Genetic architecture of early pre-inflammatory stage transcription signatures of autoimmune diabetes in the pancreatic lymph nodes of the NOD mouse reveals significant gene enrichment on chromosomes 6 and 7.

Beatrice Regnault, Evie Melanitou

HAL Id: pasteur-01441051

https://pasteur.hal.science/pasteur-01441051

Submitted on 19 Jan 2017

Genetic architecture of early pre-inflammatory stage transcription signatures of autoimmune diabetes in the pancreatic lymph nodes of the NOD mouse reveals significant gene enrichment on chromosomes 6 and 7

Beatrice Regnault [a], Evie Melanitou [b],*

[a] *Department of Genomes and Genetics, Eukaryote Genotyping Platform, Institut Pasteur, 25-28 rue du Dr Roux, 75725 Paris, Cedex 15, France*
[b] *Department of Parasites and Insect-Vectors, Institut Pasteur, 25-28 rue du Dr Roux, 75725 Paris, Cedex 15, France*

A B S T R A C T

Autoimmune diseases are characterized by the stimulation of an excessive immune response to self-tissues by inner and/or outer organism factors. Common characteristics in their etiology include a complex genetic predisposition and environmental triggers as well as the implication of the major histocompatibility (MHC) locus on human chromosome 6p21. A restraint number of non-MHC susceptibility genes, part of the genetic component of type 1 diabetes have been identified in human and in animal models, while the complete spectrum of genes involved remains unknown. We elaborate herein patterns of chromosomal organization of 162 genes differentially expressed in the pancreatic lymph nodes of Non-Obese Diabetic mice, carefully selected by early sub-phenotypic evaluation (presence or absence of insulin autoantibodies). Chromosomal assignment of these genes revealed a non-random distribution on five chromosomes (47%). Significant gene enrichment was observed in particular for two chromosomes, 6 and 7. While a subset of these genes coding for secreted proteins showed significant enrichment on both chromosomes, the overall pool of genes was significantly enriched on chromosome 7. The significance of this unexpected gene distribution on the mouse genome is discussed in the light of novel findings indicating that genes affecting common diseases map to recombination "hotspot" regions of mammalian genomes. The genetic architecture of transcripts differentially expressed in specific stages of autoimmune diabetes offers novel venues towards our understanding of patterns of inheritance potentially affecting the pathological disease mechanisms.

1. Introduction

Patterns of mRNA expression can offer important hints not only about tissue specificity and gene function but also can be indicative of chromosomal organization of transcription (Su et al., 2004). In particular in complex disorders whereas genetic, epigenetic and environmental factors influence phenotypic outcomes, transcriptional regulation may be dependent upon spatial organization and chromosomal localization of the underlying the disorder genes.

Autoimmune diabetes or type 1 diabetes (T1D) is an inherited condition, classified in the complex diseases concurred by a multitude of factors including genetic and environmental. It affects a continuously increasing number of individuals in industrialized countries and it appears to follow an incidence reverse to the prevalence of infectious diseases in the world (Airaghi and Tedeschi, 2006; Bach and Chatenoud, 2012).

The heritable component of T1D includes MHC alleles together with multiple weak loci carrying non-MHC genes that influence the genetic risk to develop the disease and may contribute to the final disease phenotype by genetic interactions (Ridgway et al., 2008).

Few of the non-MHC genes have been identified by genetic analysis studies however a large number remains hidden mainly due to small effects on the disease incidence. Moreover the increasing number of SNPs identified outside gene regions and significantly linked with T1D, indicates that chromosomal regions might influence transcription of genes not necessarily located nearby these polymorphic regions (Barrett et al., 2009; Pociot et al., 2010; Torn et al., 2015). Intergenic disease-associated genetic loci (IDAGL) carrying disease associated polymorphisms (SNPs) were found to be frequently transcribed and have the potential to influence the biological behavior of human cells via non-coding RNAs (Glinskii et al., 2011a). These authors demonstrated that IDAGLs possess intrinsic regulatory functions mediated by both DNA sequences and transcribed RNA molecules. Therefore transcriptional activity of common disease-associated variants located within

* Corresponding author at: Department of Parasites and Insect-Vectors, Institut Pasteur, 25-28 rue du Docteur Roux, 75725 Paris, Cedex 15, France.
E-mail addresses: beatrice.regnault@pasteur.fr (B. Regnault), evie.melanitou@pasteur.fr (E. Melanitou).

intergenic regions of the genome may alter phenotypes and carry potential clinical significance. Moreover there is evidence that single nucleotide changes of the human genome create small regulatory RNA molecules that contribute to the pathogenesis of several common human disorders (Glinskii et al., 2009). Interestingly, such SNP-polymorphic containing segments were found to be often highly conserved in other mammals including the rat and the mouse genomes, indicating that are functionally significant (Jin et al., 2007).

We have established that Early Insulin Autoantibodies (E-IAA) are present in the Non-Obese Diabetic (NOD) mouse, and represent a landmark for early T1D development (Melanitou et al., 2004). The maternal autoimmune-prone environment influences the IAA levels of the litters (Melanitou, 2005; Melanitou et al., 2004). Thus the presence of E-IAA predisposes to early T1D and emphasizes the biological significance of this sub-phenotype as an early marker of autoimmunity.

We used the predictive value of the presence of E-IAA to select NOD mice in a systematic functional genomics approach by transcriptome analysis applied in the pancreatic lymph nodes (PLN) of NOD mice (Regnault et al., 2009). Functional genomics analysis of the PLN transcripts demonstrated the existence of genes that are highly regulated at an early stage, prior to clinical signs, other than the presence of E-IAA (Regnault et al., 2009).

Seventy four of these genes code for secreted proteins (SPGs) (Melanitou et al., 2013). The potential of secreted proteins to be recovered in the peripheral blood renders them prospective biomarkers for the early steps of the disease.

It has been previously reported the existence of windows of genes with correlated expression patterns, clustered in a locus-dependent manner (Su et al., 2004). Genes expressed in tissues and cells of the immune system showed to be organized within very specific clusters called "Regions of correlated transcription" more often than genes expressed in other tissues (Su et al., 2004). Our aim in the present study was to address the genetical genomics image of transcripts differentially expressed in the early preinflammatory stages of T1D and evaluate the venues possibly uncovered by such image, towards our understanding of the mechanisms underlying the disease. To our knowledge autoimmune diabetes-related genes identified by transcriptome analysis have not been considered previously for their genomic architecture but only in relation with genetic loci issued from genetic analysis studies (Kodama et al., 2008).

In this context, we evaluate in the present report the genomic architecture corresponding to the transcripts expressed in the PLN under the conditions of the E-IAA sub phenotype, which may indicate higher order mechanisms of transcriptional regulation. We identified that a non-random distribution of these genes is observed in the mouse genome, with a significant over representation on mouse chromosomes 6 and 7.

2. Materials and methods

2.1. Mice, tissues RNA preparations and IAA assay

The experimental setting for data collection has been previously described (Regnault et al., 2009). Briefly NOD/tac mice were purchased from Taconic Farms and maintained under specific pathogen-free barrier facilities at the Barbara Davis Center (Denver, CO). All experimental protocols were performed under conditions according to guidelines approved by the Institutional Animal Care and Use Committee of the University of Colorado as previously described (Melanitou et al., 2004; Regnault et al., 2009). Overall nine female mice at 5 weeks of age were used for these experiments: four mice E-IAA$_{pos}$ and five E-IAA$_{neg}$. IAA assays were performed in 96 well filtration plates, as previously described (Melanitou et al., 2004; Yu et al., 2000) using a standard radioimmunoassay and incorporating competition with unlabeled insulin and Protein A/G sepharose precipitation.

2.2. Microarray data analysis

Microarray data have been obtained as described (Regnault et al., 2009). Briefly six E-IAA negative and three E-IAA positive animals at 5 weeks of age have been used for isolation of PLN. One E-IAA negative sample (A36.4) was grouped according to its gene expression profiles together with the E-IAA positive group of animals, by clustering analysis. Therefore the final number of positive samples analyzed together contained four samples and the negative control group five samples.

