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 R T I C L E  I N F O

Article history:
Received 9 June 2015
Revised 11 September 2015
Accepted 18 September 2015
Available online 22 October 2015

Keywords:
Transcriptome
Type 1 diabetes
Genomics
Pancreatic lymph nodes
Polymorphisms

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 subphenotypic 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.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

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
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 (Glinkski 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 a stage of autoimmunity, 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 focus-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 generality of the genome in terms of transcripts differentially expressed in the early inflammatory stages of T1D and evaluate the venous possibly uncovered by such images, 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-IAAneg and five E-IAApos. 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 profile 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 Average (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.
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 (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/ShiLt 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.

<table>
<thead>
<tr>
<th>CHR</th>
<th>E-IAA</th>
<th>PLN</th>
<th>AFFY</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr7</td>
<td>24</td>
<td>19</td>
<td>818</td>
</tr>
<tr>
<td>chr6</td>
<td>16</td>
<td>33</td>
<td>622</td>
</tr>
<tr>
<td>chr3</td>
<td>13</td>
<td>25</td>
<td>553</td>
</tr>
<tr>
<td>chr8</td>
<td>12</td>
<td>22</td>
<td>510</td>
</tr>
<tr>
<td>chr14</td>
<td>11</td>
<td>32</td>
<td>318</td>
</tr>
<tr>
<td>Total gene n° on chr</td>
<td>76 (47%)</td>
<td>131 (25%)</td>
<td>2821 (27%)</td>
</tr>
<tr>
<td>Total n° of genes</td>
<td>162</td>
<td>529</td>
<td>10,417</td>
</tr>
</tbody>
</table>

* 529 out of 563 PLN-expressed genes have known chromosome localizations.

### 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 < 0.00001 and E-IAA vs AFFY: P < 0.0001.

<table>
<thead>
<tr>
<th>Gene set</th>
<th>SPG n° (%)</th>
<th>IPG n° (%)</th>
<th>Total gene n°</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-IAA</td>
<td>74 (46%)</td>
<td>88 (54%)</td>
<td>162</td>
</tr>
<tr>
<td>PLN</td>
<td>49 (9%)</td>
<td>480 (91%)</td>
<td>529*</td>
</tr>
<tr>
<td>AFFY</td>
<td>2031 (19.5%)</td>
<td>8386 (80.5%)</td>
<td>10,417</td>
</tr>
</tbody>
</table>

* 529 out of 563 PLN-expressed genes have known cellular 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).

![Fig. 2.](image-url) 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).
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 further 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 = 0.0030 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]), 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-IAApos 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-IAApos 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 (Riezman et al., 1992; Stevenson et al., 1994).
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 (lkgv4-53, lkgc), within a genetic interval of 0.5 cM, all down regulated in the E-IAApos 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-IAApos 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, is as 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 regulating islet-derived transcripts, all highly up-regulated in the E-IAApos 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 tool to prevent islet duct cell by releasing cytokines in pDC. Moreover, on proximal and distal mouse chr6 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, NOD, SJLJ 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/Shi versus C57Bl/6, C3H, DBA/2, SJL/J and NON/Shi. (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).

