

European Journal of Immunology Highlights

FRONTLINE:

Single nucleotide polymorphisms as a prerequisite for autoantigens

Michael B. Stadler, Diana Arnold, Sara Frieden, Sarah Luginbühl and Beda M. Stadler

Institute of Immunology, University of Bern, Bern, Switzerland

It is still elusive why certain self proteins induce an autoimmune response. One immunological hypothesis is that only modified or altered self-proteins may become a target. Thus, we asked whether such alterations may actually be genetic polymorphisms that can be revealed by analyzing sequence variability in the known human autoantigens. Indeed, we found autoantigens to contain significantly more single nucleotide polymorphisms (SNP) than other human genes do. Our finding may offer an explanation for autoimmune responses through allogeneic exposure. Besides other contributing factors in autoimmunity, SNP may represent an essential prerequisite for the primary induction of an autoimmune response.

Received 14/7/04

Revised 14/10/04

Accepted 5/11/04

[DOI 10.1002/eji.200425481]

Key words:

Autoantigens

· Alloantigens

· Autoimmunity



See accompanying Commentary: <http://dx.doi.org/10.1002/eji.200425888>



Supporting Information for this article is available at http://www.wiley-vch.de/contents/jc_2040/2005/25481_s.pdf or from the author.

Introduction

Although most foreign antigens such as bacterial and viral proteins induce immune responses, self proteins (autoantigens) normally do not amplify an immune response. Nevertheless, humoral and cell-mediated

immune reactions against self-constituents occur in patients with autoimmune diseases [1]. Autoantigens are a diverse class of molecular structures consisting mostly of proteins, but also including carbohydrates and polynucleotides. We focused our study on the analysis of protein autoantigens, representing the largest and best-characterized group among autoantigens.

Why only a small group of distinct proteins without obvious relation is targeted by autoimmune responses is not understood [2]. Interestingly, autoantigens recognized in mouse models of autoimmunity, both of spontaneous and manipulated origin, are similar to those in human diseases (reviewed in [2]). Moreover, the range of autoantibodies generated in a number of transgenic mouse strains with modified immune pathways resembles the repertoire of autoantibodies present

The first two authors contributed equally to this work.

Correspondence: Michael B. Stadler, Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Ave., 68-211, Cambridge, MA 02139-4307, USA

Fax: +1-617-452-2936

e-mail: stadler@mit.edu

Abbreviations: **GO:** Gene ontology · **OMIM:** Online Mendelian Inheritance in Man · **Ro52:** 52-kDa Ro protein · **SNP:** Single nucleotide polymorphism · **UTR:** Untranslated region

in spontaneous human autoimmune disorders (reviewed in [2]). Taken together, these results support a concept of common features in autoantigens.

Various hypotheses have been proposed to describe the restricted set of autoantigens, and various features have been observed. In the early 1990s, sequence analysis of known autoantigens revealed a higher frequency of autoantigens with charged surfaces and in particular coiled-coil structures – a pair of α -helices with high surface charge that interact along their length – compared with randomly selected human proteins [3, 4]. Many systemic autoantigens become clustered in surface blebs of apoptotic cells [5] and are processed by the protease granzyme B [6]. Recently, it was shown that failure to remove apoptotic lymphocytes from germinal centers may lead to autoimmunity in mice [7]. The high number of nucleic-acid-binding proteins amongst autoantigens has been associated with the ability of nucleic acids to stimulate Toll-like receptors, thereby activating specific B cells [8]. Although all of these observed features are likely to play a role in the ontogeny of autoimmunity, they only account for a subset of known autoantigens and are not a common denominator.

Many autoimmune diseases show an increased prevalence in women [9], and it is often reported that sex steroids have an effect on the immune system [10]. Maternal cells that remain present in the offspring after pregnancy, or persisting fetal cells in the mother (situations referred to as microchimerism [11, 12]), have also been proposed as an initial trigger. Such exposure to allogeneic proteins may represent the link between autoimmunity and genetic polymorphisms, as described in post-transfusion purpura [13]. Other autoimmune diseases favoring this hypothesis and displaying a sex bias are postpartum thyroiditis [14], systemic lupus erythematosus [15] and peripartum cardiomyopathy [16]. Nevertheless, women without

pregnancies, men, and children are also likely to be exposed to alloantigens, as maternal cells also traffic into the fetal circulation. Furthermore, microchimerisms may also result from transplantation or blood transfusion, or occur from *in utero* transfer between twins [12]. Given that a low level of autoreactivity is even found in healthy individuals [17], and it is well documented that antigen recognition by B lymphocytes [18] and T lymphocytes [19] is promiscuous, it may be assumed that lymphocyte clones stimulated by an alloantigen may also react with a similar host allele.