The MG_U74A_version 2 arrays (Affymetrix, Santa Clara, Ca) were used containing 12,486 probe sets. All the initial data, from which we extracted this novel analysis, were deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) with the GEO series accession number GSE15582 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15582). Data were processed as previously described (Regnault et al., 2009). Briefly, raw data were preprocessed using the Robust Multiarray Averaging (RMA) normalization method for individual probe values and for summary values for each probe set. Statistical analysis of samples was performed using the Local Pool Error (LPE) test, an algorithm dedicated to small number of samples (Jain et al., 2003) and the P values adjusted by the Benjamini–Hochberg multiple testing correction. The dChip software was used for hierarchical clustering with Euclidean distance and average as a linkage method (Li and Wong, 2001). Heatmap was constructed by taking into consideration the gene expression patterns (Pelizzola et al., 2006).

Functional and cellular localization annotations were evaluated by using the DAVID NIAID/NIH online annotation tool (http://david.niaid.nih.gov/david/), the Panther classification system (http://www.pantherdb.org/) and the KEGG database (http://www.genome.jp/kegg/). Annotation terms were obtained taking into consideration the significance of P values for GO terms (Dennis et al., 2003). The GO Browser tool of NetAffx was used to confirm independently the GO annotations. Further evaluations of functional annotations and pathways were performed using the Ingenuity Pathway Analysis software (Ingenuity Systems) and the iReport data mining system (http://www.ingenuity.com/products/ireport).

For functional predictions of gene networks and gene co-expression search the GeneMANIA Cytoscape plug-in online tool (http://www.genemania.org/) has been used (Warde-Farley et al., 2010).

2.3. Chromosomal enrichment

Chromosomal enrichment was performed using the DAVID NIAID/NIH online annotation tool (http://david.niaid.nih.gov/david/) for chromosome localization.
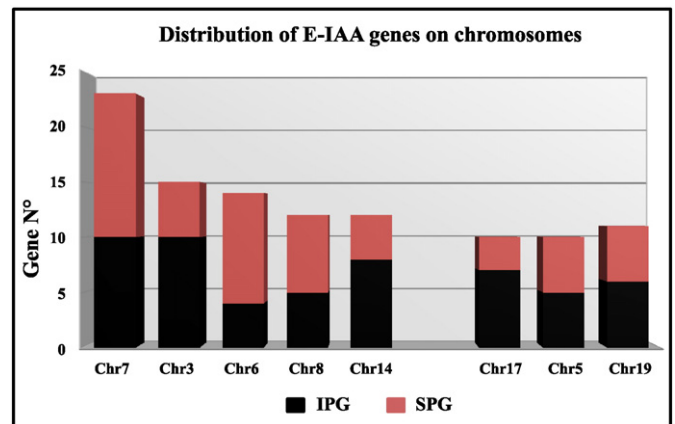

Fig. 1. *Chromosomal assignment of E-IAA PLN transcripts.* Representation of the five chromosomes enriched with E-IAA PLN genes (SPGs and IPGs). Only chromosomes containing >5 genes are shown. chr5, 17 and 19 are shown merely as representative for the rest of the genome.
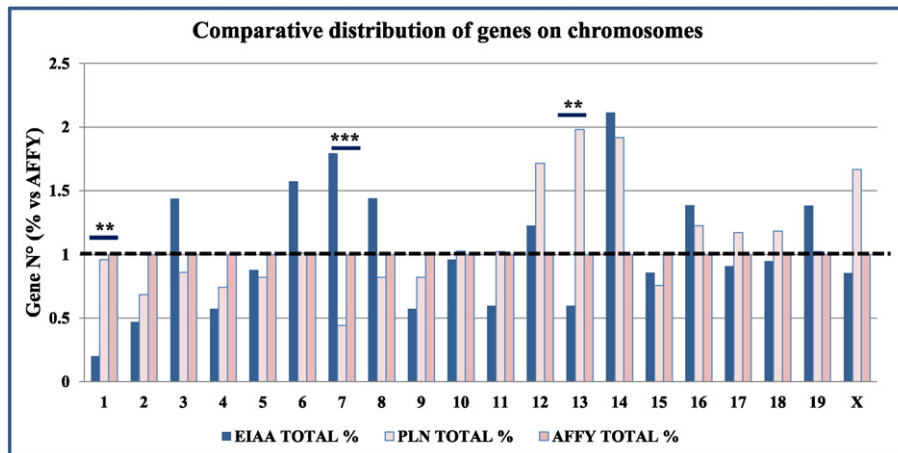
Fig. 2. *Chromosome distribution of E-IAA, PLN-expressed and Affy genes.* Significant gene enrichment is observed for the E-IAA genes on chr7 compared to the PLN-expressed genes (P < 0.0001) or to the Affy genes (P = 0.003). In contrast enrichment of the E-IAA genes on chr3, chr6 and chr8 was not significant. Significant under-representation of the E-IAA genes was observed for chr1 (PLN-expressed or Affy vs E-IAA: P = 0.014). Total number of genes: E-IAA: 162; PLN-expressed: 529; Affy: 10,417. Data for each set (E-IAA and PLN) are presented relative to Affy gene numbers for each chromosome (=1).

P-values for each chromosome output were calculated according to the number of E-IAA differentially expressed transcripts attributed to each chromosome versus i) the number of unique genes found to be expressed to all nine PLN samples with known chromosome localizations (P calls, gene n° = 560; 529 genes on chromosomes) (Table S3), considered to be PLN expressed at least in our conditions and ii) the total number of genes present on the MG_U74A_version 2 arrays (10,913 unique genes, 10,417 genes with known chromosome localizations). Chromosomal loci for insulin dependent diabetes identified by genetic analysis in mouse (*Idd*), in rat (*Iddm*) and in human (*IDDM*) were retrieved by search of the literature and from the T1D Database (http://www.t1dbase.org/) (Burren et al., 2011).

P-values between gene sets were calculated by a two-tailed parametric test.

2.4. SNPs identification

We searched for SNPs, their genomic positions and associated genes when applied, between the NOD/*Shi*LtJ strain and strains C57Bl/6, C3H, DBA/2, SJL and NON in the public data made available in the Mouse Genome Informatics web site (Blake et al., 2014) (http://www.informatics.jax.org/strains_SNPs.shtml).

2.5. Statistical analysis

Statistical analyses were performed by the XLSTAT Software. Comparative chromosome enrichments of transcripts were evaluated using a parametric z test for two proportions, with 95% confidence interval for one-tailed P-values.

Table 1
Comparison of chromosome enrichment between gene sets. Five chromosomes are over represented in the E-IAA transcriptome. Significance for E-IAA vs PLN: P < 0.0001 and E-IAA vs AFFY: P < 0.0001.

CHR	E-IAA	PLN	AFFY
chr7	24	19	818
chr6	16	33	622
chr3	13	25	553
chr8	12	22	510
chr14	11	32	318
Total gene n° on chr	**76 (47%)**	**131 (25%)**	**2821 (27%)**
Total n° of genes	162	529[a]	10,417

[a] 529 out of 563 PLN-expressed genes have known chromosome localizations.

3. Results and discussion

3.1. Genomics of E-IAA PLN gene expression

The E-IAA sub-phenotype is a quantal phenotype and carries genetic complexity as it does the final T1D phenotype (Melanitou et al., 2004). This genetic complexity is composed from the multiple implicated genes and from possible interactions among these genes and/or loci with environmental variables.

Taking into consideration that i) gene expression is the key mechanism by which information encoded by the genome is converted into physiological or pathological phenotypes and ii) that transcript levels are under the influence of environmental or inherent to the organism perturbations (in our case the E-IAA status of the animal), we examined the relevance of the variation of the identified E-IAA PLN transcripts versus the genetics of the corresponding loci and evaluated the potential of these genes as candidates in genetic analysis studies in the mouse and in human. In this respect, genetic polymorphisms in relation with the transcription patterns of genes within polymorphic regions carry potential information for the disease.

