Journal of Autoimmunity xxx (2018) 1–11

Contents lists available at ScienceDirect

Journal of Autoimmunity

journal homepage: www.elsevier.com/locate/jautimm

T cells are influenced by a long non-coding RNA in the autoimmune associated *PTPN2* locus

Miranda Houtman [a, *], Klementy Shchetynsky [a], Karine Chemin [a], Aase Haj Hensvold [a], Daniel Ramsköld [a], Karolina Tandre [b], Maija-Leena Eloranta [b], Lars Rönnblom [b], Steffen Uebe [c], Anca Irinel Catrina [a], Vivianne Malmström [a], Leonid Padyukov [a]

[a] Rheumatology Unit, Department of Medicine, Karolinska University Hospital Solna, Karolinska Institutet, Stockholm, Sweden
[b] Department of Medical Sciences, Rheumatology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden
[c] Institute of Human Genetics, University of Erlangen-Nuremberg, Erlangen, Germany

ARTICLE INFO

Article history:
Received 5 December 2017
Received in revised form
10 January 2018
Accepted 11 January 2018
Available online xxx

Keywords:
RA
Genome-wide association
methQTL
eQTL
Long non-coding RNA
T cell

ABSTRACT

Non-coding SNPs in the protein tyrosine phosphatase non-receptor type 2 (*PTPN2*) locus have been linked with several autoimmune diseases, including rheumatoid arthritis, type I diabetes, and inflammatory bowel disease. However, the functional consequences of these SNPs are poorly characterized. Herein, we show in blood cells that SNPs in the *PTPN2* locus are highly correlated with DNA methylation levels at four CpG sites downstream of *PTPN2* and expression levels of the long non-coding RNA (lncRNA) *LINC01882* downstream of these CpG sites. We observed that *LINC01882* is mainly expressed in T cells and that anti-CD3/CD28 activated naïve $CD4^+$ T cells downregulate the expression of *LINC01882*. RNA sequencing analysis of *LINC01882* knockdown in Jurkat T cells, using a combination of antisense oligonucleotides and RNA interference, revealed the upregulation of the transcription factor *ZEB1* and kinase *MAP2K4*, both involved in IL-2 regulation. Overall, our data suggests the involvement of *LINC01882* in T cell activation and hints towards an auxiliary role of these non-coding SNPs in autoimmunity associated with the *PTPN2* locus.

1. Introduction

Deregulation of immune response is a hallmark of autoimmune diseases, which likely involves multiple immune mechanisms. T cell-related regulation is often considered to be important for autoimmune diseases, although it is hard to distinguish causality in a sequence of T and B cell-related immunological events during the course of disease development. Genetic association studies inform about possible causal germ line variations that regulate cellular phenotypes and could precede development of clinical symptoms. Therefore, the study of functional consequences of disease-associated genetic variants is important for understanding disease mechanisms and for discovering new drug targets. However, the annotation of signals to the closest gene in genome-wide association (GWA) studies has often been proven inaccurate and sometimes even misleading [1–4]. It requires broad assessment of the function of genes sometimes relatively distant from the association hit. This is in part due to linkage disequilibrium between genetic markers, but is also due to complex relation between genomic sequences and chromatin structure.

Several diseases associate with genetic variations at the locus where the protein tyrosine phosphatase 2 (*PTPN2*) is located, including type 1 diabetes (T1D) [5], juvenile idiopathic arthritis (JIA) [6], inflammatory bowel disease (IBD) [7], and rheumatoid arthritis (RA) [8]. PTPN2 has been shown to act on several phosphorylated proteins to regulate cytokine signaling [9]. The most strongly associated SNPs in the *PTPN2* locus in RA are the intronic SNPs rs2542151, rs1893217, rs62097857, rs7234029, and rs8083786 [8,10–15]. In addition, the intronic SNPs rs657555 and rs2847297 were associated with RA in a Korean and Japanese population, respectively [16,17]. Even though multiple SNPs within the *PTPN2* locus have been discovered, the functional consequences of these variations in the pathogenesis of RA remain unclear.

The majority of associated variants identified by GWA studies lie within non-coding regions [18]. Various studies have shown that

* Corresponding author. Rheumatology Unit, CMM L8:04, Karolinska University Hospital, 171 76, Stockholm, Sweden.
E-mail address: miranda.houtman@ki.se (M. Houtman).

https://doi.org/10.1016/j.jaut.2018.01.003

genetic variations within non-coding regions may cause different consequences for regulating genes. Intronic variations might be affecting alternative splicing, chromatin modifications, gene expression, and DNA methylation levels [2,19,20]. Also in many cases, the associated SNPs seem to act through the disruption of transcription factor binding sites [21].

In this study, we investigated the autoimmune disease associated *PTPN2* locus. Our results show that SNPs in the *PTPN2* locus are associated with changes in levels of DNA methylation at four CpG sites 7.5 kb downstream of *PTPN2* and changes in expression of the long non-coding RNA (lncRNA) *LINC01882*. Further, silencing of this lncRNA suggested a transcription regulation pathway in T cells that could be involved in RA and other autoimmune diseases where *PTPN2* SNPs are known to be associated, thus providing a functional link between GWA hits and disease.

2. Materials and methods

2.1. Study approval

This study was performed with ethical approval from the local Ethical Boards according to the Swedish law and informed consent was received from participants included in the study. All RA cases met the American College of Rheumatology 1987 criteria for RA [22].

