Human Molecular Genetics, 2012, Vol. 21, No. 2 **322–333**
doi:10.1093/hmg/ddr468
Advance Access published on October 11, 2011

Long-range DNA looping and gene expression analyses identify *DEXI* as an autoimmune disease candidate gene

Lucy J. Davison[1,*], Chris Wallace[1], Jason D. Cooper[1], Nathan F. Cope[3], Nicola K. Wilson[2], Deborah J. Smyth[1], Joanna M.M. Howson[1], Nada Saleh[1], Abdullah Al-Jeffery[1], Karen L. Angus[1], Helen E. Stevens[1], Sarah Nutland[1], Simon Duley[1], Richard M.R. Coulson[1], Neil M. Walker[1], Oliver S. Burren[1], Catherine M. Rice[4], Francois Cambien[5], Tanja Zeller[6], Thomas Munzel[7], Karl Lackner[8], Stefan Blankenberg[6], The Cardiogenics Consortium[†], Peter Fraser[3], Berthold Gottgens[2] and John A. Todd[1]

[1]Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics and [2]Haematopoetic Stem Cell Lab, Cambridge Institute for Medical Research (CIMR), NIHR Biomedical Research Centre, University of Cambridge, Cambridge, UK, [3]Nuclear Dynamics Laboratory, Babraham Institute, Cambridge, UK, [4]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK, [5]INSERM UMRS 937, Pierre and Marie Curie University and Medical School, Paris, France, [6]University Heart Center Hamburg, Clinical for General and Interventional Cardiology, 20246 Hamburg, Germany, [7]Medizinische Klinik und Poliklinik, Johannes-Gutenberg Universität Mainz, Germany and [8]Department of Clinical Chemistry and Laboratory Medicine, Johannes-Gutenberg Universität Mainz, Germany

Received July 18, 2011; Revised September 27, 2011; Accepted October 6, 2011

The chromosome 16p13 region has been associated with several autoimmune diseases, including type 1 diabetes (T1D) and multiple sclerosis (MS). *CLEC16A* has been reported as the most likely candidate gene in the region, since it contains the most disease-associated single-nucleotide polymorphisms (SNPs), as well as an imunoreceptor tyrosine-based activation motif. However, here we report that intron 19 of *CLEC16A*, containing the most autoimmune disease-associated SNPs, appears to behave as a regulatory sequence, affecting the expression of a neighbouring gene, *DEXI*. The *CLEC16A* alleles that are protective from T1D and MS are associated with increased expression of *DEXI*, and no other genes in the region, in two independent monocyte gene expression data sets. Critically, using chromosome conformation capture (3C), we identified physical proximity between the *DEXI* promoter region and intron 19 of *CLEC16A*, separated by a loop of >150 kb. In reciprocal experiments, a 20 kb fragment of intron 19 of *CLEC16A*, containing SNPs associated with T1D and MS, as well as with *DEXI* expression, interacted with the promotor region of *DEXI* but not with candidate DNA fragments containing other potential causal genes in the region, including *CLEC16A*. Intron 19 of *CLEC16A* is highly enriched for transcription-factor-binding events and markers associated with enhancer activity. Taken together, these data indicate that although the causal variants in the 16p13 region lie within *CLEC16A*, *DEXI* is an unappreciated autoimmune disease candidate gene, and illustrate the power of the 3C approach in progressing from genome-wide association studies results to candidate causal genes.

[*]To whom correspondence should be addressed at: JDRF/WT Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, Hills Road, Cambridge CB2 0XY, UK. Tel: +44 1223762104; Fax: +44 1223762102; Email: lucy.davison@cimr.cam.ac.uk
[†]Full list of Cardiogenics contributors is given in Supplementary Material.

INTRODUCTION

The chromosome 16p13 region of the human genome has gained increasing attention since it was first associated with risk of type 1 diabetes (T1D) by genome-wide association study (GWAS) and fine mapping in 2007 (1–3). Subsequent association and candidate gene studies, in other autoimmune diseases such as multiple sclerosis (MS) (4–6), Addison's disease (7), primary biliary cirrhosis (8) and systemic lupus erythematosus (SLE) (9,10), have also demonstrated association of this region with disease risk, implying that the 16p13 region contains a key regulator of the self-reactive immune response.

The 16p13 region is dominated by the large, 238 kb CLEC16A gene (previously known as *KIAA0350*). The most highly disease-associated single-nucleotide polymorphisms (SNPs) lie predominantly within the 60 kb intron 19 of *CLEC16A* (1,2,4,5,10,11), as well as within intron 10 of the same gene. The most commonly cited disease-associated SNPs within these respective introns are rs12708716 and rs8062322, which are in high linkage disequilibrium (LD) and are, therefore, likely to be tagging the same signal. There is also evidence of a second T1D association signal in the 16p13 region (3,12), 3′ of *CLEC16A*, within the C16orf75 gene. This second disease-association signal was detected in the absence of the intron 19/intron 10 signal in celiac disease (13,14), and is likely to be related to an independent causal variant common to both celiac disease and T1D. The complexity of the region is also highlighted by a recent report that the 16p13 region harbours three independent MS-associated loci (6).

CLEC16A is regarded as a potential causal gene, since it contains an imunoreceptor tyrosine-based activation motif (ITAM) (3) and is widely expressed in cells of the immune system; we note that the C-lectin-binding function implied by its name remains questionable since it is only 22 amino acids long. A recent study suggested that MS-associated SNPs in *CLEC16A* were correlated with relative expression of two *CLEC16A* isoforms in thymus but not in peripheral blood (15). However, there are several genes in the 16p13 region, in addition to *CLEC16A*. These include some attractive candidate genes for T1D and/or MS risk, such as MHC class II transactivator (*CIITA*) and suppressor of cytokine signalling 1 (*SOCS1*) (Fig. 1), as well as *DEXI*, a gene of unknown function, which has not previously been regarded as a strong autoimmune candidate gene.

Identification of the causal variant(s) and gene(s) presents a particular challenge when there is a high degree of LD in a disease-associated region and also when there are multiple genes in a region, some of which have an unknown function. If the expression of a particular gene is correlated with disease-associated SNPs, this strengthens the candidacy of that gene for an active role in the pathogenesis of disease. We used published and unpublished human monocyte gene expression data to examine expression quantitative trait loci (eQTLs) within the 16p13 region of autoimmune disease association. Interrogation of the recently published ChIP-Seq ENCODE data sets was undertaken to identify enhancer, transcription-factor-binding and RNA-polymerase-II-binding marks in the 16p13.13 region, which were consistent with the role of intron 19 of *CLEC16A* as a regulatory sequence. Gene expression analysis was followed by chromosome conformation capture (3C) experiments to evaluate candidate long-range DNA interactions in the 16p13 region. This indicated the mechanism by which disease-associated SNPs within one gene, *CLEC16A*, might influence the expression of the neighbouring gene, *DEXI*.

RESULTS

eQTL analysis in human monocytes

The two independent data sets interrogated were generated previously from normal human monocytes purified from fresh blood samples from 1370 individuals [Gutenberg Health Study (GHS) (16)] and 753 individuals [Cardiogenics Project (CGP) (17), http://www.cardiogenics.eu/web/], respectively, and subjected to genome-wide genotyping and microarray gene expression analysis. In both data sets, we found evidence of a single eQTL in the region. Expression of the DEXI gene was correlated with several chromosome 16p13 SNPs in high LD within *CLEC16A*, including rs12708716 (Fig. 1) and rs8062322. Since the majority of the *DEXI* eQTL SNPs in intron 10 and intron 19 of the CLEC16A gene are in high LD, a single SNP most correlated with *DEXI* expression could not confidently be determined. The *CLEC16A* alleles that confer protection from T1D and MS were only associated with increased expression of *DEXI* ($P = 3.8 \times 10^{-38}$ in GHS data set; $P = 1.8 \times 10^{-7}$ in CGP data set), and not with expression of any other genes in the region, in both monocyte expression data sets.

