

Enrichment of Genetic Variants for Rheumatoid Arthritis within T-Cell and NK-Cell Enhancer Regions

Jan Freudenberg,^{*} Peter Gregersen, and Wentian Li

Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, North Shore-LIJ Health System, Manhasset, New York, United States of America; ^{*}current affiliation: Regeneron Genetics Center, Tarrytown, NY, United States of America

To identify disease-causative variants, we intersected the published results of a metaanalysis of genome-wide association studies (GWAS) for rheumatoid arthritis (RA) with the set of enhancer regions for 71 primary cell types that was provided by the FANTOM consortium. We first retrieved all single nucleotide polymorphisms (SNPs) that are associated ($P < 5 \times 10^8$) with RA in the GWAS meta-analysis and that are located in any of these enhancer regions. After excluding the major histocompatibility complex (MHC) region, we identified 50 such RA-associated SNPs that are located in enhancer regions. Enhancer sets from different cell types were then compared with each other for their number of RA-associated SNPs by permutation analysis. This analysis showed that RA-associated SNPs are preferentially located in enhancers from several immunological cell types. In particular, we see a strong relative enrichment in enhancer regions that are active in T cells ($P < 0.001$) and NK cells ($P < 0.001$). Several loci display multiple RA-associated SNPs in tight linkage disequilibrium that are located within the same or neighboring enhancers. These haplotypes may have a greater likelihood to influence enhancer activity than any SNP on its own. Taken together, these results support the hypothesis that RA-causative variants often act through altering the activity of immune cell enhancers. The enrichment in T-cell and NK-cell enhancer regions indicates that expression changes in these cell types are particularly relevant for the pathogenesis of RA. The specific SNPs that account for this enrichment can be used as a basis for focused genotype-phenotype studies of these cell types.

Online address: <http://www.molmed.org>

doi: 10.2119/molmed.2014.00252

INTRODUCTION

Several genome-wide association studies (GWAS) have been undertaken to investigate the genetic basis of rheumatoid arthritis (RA) (1–7). These studies have been recently combined into a transcontinental metaanalysis (8), which most comprehensively evaluates the influence of common single nucleotide variants on RA susceptibility in populations with European and Asian ancestry. On the basis of the imputation into the 1000 Genomes data set, this

metaanalysis has generated association statistics for nearly 10 million single nucleotide polymorphisms (SNPs) across the human genome. The implicated genes were then further tested for their enrichment in molecular pathways. However, for many loci, it remains unclear which SNPs constitute the actual causal variants. In the present report, we examine this question by combining the RA GWAS metaanalysis results with a set of enhancer regions for 71 primary cell types, which was recently pub-

lished by the FANTOM consortium (9). These enhancer regions were detected by their noncoding RNA expression signature, that is, their localized expression of noncoding capped bidirectional short transcripts. Given that susceptibility variants typically exert their effect through modifying gene expression (10), this large enhancer set can be highly valuable to search for causal variants in current GWAS results. Moreover, using this set of enhancer regions to interpret RA GWAS results can point to cell types where gene activity may be altered through RA susceptibility variants and that therefore have important roles in RA pathogenesis.

MATERIALS AND METHODS

Transcontinental RA GWAS summary statistics were downloaded from the RIKEN genome center (<http://plaza.umin.ac.jp/~yokada/datasource/software.htm>; August 2014). Enhancer regions that are present across 71 human primary cell

Address correspondence to Jan Freudenberg, The Feinstein Institute for Medical Research, 350 Community Drive, Manhasset, NY 11030. Phone: 516-562-1542; Fax: 516-562-1153; E-mail: jan.freudenberg@nslj-genetics.org.

Submitted December 11, 2014; Accepted for publication March 14, 2015; Published Online (www.molmed.org) March 16, 2015.

The Feinstein Institute
for Medical Research 

Empowering Imagination. Pioneering Discovery.®

Table 1. Cell types of enhancer regions being enriched ($P < 0.05$) for RA-associated SNPs.