Chromosomal distribution of the E-IAA genes revealed a non-random pattern of gene representation on chromosomes. Indeed a high number of genes were located on five chromosomes: chr7 (24 genes), chr6 (16 genes), chr3 (13 genes), chr8 (12 genes) and chr14 (11 genes) (Fig. 1).

In order to exclude a bias from a possible over representation of the transcripts on the arrays on these chromosomes, we compared the chromosomal distribution of the E-IAA PLN genes with the distribution: i) of the total number of unique genes with known genome positions present on the mouse chip (Affy MG_U74A), corresponding to 10,417 genes (out of the 12,488 probe sets) and ii) of the total number of genes expressed in the PLN of all nine mice used in our study, independently of the phenotype, with known chromosomal localizations (529 genes, Table S2).

Table 2
SPG and IPG representation in gene sets. Secreted protein genes are over represented in the E-IAA transcriptome. Significance for E-IAA vs PLN: P < 0.0001 and E-IAA vs AFFY: P < 0.0001.

Gene set	SPG n° (%)	IPG n° (%)	Total gene n°
E-IAA	74 (46%)	88 (54%)	162
PLN	49 (9%)	480 (91%)	529[a]
AFFY	2031 (19.5%)	8386 (80.5%)	10,417

[a] 529 out of 563 PLN-expressed genes have known cellular localizations.

These genes may be considered as PLN-expressed genes at least in the NOD mouse at 5 weeks of age. The expression of these genes was not however solely confined in the PLN (Fig. S2). The purpose of the comparison with the PLN-expressed genes is to farther evaluate phenotype-specificity rather than a tissue-expression bias of this non-random chromosome segregation.

Overall a significantly higher number of E-IAA genes than expected was attributed on these five chromosomes in comparison with the distribution of the array genes (47% E-IAA vs 27% Affy genes, P < 0.0001, Fig. 2 and Table 2). A similar distribution was also observed versus the PLN-expressed genes (47% vs 25%, P < 0.0001, Fig. 2 and Table 1). In particular, high E-IAA gene enrichment was observed systematically for chr7 in comparison with the total number of genes in the arrays (10,417 genes, P = 0.003) or in the PLN-expressed set (529 genes on chromosomes; P < 0.0001) (Fig. 2). In contrast, while E-IAA genes were over represented on chr3, chr6 and chr8 differences did not attain significance for these chromosomes. The remaining genes were randomly distributed on the genome with the exception of an under-representation of E-IAA genes on chr1 and chr13 (P = 0.014 and P = 0030 respectively) (Fig. 2).

Interestingly the SPGs contribute mainly to this non-random distribution (for chr7: E-IAA vs AFFY: P < 0.001; vs PLN: P = 0.002 and for chr6 (E-IAA vs PLN P = 0.001), (Fig. S1A), while intracellular protein coding genes (IPGs) were over-represented also on chr14 (chr14: E-IAA vs AFFY: P = 0.032), in addition to chr7 (chr7: E-IAA vs PLN: P = 0.024), (Fig. S1B).

Finally, the total number of SPGs over-represented on three chromosomes is 47% (35/74 on chr6, 7 and 8) (Table S1B), whereas the remaining genes of this category are dispersed on 13 chromosomes, each carrying fewer than 5 genes per chromosome (<7% of the total SPG number). Four chromosomes (chr1, chr13, chr18 and chrX) do not carry any of the SPGs, in agreement with the non-random distribution observed (Fig. S1A).

In contrast to the E-IAA SPG genes, genome distribution of the array genes coding for secreted proteins (19.5%, Table 2), are randomly distributed in the mouse genome. In particular, genes mapped on chromosomes 7, 6 and 8 represent together 19% of the total number of the transcripts (8%, 6% and 5% respectively, Fig. S1A). Similar random chromosome distribution was also observed for the PLN-expressed genes.

3.2. Regions of correlated transcription

It was previously reported that proteins, encoded by linked genes within genomic regions associated with immune-mediated diseases, physically interact and such interactions suggest underlying biology (Rossin et al., 2011). In this context we investigated the existence of regions containing clusters of genes that have shown, at least in our experimental design, correlated transcription on chr6 and chr7.

Segregation of the identified E-IAA SPG transcripts is observed on 4 regions on chr7 while four IPGs confine a region spanning 29 Mb (Fig. 3). The locus II on chr7 (28.3 cM) spans 0.3 cM (approximately 480 kb) and contains a highly conserved multigene family (Evans et al., 1987) encoding several serine proteases that are essential to many biological processes including inflammation (Bhoola et al., 1992a; Bhoola et al., 1992b). Several of these genes are up regulated in the E-IAA$_{pos}$ samples (Table S1A). Expression differences vary from 2 to 37 fold and in particular three genes: *Klk1b5*, *Klk1b9*, *Klk1b5r* are highly expressed in the E-IAA$_{pos}$ PLN (37, 19 and 17 fold respectively, Table S1A). The syntenic human KLK locus contains 14 clustered genes within 300 kb, on CHR19q13.3–13.4 (Riegman et al., 1992; Stevenson
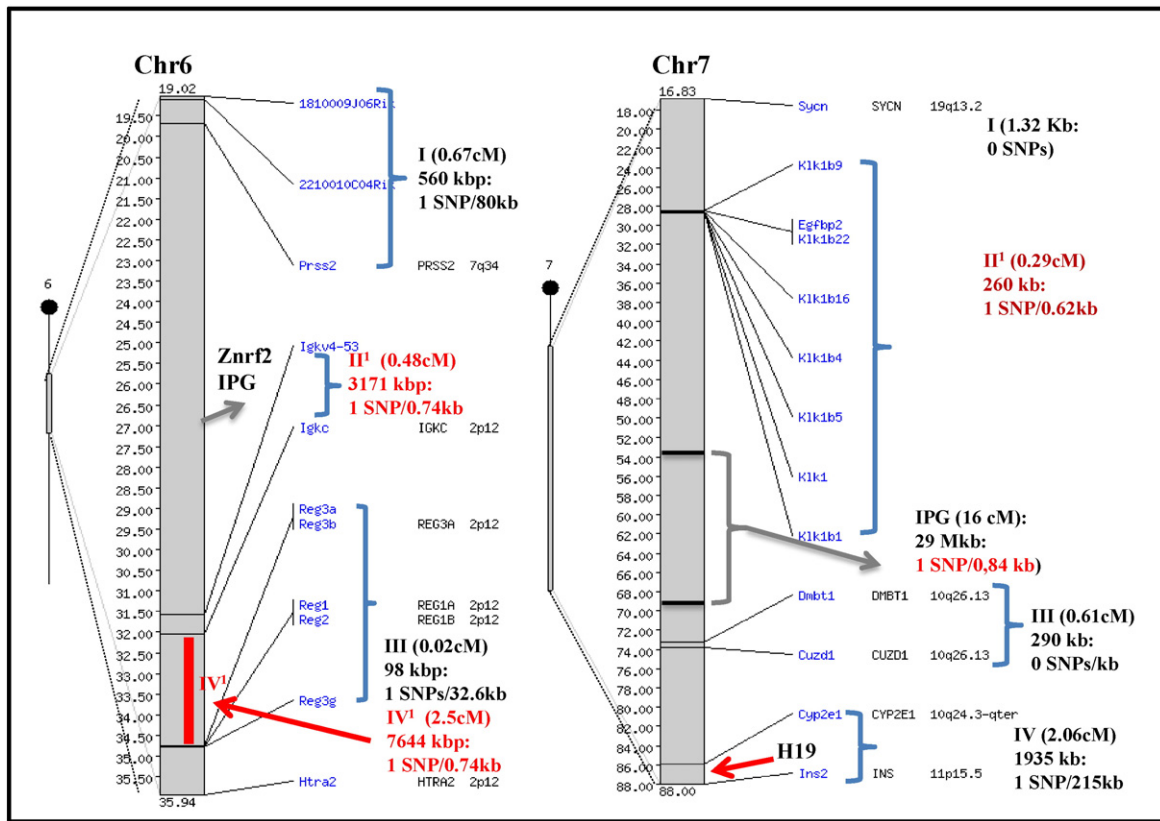

Fig. 3. *Schematic representation of E-IAA PLN gene enrichment on chromosomes 6 and 7.* Loci are shown for regions defined by the SPGs (Regions I–IV) and one locus where several IPGs are mapped. A. chr6 and B. chr7. Regions are numbered and genetic sizes are in parenthesis. Numbers of SNP polymorphisms identified are as indicated for each locus. [1]High SNP-content region.

et al., 1986). Conservation of these genes between human, mouse and rat indicate that they play an important function in these species.