Importantly, there was no correlation between expression of *DEXI* and SNPs in the proposed second region of T1D association, near *C16orf75* ($P > 0.05$). In addition, we noted specifically that no eQTLs were detected for expression of any other gene within the 16p13 region, including *CLEC16A* and *SOCS1* in either data set (Supplementary Material, Fig. S1). In addition, in contrast to a recent report evaluating gene expression in the 16p13 region in lymphoblastoid cell lines, no correlation between expression of *SOCS1*, *DEXI* and *CLEC16A* was detected (6).

We identified confirmatory evidence supporting an eQTL within intron 19 of *CLEC16A* for *DEXI* expression within the supplementary information of two further, recent independent data sets (18,19). The first was an investigation of eQTLs in Epstein Barr virus (EBV)-transformed lymphoblastoid cell lines (18), and the second, a study of eQTLs in human primary bone cells treated with PGE_2 (19). Both studies correlated *DEXI* expression by microarray (Supplementary Material, Fig. S2), with genotype at eQTL SNPs within intron 19 of *CLEC16A*, and neither study reported eQTLs in the 16p13 region for expression of *CLEC16A*, *SOCS1* or *CIITA*.

Evaluation of candidate long-range DNA interactions by 3C

The *DEXI* promoter region and the SNP most associated with T1D in *CLEC16A* (rs12708716) are separated by a distance of ~160 kb. If one region is to influence the other, this would suggest that a DNA loop is formed during transcription of *DEXI*, allowing the two regions to be in close physical proximity. The 3C technique has been used to examine long-

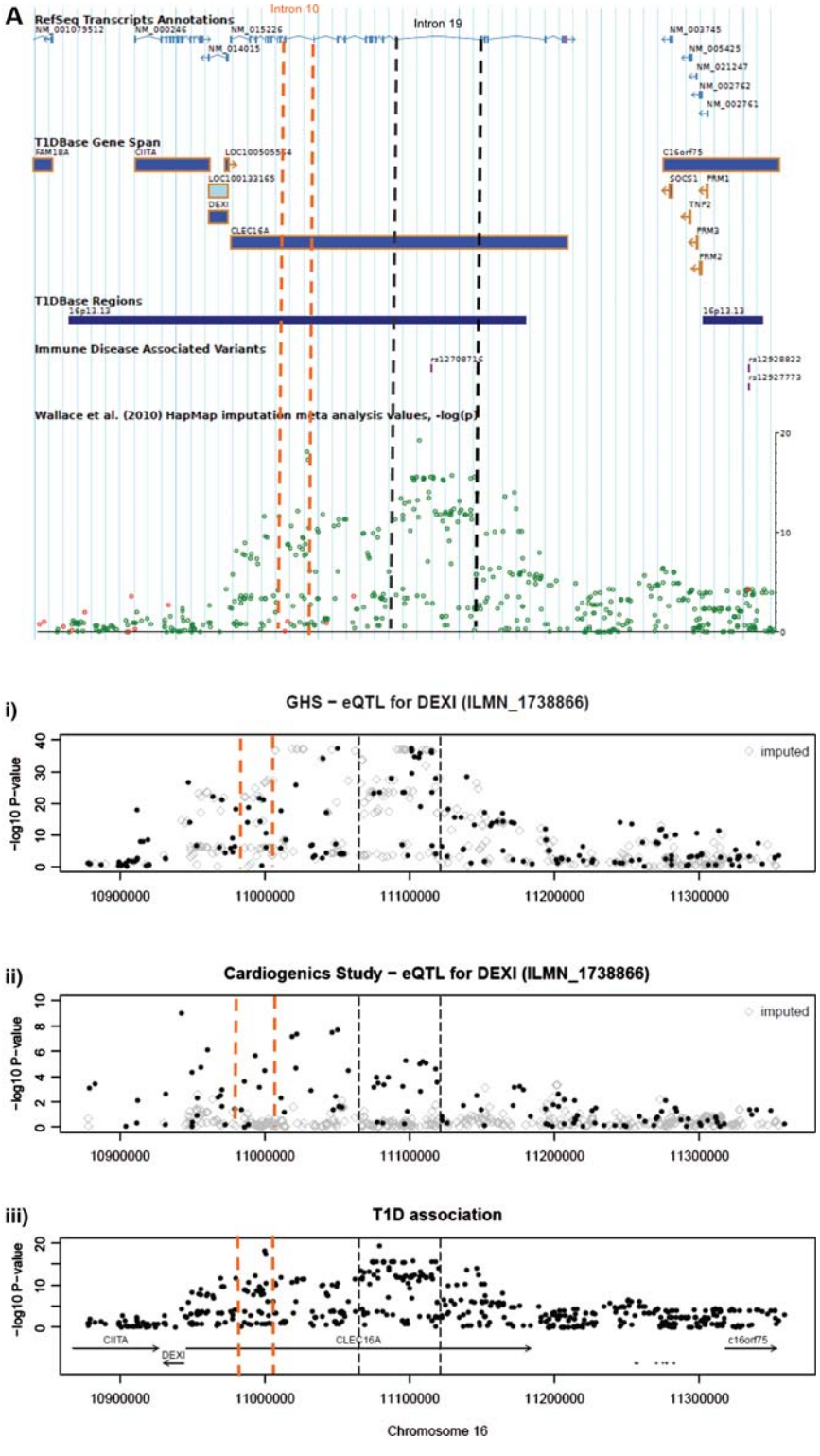

Figure 1. Expression of the DEXI gene, contained within the 16p13 region of T1D association (**A**) by microarray using Illumina probe set ILMN_1738866, is correlated with the genotype at SNPs within the CLEC16A gene in two independent data sets. The first data set (i) was generated from purified human monocytes ($n = 1370$ individuals) collected during GHS and the second (ii) from purified human monocytes ($n = 753$ individuals) collected during the CGP Project. SNPs associated with T1D are illustrated on the same scale (iii) to illustrate the co-localization of the eQTL and T1D association signals. Boundaries of intron 10 and intron 19 of *CLEC16A* contain the most associated SNPs for T1D and are contained within orange and black dotted lines, respectively. (**C**) DEXI expression by genotype at the T1D-associated SNP rs12708716 within the published GHS monocyte data set, where the minor (G) allele is protective, and (**D**) DEXI expression by genotype at the T1D-associated SNP rs725613 in the unpublished CGP monocyte data set.

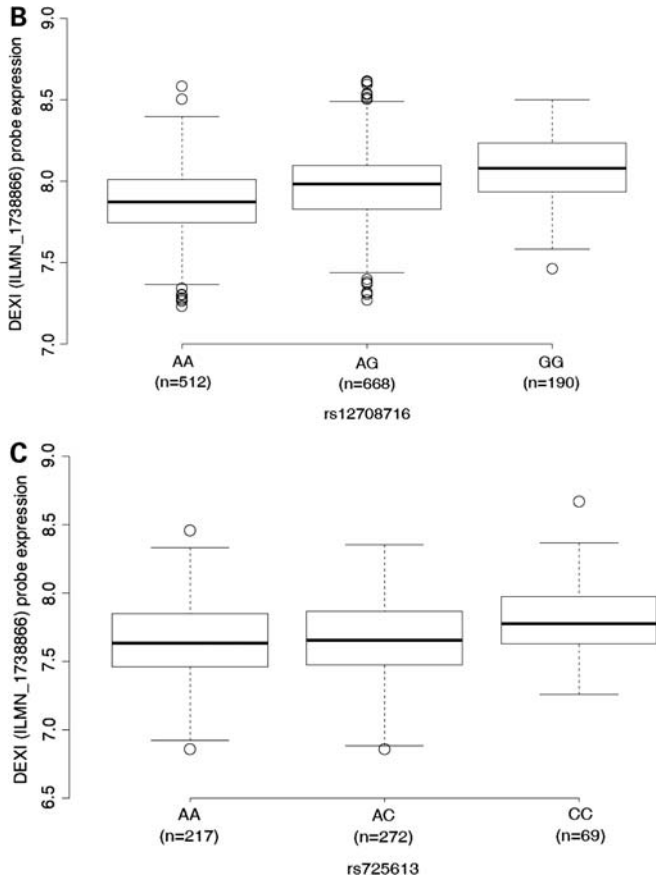

Figure 1. *Continued*

distance chromosomal interactions (20,21) and was employed here to test our hypothesis that the *DEXI* promoter region and intron 19 of *CLEC16A* are in close physical proximity during gene transcription. The *Bgl*II restriction enzyme was chosen for use in 3C since it makes two cuts in the 2 kb region between the promoter regions of *CLEC16A* and *DEXI*, allowing fragments containing these regions to be evaluated separately, and it also cuts seven times throughout intron 19 of *CLEC16A*. Three human cell lines were selected for these experiments—a monocyte cell line (THP-1), since we identified the eQTL originally in monocytes, a lung epithelial cell (A549), since *DEXI* expression was first reported in this cell line (22) and a human EBV-transformed B-cell line, since evidence exists for the *DEXI* eQTL in *CLEC16A* in this cell type in the literature (18).