Cell type	Number of RA-associated SNPs	Number of loci	Cumulative size of enhancers/bp	RA-SNPs/Mb	<i>P</i>
T cell	33	15	1,954,612	16.88	0.0001
Natural killer cell	24	10	1,609,644	14.91	0.0004
Lymphocyte B lineage	11	6	876,029	12.56	0.008
Neutrophil	13	6	1,224,325	10.62	0.013
Basophil	18	8	2,224,115	8.09	0.018
Dendritic cell	19	10	2,457,374	7.73	0.036
Intestinal epithelial	6	2	407,986	14.71	0.039
Mast cell	15	8	1,792,867	8.37	0.044

bp, Base pair.

types were retrieved from the FANTOM website (http://enhancer.binf.ku.dk/presets/acet_expressed_enhancers.tgz; August 2014), and enhancer regions were defined by the genomic coordinates provided. Note that these enhancers constitute a superset of the cell type-specific enhancers that are displayed in the genome browser on this website. Out of the 9,739,303 (0.3%) SNPs being genotyped or imputed in the RA GWAS, 29,632 fall into any such enhancer. RA-associated SNPs were defined as all SNPs that obtained statistical significance ($P < 5 \times 10^{-8}$) in the transcontinental GWAS analysis or any of the two population-specific GWAS analyses (European or Asian). Excluding the MHC region, a total of 3,581 SNPs are associated with RA under this criterion (8). We excluded the MHC region, because the extensive linkage disequilibrium (LD) and the strong effects of specific coding variants make it difficult to attribute any role to other variants. Importantly, recent analyses indicate that coding region differences within the classical human leukocyte antigen (HLA) loci account for the vast majority of the association signals in the MHC region (12).

SNP genotypes for estimation of LD were downloaded from the 1000 Genomes consortium website (<http://www.1000genomes.org>; August 2014). Gene annotations were downloaded from the UCSC genome browser (<http://genome.ucsc.edu>; August 2014). GWAS metaanalysis results for schizo-

phrenia were used as a negative control and downloaded from the Psychiatric Genomics Consortium (<http://www.med.unc.edu/pgc/downloads>; August 2014). Haplotypes were estimated with the program snpnap as obtained from <http://www.gene.cimr.cam.ac.uk/staff/clayton/software> (August 2014). Pairwise LD was estimated with the program Haploview as obtained from <http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/downloads> (August 2014).

To test the enrichment of significant SNPs in enhancer regions across the different cell types, we randomly permuted 10,000 times the assignment of cell types to enhancers. Thus, each cell type was assigned a random set of enhancer regions, with its number of enhancers kept constant. We then counted the number of permutations for which an equal or greater number of significant SNPs was observed in the random enhancer set than in the actual enhancer set of a cell type. If multiple SNPs are located in enhancers at a same locus, each such SNP may increase the chance for an influence of genetic variation on enhancer activity, and it is therefore counted separately. This strategy compares the enrichment of RA-associated SNPs in different cell types relative to each other while accounting for the differences in the number of enhancers. The number of enhancer regions may differ across cell types because of bio-

logical factors or technical issues. Note that we do not intend to show here that RA-associated SNPs are generally enriched in FANTOM enhancers compared with the rest of the genome, which would require a different testing strategy. Thus, the null hypothesis represented by this simulation is that RA-associated SNPs are distributed randomly across the enhancers from different cell types. The statistical analysis was performed with the R software package (<http://www.r-project.org>; August 2014).

RESULTS

The FANTOM data set provides a list with 29,202 enhancer regions across 71 primary cell types, which jointly covers 9.1 Mb (~0.3 % of the genome). To find disease-causing enhancer variants, we retrieved all enhancer SNPs that are associated with RA ($P < 5 \times 10^{-8}$) in the transcontinental GWAS analysis or any of the two population-specific GWAS analyses (European or Asian). In total, we found 50 RA-associated SNPs that fall into enhancer regions, which are distributed over 21 different gene loci.

We then evaluated whether RA-associated SNPs are overrepresented in the enhancer regions of any specific cell types compared with other cell types. Notably, RA-associated SNPs are enriched in enhancer regions that are preferentially active in immunological cells (Table 1, Supplementary Table S1). In particular, there is an enrichment of RA-associated SNPs in T-cell enhancers (16.9 SNPs/Mb, $P < 0.001$) and NK-cell enhancers (14.9 SNPs/Mb, $P < 0.001$). RA-associated SNPs also tend to be enriched in B-cell enhancers (12.6 SNPs/Mb, $P = 0.008$), neutrophil enhancers (10.6 SNPs/Mb, $P = 0.013$), basophil cell enhancers (8.1 SNPs/Mb, $P = 0.018$), dendritic cell enhancers (7.7 SNPs/Mb, $P = 0.036$), intestinal epithelial enhancers (14.7 SNPs/Mb, $P = 0.039$) and mast cell enhancers (8.4 SNPs/Mb, $P = 0.044$). This enrichment in immunological cell types supports the notion