Chr6, similarly to chr7, contains also a high number of E-IAA genes especially for the secreted proteins coding genes (E-IAA vs AFFY: P = 0.001, Fig. 3). Interestingly enrichment of SPGs on chr6 was also observed for the PLN-expressed genes therefore a tissue-bias of this enrichment cannot be excluded (Fig. S1A).

Region II contains genes coding for immunoglobulins (*Igkv4*-53, *Igkc*), within a genetic interval of 0.5 cM, all down regulated in the E-IAA$_{pos}$ samples (Fig. 3). In contrast several *Reg* genes map *in tandem* in region III of chr6, are all up-regulated and form a cluster delineating a smaller genetic interval of 0.02 cM with the quasi totality of the genes been highly expressed (28 to 58 fold) in the E-IAA$_{pos}$ samples (Fig. 3 and Table S1). One additional region, (I) spans 0.57 cM and contains one gene (*Prss2*) and two Ests (*1810009J06Rik* and *2210010C04Rik*) with levels of gene expression 52 and 3.96 fold respectively (Table S1A). The other regions on these chromosomes are limited by one or two genes.

Several imprinted regions are located on the proximal part of mouse chromosome 6 (Beechey, 2004), while parental imprinting characterizes large regions of chromosome 7, dispersed throughout the entire chromosome, including the insulin 2 gene locus on the distal part (Shmela and Gicquel, 2013). The implication of parental imprinting, exerting chromatin remodeling and consequently influencing transcriptional activity on specific chromosomal regions on chr7, as is the case of the H19 locus (Murrell et al., 2004), represents an attractive hypothesis for the simultaneous expression of closely linked genes, especially taking in consideration the correlation of the E-IAA sub phenotype to the maternal status of IAA (Melanitou et al., 2004). Windows of genes with correlated expression patterns (RCTs) on specific chromosomal regions have been previously reported, particularly enriched in samples from purified immune cell populations (Su et al., 2004). These authors showed that several of these RCTs were conserved, including a cluster of pancreas-specific genes mapping on the mouse chr6 (region III, Fig. 3) and its syntenic region on human chromosome 2. The mouse cluster contains several regenerating islet-derived transcripts, all highly up-regulated in the E-IAA$_{pos}$ samples (*Reg1*, *Reg2*, *Reg3a*, *Reg3b* and *Reg3g*) (Table S3). The human cluster comprises the corresponding orthologs that are highly conserved between the two species suggesting that these genes participate to similar and important functions in both of these mammals. Noticeably the *Reg* genes play a role in inflammation and tissue regeneration (Yuan et al., 2005). Reg proteins are involved in mechanisms conferring morphogenetic plasticity of the pancreas, as shown in the adult human islets (Jamal et al., 2005). Such plasticity may contribute to pancreatic carcinogenesis and islet neogenesis. Islet neogenesis and tissue preservation by inherent factors could be a valuable mechanism against β-cell destruction and may take place prior or simultaneously to the early inflammatory state, while active self-antigen presentation concurs with the enhanced autoimmune response. The *Reg* genes may have an impact on this phenomenon.

Correlated transcription of the genes within these regions, as illustrated with the example of the *Reg* genes on chr6 is likely mediated through either common promoter elements (due to gene duplication) or through site-specific chromatin remodeling. The impact of a miRNA, encoded by another region and acting in *trans* cannot be excluded. Similarly classes of sncRNAs (small non-coding RNAs) and lncRNAs (long non-coding RNAs) were associated by experimental and clinical observations with a broad spectrum of common polygenic diseases (Glinskii et al., 2009; Glinskii et al., 2011a; Martin and Chang, 2012). A disease phenocode was previously defined as the pattern of associations between disease-linked polymorphisms, microRNAs and mRNAs (Glinsky, 2008b). Therefore the possibility of such transcripts coding or non-coding and affected by single nucleotide polymorphisms may also have an impact on gene expression variation.

Considering that selection of variants acts on specific regions affecting a trait, segregation of 47% of the E-IAA transcripts on five chromosomes

indicates that this set of genes and probably neighbor loci have functional effects and possibly participate in QTLs. Indeed several QTLs were identified within or nearby the loci defined by the E-IAA genes (Table S3) (http://www.t1dbase.org/), (Burren et al., 2011). A similar phenomenon was shown in *Caenorhabditis elegans*, whereas selection acting on specific loci and their neighboring adjacent sequences, affects a quantitative trait (Rockman et al., 2010). It was shown that QTLs were not distributed evenly across the genome but were highly enriched on specific chromosome regions. Interestingly on proximal mouse chr7 a QTL was mapped corresponding to a locus controlling the regulation of the size of the plasmacytoid dendritic cell (pDC) compartment (Pelletier et al., 2012). DCs are innate effectors with the ability to control T-cell immunity by presenting antigens to T-cells, which is essential for T-cell activation and expansion and by secreting cytokines influencing the nature of the T-cell responses (Steinman, 2007). pDCs in particular, have been proposed to be a new player in regulating β-cell-specific T cells reactivity (Tisch and Wang, 2009).

We evaluate below the correlation of the mapping positions of the identified genes with the presence of *Idd* loci in these regions.

3.3. Correlation of mapping positions on chr6 and chr7 with known T1D loci

Taking into consideration the two most enriched chromosomes for the E-IAA transcripts (chr6 and chr7) we searched for the existence of T1D loci, identified previously by genetic analysis in rodents (mouse *Idd* and rat *Iddm*) and in human families (IDDM).
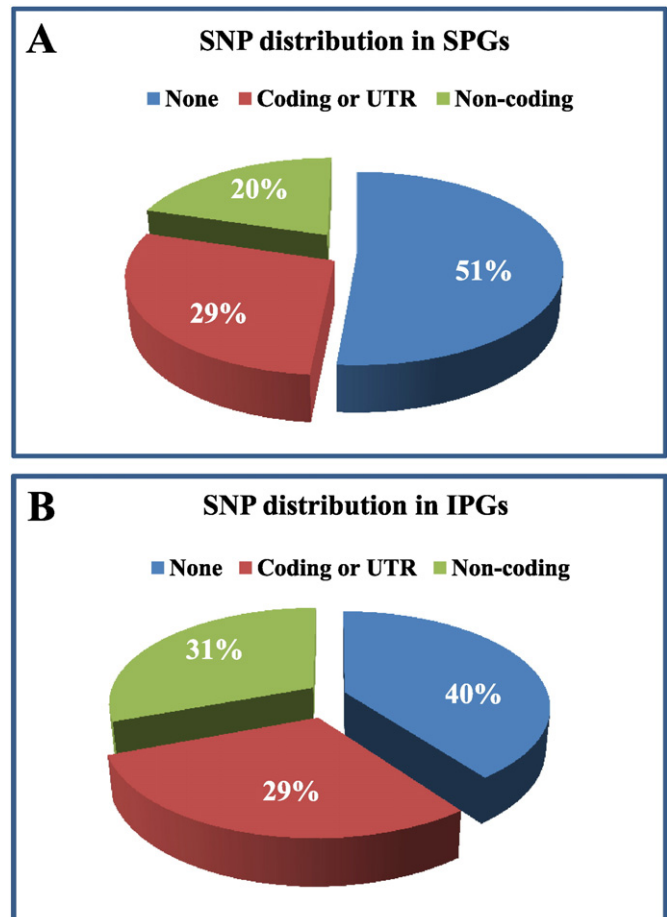

Fig. 4. *Comparison of variants distribution in SPG and IPG.* A. Extracellular proteins coding genes (SPGs) and B. intracellular (IPGs) protein coding genes. SNPs were searched between C57Bl/6J, C3H, NON, SJL/J and the NOD/*ShiLtJ* strains. Percentages are represented for genes with none SNPs, with polymorphisms detected within the coding or mRNA-UTR regions and in non-coding intergenic regions. Data are retrieved from MGI bioinformatics and see Table S3 (http://www.informatics.jax.org/strains_SNPs.shtml).