The 3C technique generates ligated DNA fragments through cross-linking of distal DNA sequences based on their physical proximity in living cells. Using a 'bait' reverse primer next to a *Bgl*II restriction site located in the region of the *DEXI* promoter, quantitative polymerase chain reactions (qPCRs) were initially undertaken using 13 different forward primers in candidate interaction *Bgl*II fragments, throughout the 16p13 region. In all three cell lines, a specific association was detected between a fragment representing the *DEXI* promoter region and a single specific *Bgl*II fragment of ~20 kb within intron 19 of *CLEC16A* (Figs 2 and 3). This finding was confirmed by gel electrophoresis and direct sequencing of the qPCR products generated (data not shown). The additional interaction detected between the *DEXI* promoter fragment and an adjacent fragment containing the *CLEC16A* promoter region is related to the proximity of these two regions in the genome, and reflects random collisions between the chromatin fibres, commonly seen in 3C experiments (23).

The interaction between the *DEXI* promoter region and intron 19 of *CLEC16A* was confirmed using several different primer sets. First, using alternative *DEXI* promoter primers and a more comprehensive set of primer pairs for *Bgl*II sites within intron 19 of *CLEC16A*, the original region of interaction was confirmed as the only interacting fragment within intron 19. Second, reciprocal qPCRs were undertaken using a primer in the interacting intron 19 fragment as 'bait'. As well as confirming the interaction with the *DEXI* promoter, these experiments also demonstrated that this specific region of intron 19 did not interact with any other candidate sites (Supplementary Material, Table S1) within the region. We also conducted further 3C/qPCR analysis of potential interactions involving intron 10 *Bgl*II fragments, to evaluate the possibility of a double loop, involving both intron 10 and intron 19 of *CLEC16A* since disease-associated SNPs have also been reported within intron 10. We used both the *DEXI* promoter region and the interacting region of intron 19 as bait, but found no evidence of further interactions by qPCR (data not shown). Finally, additional qPCRs were undertaken using the interaction region in intron 19 as a 'bait' to test for DNA interactions within intron 19 itself. No further sequencing-confirmed interactions were detected in any of these experiments (data not shown), but the qPCR product representing the *DEXI* promoter region—intron 19 interaction was consistently present.

The challenge of progressing from a disease-associated region identified in GWAS to convincing evidence for which gene(s) are causal candidates in the disease pathogenesis is considerable. This is, in part, due to the fact that a large region of LD can contain many genes, in addition to the strong possibility that the causal variant, lying within the LD region, affects the expression of genes outside the LD block in long-range functional interactions. We note that rs12708716, reported in several studies as the SNP most associated with T1D and MS, and shown here to correlate with *DEXI* expression, lies within the 20 kb fragment that we have shown to interact with the *DEXI* promoter region (Fig. 3).

Evaluation of epigenetic markers associated with enhancer activity and transcription factor binding

Non-coding regions of DNA may impact upon expression of distant genes by acting as enhancers, requiring looping of DNA and interaction of remote regions via protein–protein contacts (24). Although the exact mechanisms by which enhancers or suppressors mediate their effects on gene transcription are unknown, many enhancer regions contain binding sites for transcription factors and are identifiable by the presence of epigenetic histone modifications such as H3K4me1 and H3K27Ac (25,26). In addition, recent evidence suggests that some enhancers in non-coding parts of the genome are transcribed, by RNA polymerase II, into RNA with regulatory functions (27–29). We, therefore, made use of publicly available genome-wide ChIP-Seq data sets from the ENCODE project (30,31) to

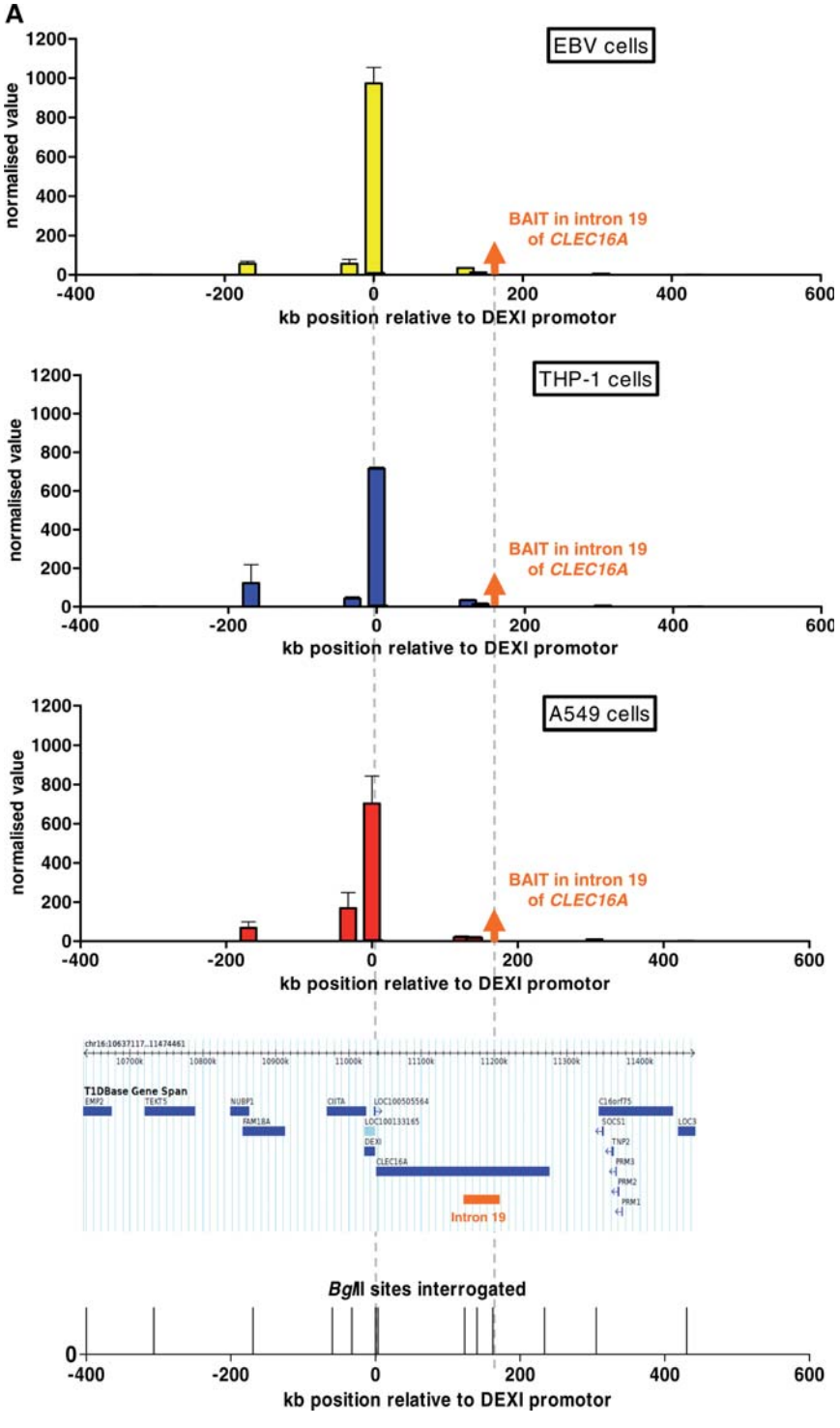

Figure 2. A long-range interaction was detected between the *DEXI* promotor region and intron 19 of *CLEC16A* in an EBV-transformed B cell line (yellow), the monocyte-like line THP-1 (blue) and the lung epithelial cell line A549 (red). Chromatin was cross-linked and digested by *Bgl*II and re-ligated. The interaction frequency between (**A**) the *DEXI* promotor fragment bait or (**B**) the 20 kb intron 19 region bait containing rs12708716 and distal candidate fragments was determined by qPCR and normalized to control template interactions generated using PCR-digested and ligated PCR products from genomic DNA. Error bars represent the standard error of three independent PCR reactions and peaks were confirmed by sequencing of the qPCR products. The two grey dashed lines represent the location of the *DEXI* promotor region and the region of intron 19 shown to interact with the *DEXI* promotor region.