2.2. DNA methylation

Methylation data for whole blood genomic DNA was obtained from our previous study (GSE42861) [23]. In short, bisulphite converted DNA from whole blood from 335 healthy controls and 354 ACPA-positive RA patients were analyzed using the Illumina HumanMethylation450 BeadChip Array. These samples were genotyped with the Illumina HumanHap300 Array [24]. SNP imputation was done with the 1000G project data as a reference. Only SNPs with $\geq$95% posterior probability, MAF $\geq$0.01, missing rate $\leq$0.05 and HWE *P*-value $\geq$ .001 were used for the analyses. 417 normalized CpG sites 1 Mb up- and downstream of the *PTPN2* gene (11,785,478−13,929,643 bp, GRCh38) were tested for association with 37 SNPs (peak signals (rs1893217, rs11875687, rs62097857, rs2847293, rs12971201, rs8083786, and rs2542151) and SNPs in high LD; $r^2 > 0.8$) from the *PTPN2* locus. The R package GEM (release 3.5) [25] was used to identify methQTLs using the covariates gender, age, and smoking status.

2.3. eQTL

Expression data for PBMCs was obtained from our previous COMBINE study [26]. In short, RNA of three patient groups and healthy individuals was purified using isopropanol extraction and sequenced using the TruSeq RNA sample preparation kit (2 × 100 bp) on an Illumina HiSeq 2000 sequencer (Aros Applied Biotechnology center). After quality filtering of reads and adaptor removal, TopHat2 (version 2.0.10) was used to align reads to the hg19 assembly. Gene expression was quantified using HTSeq followed by TMM-normalization, mean-scaling and log2 transformation. Genotyping was performed using the HumanOmniExpress BeadChip Kit and Illumina OmniExpress arrays (12v1). A linear mixed model (nlme package (version 3.1−13.1)) was used to calculate the associations of SNPs and gene expression levels, with gender (and patient groups) as a fixed effect and study individuals as a random effect.

2.4. Gene-gene interactions

Gene-gene interactions were assessed by testing for additive interactions (departure from additivity) using the GEIRA algorithm (http://www.epinet.se) [27] in R (version 3.2.0). SNPs in *PTPN2* locus (including 5 kb flanking regions (12790478−12889338 bp, GRCh38)) were extracted from EIRA [24] and NARAC GWA data [28]. Interactions were investigated by calculating the attributable proportion (AP) due to interaction with 95% confidence interval (95% CI) between 11 SNPS in the *PTPN2* region and *HLA-DRB1* SE using a dominant genetic model [27].

2.5. Sorting of PBMCs

PBMCs from healthy individuals (Uppsala Biobank) were isolated on a Ficoll gradient (GE Healthcare). CD4+ T cells, CD8+ T cells, CD14+ monocytes, and CD19+ B cells were isolated using magnetic beads on the autoMACS Pro Separator (Miltenyi Biotec). RNA was extracted using the RNeasy Mini kit (Qiagen). Sequencing libraries were prepared using the Illumina TruSeq stranded total RNA sample preparation kit with ribosomal depletion using RiboZero (2 × 125 bp) and analyzed on an Illumina HiSeq 2500 sequencer (SNP&SEQ Technology Platform, Uppsala, Sweden). At least 20 million reads were produced per sample. Pre-filtering on quality of reads using cutadapt (version 1.9.1) was applied (-q 30 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -A AGATCGGAAGA-GCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT -m 40). Filtered reads were aligned to the hg38 assembly and quantified using STAR (version 2.5.1b) [29] with default settings, followed by TMM-normalization, mean-scaling and log2 transformation.

2.6. Activation of naïve $CD4^+$ T cells

Expression data for naïve $CD4^+$ T cells and CD3/CD28 activated naïve $CD4^+$ T cells was obtained from GSE94859. In short, naïve $CD4^+$ T cells ($CD4^+$, CD45RA+, CCR7+) were sorted from PBMCs of healthy individuals using an influx sorter (BD) and activated using anti-CD3/CD28 beads (Dynabeads) for 16 h. RNA was extracted using the RNeasy Mini kit (Qiagen). Sequencing libraries were prepared using the Illumina TruSeq mRNA kit (poly(A) selection) on the Illumina HiSeq 2500 (SciLifeLab NGI Stockholm, Sweden). Reads were aligned against the hg38 genome using STAR and quantified using rpkmforgenes.py (version 13 March 2015).

2.7. Knockdown of *LINC01882*

Jurkat cells (LGC Standards) were kept in RPMI-1640 medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum (Sigma-Aldrich), L-glutamine (Gibco) and penicillin/streptomycin (Gibco) in a 37 °C incubator with 5% CO_2. Jurkat cells were transfected with a mixture (50 nM) of a dsiRNA (Integrated DNA Technologies) and an antisense LNA™ GapmeR (Exiqon) targeting *LINC01882* using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. The following target sequences were used: dsiRNA (5′-AAATAAAGGATTCTAGTTATGTGAA-3′) and antisense LNA™ GapmeR (5′-ACTGATCTGAGGAGAA-3′). A mixture of the dsiRNA NC1 and the antisense LNA™ GapmeR (5′-AACACGTCTATACGC-3′) was used as negative control. Cells were harvested 24 and 48 h after transfection. Total RNA was extracted with the RNeasy Mini kit (Qiagen) according to manufacturer's instructions. Samples were treated with DNase (Qiagen) for 20 min at room temperature to avoid contamination with genomic DNA. RNA was converted into cDNA using the iScript cDNA Synthesis Kit (Bio-Rad). Efficiency of knockdown was evaluated by qPCR.

2.8. Quantitative real-time PCR

qPCR quantification was carried out using iQ SYBR Green

Please cite this article in press as: M. Houtman, et al., T cells are influenced by a long non-coding RNA in the autoimmune associated *PTPN2* locus, Journal of Autoimmunity (2018), https://doi.org/10.1016/j.jaut.2018.01.003