Table 3

Summary of SNPs identified within loci on chromosomes 6 and 7. The identified polymorphisms are between strains NOD/ShiJ versus C57Bl/6, C3H, DBA/2, SJL/J and NON/ShiJ. (from MGI 5.18 update: dbSNP build 137, (Blake et al., 2014). Highlighted in red are highly polymorphic regions and genes containing several SNPs (from MGI: dbSNP Build 137).

Genomic region	Genome Coordinates (bp) (region size in Kbp)	Gene n° within locus	SNP in Lab. vs all strains (n°)	SNP density1 (Lab strains)	SNP2 (N°/Total N° of SNPs)	Genes within segment
SPGs						
chr6, locus I	40914771–41475078 (560.3)	60 & 9 QTLs=51	7/1189	80 kb	1/7	T cell receptor beta variable chains locus (Trbv genes) & Protease serine (trypsin, *Prss1, Prss2, Prss3, Try10, Try5, Try4* & QTL Diabetes Dbts1)
chr6, locus II	67505630–70676748 (3,171.1)	191	4243/7503	0.74 kb	0/4243	Immunoglobulin kappa chain complex (Igk) genes, *Igkj1*, *Igkv4–5* 3
chr6, locus IV interregion	70676748 – 78320879 (7,644. 1)	122	10222/25350	0.74 kb	347/10222	*Rpia, Eif2ak3, 1700011F03Rik, Thnsl2, Fabp1, Smyd1, Cd8b1, Rmnd5a, Rnf103, Chmp3, Kdm3a, Gm8933, Rpl19–ps9, Reep1, Mrpl35, Immt, Gm17825, Ptcd3, Polr1a, Gm19078, Gm16601, St3gal5, Atoh8, Sftpb, Usp39, Rnf18, Tmem150a, Vamp5,Vamp8, Sh2d6, Cap, Ggcx, Elmod3, Retsat, Tgoln1, Gm20536, Tcf7l1, Gm15401, Kcmf1, Gm17554, Tmsb10, Dnahc6, Suclg1, Ctnna2, Gm5576*
chr6, locus III	78320879–78418868 (98)	6	3/315	32.6 kb	1/3:*Reg3g* (mRNA–UTR)	*Reg3b*, *Refg3d*, *Reg3a*, *Reg1*, *Reg3g*, *Reg2*
chr7, locus I (16.83 cM)	2854088–28542210 (1.32)	1	0/2	No SNPs	None	*Sycn*
chr 7, locus II (28.45–28.74 cM)	4423612–43976061 (260)	27	414/819	0.62 Kb	36/414	*Klkb1*, *klk1b9*, Klk1b10–ps, Klk1b11, *Klk1b12–ps*, klk1b26, Klk1b18–ps, klk1b27, Klk1b14–p, Klk1b15–ps, klk1b21, klk1b22, Klk1b23–ps, *Klk1b16*, Klk1b24, Klk1b3, Klk1b4, Klk1b5, Gm10109, Klk1, *Klk6*, 1700028J19Rik & QTLs: Drsi (Drsi related spp1 induction), SLE19
chr 7, locus III (73.2–73.81 cM)	131032076–131322292 (290)	2	3/841	No SNPs	None	*Dmbt1*, Fgfr2, *Cuzd1*
chr 7, locus IV (85.94–88 cM)	140763832–142699510 (1935)	95	9/5386	215 Kb	1/9	*Cyp2e1*, Syce1, Zfp941, 1190003J15Rik, C330022C24Rik, Scgb1c1, Odf3, Bet1l, Ric8, Sirt3, Psmd13, Gm15542, Cox8b, BC024386, Nlrp6, Athl1, Ifitm5, Ifitm1 , Ifitm2, Gm7514 , Ifitm3, B4galnt4,Pkp3,Sigirr, Ano9, Ptdss2, Rnh1, Hras1, 1600016N20Rik, Rassf7, Phrf1, Irf7, Cdhr5, Drd4, Deaf1, Eps8l2, Tmem80, B230206H07Rik, Taldo1, Pddc1, Cend1, Slc25a22, Rplp2, Pnpla2, Cd151, Polr2l, Tspan4, Chid1, Ap2a2, Muc6, Muc5ac, Muc5ab, Tollip, Gm6449, Brsk2, Mob2, Gm2391, Gm9340, Gm4552, Gm4553, Krtap5–2, Gm10153, Krtap5–3, Gm7579, Gm19094, Krtap5–5, Gm7288, Gm4559, B130016D09Rik, Krtap5–1, Krtap5–4, 6330512M04Rik, Ctsd, Syt8, Tnni2, Lsp1, Tnnt3, Mrpl23, H19, Igf2, Igf2as, *Ins2* & Mir210 (microRNA 210), Mir675 (microRNA 675), Mir483
IPGs						
chr 7 IPG locus (53.49–69.21 cM)	97417730–126627010 (29209)	969 genes, pseudogenes, QTLs, miRNAs	34596/94480	0.84 Kb	Total in coding regions: 1938/345963	Idd27, Heal6, wound healing/regeneration 6, Ssial3, susceptibility to sialadenitis 3, Eae26, experimental allergic encephalomyelitis susceptibility 26, *Thrsp*, *Hbb–b1*, *Hbb–b2*, Tsiq1, *Rps3*, Tcell secretion of IL4 QTL 1, Il4ppq, IL–4 producing potential QTL & Mir3102, microRNA 3102, Mir139, microRNA 139, Mir326, microRNA 326, *Nupr1*

Notes to Table 3
[1] Distance (kb) between SNPs (Lab strains only).
[2] Number of SNPs in coding regions (mRNA-UTR, nsSNPs, sSNPs) only.
[3] Coding non-syn: 435, coding syn: 788, mRNA-UTR: 715, splice sites: 116.

A "genetical genomics" image summarizes the correspondence of diabetes-related loci and the relevant to these loci associated E-IAA-expressed genes (Table S3).

Several E-IAA genes are located nearby or within regions defined by the presence of *Idd* loci identified in mouse crosses. This is the case for *Idd6* and *Idd19* on chr6 (Table S3) (Melanitou et al., 1998). Similarly, *Iddm14* and *Iddm11* (on chr6) (Mordes et al., 2009; Martin et al., 1999), *Iddm10* (on chr7) correspond (Klöting et al., 1998) to syntenic rat *Iddm* QTLs (Table S3).

Furthermore on mouse chr7 two human syntenic regions were reported to contain IDDM loci on the distal part (88 cM): IDDM2 on CHR 11p15.5 and IDDM4 (CHR11q13) located at 54 cM and overlapping with E-IAA genes (Bell et al., 1981; Eisenbarth et al., 1998; Verge et al., 1998) (Table S3). IDDM2 corresponds to the insulin gene that was indeed shown to be the candidate gene for this locus (Bell et al., 1984) and is associated with the E-IAA$_{pos}$ sub phenotype in the PLN, in our study. Consideration should be given that variants might confer larger effects on cellular processes than on disease susceptibility. Indeed immune disease risk genotypes were shown to profoundly affect the homeostasis of proximally located genes (Cotsapas and Hafler, 2013). One example is the effect of T1D-related sequence variations on the *IL2RA* locus (mouse chr2, human IDDM10) on the expression patterns of T cells and monocytes (Dendrou et al., 2009).

3.4. Structural variation of the primary sequences of the E-IAA transcripts and the corresponding loci on chr6 and chr7

Genetic analysis studies are generally designed to identify loci implicated in the final onset of T1D, thus discovery of genes implicated in the early steps may be hampered. Moreover SNP polymorphisms affecting gene expression are difficult to predict, contrary to gene expression levels that were shown to be highly heritable in several organisms (Dixon et al., 2007; Emilsson et al., 2008; Goring et al., 2007).