that some of these SNPs may be causal variants that modify cell-specific expression regulation of nearby genes. In total, there were 33 SNPs located in T-cell enhancer regions, and out these, 22 SNPs were also located in NK-cell enhancer regions. Thus, an overlapping set of variants accounts for the enrichment of RA-associated SNPs in T-cell and NK-cell enhancer regions.

As a negative control, we next used the set of 56 non-MHC SNPs that fall into any FANTOM enhancer region and that display a P value $< 5 \times 10^{-6}$ in a recent GWAS metaanalysis of schizophrenia (11). None of the enhancer regions of any immune cell types are enriched for these schizophrenia SNPs ($P > 0.25$ for all cell types listed in Table 1). Given that schizophrenia is not a classical autoimmune disease such as RA, no such enrichment of potential risk variants in immune cell enhancers would be expected.

We next looked at which RA-associated SNPs account for the observed enrichment of RA-associated SNPs in immune cell enhancer regions (Table 2). At 9 loci (*ANKRD55*, *MTF1*, *LBH*, *CD28*, *NFKBIE*, *BLK*, *RAD51B*, *RASGRP1*, *IRF8*), we found exactly one RA-associated SNP falling into an enhancer. At *ANKRD55*, the published RA lead SNP (rs7731626) itself is located in an enhancer, which is specific to dendritic cells. At the remaining 8 loci with only one RA SNP in enhancers, a tight correlation (as measured by r^2) exists between the observed enhancer SNP and the respective lead SNP from the GWAS metaanalysis (8). Therefore, all these enhancer SNPs can be viewed as strong candidates to act as causal variants that contribute to the reported genetic association signals at these gene loci.

At the remaining 12 gene loci for which RA-associated SNPs fall into FANTOM enhancers (*C4orf52*, *TNEAIP3*, *CCR6*, *TRAF1/C5*, *IL2RA*, *PRKCQ*, *GATA3*, *CXCR5*, *ETS1*, *LOC145837*, *MED1*, *PTPN2*), at least two RA-associated SNPs are located in enhancer regions. At most

loci, these enhancer SNPs are in tight LD with the published RA lead SNP (Table 2). Of particular interest might be the enhancer haplotypes near the genes *PRKCQ* and *GATA3*, which are both close to the telomere of the short arm of chromosome 10 (Figure 1). At *PRKCQ*, two closely neighboring RA-associated SNPs in tight LD fall into a same enhancer region. Thus, these two SNPs may act together to alter the activity of this enhancer, which is active in T cells, NK cells, dendritic cells neutrophils, and monocytes. A similar pattern can be observed at the *GATA3* locus, where r^2 is close to 1 across three RA-associated SNPs. These three SNPs in close physical proximity fall into a same enhancer that is specific for T cells and NK cells. Again, enhancer activity may be altered by variation at multiple nucleotide sites. Thus, at *PRKCQ* and *GATA3*, the described multimarker haplotypes are particularly good candidates to act as functional variants that may explain the genetic associations at these loci. Importantly, we did not see any long-range LD across the two associated haplotypes at *PRKCQ* and *GATA3*, which are spaced 1.5 Mb apart.

DISCUSSION

Our analysis combines the recently published RA GWAS metaanalysis (8) with a recently published set of enhancer regions in the human genome (9). We notice the strongest enrichment of RA-associated SNPs in enhancer regions with T-cell specificity. This finding confirms earlier results that found preferential expression in T cells for genes near RA-associated SNPs (13). In addition, our analysis indicates a strong enrichment of RA-associated SNPs in NK-cell enhancers, suggesting that the possible role of NK cells in RA pathogenesis may deserve more attention (14).

