskirchen, J.D. Fleming, F. Grubert, L. Habegger, M. Hariharan, A. Harman, S. Iyengar, V.X. Jin, K.J. Karczewski, M. Kasowski, P. Lacroix, H. Lam, N. Larnar-Vincent, J. Lian, M. Lindahl-Allen, R. Min, B. Miotto, H. Monahan, Z. Moqtaderi, X.J. Mu, H. O'Geen, Z. OuYang, D. Patacsil, D. Raha, L. Ramirez, B. Reed, M. Shi, T. Slifer, H. Witt, L. Wu, X. Xu, K.K. Yan, X. Yang, K. Struhl, S.M. Weissman, S.A. Tenebaum, L.O. Penalva, S. Karmakar, R.R. Bhanu, A. Choudhury, M. Domanus, L. Ma, J. Moran, A. Victorsen, T. Auer, L. Centarin, M. Eichenlaub, F. Gruhl, S. Heerman, B. Hoekendorf, D. Inoue, T. Kellner, S. Kirchmaier, C. Mueller, R. Reinhardt, L. Schertel, S. Schneider, R. Sinn, B. Wittbrodt, J. Wittbrodt, G. Jain, G. Balasubramanian, D.L. Bates, R. Byron, T.K. Canfield, M.J. Diegel, D. Dunn, A.K. Ebersold, T. Frum, K. Garg, E. Gist, R.S. Hansen, L. Boatman, E. Haugen, R. Humbert, A.K. Johnson, E.M. Johnson, T.M. Kutayavin, K. Lee, D. Lotakis, M.T. Maurano, S.J. Neph, F.V. Neri, E.D. Nguyen, H. Qu, A.P. Reynolds, V. Roach, E. Rynes, M.E. Sanchez, R.S. Sandstrom, A.O. Shafer, A.B. Stergachis, S. Thomas, B. Vernot, J. Vierstra, S. Vong, M.A. Weaver, Y. Yan, M. Zhang, J.A. Akey, M. Bender, M.O. Dorschner, M. Groudine, M.J. MacCoss, P. Navas, G. Stamatoyannopoulos, J.A. Stamatoyannopoulos, K. Beal, A. Brazma, P. Flicek, N. Johnson, M. Lusk, N.M. Luscombe, D. Sobral, J.M. Vaquerizas, S. Batzoglou, A. Sidow, N. Hussami, S. Kyriazopoulou-Panagiotopoulou, M.W. Libbrecht, M.A. Schaub, W. Miller, P.J. Bickel, B. Banfai, N.P. Boley, H. Huang, J.J. Li, W.S. Noble, J.A. Bilmes, O.J. Buske, A.O. Sahu, P.V. Kharchenko, P.J. Park, D. Baker, J. Taylor, L. Lochovsky, An integrated encyclopedia of DNA elements in the human genome, *Nature* 489 (2012) 57–74.
- [15] E.E. Schadt, S. Woo, K. Hao, Bayesian method to predict individual snp genotypes from gene expression data, *Nat. Genet.* 44 (2012) 603.
- [16] C. Giambartolomei, D. Vukcevic, E.E. Schadt, A.D. Hingorani, C. Wallace, V. Plagnol, Bayesian test for colocalisation between pairs of genetic association studies using summary statistics, *ArXiv e-prints* (2013) 1305.
- [17] L.D. Ward, M. Kellis, Haploreg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants, *Nucleic Acids Res.* 40 (2012) D930–D934.
- [18] K. Cartharius, K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein, T. Werner, MatInspector and beyond: promoter analysis based on transcription factor binding sites, *Bioinformatics* 21 (2005) 2933–2942.
- [19] A.J. Smith, P. Howard, S. Shah, P. Eriksson, S. Stender, C. Giambartolomei, L. Folkersen, A. Tybjaerg-Hansen, M. Kumari, J. Palmen, A.D. Hingorani, P.J. Talmud, S.E. Humphries, Use of allele-specific fairs to determine functional regulatory polymorphism using large-scale genotyping arrays, *PLoS Genet.* 8 (2012) e1002908.
- [20] A.J. Smith, S.E. Humphries, Characterization of DNA-binding proteins using multiplexed competitor emsa, *J. Mol. Biol.* 385 (2009) 714–717.
- [21] B. Lu, A. Moser, J.K. Shigenaga, C. Grunfeld, K.R. Feingold, The acute phase response stimulates the expression of angiotensin like protein 4, *Biochem. Biophys. Res. Commun.* 391 (2010) 1737–1741.
- [22] T.E. Reddy, J. Gertz, F. Pauli, K.S. Kucera, K.E. Varley, K.M. Newberry, G.K. Marinov, A. Mortazavi, B.A. Williams, L. Song, G.E. Crawford, B. Wold, H.F. Willard, R.M. Myers, Effects of sequence variation on differential allelic transcription factor occupancy and gene expression, *Genome Res.* 22 (2012) 860–869.
- [23] J.F. Degner, A.A. Pai, R. Pique-Regi, J.B. Veyrieras, D.J. Gaffney, J.K. Pickrell, S. De



- Leon, K. Michelini, N. Lewellen, G.E. Crawford, M. Stephens, Y. Gilad, J.K. Pritchard, Dnase i sensitivity qtls are a major determinant of human expression variation, *Nature* 482 (2012) 390–394.
- [24] Q. Ding, Y.K. Lee, E.A. Schaefer, D.T. Peters, A. Veres, K. Kim, N. Kuperwasser, D.L. Motola, T.B. Meissner, W.T. Hendriks, M. Trevisan, R.M. Gupta, A. Moisan, E. Banks, M. Friesen, R.T. Schinzel, F. Xia, A. Tang, Y. Xia, E. Figueroa, A. Wann, T. Ahfeldt, L. Daheron, F. Zhang, L.L. Rubin, L.F. Peng, R.T. Chung, K. Musunuru, C.A. Cowan, A talen genome-editing system for generating human stem cell-based disease models, *Cell Stem Cell* 12 (2013) 238–251.
- [25] F.A. Ran, P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, F. Zhang, Genome engineering using the crispr-cas9 system, *Nat. Protoc.* 8 (2013) 2281–2308.