Therefore due to the implication of the identified genes at the early pre-inflammatory stages of T1D it was interesting to search for the presence of variants, by referring to the existence of the available unique single nucleotide polymorphic loci (SNPs) of the mouse genome (Frazer et al., 2007). We compared strains that were used as mating partners in crosses with the NOD mouse in genetic studies (Leiter et al., 1987; McClive et al., 1994; Peterson et al., 1994; Suzuki et al., 2008) and assessed for polymorphisms between these strains (C57Bl/6, C3H, DBA/2, SJL/J and NON/*Lt*J), with as reference, the NOD/*ShiLt*J.

As expression QTLs (eQTLs) are usually highly enriched outside the coding regions, especially at the transcription start (TSS) or end sites (TES) (Veyrieras et al., 2008), our search spanned sequences including 2 kb upstream and 2 kb downstream regions for each gene. Few SNPs were identified within sequences covering each of the PLN genes (Table S4). While the majority (51%) of the SPGs have none allelic variants between the strains, only 40% of the IPGs were not polymorphic (Fig. 4). The infrequency of SNPs between the NOD and the other strains was however not surprising as phylogenetic distances are very limited between the *Mus domesticus* laboratory strains.

The overall allelic variation for all E-IAA genes in coding regions, including mRNA-UTRs and coding synonymous or non-synonymous SNPs, did not exceed 29% (47 genes, Fig. 4 and see Table S4). Coding non-synonymous substitutions (nsSNP) are of critical importance as they confer amino-acid changes and could potentially affect the function of the protein and subsequently alter the carrier's phenotype. Polymorphisms in untranslated mRNA regions (UTRs) may interfere with the regulatory RNA motifs influencing gene expression. Twenty-eight genes (18%) found to carry 5′ or 3′UTR single point mutations between the strains studied (see Table S4). UTRs contain motifs that are highly conserved among species and represent *cis*-acting regulatory elements that may interfere with mRNA stabilization (Boado and Pardridge, 1998). Interestingly ten transcripts out of the 21 SPGs showing several polymorphisms between the NOD and the other strains studied,

contained mRNA-UTR polymorphisms and include the *Fga*, *Cela2a*, *Spp1*, *Reg3g*, *Klk1b16*, *Klk6*, *Lpl*, *Sel1l*, *Rnase1*, and *Tff2* genes (see Table S4).

In contrast to the overall SNP infrequency observed within sequences covering genes, when entire regions were examined, three in particular, loci II on chr6 and chr7 for the SPGs and one locus on chr7 whereas IPGs are mapped, stand out as they contain unexpectedly dense sequence variations, whereas one SNP every 0.74 kb, 0.62 kb and 0.84 is detected respectively, while in comparison to other chromosomal regions, the mean frequency is less than one variation per over 100 kb (Table 3). One additional region was examined, locus IV on chr6, corresponding to the region between loci II and III on this chromosome (Fig. 3). Finally five regions contained 0 to 1 SNPs between the selected laboratory strains and the NOD/*ShiLt*J reference strain (Table 3). This seems to assent with published reports referring that SNPs, rather than distributed at random across the mammalian genomes, are usually clustered and associated preferentially with recombination hot spots (Amos, 2010; Mefford and Eichler, 2009). Small clusters of mutations occur preferentially near heterozygous sites in human and it was proposed that in mammals such patterns could be expected from gene conversion events that occur during meiosis (Tian et al., 2008). The non-random distribution of SNPs indicates that chromosomal regions that are polymorphic are auspicious to receive more mutations. This could be explained by their localization i) in hot spots of meiotic recombination or ii) in genomic regions that escape selective pressure. In this respect, these single point mutations may confer a benefit to the organism by increasing the generation of diversity at highly polymorphic loci. Highly polymorphic loci include immune-related genes, such as the major histocompatibility complex for which high genetic diversity confers an adaptive advantage to exterior organisms (microbes) (Doherty and Zinkernagel, 1975; Schwensow et al., 2007).

Alternatively, functional significance of disease-related SNPs may be dependent upon their primary sequence homology of non-coding transcripts (microRNAs) of these regions with protein coding genes, as it has been described for the inflammasome proteins NLRP1 and NLRP3 in human (Glinsky, 2008a). In this case the author concluded that SNPs variations associated with disease genes may be at the origin of genetically-defined failure of the nuclear import pathway (NIP) and inflammasome/innate immunity pathways and thus contribute to the pathogenesis of multiple common human disorders, including T1D.

In particular on locus II on chr6 maps the *Igk* gene cluster that is poised for rearrangement in pre-B cells and functional Vk genes are semi randomly selected for recombination to the Jk region (Brekke and Garrard, 2004). Diversity of the antibody repertoire is a prerequisite for efficient recognition of a broad spectrum of invading microorganisms by the immune system. Moreover incorrect or miss regulated repertoire can initiate autoimmunity (Ishida et al., 2006; Jankovic and Nussenzweig, 2003). In our data, the only SPG transcripts down regulated in the E-IAA$_{pos}$ samples, are genes coding for immunoglobulins and two of those are located within this locus (Fig. 3). It is tempting to speculate that the existence of several SNPs within this region, between the NOD and the laboratory strains examined, might confer functional significance in the implication of B cells in the autoimmune processes.

These observations convey i) that polymorphisms within the coding or non-coding regions of these genes are not expected to contribute to phenotypic differences between the NOD and other strains at least for the 45% of the E-IAA PLN transcripts for which no SNPs have been identified (Fig. 4). Therefore gene expression variations at least for these genes, might be controlled at the post-transcriptional level or by gene products acting in *cis* (mapped in proximity) or in *trans* (from other loci), including loci carrying SNP variants, or alternatively by chromatin modifications (epigenetic mechanisms). ii) Four regions (loci II and IV on chr6 and II and the IPG containing region on chr7) showed an unusual dense SNP frequency between the laboratory strains studied (Table 3). In particular the region IV on chr6 maps downstream to the chr6 locus II, that contains the immunoglobulin genes and immediate

upstream of the *Reg* gene cluster (locus III) that contains few SNPs (Fig. 3 and Table 3) and it may represent a recombination hot spot of chr6. It is likely that the existence of molecular perturbations created by associated polymorphic alleles influence the expression of these genes.

These observations elicit the hypothesis that a possible key-regulatory functional significance may be attributed to these loci and may represent potential "expression variants" hotspots, possibly controlling 'local' or 'distant' immune or cellular mechanisms implicated in the initiation of T1D. Additional studies, functional or NGS (next generation sequencing) are required in order to assert the relevance of these data and observations.

4. Conclusions

Here we report a complete genomic image of a set of genes regulated upon the presence or absence of insulin auto antibodies in the early stages of T1D in the NOD mouse. The genetic architecture of these transcripts and the modes of their expression in the PLN converge towards a complex interplay of factors in the genetics of T1D. SNP polymorphisms, chromosome environment within gene clusters, together with genetic modifiers controlling gene expression in *cis-* or *trans-* sway the variation of gene expression at a specific disease steady state, translated in our study by the E-IAA sub phenotype.

Further studies taking in consideration the different levels of possible gene regulation, including epistatic interactions between genes and regulation by various classes of RNA molecules including transRNAs harboring SNP sequences and miRNAs, may lead to the discovery that a smaller number of genes than was initially estimated are indeed implicated in disease pathogenesis. Moreover epistatic interactions may prompt our understanding of the epigenetic control throughout the influence of pathogenic microorganisms as environmental triggers, in an autoimmune predisposed genetic background.

Conflict of interests

Acknowledgments

References

Airaghi L, Tedeschi A (2006) Negative association between occurrence of type 1 diabetes and tuberculosis incidence at population level. Acta Diabetol. 43: 43–45

Amos, W., 2010. Even small SNP clusters are non-randomly distributed: is this evidence of mutational non-independence? Proc. Biol. Sci. 277, 1443–1449.