On the basis of our hypothesis, autoantigens should display a greater polymorphism, as this would increase the probability of allogeneic exposure. Indeed, we were able to demonstrate that single nucleotide polymorphisms (SNP) within coding regions are more frequent in autoantigens than in other human genes, suggesting that these polymorphisms are a crucial prerequisite for human proteins to be autoantigens.

Results and discussion

Database of known human autoantigens

To collect known autoantigens and their corresponding sequences, we searched biomedical literature (as indexed in Medline) and major public sequence databases, using keywords related to autoimmunity, known autoantigens and autoimmune diseases. As we aimed to analyze genetic variability, we omitted autoantigens whose corresponding genes could not be identified unambiguously. A total of 348 autoantigen sequences were collected by this procedure. A complete list of the autoantigens with corresponding database identifiers is available online (See Supporting Information, Supporting Table 1).

Table 1. Publication bias in human genes

Group	Number of genes	GO annotations ^{a)} (% high quality)	Coding SNP per gene	Coding SNP per base pair
Autoantigen genes	348	6.65 (37.5%)	7.20 ^{b)}	4.15×10^{-3} ^{c)}
Other human genes ^{d)}	14,881	4.95 (30.5%)	3.63	2.82×10^{-3}
Control genes ^{e)}	643	6.43 (61.1%)	3.67	2.84×10^{-3}
Disease-associated genes ^{f)}	2,161	6.54 (37.3%)	5.55	3.94×10^{-3}
All	15,229	5.00 (30.8%)	3.72	2.85×10^{-3}

^{a)} The average number of GO nodes annotated per gene is given, with the percentage of high-quality annotations (with evidence tags 'inferred from direct assay' or 'traceable author statement') amongst all annotations for genes in the group in parentheses.

^{b), c)} Significantly different from other human genes as well as from control genes, $p < 0.0001$.

^{d)} Known human genes except autoantigens (including 'control genes' and 'disease-associated genes').

^{e)} Genes without disease association (not contained in OMIM database), with at least as many high quality GO annotations (evidence tags 'inferred from direct assay' or 'traceable author statement') as autoantigen genes have.

^{f)} Genes with well-characterized disease association (contained in OMIM database, status 'confirmed'), except autoantigens.

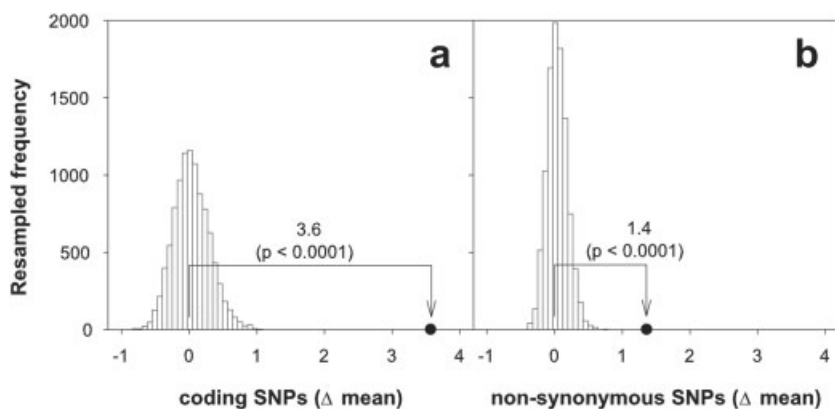


Fig. 1. Autoantigen genes contain more SNP than other human genes do. Mean numbers of coding (a) and non-synonymous (b) SNP for autoantigen genes ($n=348$) and other human genes from dbSNP ($n=14,881$) were calculated. The real differences of means (Δ mean) are indicated (closed circles), with statistical significances given in parentheses. Histograms of 10,000 random Δ means generated by resampling are shown by white bars.

Increased number of polymorphisms in autoantigens

We first analyzed coding SNP (*i.e.* SNP in the translated region of a gene), as well as the subclass of non-synonymous SNP (*i.e.* SNP that lead to an amino acid substitution in the expressed protein), as they are the most likely to cause alterations in the expressed protein recognizable by the immune system. On average, 7.2 coding and 3.1 non-synonymous SNP were found per autoantigen gene, whereas other human genes only contained 3.6 coding and 1.8 non-synonymous SNP.