Our study differs from many earlier pathway analyses of GWAS in the sense that we do not attempt to link implicated genes to pathways, as it is typically done in the network-assisted pri-

oritization of GWAS results (15). Instead, we test the enrichment of RA-associated SNPs in different sets of enhancers that are active in particular cell types. We thereby connect genetic findings to tissue biology without considering the role of genes and gene functions. This approach will likely be useful for understanding genetic variants involved in other diseases too. The approach has become possible through the availability of the FANTOM data set and is principally applicable to any phenotype with a sufficiently large number of GWAS hits. In our analysis of RA, this scenario allows us to point out SNPs at several loci that are likely to exert causal effects instead of being in LD with some unknown causal variant. Interestingly, there are several loci where multiple enhancer SNPs occur in tight LD. These haplotypes may have a greater chance of affecting the activity of enhancers by altering more than one nucleotide site. This observation is consistent with the proposed “multiple enhancer variant hypothesis” of complex disease associations (16). Two notable examples are the loci *PRKCQ* and *GATA3*, where multiple variants in nearly perfect LD could account for the GWAS association signals. Interestingly, a genetic interaction between *GATA3* (rs2275806) and *PRKCQ* (rs947474) has been reported (17), with one interacting SNP being part of our *PRKCQ* enhancer haplotype and the other interacting SNP being in tight LD with the SNPs from our *GATA3* enhancer haplotype ($r^2 = 0.9$ in Europeans and $r^2 = 0.98$ in Asians). Altered expression of *GATA3* and *PRKCQ* in T cells and NK cells could be caused by these haplotypes, which genetically interact to promote the development of RA (17). *GATA3* mediates *PRKCQ*-induced gene expression (18), and *GATA3* is a master regulator of Th2-cell-specific gene expression (19). It is therefore conceivable that expression levels of *GATA3* could influence the RA phenotype through modifying T-cell activity. The expectation would be that genetic variants that influence *PRKCQ*

Table 2. RA-associated SNPs that are located in cell type-specific FANTOM enhancer regions.