Bach, J.F., Chatenoud, L., 2012. The hygiene hypothesis: an explanation for the increased frequency of insulin-dependent diabetes. Cold Spring Harb. Perspect Med. 2 (a007799).

Barrett, J.C., Clayton, D.G., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H.A., Julier, C., Morahan, G., Nerup, J., Nierras, C., Plagnol, V., Pociot, F., Schuilenburg, H., Smyth, D.J., Stevens, H., Todd, J.A., Walker, N.M., Rich, S.S., 2009. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat. Genet. 41, 703–707.

Beechey, C.V., 2004. A reassessment of imprinting regions and phenotypes on mouse chromosome 6: Nap1l5 locates within the currently defined sub-proximal imprinting region. Cytogenet. Genome Res. 107, 108–114.

Bell, G.I., Karam, J.H., Rutter, W.J., 1981. Polymorphic DNA region adjacent to the 5′ end of the human insulin gene. Proc. Natl. Acad. Sci. U. S. A. 78, 5759–5763.

Bell, G.I., Horita, S., Karam, J.H., 1984. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes 33, 176–183.

Bhoola, K.D., Elson, C.J., Dieppe, P.A., 1992a. Kinins—key mediators in inflammatory arthritis? Br. J. Rheumatol. 31, 509–518.

Bhoola, K.D., Figueroa, C.D., Worthy, K., 1992b. Bioregulation of kinins: kallikreins, kininogens, and kininases. Pharmacol. Rev. 44, 1–80.

Blake, J.A., Bult, C.J., Eppig, J.T., Kadin, J.A., Richardson, J.E., 2014. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. Nucleic Acids Res. 42, D810–D817.

Boado, R.J., Pardridge, W.M., 1998. Ten nucleotide cis element in the 3′-untranslated region of the GLUT1 glucose transporter mRNA increases gene expression via mRNA stabilization. Brain Res. Mol. Brain Res. 59, 109–113.

Brekke, K.M., Garrard, W.T., 2004. Assembly and analysis of the mouse immunoglobulin kappa gene sequence. Immunogenetics 56, 490–505.

Burren, O.S., Adlem, E.C., Achuthan, P., Christensen, M., Coulson, R.M., Todd, J.A., 2011. T1DBase: update 2011, organization and presentation of large-scale data sets for type 1 diabetes research. Nucleic Acids Res. 39, D997–1001.

Cotsapas, C., Hafler, D.A., 2013. Immune-mediated disease genetics: the shared basis of pathogenesis. Trends Immunol. 34, 22–26.

Dendrou, C.A., Plagnol, V., Fung, E., Yang, J.H., Downes, K., Cooper, J.D., Nutland, S., Coleman, G., Himsworth, M., Hardy, M., Burren, O., Healy, B., Walker, N.M., Koch, K., Ouwehand, W.H., Bradley, J.R., Wareham, N.J., Todd, J.A., Wicker, L.S., 2009. Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. Nat. Genet. 41, 1011–1015.

Dennis Jr., G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4, P3.

Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M., Lathrop, G.M., Abecasis, G.R., Cookson, W.O., 2007. A genome-wide association study of global gene expression. Nat. Genet. 39, 1202–1207.

Doherty, P.C., Zinkernagel, R.M., 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256, 50–52.

Edgar, R., Domrachev, M., Lash, A.E., 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 30, 207–210.

Eisenbarth, G.S., Gianani, R., Yu, L., Pietropaolo, M., Verge, C.F., Chase, H.P., Redondo, M.J., Colman, P., Harrison, L., Jackson, R., 1998. Dual-parameter model for prediction of type I diabetes mellitus. Proc. Assoc. Am. Physicians 110, 126–135.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., Mouy, M., Steinthorsdottir, V., Eiriksdottir, G.H., Bjornsdottir, G., Reynisdottir, I., Gudbjartsson, D., Helgadottir, A., Jonasdottir, A., Styrkarsdottir, U., Gretarsdottir, S., Magnusson, K.P., Stefansson, H., Fossdal, R., Kristjansson, K., Gislason, H.G., Stefansson, T., Leifsson, B.G., Thorsteinsdottir, U., Lamb, J.R., Gulcher, J.R., Reitman, M.L., Kong, A., Schadt, E.E., Stefansson, K., 2008. Genetics of gene expression and its effect on disease. Nature 452, 423–428.

Evans, B.A., Drinkwater, C.C., Richards, R.I., 1987. Mouse glandular kallikrein genes. Structure and partial sequence analysis of the kallikrein gene locus. J. Biol. Chem. 262, 8027–8034.

Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Sun, W., Wang, H., Wang, Y., Xiong, X., Xu, L., Waye, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Song, Y.Q., Tam, P.K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Schaffner, S.F., Sham, P.C., Varilly, P., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Yakub, I., Birren, B.W., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archeveque, P., Bellemare, G., Saeki, K., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., JE, M.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., Stewart, J., 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851–861.

Glinskii, A.B., Ma, J., Ma, S., Grant, D., Lim, C.U., Sell, S., Glinsky, G.V., 2009. Identification of intergenic trans-regulatory RNAs containing a disease-linked SNP sequence and targeting cell cycle progression/differentiation pathways in multiple common human disorders. Cell Cycle 8, 3925–3942.

Glinskii, A.B., Ma, S., Ma, J., Grant, D., Lim, C.U., Guest, I., Sell, S., Buttyan, R., Glinsky, G.V., 2011a. Networks of intergenic long-range enhancers and snpRNAs drive castration-resistant phenotype of prostate cancer and contribute to pathogenesis of multiple common human disorders. Cell Cycle 10, 3571–3597.

Glinsky, G.V., 2008a. Disease phenocode analysis identifies SNP-guided microRNA maps (MirMaps) associated with human "master" disease genes. Cell Cycle 7, 3680–3694.

Glinsky, G.V., 2008b. An SNP-guided microRNA map of fifteen common human disorders identifies a consensus disease phenocode aiming at principal components of the nuclear import pathway. Cell Cycle 7, 2570–2583.

Goring, H.H., Curran, J.E., Johnson, M.P., Dyer, T.D., Charlesworth, J., Cole, S.A., Jowett, J.B., Abraham, L.J., Rainwater, D.L., Comuzzie, A.G., Mahaney, M.C., Almasy, L., MacCluer, J.W., Kissebah, A.H., Collier, G.R., Moses, E.K., Blangero, J., 2007. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat. Genet. 39, 1208–1216.

Ishida, D., Su, L., Tamura, A., Katayama, Y., Kawai, Y., Wang, S.F., Taniwaki, M., Hamazaki, Y., Hattori, M., Minato, N., 2006. Rap1 signal controls B cell receptor repertoire and generation of self-reactive B1a cells. Immunity 24, 417–427.

Jain, N., Thatte, J., Braciale, T., Ley, K., O'Connell, M., Lee, J.K., 2003. Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays. Bioinformatics 19, 1945–1951.

Jamal, A.M., Lipsett, M., Sladek, R., Laganiere, S., Hanley, S., Rosenberg, L., 2005. Morphogenetic plasticity of adult human pancreatic islets of Langerhans. Cell Death Differ. 12, 702–712.

Jankovic, M., Nussenzweig, M.C., 2003. OcaB regulates transitional B cell selection. Int. Immunol. 15, 1099–1104.

Jin, Y., Mailloux, C.M., Gowan, K., Riccardi, S.L., LaBerge, G., Bennett, D.C., Fain, P.R., Spritz, R.A., 2007. NALP1 in vitiligo-associated multiple autoimmune disease. N. Engl. J. Med. 356, 1216–1225.

Klöting, I., Schmidt, S., Kovacs, P., 1998. Mapping of novel genes predisposing or protecting diabetes development in the BB/OK rat. Biochem. Biophys. Res. Commun. 245, 483–486.

Kodama, K., Butte, A.J., Creusot, R.J., Su, L., Sheng, D., Hartnett, M., Iwai, H., Soares, L.R., Fathman, C.G., 2008. Tissue- and age-specific changes in gene expression during disease induction and progression in NOD mice. Clin. Immunol. 129, 195–201.