To analyze the significance of the high genetic variability observed amongst autoantigens, we studied SNP contents of randomly selected human genes. All human genes including autoantigens ($n=15,229$) were randomly split into two samples with sizes corresponding to the number of autoantigen and non-autoantigen genes. Comparison of these randomly generated samples allowed the calculation of a Δ mean to be expected for randomly selected genes. A histogram of 10,000 randomly generated Δ means is shown in Fig. 1 (white bars), together with the real Δ mean observed if comparing human autoantigen with non-autoantigen genes (closed circles). For both coding (Fig. 1a) and non-synonymous SNP (Fig. 1b), a statistically significant higher number of SNP was found in autoantigens.

No autoantigen clusters in the human genome

The observed difference could be explained by a genomic clustering of autoantigens at a site of high SNP density, such as the HLA locus [20]. However, Fig. 2 shows that the genomic locations of autoantigens are randomly distributed in the human genome, and the number of autoantigen genes correlates well with the total number of genes on each chromosome (correlation = 0.91). Accordingly, less than ten autoantigen genes are found on the gene-poor chromosomes 13, 18, 21 and 22. None of the analyzed autoantigens mapped to

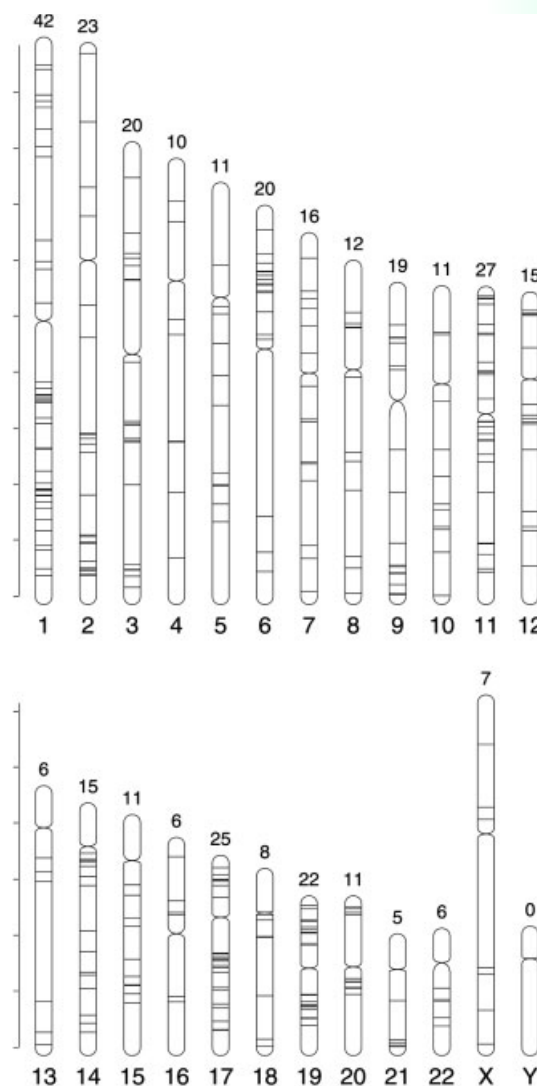


Fig. 2. Genomic distribution of autoantigen genes. Schematic representation of human chromosomes with loci of 348 autoantigens depicted as horizontal lines. Chromosomes are drawn to scale (one vertical unit equals 2.5×10^7 base pairs), with the total number of autoantigens per chromosome indicated at the top.