Chromosome	Position/bp (NCBI37)	SNP identifier	T cell	NK cell	Dendritic cell	Basophil	Mast cell	Neutrophil	B cell	Intestinal epithelial	N cell types	Locus name/ gene symbol	r^2 with GWAS lead SNP, European	r^2 with GWAS lead SNP, Asian
1	38291337	rs67704103	1	0	0	0	0	0	0	0	2	<i>MTF1/INPP5B</i>	0.822	0.941
2	30448344	rs906868	1	0	0	0	0	0	0	0	1	<i>LBH</i>	0.962	0.985
2	204647261	rs749933	0	0	0	0	0	0	0	0	6	<i>CD28</i>	0.809	0.052
4	26090375	rs7441808	1	1	1	1	0	0	0	0	15	<i>C4orf52</i>	0.874	1
4	26095314	rs35944082	1	0	1	0	0	0	0	0	2		0.892	1
5	55444683	rs7731626	0	0	1	0	0	0	0	0	1	<i>ANKRD55</i>	(1)	(1)
6	44231479	rs3799963	1	1	1	1	1	1	1	0	13	<i>NFKBIE</i>	0.969	1
6	138132123	rs58905141	0	0	0	0	0	1	0	0	2	<i>TNFAIP3</i>	0.936	NA
6	138132516	rs10499197	1	1	1	1	1	1	1	0	20		0.936	NA
6	138168028	rs111883038	0	1	1	1	1	0	0	0	6		1	0.543
6	138168083	rs80126770	0	1	1	1	1	0	0	0	6		1	0.566
6	167363948	rs3798307	1	0	1	0	0	0	0	0	3	<i>CCR6</i>	0.121	0.213
6	167372978	rs2149092	1	1	0	1	1	1	1	0	9		0.375	0.176
6	167411788	rs239935	0	0	0	0	0	0	0	0	3		0.373	0.405
6	167504880	rs2239826	0	0	1	1	0	0	0	0	4		0.695	0.435
6	167527644	rs9347175	1	1	0	0	0	0	1	0	3		0.683	0.426
6	167527818	rs6456157	1	1	0	0	0	0	1	0	3		0.683	0.426
6	167533467	rs1331299	1	0	0	0	0	0	0	0	1		0.693	0.524
6	167534290	rs3093023	0	0	1	0	0	0	1	0	2		0.957	0.831
8	11337587	rs2409780	0	0	0	0	0	0	0	0	9	<i>BLK</i>	0.979	0.931
9	123652898	rs881375	1	0	0	0	1	0	0	0	3	<i>TRAF1/C5</i>	0.719	0.388
9	123683106	rs2109896	0	0	1	0	1	0	0	0	4		0.729	0.366
9	123696871	rs10739581	0	0	1	0	0	0	0	0	2		0.968	0.965
10	6102757	rs3118471	1	0	0	0	0	0	0	0	1	<i>IL2RA</i>	0.616	0.277
10	6114856	rs10905718	1	0	1	0	0	0	0	0	3		0.556	0.25
10	6390192	rs2387397	1	1	1	1	1	1	0	0	8	<i>PRKCQ</i>	0.83	0.968
10	6390450	rs947474	1	1	1	1	1	1	0	0	8		(1)	(1)
10	8103980	rs374641	1	1	0	0	0	0	0	0	2	<i>GATA3</i>	0.962	0.888
10	8103986	rs386680	1	1	0	0	0	0	0	0	2		0.967	0.888
10	8104016	rs371668	1	1	0	0	0	0	0	0	2		0.967	0.888
11	118740864	rs4936443	1	1	1	1	1	1	1	1	14	<i>CXCR5</i>	0.599	0.473
11	118740931	rs4938572	1	1	1	1	1	1	1	1	14		0.599	0.344
11	118741157	rs7117261	1	1	1	1	1	1	1	1	14		0.607	0.473
11	118743286	rs12365699	1	0	0	0	0	0	0	0	1		0.501	0.184
11	118743338	rs4936444	1	0	0	0	0	0	0	0	1		0.592	0.473
11	128498760	rs7103484	1	1	0	0	0	0	0	0	2	<i>ETS1</i>	0.152	0.105
11	128498762	rs7119766	1	1	0	0	0	0	0	0	2		0.18	0.105
14	68749927	rs3784099	0	0	0	0	0	0	1	0	1	<i>RAD51B</i>	0.97	1
15	38903672	rs16967112	1	0	0	0	0	0	0	0	1	<i>RASGRP1</i>	0.553	0.012
15	70011941	rs71393457	1	1	0	1	1	1	0	0	7	<i>LOC145837</i>	0.031	0
15	70012293	rs71393458	1	1	0	1	1	1	0	0	7		0.031	0
15	70027568	rs36124936	0	0	0	1	0	0	0	0	3		0.031	0
15	70048116	rs10152590	1	1	0	1	0	1	0	0	5		0.031	0
15	70048157	rs10152591	1	1	0	1	0	1	0	0	5		0.031	0
16	86016401	rs9927316	0	0	0	1	0	0	0	0	3	<i>IRF8</i>	0.986	0.958
17	37912377	rs12946510	1	1	1	0	0	0	1	0	4	<i>MED1</i>	0.083	0.096
17	37970149	rs9909593	1	1	0	0	1	0	0	0	3		0.076	0.095
18	12777573	rs2542148	0	0	0	0	0	0	0	1	7	<i>PTPN2</i>	0.326	0.233
18	12777603	rs2847274	0	0	0	0	0	0	0	1	7		0.312	0.233
18	12777607	rs11876290	0	0	0	0	0	0	0	1	7		0.097	0.128

Cell types with active enhancers are marked by 1. The column "N cell types" denotes the number of cell types out of the total of 71 FANTOM cell types where an enhancer is active. bp, Base pair.