Leiter, E.H., Prochazka, M., Shultz, L.D., 1987. Effect of immunodeficiency on diabetogenesis in genetically diabetic (db/db) mice. J. Immunol. 138, 3224–3229.

Li, J., Wong, L., 2001. Emerging patterns and gene expression data. Genome Inform. 12, 3–13.

Martin, L., Chang, H.Y., 2012. Uncovering the role of genomic "dark matter" in human disease. J. Clin. Investig. 122, 1589–1595.

Martin, A.M., Blankenhorn, E.P., Maxson, M.N., Zhao, M., Leif, J., Mordes, J.P., Greiner, D.L., 1999. Non-major histocompatibility complex-linked diabetes susceptibility loci on chromosomes 4 and 13 in a backcross of the DP-BB/Wor rat to the WF rat. Diabetes 48, 50–58.

McClive, P.J., Baxter, A.G., Morahan, G., 1994. Genetic polymorphisms of the non-obese diabetic (NOD) mouse. Immunol. Cell Biol. 72, 137–142.

Mefford, H.C., Eichler, E.E., 2009. Duplication hotspots, rare genomic disorders, and common disease. Curr. Opin. Genet. Dev. 19, 196–204.

Melanitou, E., 2005. Functional genomics in early autoimmunity. Ann. N. Y. Acad. Sci. 1050, 64–72.

Melanitou, E., Devendra, D., Liu, E., Miao, D., Eisenbarth, G.S., 2004. Early and quantal (by litter) expression of insulin autoantibodies in the nonobese diabetic mice predict early diabetes onset. J. Immunol. 173, 6603–6610.

Melanitou, E., Joly, F., Lathrop, M., Boitard, C., Avner, P., 1998. Evidence for the presence of insulin-dependent diabetes-associated alleles on the distal part of mouse chromosome 6. Genome. Res. 8, 608–620.

Melanitou, E., Tekaia, F., Yeramian, E., 2013. Investigation of secreted protein transcripts as early biomarkers for type 1 diabetes in the mouse model. Gene 512, 161–165.

Mordes, J.P., Cort, L., Norowski, E., Leif, J., Fuller, J.M., Lernmark, A., Greiner, D.L., Blankenhorn, E.P., 2009. Analysis of the rat Iddm14 diabetes susceptibility locus in multiple rat strains: identification of a susceptibility haplotype in the Tcrb-V locus. Mamm. Genome. 20, 162–169.

Murrell, A., Heeson, S., Reik, W., 2004. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat. Genet. 36, 889–893.

Pelizzola, M., Pavelka, N., Foti, M., Ricciardi-Castagnoli, P., 2006. AMDA: an R package for the automated microarray data analysis. BMC Bioinformatics 7, 335.

Pelletier, A.N., Guimont-Desrochers, F., Ashton, M.P., Brodnicki, T.C., Lesage, S., 2012. The size of the plasmacytoid dendritic cell compartment is a multigenic trait dominated by a locus on mouse chromosome 7. J. Immunol. 188, 5561–5570.

Peterson, J.D., Pike, B., McDuffie, M., Haskins, K., 1994. Islet-specific T cell clones transfer diabetes to nonobese diabetic (NOD) F1 mice. J. Immunol. 153, 2800–2806.

Pociot, F., Akolkar, B., Concannon, P., Erlich, H.A., Julier, C., Morahan, G., Nierras, C.R., Todd, J.A., Rich, S.S., Nerup, J., 2010. Genetics of type 1 diabetes: what's next? Diabetes 59, 1561–1571.

Regnault, B., Osorio, Y.F.J., Miao, D., Eisenbarth, G., Melanitou, E., 2009. Early over expression of messenger RNA for multiple genes, including insulin, in the pancreatic lymph nodes of NOD mice is associated with islet autoimmunity. BMC Med. Genet. 2, 63.

Ridgway, W.M., Peterson, L.B., Todd, J.A., Rainbow, D.B., Healy, B., Burren, O.S., Wicker, L.S., 2008. Gene-gene interactions in the NOD mouse model of type 1 diabetes. Adv. Immunol. 100, 151–175.

Riegman, P.H., Vlietstra, R.J., Suurmeijer, L., Cleutjens, C.B., Trapman, J., 1992. Characterization of the human kallikrein locus. Genomics 14, 6–11.

Rockman, M.V., Skrovanek, S.S., Kruglyak, L., 2010. Selection at linked sites shapes heritable phenotypic variation in C. elegans. Science 330, 372–376.

Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., Cotsapas, C., Daly, M.J., 2011. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 7, e1001273.

Schwensow, N., Fietz, J., Dausmann, K.H., Sommer, S., 2007. Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. Heredity (Edinb) 99, 265–277.

Shmela, M.E., Gicquel, C.F., 2013. Human diseases versus mouse models: insights into the regulation of genomic imprinting at the human 11p15/mouse distal chromosome 7 region. J. Med. Genet. 50, 11–20.

Steinman, R.M., 2007. Dendritic cells: understanding immunogenicity. Eur. J. Immunol. 37 (Suppl 1), S53–S60.

Stevenson, B.J., Hagenbuchle, O., Wellauer, P.K., 1986. Sequence organisation and transcriptional regulation of the mouse elastase II and trypsin genes. Nucleic Acids Res. 14, 8307–8330.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M.P., Walker, J.R., Hogenesch, J.B., 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. U. S. A. 101, 6062–6067.

Suzuki, T., Moritani, M., Yoshino, M., Kagami, M., Iwasaki, S., Nishimura, K., Akamatsu, M., Kobori, M., Matsushime, H., Kotoh, M., Furuichi, K., Itakura, M., 2008. Diabetic modifier QTLs in F(2) intercrosses carrying homozygous transgene of TGF-beta. Mamm. Genome 19, 15–25.

Tian, D., Wang, Q., Zhang, P., Araki, H., Yang, S., Kreitman, M., Nagylaki, T., Hudson, R., Bergelson, J., Chen, J.Q., 2008. Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. Nature 455, 105–108.

Tisch, R., Wang, B., 2009. Role of plasmacytoid dendritic cells in type 1 diabetes: friend or foe? Diabetes 58, 12–13.

Torn, C., Hadley, D., Lee, H.S., Hagopian, W., Lernmark, A., Simell, O., Rewers, M., Ziegler, A., Schatz, D., Akolkar, B., Onengut-Gumuscu, S., Chen, W.M., Toppari, J., Mykkanen, J., Ilonen, J., Rich, S.S., She, J.X., Steck, A.K., Krischer, J., 2015. Role of type 1 diabetes-associated SNPs on risk of autoantibody positivity in the TEDDY study. Diabetes 64, 1818–1829.

Verge, C.F., Stenger, D., Bonifacio, E., Colman, P.G., Pilcher, C., Bingley, P.J., Eisenbarth, G.S., 1998. Combined use of autoantibodies (IA-2 autoantibody, GAD autoantibody, insulin autoantibody, cytoplasmic islet cell antibodies) in type 1 diabetes: combinatorial islet autoantibody workshop. Diabetes 47, 1857–1866.

Veyrieras, J.B., Kudaravalli, S., Kim, S.Y., Dermitzakis, E.T., Gilad, Y., Stephens, M., Pritchard, J.K., 2008. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet. 4, e1000214.

Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G.D., Morris, Q., 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 38, W214–W220.

Yu, L., Robles, D.T., Abiru, N., Kaur, P., Rewers, M., Kelemen, K., Eisenbarth, G.S., 2000. Early expression of antiinsulin autoantibodies of humans and the NOD mouse: evidence for early determination of subsequent diabetes. Proc. Natl. Acad. Sci. U. S. A. 97, 1701–1706.

Yuan, R.H., Jeng, Y.M., Chen, H.L., Hsieh, F.J., Yang, C.Y., Lee, P.H., Hsu, H.C., 2005. Opposite roles of human pancreatitis-associated protein and REG1A expression in hepatocellular carcinoma: association of pancreatitis-associated protein expression with low-stage hepatocellular carcinoma, beta-catenin mutation, and favorable prognosis. Clin. Cancer Res. 11, 2568–2575.