Table 2. Exemplary autoantigens with known epitopes and SNP

Autoantigen (accession #)	Sequence ^{a)}	Epitope reference
Insulin (P01308)	1 malwmrl ^{pl} l lallalwgp ^d ^{pa} aafVNQHL CGSHLVE ^{ly} LY LVCGER ^g FFY 51 <u>TPKTrreaed</u> lqv ^g qvel ^g g gpg ^g agsl ^q pl <u>alegsl^qkr^g</u> IVEQCCTSIC 101 SL ^Y QL ^E ENYCN	[22]
52-kDa Ro protein (P19474)	1 MASAARLTMM WEEVTCPICL DPFVEPV ^S IE CGHSFCQECI SQVGKGGGSV 51 ^C E ^V CRQR ^F ELL KNL ^R PNRQLA NMVNNLKEIS QEAREGTQGE RCAVH ^S ERLH 101 LFCEKDGKAL CWVCAQSRKH RDHAMVPLEE AA ^Q EYQEKLQ VALGELRRKQ 151 ELAEKLEVEI AIKRADWKKT VETQKSRIHA BFVQQKNFLV <u>EEEQRQLQEL</u> 201 <u>EKDEREQLRI</u> LGEKEAKLAQ <u>QSQALQELIS</u> <u>QLDRRCHSSA</u> LELLQETGVS 251 NSLSPRSESW NLKDLDTISP ELRSVCHVPG LKKMLR ^T CAV HITLDPDTAN 351 PWLILSEDRR QVRLGDTQQS IPGNEERFDS YPMVLGAQHF HSGKHYWEVD 401 VTGKEAWDLG VCRDSVRRKG HFLSSKSGF WTIWLWNKQK YEAGTYPQTP 451 LHLQVPPCQV GIFLDYEAGM VSFYNITDHG SLIYSFSECA FTGPLRPFFS 501 PGFNDGGKNT APLTLCPLNI GSQGSTDY	[23]
DNA topoisomerase 1 (P11387)	1 MSGDHLHNS QIEADFR ^L ND SHKHKDKHKD REHRHKEHKK EKDR ^E SKSHS 51 NSEHKDSEKK HKEKEKTKHK DGSSEKHKDK HKDRDKEKRK EEKVRASGDA 101 KIKKEKENG ^F SSPPQIKDEP EDDGYFVPPK EDIKPLKRPR DEDDADYKPK 151 KIKTEDTKKE KKRKLEEEED GK ^L KKPKNKD KDKKVPEDN KKKKPKKEEE 201 QKW ^W WEEER ^Y PE ^S IKW ^K FL EHKG ^P VFAPP YEPLPENVKF YYDGKVMKLS 251 PKAE ^E VATFF AKMLDHEYTT KEIFRNFFK DWRKEMTNEE KNIITNLSKC 301 DFTQMSQYFK AQTEARKQMS KEEKLKIKEE NEKLLKEYGF CIMDNHKE ^R I 351 <u>ANFKIEPPGL</u> FRGRGNH ^P KM GMLKRRIMPE DIIINCSKDA KVPSPPPGHK 401 <u>WKEVRHDNKV</u> TWLVSWT ^E NI QGSIKYIMLN PSSRIKGEK WQKYETARRL 451 KKCVDKIRNQ YREDWKS ^K EM KVRQRAVALY FIDKLALRAG NEKEEGETAD 501 TVGCCSLRVE HINLH ^F LDG QEYVVEFD ^F L GKDSIRYYNK VPVEKRVFKN 551 LQLFMENKQP EDDLFDRLNT GILNKH ^L QDL MEGLTAKVFR TYNASITLQQ 601 QLKELTAPDE NIPAKILSYN RANRAVAILC NHQRAPPKTF EKSMN ^L QTK 651 IDAKKEQLAD ARDLKSAK DAKVMKDAKT KKVVESK ^K KA VQRLEEQLMK 701 LEVQATDREE NKQIALGTSK LNYLDPRITV AWCKK ^W GVPI EKIYNKTQRE 751 KFAWAIDMAD EDYEF	[24]

^{a)} Unprocessed precursors proteins are shown, with synonymous and non-synonymous SNP shaded in gray and black, respectively. Regions typed in lower-case correspond to amino acids not present in the mature protein, and underlined amino acids are part of known epitopes recognized by T or B cells.

chromosome Y, which contains the least number of genes of all human chromosomes.

Gene length does not explain increased polymorphisms

Differences in gene length and structure between autoantigens and other genes are likely to have an influence on the mean amount of genetic variability. For instance, long genes may tend to contain higher numbers of SNP in comparison with short genes. We thus compared the average lengths of three functional regions [coding regions, intron regions and untranslated regions (UTR)], as well as the average number of splice sites in the groups of autoantigens and other human genes.