investigate histone modification marks and RNA polymerase binding throughout the 16p13 region and found intron 19 of *CLEC16A* to be particularly enriched for enhancer marks in human cells (Supplementary Material, Fig. S3A). Further evidence for a regulatory function of intron 19 of *CLEC16A* is provided by the observation that its sequence is highly conserved

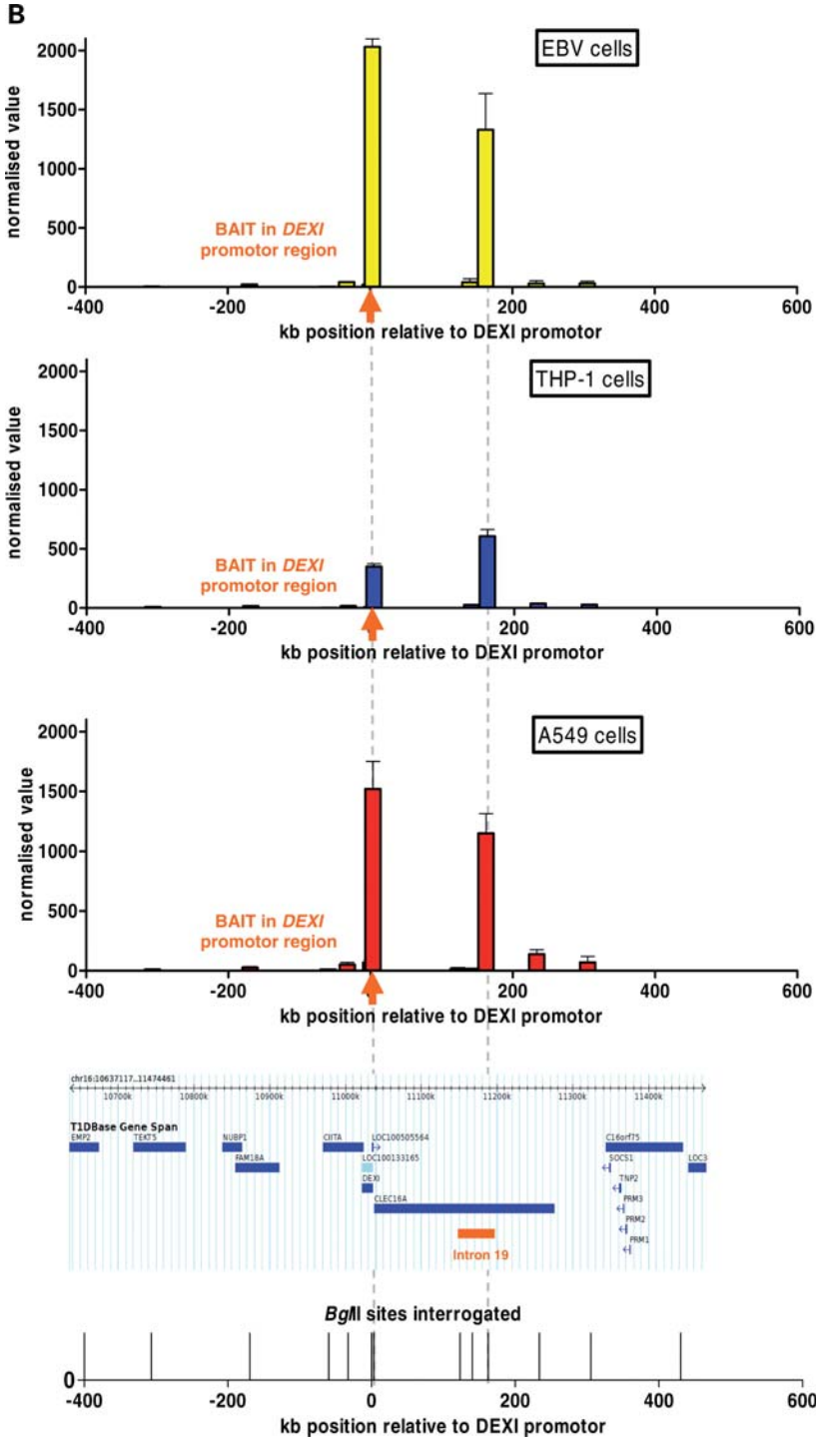

Figure 2. *Continued*

between mouse and humans, consistent with an important regulatory function in both species. It is bound by multiple transcription factors in a murine haematopoietic progenitor cell line (32) (Supplementary Material, Fig. S3B) and human cell lines (N.K.W., unpublished data). Many of these transcription-factor-binding events occur towards the 3′ end of the fragment that we have shown to interact with the *DEXI* promoter region in human cells. We also note that, according to ENCODE ChIP-Seq data from human cell lines, both the interacting region in intron 19 and the *DEXI* promoter region share binding sites for several transcription factors (Fig. 3A and B) including NF-κB, JunD and c-Myc. Furthermore, as might be predicted by recent work (29), clear RNA-polymerase-II-binding peaks are evident within the fragment of intron 19 of *CLEC16A* containing the putative enhancer region (Fig. 3B).