<table>
<thead>
<tr>
<th>Genomic region</th>
<th>Genome Coordinates (bp) (region size in Kbp)</th>
<th>Gene n° within locus</th>
<th>SNP in Lab. vs all strains (n°)</th>
<th>SNP density1 (Lab strains)</th>
<th>SNP2 (N°/Total N° of SNPs)</th>
<th>Genes within segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr6, locus I</td>
<td>40914771–41475078 (560.3)</td>
<td>60 &amp; 9 QTLs=51</td>
<td>7/1189</td>
<td>80 kb</td>
<td>1/7</td>
<td>T cell receptor beta variable chains locus (Trvb genes) &amp; Protease serine (trypsin, Prss1, Prss2, Prss3, Try10, Try5, Try4 &amp; QTL Diabetes Dblt1)</td>
</tr>
<tr>
<td>chr6, locus II</td>
<td>67505630–70676748 (3.171.1)</td>
<td>191</td>
<td>4243/7503</td>
<td>0.74 kb</td>
<td>0/4243</td>
<td>Immunoglobulin kappa chain complex (igk genes, Jkbl1, Igkv4-5-3)</td>
</tr>
<tr>
<td>chr6, locus IV interregion</td>
<td>70676748 – 73208789 (7.644.1)</td>
<td>122</td>
<td>10222/25350</td>
<td>0.74 kb</td>
<td>347/10222</td>
<td>Rpia, Eif2ak3, Sry, Rpl19–ps9, Reep1, Mrpl35, Immn, Gm17854, Tim3b10, Dnak6, Surl1, Cima2, Gm5576</td>
</tr>
<tr>
<td>chr6, locus III</td>
<td>78320879–78418868 (98)</td>
<td>6</td>
<td>3/315</td>
<td>32.6 kb</td>
<td>1/3: Reg3g</td>
<td>Rpia, Reg3d, Reg3a, Reg3c, Reg3b, Reg3a</td>
</tr>
<tr>
<td>chr7, locus I (16.83 cM)</td>
<td>2854088–28542210 (1.32)</td>
<td>1</td>
<td>0/2</td>
<td>No SNPs</td>
<td>None</td>
<td>Sycn</td>
</tr>
<tr>
<td>chr7, locus II (28.45–28.74 cM)</td>
<td>4423612–43976661 (260)</td>
<td>27</td>
<td>414/819</td>
<td>0.62 kb</td>
<td>36/414</td>
<td>Klk1, Klk1b5, Klk1b10–ps, Klk1b18–ps, Klk1b27, Klk1b14–ps, Klk1b121, Klk1b22, Klk1b23–ps, Klk1b15, Klk1b124, Klk1b103, Klk1b14, Klk1b50, Gm10109, Gm1k1, Gm6002b, Gm10002b, Gm10003b, QTLs: Drs1, SLE19</td>
</tr>
<tr>
<td>chr7, locus III (73.2–73.81 cM)</td>
<td>131032076–131322292 (290)</td>
<td>2</td>
<td>3/841</td>
<td>No SNPs</td>
<td>None</td>
<td>Gm1012, Fglr2, Cazd1</td>
</tr>
<tr>
<td>chr7, locus IV (85.94–88 cM)</td>
<td>140763832–142699510 (1935)</td>
<td>95</td>
<td>9/5386</td>
<td>215 Kb</td>
<td>1/9</td>
<td>Gycp2e1, Sycn, Zfp941, Trf0009c368, C310022C248k, Scgb1c1, Dcl1, Bhlh, Rich, Sst3, Pomp13, Gm15542, Cosbb, Kcgu2a86, Nitr6, Abhl1, Ifitm5, Ifitm1, Ifitm2, Gm7514, Ifitm3, Tgodm, Kpg3, Siger, An09, Ped2a, Roh1, Hras1, 1600016N02Rik, Rasfl, Pbrf1, Igf1, Cadh5, Drd3, Deaf1, Eps8d2, Tmem80, B230006H08Rik, Taldo1, Pdcl1, Cerd1, Slc25a22, Rplp2, Pmp2a, Cd915, Pehl2, Eparp, Chd1b, Ap2a2, Musc, Musc3, Nucub, Tenp, Gm6449, Brk2, Mob2, Gm3291, Gm9340, Gm4552, Gm4553, Krtap5–2, Gm19053, Krtap5–3, Gm7579, Gm19094, Krtap5–5, Gm7208, Gm4559, B1000160D09Rik, Krtap5–1, Krtap5–6, 630012M04Rik, Cctb, Syt8, Tms2, Lgpl1, Lgpl3, Mrpl31, H19, Igf2, Igf2a, Igf2b, Igf3, Igf5, Igf6, Mir210 &amp; Mir675 (microRNA 210), Mir675 (microRNA 675), Mir483</td>
</tr>
<tr>
<td>chr7 IPG locus (53.49–69.21 cM)</td>
<td>97417730–126627010 (29209)</td>
<td>969 genes, pseudogenes, QTLs, miRNAs</td>
<td>34596/94480</td>
<td>0.84 Kb</td>
<td>Total in coding regions: 1938/345963</td>
<td>Idd27, Health, wound healing/regeneration 6, Ssa13, susceptibility to sialadenitis 3, Eae26, experimental allergic encephalomyelitis susceptibility 26, Idrpp, Hbb–b1, Hbb–b2, Tm31, Ins2, Cebpa, Hsehi, Ifer, Cyclin 1, Ifer, Cebpa, Krtap5–2, Gm19053, Krtap5–3, Gm7579, Gm19094, Krtap5–5, Gm7208, Gm4559, B1000160D09Rik, Krtap5–1, Krtap5–6, 630012M04Rik, Cctb, Syt8, Tms2, Lgpl1, Lgpl3, Mrpl31, H19, Igf2, Igf2a, Igf3, Igf5, Igf6, Mir210 &amp; Mir675 (microRNA 210), Mir675 (microRNA 675), Mir483</td>
</tr>
</tbody>
</table>

Notes to Table 3
1 Distance (kb) between SNPs (Lab strains only).
2 Number of SNPs in coding regions (mRNA-UTR, nsSNPs, sSNPs) only.

---

B. Regnault, E. Melanitou / Meta Gene 6 (2015) 96–104
A “genetical genomics” image summarizes the correspondence of diabetes-related loci and the relevant to these loci associated E-I AA- expressed genes (Table S3).

Several E-I AA genes are located nearby or within regions defined by the presence of ldd loci identified in mouse crosses. This is the case for ldd6 and ldd19 on chr6 (Table S3) (Melanitou et al., 1998). Similarly, lddm14 and lddm11 (on chr7) (Mordes et al., 2009; Martin et al., 1999), lddm10 (on chr7) correspond (Kloting et al., 1998) to synthetic rat lddm 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-I AA 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-I AApos 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 Halder, 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-I AA 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; McClave et al., 1994; Peterson et al., 1994; Suzuki et al., 2008) and assessed for polymorphisms between these strains (C57Bl/6, C3H, DBA/2, SJLJ and Non/LtJ), with as reference, the NOD/ShiLtJ.

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-I AA 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, Sel11, Rnase1, and Tfj2 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/ShiLtJ 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; Schwenso 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 (Breker 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-I Aapos 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-I AA 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 set of genes involved in the disease steady state, translated in our study by the E-IAA sub phenotype. Further studies taking into consideration the different levels between possible gene regulation, including epistatic interactions between genes and partially sequenced arrays of RNA molecules including microRNAs harboring SNP sequences and miRNAs, may lead to the discovery that a smaller number of genes than that 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
The authors declare no conflict of interest.

Supplementary data
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jgene.2015.09.003.

Acknowledgments
This work has been supported by the Institut Pasteur, Paris, France and the Barbara Davis Center for Childhood Diabetes, University of Colorado, Denver, CO, USA.

References