The two groups were found to vary considerably for all four of these indicated values (absolute values and significance levels are available in Supporting Table 2);

thus we normalized the number of SNP in a gene with the corresponding length of the functional region or the number of splice sites in that gene. Fig. 3 illustrates the percentage variability in autoantigens before and after normalization, setting the variability in the other human genes to 100%. Autoantigen variability in introns, UTR and splice sites, although not equal to 100%, is not significantly different from other genes. The apparent differences when compared with other genes are generated by a few genes with extremely high numbers of SNP that strongly influence the mean of the group (data not shown). Interestingly, the number of coding SNP (both synonymous and non-synonymous) is significantly higher in autoantigens. This is true both for the non-normalized values (Fig. 3, black bars), as well as for the normalized values (Fig. 3, open bars). We thus concluded that the increased numbers of coding SNP in autoantigens cannot simply be explained by a greater length of the corresponding coding sequences.

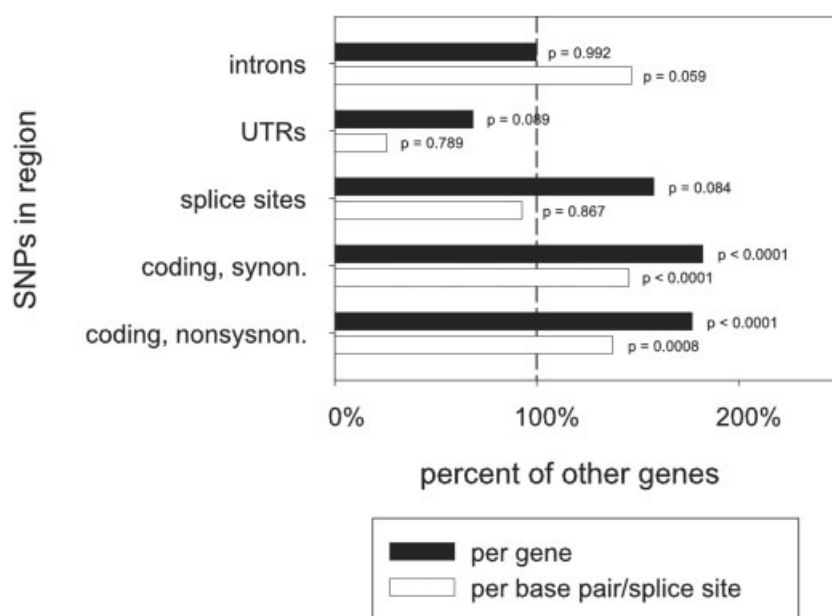


Fig. 3. Normalization of SNP counts in autoantigens. For selected functional regions of human genes, SNP were counted directly (per gene, filled bars), or with normalization (per base pair / per splice site, open bars). Values are shown for autoantigen genes, expressed as percent of the corresponding value in other human genes. The *p*-values were calculated by resampling as described in 'Materials and methods'.

Investigation bias does not explain increased polymorphisms

It has been estimated that only about 44% of SNP with minor allele frequencies greater than 5% are represented in public databases [21]. It is thus likely that the SNP numbers used in our study underestimate the true amount of genetic variability, especially for poorly investigated genes. Therefore, autoantigens may bias the results of our analysis because they have been investigated in greater detail.

To assess the number of investigations in the context of a group of human genes, we calculated the mean number of gene ontology (GO) terms associated per gene as well as the percentage of these annotations that are explicitly documented (Table 1). Indeed, autoantigens have been studied to a greater extent than other human genes have: on average, more GO terms are annotated per autoantigen, and a greater percentage of these annotations has a solid experimental basis. As a suitable control group, we thus collected a subset of well-characterized genes from the group of other human genes. Only genes that were better-characterized than autoantigens are (equal or more annotated GO terms with high-quality evidence) and with no known disease association were included in this group (Table 1, control genes). These control genes contained on average very similar numbers of SNP as the whole group of other human genes did, and genetic variability of autoantigens was also significantly higher when compared with the well-characterized human genes. This indicated that there is no investigation bias that would have provoked our observed difference in SNP numbers.

SNP and autoantigen epitopes

We have chosen three arbitrary autoantigens [insulin, 52-kDa Ro protein (Ro52) and DNA topoisomerase 1] to exemplify the effect of SNP on known T and B cell epitopes. Insulin is one of the major antigens associated with type 1 diabetes mellitus, an autoimmune disease believed to be primarily mediated by T cells [22]. As illustrated in Table 2, human preproinsulin contains seven synonymous and two non-synonymous SNP. Interestingly, the single non-synonymous SNP located in mature insulin (position 38 of preproinsulin) is contained in two of the six known class-II-restricted T cell epitopes (33–47, 35–51), and directly flanking the only known class-I-restricted epitope (39–47) in human insulin ([22] and references therein).