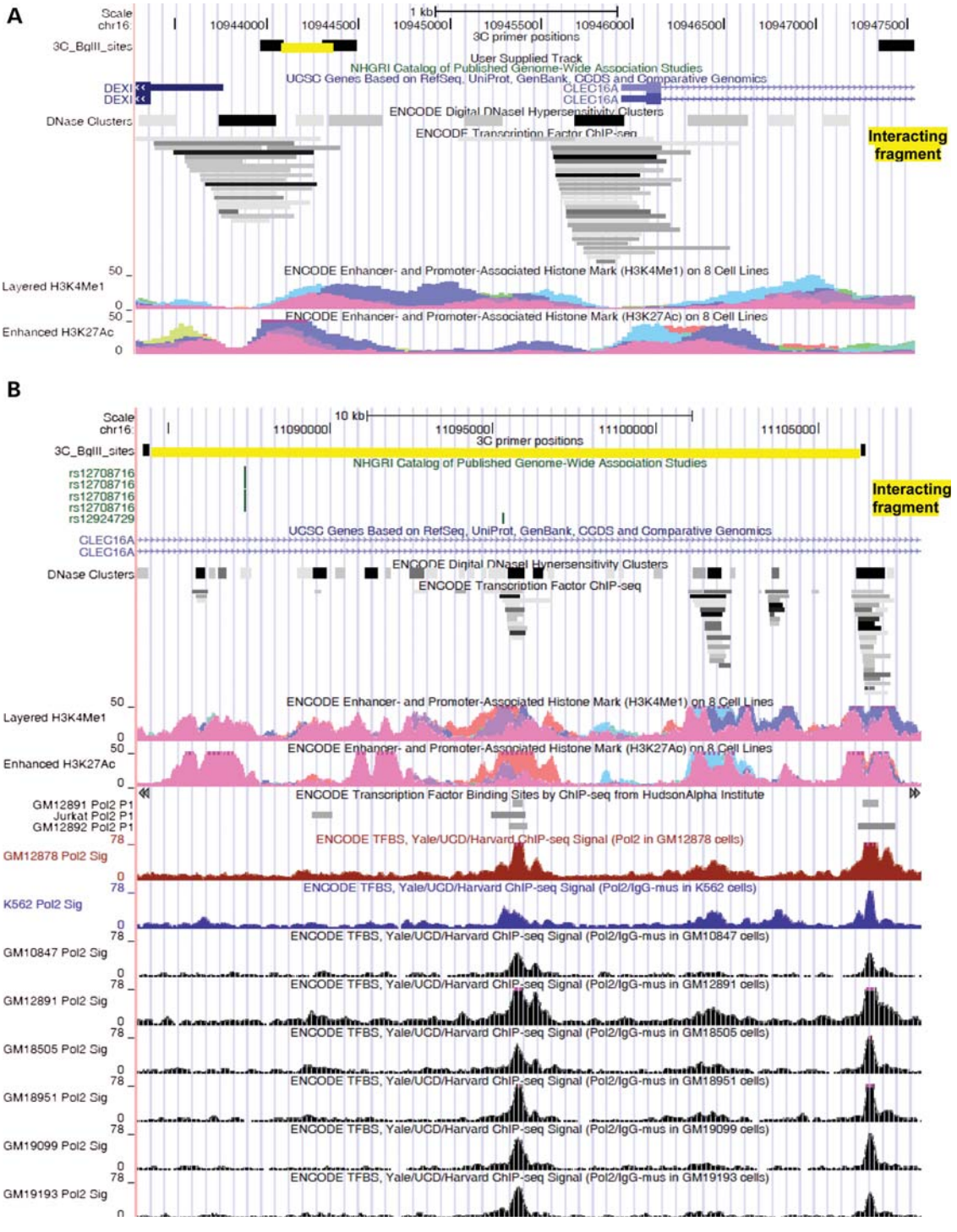

Figure 3. The interaction sites detected by 3C at the *DEXI* promotor region (**A**) and intron 19 of *CLEC16A* (**B**) are enriched for transcription-factor-binding sites (marked by grey and black horizontal lines) and enhancer-associated histone modifications (ENCODE project, displayed using the UCSC Genome browser). Intron 19 of *CLEC16A* also contains RNA polymerase II binding peaks. The 3C interaction fragment, between two *Bgl*II restriction enzyme digestion sites, is highlighted with a yellow line.

The gene expression and 3C data provide strong support for *DEXI*'s candidacy as a new autoimmunity gene. Apart from a pseudogene on chromosome 15, which we have shown to produce little or no RNA (Supplementary Material, Fig. S4), *DEXI* has no clear paralogue in the human genome. Although *CLEC16A* itself appears to be the focus of many investigators studying the 16p13 region, our results imply an immediate additional need to investigate and understand the biology of DEXI and its role in the development of autoimmune disease.

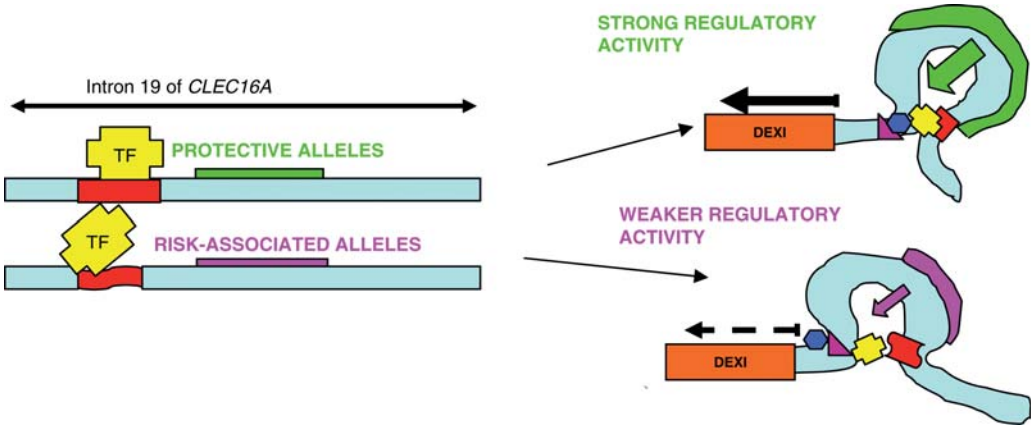

Figure 4. The eQTL and 3C data allow a model to be proposed in which a DNA loop is formed between the *DEXI* promotor region and intron 19 of *CLEC16A*. In this model, the T1D-associated SNPs within the loop affect gene transcription by influencing the binding of transcription factors and other proteins at or near the interacting site in intron 19. A more efficient transcription complex is formed in the presence of protective T1D alleles at 16p13, allowing more efficient *DEXI* transcription. It is also possible that a converse model is true, with protective alleles being associated with reduced binding of suppressor/silencer proteins.

DISCUSSION

This study highlights *DEXI* as an autoimmune candidate gene in 16p13, based on an eQTL for *DEXI* in monocytes that appears to co-localize with the T1D-association signal and demonstration of close physical proximity of regulatory sequences in intron 19 of *CLEC16A* with the *DEXI* promoter region by the 3C technique. *DEXI* was originally identified in 2001 as a transcript which was differentially expressed in lung tissue of patients with emphysema compared with normal lung tissue (22). The function of DEXI is unknown; however, the gene was named dexamethasone-induced transcript since dexamethasone treatment of the A549 cell line, used here for the 3C experiments, was reported to increase *DEXI* mRNA expression using a semi-quantitative blotting method. More recently reported RNA-Seq and microarray results using dexamethasone-treated cell lines do not support this claim (19,33) and, therefore, we suggest that this name may be a misnomer.

Historically, eQTLs have been classified into *cis* and *trans (*or *distant)*, where *cis* eQTLs affect expression of a gene nearby and *trans* act at greater distances (usually on another chromosome) in a more indirect way, such as affecting the expression of a regulator that has a subsequent impact of the *trans*-controlled gene. The findings presented here represent a biomedically important finding at an intermediate distance level, rather than a classical *cis* or *trans* effect. The eQTLs act on the same chromosome, but at a distance of >150 kb, and importantly do not, according to our current results, affect the nearest gene, but the next one along chromosome 16p13. It is likely that this will not be an isolated event in future studies of common disease, hence this finding is very relevant to the wider field of GWAS follow-up. In addition, although the interaction fragment in intron 19 contains several disease-associated SNPs, the question of exactly how allele-specific expression is driven is raised, and how this process is made gene-specific. Since the other genes in the region are expressed, their promoters are accessible, so it is unlikely that specificity is simply related to promoter accessibility. Instead, we hypothesize that specific transcription factor combinations bound to the intron 19 region and *DEXI* promoter are important and the mechanism by which gene expression is controlled might involve transcription of intron 19 and its function as a non-coding regulatory RNA. Recent evidence of a strong genetic component for allele-specific differences at the level of transcription factor binding and chromatin structure has been reported (34). In addition, the recruitment of cohesin and the multi-protein Mediator complex has been associated with chromosome looping and regulation of gene expression via enhancer sequences (35). We postulate that the causal variant(s) in the region affect enhancer activity and that this could arise because of an allele-specific effect on chromatin structure and/or recruitment of a multi-protein transcriptional co-activator complex (Fig. 4).