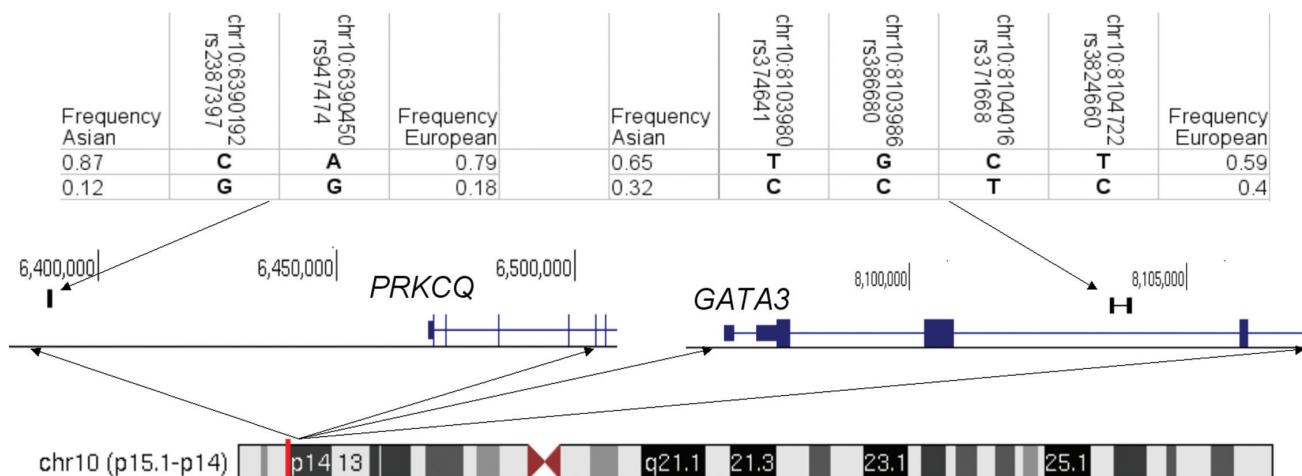


Figure 1. Haplotypes of RA-associated SNPs that are in near-perfect LD and fall into enhancer regions at *PRKCQ* and *GATA3*. Enhancer regions are displayed as black bars and genes are displayed as blue bars.

expression in cis would influence *GATA3* expression in trans while interacting with other variants that influence *GATA3* expression in cis.

CONCLUSION

Detailed functional studies are now needed to test this hypothesis and find out how these and other enhancer variants alter cellular function and cause RA susceptibility. Our analysis indicates potentially functional enhancer variants for 21 RA loci, but it does not make any prediction about causal variants at the remaining RA GWAS loci. Given that coding SNPs play a role only in a minority of GWAS loci, the search for causal variants may require even more comprehensive functional annotations as well as the analysis of rare variants, insertion/deletion variants and copy number variants.

ACKNOWLEDGMENTS

J Freudenberg was supported by a National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS)/National Institutes of Health (NIH) grant (R03 AR063340).

DISCLOSURE

After completion of this work, J Freudenberg became a full time em-

ployee of Regeneron Genetics. This did not have any influence on the results in this study or their presentation.

REFERENCES

- Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 447:661–78.
- Plenge RM, et al. (2007) TRAF1–C5 as a risk locus for rheumatoid arthritis: a genome-wide study. *N. Engl. J. Med.* 357:1199–209.
- Barton A, et al. (2008) Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nat. Genet.* 40:1156–9
- Gregersen PK, et al. (2009) REL, encoding a member of the NF- κ B family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat. Genet.* 41:820–3.
- Kochi Y, et al. (2010) A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. *Nat. Genet.* 42:515–9.
- Freudenberg J, et al. (2011) Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. *Arthritis Rheum.* 63:884–93.
- Terao C, et al. (2011) The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. *Hum. Mol. Genet.* 20:2680–5.
- Okada Y, et al. (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 506:376–81.
- Andersson R, et al. (2014) An atlas of active enhancers across human cell types and tissues. *Nature*. 507:455–61.
- Hindorf LA, et al. (2009) Potential etiologic and functional implications of genome-wide associa-
- tion loci for human diseases and traits. *Proc. Natl. Acad. Sci.* 106:9362–7.
- Ripke S, et al. (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 511:421–7.
- Raychaudhuri S, et al. (2012) Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature Genet.* 44:291–6.
- Hu X, et al. (2011) Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am. J. Hum. Genet.* 89:496–506.
- Ahern DJ, Brennan FM. (2011) The role of Natural Killer cells in the pathogenesis of rheumatoid arthritis: major contributors or essential homeostatic modulators? *Immunol. Lett.* 136:115–21.
- Jia P, Zhao Z. (2014) Network-assisted analysis to prioritize GWAS results: principles, methods and perspectives. *Human Genet.* 133:125–38.
- Corradin O, et al. (2014) Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* 24:1–13.
- Eyre S, et al. (2012) High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.* 44:1336–40.
- Stevens L, et al. (2012) Involvement of GATA3 in protein kinase C θ -induced Th2 cytokine expression. *Eur. J. Immunol.* 36:3305–14.
- Sasaki T, et al. (2013) Genome-wide gene expression profiling revealed a critical role for GATA3 in the maintenance of the Th2 cell identity. *PLoS One.* 8:e66468.

Cite this article as: Freudenberg J, Gregersen P, Li W. (2015) Enrichment of genetic variants for rheumatoid arthritis within T-cell and NK-cell enhancer regions. *Mol. Med.* 21:180–4.