Our study illustrates that increased expression of *DEXI* in monocytes is associated with SNP alleles that are protective from autoimmune disease. Monocytes are already known to be important in the pathogenesis of T1D, giving rise to macrophages, and the dendritic cells that are central in 'priming' the islets and establishing an inflammatory milieu prior to the destruction of islets by antigen-specific T cells (36). In addition, monocytes, macrophages and microglia have a central role in the central nervous system (CNS) inflammation of MS. During MS attacks, T lymphocytes and monocyte-derived macrophages gain entry to the CNS and form peri-vascular infiltrates, a process which is accompanied by enhanced permeability of the blood–brain barrier (37,38). Although monocytes play a role in disease risk, we note that the *DEXI* transcript is found in higher abundance in $CD4^+$ and natural killer cells compared with monocytes, and additional studies have provided evidence for a *DEXI* eQTL in intron 19 of *CLEC16A* in cells other than monocytes, including primary bone cells and EBV-transformed lymphoblastoid cell lines (18,19). Therefore, it is possible that *DEXI*'s role in autoimmunity is related to genotype-regulated expression in a cell subset other than, or in addition to, monocytes and might also be specific to a particular time in development. We note that *DEXI* expression was increased in macrophages, generated in culture from a subset of the CGP samples, compared with resting monocytes (Supplementary Material,

Fig. S5), but that the *DEXI* eQTL was not preserved in macrophages (data not shown). We postulate that this may be the result of a stricter (genotype-related) control of basal expression in monocytes.

The DEXI gene is conserved across many species including the mouse, rat, dog, elephant, zebrafish and chicken, but is not found in *Caenorhabditis elegans* or *Drosophila* (Supplementary Material, Fig. S6). It is predicted by some, but not all, software packages to contain a trans-membrane domain, a sequence with a repeating leucine motif (Supplementary Material, Fig. S6) and a predicted casein kinase phosphorylation site (22). Expression is most strongly detectable by microarray in liver, brain, heart and lung tissue, as well as in some cells of the immune system (http://biogps.gnf.org/#goto=genereport&id=28955 and Supplementary Material, Fig. S7). Although not originally an obvious T1D candidate gene at 16p13, a previous survey of allelic expression using EST mining lists *DEXI* as one of 40 genes whose expression is in allelic imbalance by *in silico* analysis (39). It is also interesting to note that *DEXI* expression is reported to be affected by exposure to cigarette smoke (16) ($P = 8 \times 10^{-8}$, data not shown), perhaps explaining in part the original report of increased *DEXI* expression seen in the lungs of emphysema patients (22). Additional opportunities for the control of *DEXI* expression within cells exist at the level of microRNAs, since binding sites for microRNAs, including miR-137 and miR-30−5p, are predicted in the 3′UTR of the gene, using TargetScan (40).

To our knowledge, this is one of only a small number of reports of successful application of 3C to the dissection of GWAS SNPs and target genes (41–44). In contrast to previous publications, however, the present study is the first to report GWAS SNP alleles within an intron region appearing to regulate expression of a neighbouring gene. These data provide an obvious rationale to search for long-distance regulatory sequences as an explanation for a proportion of GWAS SNP associations. The experiments reported here were able to identify a long-distance interaction using 3C, based on a hypothesis generated using eQTL data. Further evaluation of the region using more exhaustive interrogation methods for interacting fragments in a variety of tissues may yet yield evidence of further interactions. Taken together, there is now clear justification not only for the evaluation of the mechanism(s) by which DEXI might protect against autoimmune disease, but also for the development and use of 3C techniques involving next-generation sequencing, on a genome-wide scale, such as ChIA-PET (45), HiC (46) and 5C (47) in multiple cell types. Based on the findings presented here, we propose that the small DEXI gene represents the key to a novel and important pathway in the pathogenesis of T1D, MS and other immune-mediated diseases.

MATERIALS AND METHODS

Ethics statement

All research involving human participants has been approved by Fenland and Peterborough Local Research Ethics Committee.

Monocyte purification and preparation for microarray

Gutenberg Health Study. This protocol has been previously described (16,17) but briefly, GHS is a community-based, prospective, observational single-centre cohort study in the Rhein-Main region in Germany. Separation of monocytes was conducted within 60 min of collection of 8 ml of blood in patients recruited as part of this study, and RNA was extracted the same day. Monocytes were separated by negative selection, using the Vacutainer CPT Cell Preparation Tube System (BD, Heidelberg, Germany) with 400 μl Rosette Sep Monocyte Enrichment Cocktail (StemCell Technologies, Vancouver, Canada) added immediately after blood collection. After separation, cells were washed twice in ice-cold phosphate-buffered saline (PBS) buffer containing 2 mM EDTA and the purity in selected samples was confirmed by flow cytometry. Cells were resuspended in 1.5 ml TRIzol reagent (Invitrogen, Karlsruhe, Germany) and RNA extraction was performed within 5 h using chloroform extraction and the RNeasy Mini Kit (Qiagen, Hilden, Germany), including DNAse digestion. Genome-wide expression analysis was performed on monocyte RNA samples using the Illumina HT-12 v3 BeadChip (http://www.Illumina.com), in batches of 96 samples. Genotyping was performed using the Affymetrix (Santa Clara, CA, USA) Genome-Wide Human SNP Array 6.0 and the Genome-Wide Human SNP *Nsp*I/*Sty*I 5.0 Assay Kit.

Cardiogenics project. This protocol has previously been described (17), but briefly, the multi-centre Cardiogenics (CGP) study includes 363 patients with coronary artery disease or myocardial infarction and 395 healthy individuals of European descent (http://www.cardiogenics.eu). Blood samples (30 ml) from fasting subjects were collected into EDTA blood tubes and monocytes were isolated by positive selection, using CD14 microbeads and AutoMACS/AutoMACS Pro (Miltenyi - Bergisch Gladbach, Germany) according to the manufacturer's instructions. Monocyte purity was measured as the percentage of CD14+ve cells analysed by flow cytometry. Isolated monocytes were lysed in TRIzol reagent and RNA was extracted by a method similar to that described above for GHS. Whole-genome genotyping was carried out at the Wellcome Trust Sanger Institute, using two arrays, the Sentrix Human Custom 1.2M array and the Human 610 Quad Custom array (Illumina). Gene expression profiling was performed using Human Ref-8 Sentrix Bead Chip arrays (Illumina).

eQTL analysis

Normalized expression probes in 16p13 were tested for association in the region, assuming an additive model, using the 'R' software package.

3C and qPCR

The method for 3C has been previously reported in detail (21) by co-authors N.F.C. and P.F., so is described briefly here. The lung epithelial cell line A549 and the monocyte-like cell line THP-1 were obtained from ATCC (http://www.lgcstandards-atcc.

org/ATCCCulturesandProducts/CellBiology/) and grown according to standard tissue culture protocols, in addition to an EBV-transformed B-cell line from the *CEPH* collection obtained from ECACC (http://www.hpacultures.org.uk/collections/ecacc.jsp).

Fifty million cells at ~80% confluence from each cell line were harvested and washed in ice-cold PBS before being passed through a cell strainer. Following a second PBS wash, cells were re-suspended in 45 ml high-glucose DMEM medium + L-glutamine (Invitrogen) and 10% fetal bovine serum. Cross-linking was performed by adding 2.7 ml of 37% formaldehyde (Merck) for 5 min at room temperature with rocking. The reaction was quenched with 3 ml of 2 M glycine, followed by washing with 50 ml ice-cold PBS. The pellet for each cell line was re-suspended in 50 ml of permeabilization buffer [10 nM Tris–HCl, pH8, 10 mM NaCl, 0.2% igepal and one complete protease inhibitor tablet, EDTA-free (Roche Diagnostics, Penzberg, Germany)] and incubated at 4°C on ice with mixing. Following a centrifugation step, and manual cell count, each 1×10^7 nuclei from each cell line were resuspended in 100 μl of NEB Buffer 3 (New England BioLabs, Ipswich, MA, USA). One 1×10^7 nuclear pellet from each cell line was treated with 7.5 μl of 20% SDS for 1 h at 37°C, 950 r.p.m. on a Thermomixer (Eppendorf, Histon, UK), to remove any non-cross-linked proteins from the DNA. This was followed by the addition of 50 μl of 20% Triton-X100 (Sigma-Aldrich) for 1 h at 37°C, 950 r.p.m., to sequester SDS and allow subsequent digestion. *Bgl*II enzyme (1500 IU) (NEB) was added to each tube, followed by incubation overnight at 37°C and 950 r.p.m.

A 5 μl of aliquot of digested DNA was assessed for completeness of digestion using gel electrophoresis. The *Bgl*II enzyme was inactivated by the addition of 40 μl of 20% SDS for 25 min at 65°C at 950 r.p.m. The incubated samples were added to 7 ml of 1.1× ligation buffer (NEB) and 375 μl of Triton-X100 for 1 h at 37°C, with mixing. The cross-linked digested DNA was re-ligated by the addition of 800 IU NEB T4 DNA ligase for 4 h at 16°C and then 30 min at room temperature. Then 900 μg of Proteinase K (Roche Diagnostics) was added followed by an overnight incubation at 65°C.

The samples were cooled to room temperature and 300 μg of RNAse A (Sigma) added for 1 h at 37°C. This was followed by phenol–chloroform extraction and ethanol–acetate precipitation of DNA, which was quantified by PicoGreen assay.

3C interaction products were detected by PCR using candidate primer pairs with and Qiagen HotStar Taq Polymerase and 250 ng of DNA per reaction, followed by agarose gel electrophoresis. Quantification of interaction products was undertaken using qPCR and 2× SYBR Green Mastermix (Applied BioSystems, Final 1×) and candidate forward and reverse primers at a final concentration of 400 nM. Samples were tested in a 96-well format in triplicate, using the ABI 7000 sequence detection system. Quantification was achieved using serial dilutions of a 3C-positive control template on each plate. This was generated by synthesis of all possible PCR products using the available primers (Supplementary Material, Table S1), followed by gel extraction and purification. PCR products underwent *Bgl*II digestion before being mixed in equimolar concentrations and ligated with T4 ligase (NEB) to generate a pool of potential interaction products, which was purified by phenol–chloroform extraction and ethanol precipitation. The control template was mixed with genomic DNA that had undergone digestion and random ligation so that PCR efficiency was not affected by the total amount of DNA present (similar to the 300 ng/reaction for the real 3C samples) (16). Quantification of PCR products was achieved by comparison with the standard curve on each plate and in addition, values were normalized for each experiment using the result obtained from the most 3′ interacting primer and the bait within each cell line.

Sequencing of qPCR products at candidate interaction sites was undertaken following gel purification and PCR of purified products with the original 3C primers and the BigDye Terminator v3.1 sequencing kit. Samples were analysed in triplicate from each cell line, using the 3730xl DNA Analyzer (Applied BioSystems) and sequencing traces were visually checked using the Peak Picker™ software.

AUTHORS' CONTRIBUTIONS

Access to GHS and CGP data was provided by S.B. and F.C., with the written permission of their respective data management groups. eQTL and co-localization analysis was developed and performed by C.W. and J.D.C.

L.J.D., C.W. and J.A.T. participated in the conception and overall design of the study. The manuscript was prepared by L.J.D., J.A.T. and J.D.C. with helpful comments from N.F.C., P.F., B.G., F.C., S.B., J.M.M.H., D.J.S., N.K.W. and P.F. Supplementary Material, Figure S4b was prepared by N.K.W.

3C experiments were designed by L.J.D., J.A.T., N.F.C., P.F., N.K.W. and B.G. and performed by L.J.D., with the guidance of N.F.C. and assistance with tissue culture from S.D.

DNA collection, preparation and curation were undertaken by S.N., D.J.S., H.E.S. with database management by N.M.W. and statistical assistance by J.M.M.H.

S.B., T.Z., T.M. and K.L. participated in the Gutenberg Health Study.

F.C. and C.M.R. participated in the Cardiogenics Study.

Bioinformatic and T1DBase support was provided by O.S.B. and R.M.R.C.

Quantitative PCR for *DEXI* and *PseudoDEXI* was undertaken by L.J.D., K.L.A., N.S. and A.A.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from http://www.wtccc.org.uk/. We also thank P. Clarke, G. Coleman, D. Harrison, S. Hawkins, M. Maisuria-Armer, T. Mistry and N. Taylor for the preparation of DNA samples. We thank Bing Ge and Tomi Pastinen for sharing

their PeakPicker software. We are grateful to Randy Read for helpful discussion.

Conflict of Interest statement. None declared.

FUNDING

We acknowledge the use of DNA from The UK Blood Services collection of Common Controls (UKBS collection), funded by the Wellcome Trust grant 076113/C/04/Z, by the Wellcome Trust/Juvenile Diabetes Research Foundation grant 061858 and by the National Institute of Health Research of England. The collection was established as part of the Wellcome Trust Case-Control Consortium. Funding for the project was provided by the Wellcome Trust under award 076113. L.J.D. is supported by a Wellcome Trust Intermediate Clinical Fellowship (082549/Z/07/Z). C.W. is supported by a Wellcome Trust Career Development Fellowship (089989/Z/09/Z). K.L.A. is supported by a 4-year Wellcome Trust CIMR PhD studentship. N.K.W. is supported by the Medical Research Council. N.S. and A.A. are supported by King Saud bin Abdulaziz University of Health Sciences and the Saudi Ministry for Higher Education. This work was funded by the Juvenile Diabetes Research Foundation International, the Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Centre. The Cambridge Institute for Medical Research (CIMR) is in receipt of a Wellcome Trust Strategic Award (079895). The Gutenberg Health Study is funded through the government of Rheinland-Pfalz (Stiftung Rheinland Pfalz für Innovation, contract number AZ 961-386261/733), the research programs Wissen schafft Zukunft and Schwerpunkt Vaskuläre Prävention of the Johannes Gutenberg-University of Mainz and its contract with Boehringer Ingelheim and PHILIPS Medical Systems including an unrestricted grant for the Gutenberg Health Study. Specifically, the research reported in this article was supported by the National Genome Network NGFNplus (contract number project A3 01GS0833) by the Federal Ministry of Education and Research, Germany. Funding to pay the Open Access publication charges for this article was provided by the Wellcome Trust.

REFERENCES

1. Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
2. Hakonarson, H., Grant, S.F., Bradfield, J.P., Marchand, L., Kim, C.E., Glessner, J.T., Grabs, R., Casalunovo, T., Taback, S.P., Frackelton, E.C. *et al.* (2007) A genome-wide association study identifies *KIAA0350* as a type 1 diabetes gene. *Nature*, **448**, 591–594.
3. Todd, J.A., Walker, N.M., Cooper, J.D., Smyth, D.J., Downes, K., Plagnol, V., Bailey, R., Nejentsev, S., Field, S.F., Payne, F. *et al.* (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.*, **39**, 857–864.
4. The International Multiple Sclerosis Genetics Consortium. (2009) The expanding genetic overlap between multiple sclerosis and type I diabetes. *Genes Immun.*, **10**, 11–14.
5. Nischwitz, S., Cepok, S., Kroner, A., Wolf, C., Knop, M., Muller-Sarnowski, F., Pfister, H., Rieckmann, P., Hemmer, B., Ising, M. *et al.* (2010) More *CLEC16A* gene variants associated with multiple sclerosis. *Acta Neurol. Scand.*, **123**, 400–406.
6. Zuvich, R.L., Bush, W.S., McCauley, J.L., Beecham, A.H., De Jager, P.L., Ivinson, A.J., Compston, A., Hafler, D.A., Hauser, S.L., Sawcer, S.J. *et al.* (2011) Interrogating the complex role of chromosome 16p13.13 in multiple sclerosis susceptibility: independent genetic signals in the CIITA-CLEC16A-SOCS1 gene complex. *Hum. Mol. Genet.*, **20**, 3517–3524.
7. Skinningsrud, B., Husebye, E.S., Pearce, S.H., McDonald, D.O., Brandal, K., Wolff, A.B., Lovas, K., Egeland, T. and Undlien, D.E. (2008) Polymorphisms in *CLEC16A* and *CIITA* at 16p13 are associated with primary adrenal insufficiency. *J. Clin. Endocrinol. Metab.*, **93**, 3310–3317.
8. Mells, G.F., Floyd, J.A., Morley, K.I., Cordell, H.J., Franklin, C.S., Shin, S.Y., Heneghan, M.A., Neuberger, J.M., Donaldson, P.T., Day, D.B. *et al.* (2011) Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.*, **43**, 329–332.
9. Gateva, V., Sandling, J.K., Hom, G., Taylor, K.E., Chung, S.A., Sun, X., Ortmann, W., Kosoy, R., Ferreira, R.C., Nordmark, G. *et al.* (2009) A large-scale replication study identifies *TNIP1*, *PRDM1*, *JAZF1*, *UHRF1BP1* and *IL10* as risk loci for systemic lupus erythematosus. *Nat. Genet.*, **41**, 1228–1233.
10. Zhang, Z., Cheng, Y., Zhou, X., Li, Y., Gao, J., Han, J., Quan, C., He, S., Lv, Y., Hu, D. *et al.* (2011) Polymorphisms at 16p13 are associated with systemic lupus erythematosus in the Chinese population. *J. Med. Genet.*, **48**, 69–72.
11. Barrett, J.C., Clayton, D.G., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H.A., Julier, C., Morahan, G., Nerup, J., Nierras, C. *et al.* (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.*, **41**, 703–707.
12. Cooper, J.D., Smyth, D.J., Smiles, A.M., Plagnol, V., Walker, N.M., Allen, J.E., Downes, K., Barrett, J.C., Healy, B.C., Mychaleckyj, J.C. *et al.* (2008) Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat. Genet.*, **40**, 1399–1401.
13. Smyth, D.J., Plagnol, V., Walker, N.M., Cooper, J.D., Downes, K., Yang, J.H., Howson, J.M., Stevens, H., McManus, R., Wijmenga, C. *et al.* (2008) Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N. Engl. J. Med.*, **359**, 2767–2777.
14. Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G.A., Adany, R., Aromaa, A. *et al.* (2010) Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.*, **42**, 295–302.
15. Mero, I.L., Ban, M., Lorentzen, A.R., Smestad, C., Celius, E.G., Saether, H., Saeedi, H., Viken, M.K., Skinningsrud, B., Undlien, D.E. *et al.* (2011) Exploring the CLEC16A gene reveals a MS-associated variant with correlation to the relative expression of *CLEC16A* isoforms in thymus. *Genes Immun.*, **12**, 191–198.
16. Zeller, T., Wild, P., Szymczak, S., Rotival, M., Schillert, A., Castagne, R., Maouche, S., Germain, M., Lackner, K., Rossmann, H. *et al.* (2010) Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS One*, **5**, e10693.
17. Heinig, M., Petretto, E., Wallace, C., Bottolo, L., Rotival, M., Lu, H., Li, Y., Sarwar, R., Langley, S.R., Bauerfeind, A. *et al.* (2010) A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature*, **467**, 460–464.
18. Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I. and Dermitzakis, E.T. (2010) Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.*, **6**, e1000895.
19. Grundberg, E., Adoue, V., Kwan, T., Ge, B., Duan, Q.L., Lam, K.C., Koka, V., Kindmark, A., Weiss, S.T., Tantisira, K. *et al.* (2011) Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. *PLoS Genet.*, **7**, e1001279.
20. Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306–1311.
21. Cope, N.F. and Fraser, P. (2009) Chromosome conformation capture. *Cold Spring Harb. Protoc.*, **2009**, pdb prot5137.
22. Edgar, A.J., Birks, E.J., Yacoub, M.H. and Polak, J.M. (2001) Cloning of dexamethasone-induced transcript: a novel glucocorticoid-induced gene that is upregulated in emphysema. *Am. J. Respir. Cell. Mol. Biol.*, **25**, 119–124.
23. Dekker, J. (2006) The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat. Methods*, **3**, 17–21.

24. Farnham, P.J. (2009) Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.*, **10**, 605–616.
25. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W. *et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, **459**, 108–112.
26. Visel, A., Rubin, E.M. and Pennacchio, L.A. (2009) Genomic views of distant-acting enhancers. *Nature*, **461**, 199–205.
27. Mattick, J.S. (2009) The genetic signatures of noncoding RNAs. *PLoS Genet.*, **5**, e1000459.
28. Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.*, **10**, 155–159.
29. De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L. and Natoli, G. (2010) A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.*, **8**, e1000384.
30. Raney, B.J., Cline, M.S., Rosenbloom, K.R., Dreszer, T.R., Learned, K., Barber, G.P., Meyer, L.R., Sloan, C.A., Malladi, V.S., Roskin, K.M. *et al.* (2011) ENCODE whole-genome data in the UCSC genome browser (2011 update). *Nucleic Acids Res.*, **39**, D871–D875.
31. (2011) A user's guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol.*, **9**, e1001046.
32. Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schutte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E. *et al.* (2010) Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*, **7**, 532–544.
33. Wang, J.C., Derynck, M.K., Nonaka, D.F., Khodabakhsh, D.B., Haqq, C. and Yamamoto, K.R. (2004) Chromatin immunoprecipitation (ChIP) scanning identifies primary glucocorticoid receptor target genes. *Proc. Natl Acad. Sci. USA*, **101**, 15603–15608.
34. McDaniell, R., Lee, B.K., Song, L., Liu, Z., Boyle, A.P., Erdos, M.R., Scott, L.J., Morken, M.A., Kucera, K.S., Battenhouse, A. *et al.* (2010) Heritable individual-specific and allele-specific chromatin signatures in humans. *Science*, **328**, 235–239.
35. Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S. *et al.* (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, **467**, 430–435.
36. Calderon, B., Suri, A., Miller, M.J. and Unanue, E.R. (2008) Dendritic cells in islets of Langerhans constitutively present beta cell-derived peptides bound to their class II MHC molecules. *Proc. Natl Acad. Sci. USA*, **105**, 6121–6126.
37. Filion, L.G., Graziani-Bowering, G., Matusevicius, D. and Freedman, M.S. (2003) Monocyte-derived cytokines in multiple sclerosis. *Clin. Exp. Immunol.*, **131**, 324–334.
38. Slavin, A., Kelly-Modis, L., Labadia, M., Ryan, K. and Brown, M.L. (2010) Pathogenic mechanisms and experimental models of multiple sclerosis. *Autoimmunity*, **43**, 504–513.
39. Ge, B., Gurd, S., Gaudin, T., Dore, C., Lepage, P., Harmsen, E., Hudson, T.J. and Pastinen, T. (2005) Survey of allelic expression using EST mining. *Genome Res.*, **15**, 1584–1591.
40. Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell.*, **27**, 91–105.
41. Harismendy, O., Notani, D., Song, X., Rahim, N.G., Tanasa, B., Heintzman, N., Ren, B., Fu, X.D., Topol, E.J., Rosenfeld, M.G. *et al.* (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature*, **470**, 264–268.
42. Sotelo, J., Esposito, D., Duhagon, M.A., Banfield, K., Mehalko, J., Liao, H., Stephens, R.M., Harris, T.J., Munroe, D.J. and Wu, X. (2010) Long-range enhancers on 8q24 regulate c-Myc. *Proc. Natl Acad. Sci. USA*, **107**, 3001–3005.
43. Meyer, K.B., Maia, A.T., O'Reilly, M., Ghoussaini, M., Prathalingam, R., Porter-Gill, P., Ambs, S., Prokunina-Olsson, L., Carroll, J. and Ponder, B.A. (2011) A functional variant at a prostate cancer predisposition locus at 8q24 is associated with *PVT1* expression. *PLoS Genet.*, **7**, e1002165.
44. Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K.V., Li, X., Li, H., Kuperwasser, N., Ruda, V.M. *et al.* (2010) From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*, **466**, 714–719.
45. Li, G., Fullwood, M.J., Xu, H., Mulawadi, F.H., Velkov, S., Vega, V., Ariyaratne, P.N., Mohamed, Y.B., Ooi, H.S., Tennakoon, C. *et al.* (2010) ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.*, **11**, R22.
46. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
47. Vassetzky, Y., Gavrilov, A., Eivazova, E., Priozhkova, I., Lipinski, M. and Razin, S. (2009) Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. *Methods Mol. Biol.*, **567**, 171–188.