Human Ro52 is a member of the Ro/La ribonucleoprotein complex, and autoantibodies specific for Ro52 are found in various autoimmune diseases, most notably in rheumatoid arthritis, systemic lupus erythematosus and Sjögren's syndrome. Ro52 contains two synonymous and three non-synonymous SNP (Table 2), and one of the non-synonymous SNP is located in the central immunodominant region of the autoantigen [23].

Autoantibodies to DNA topoisomerase 1 are markers for systemic sclerosis and are associated with pulmonary fibrosis, a major clinical feature that increases the morbidity in these patients. Four major epitopes were identified to be recognized by anti-topoisomerase-1 sera [24]. Two of these epitopes (205–224 and 517–536) contain a non-synonymous SNP and are thus likely to be auto- and allo-epitopes at the same time.

SNP located outside immunogenic regions cannot be ruled out from having an effect on protein tertiary

structure. The above-described Ro52 is an example for non-coding SNP affecting a protein's antigenicity: in a genetic study of Ro52, an intronic SNP was identified to be strongly associated with the presence of anti-Ro52 autoantibodies in primary Sjögren's syndrome [25]. The non-coding SNP located in intron 3 leads to aberrant splicing of Ro52 mRNA, resulting in skipping of exon 4 and, thus, generation of a shortened version of the Ro52 protein [25]. As many exons are known to contain signals recognized by the splicing machinery [26], synonymous SNP may exert an effect on the protein via a similar mechanism.

SNP in disease-associated genes

For a great number of human diseases, the role of point mutations for disease predisposition and pathophysiology is well documented and available in the Online Mendelian Inheritance in Man (OMIM) database [27]. Therefore, we also analyzed the number of SNP in disease-associated genes. Similarly to autoantigens, this group displayed increased numbers of coding SNP compared with other human genes (Table 1, $p < 0.0001$). However, the number of coding SNP per gene was still significantly lower than the one observed in autoantigens ($p = 0.0006$).

Conclusions

Whereas an increased number of polymorphisms in genes associated with hereditary diseases was expected, the high number of polymorphisms in autoantigens is a surprising finding, and the consequences of SNP in autoantigens for protein structure, function and antigenicity remain to be studied. Ideally, frequencies of polymorphisms should be compared between groups of autoimmune patients and healthy individuals. Unfortunately, only very few data of this type are available in dbSNP. For populations with endemic autoimmune diseases, it would be interesting to study the correlation of polymorphisms and disease prevalence. More detailed genetic studies of autoantigens, such as the one performed on Ro52 [25], are necessary to establish the pathophysiological role of SNP for given autoimmune diseases.

Our study focused on SNP as these data are readily available. However, point mutations occurring at different times during individual development may represent the same immunological alteration as a qualified SNP. Any point mutation may be the reason for increased occurrence of autoimmune diseases in certain organs, families or racial groups [28, 29].

The same principle may account for the induction of autoantibodies by foreign proteins. In humans, even distantly related proteins (streptococcal surface enolase

[30]) or, in rats, close homologues (bovine α -tropomyosin [31]) are known to trigger an autoimmune response. This immunological cross-reactivity between proteins usually reflects their evolutionary relationship (homology), and can be explained by structural similarity. Induction of autoantibodies by foreign alleles (alloantigens) is conceptually equivalent and differs only by the source of the provoking agent within the same species. The concept is consistent with published findings on the etiology of autoimmune diseases, as many factors are known to contribute to disease development. Yet, a high number of polymorphisms may represent a general and essential prerequisite for autoantigens.

Materials and methods

External databases

The major public sequence databases used as a source of autoantigen sequences and sequence annotation were: Swiss-Prot Release 40 [32], EMBL Release 78 [33], and GenBank Release 141 [34]. The database of genetic variations in human genes was dbSNP build 120 [35]. Human gene structure was from Ensembl [36], human release 34c (accessed via Ensembl perl API and Bioperl modules [37]). For gene annotation we used the GO database [38] with human annotation release 18.0 [39]. For disease-associated genes, we used the OMIM database [27], downloaded on 04/27/2004.

Groups of genes

In all groups, only genes available in dbSNP (available variability information) with unambiguous correspondence in Ensembl (available gene structure) were included in the analysis. For identification of autoantigens, biomedical literature and major sequence databases were searched using keywords related to autoimmunity, autoantigens, and autoimmune diseases. Keyword searches of literature and major sequence databases were performed using Medline publication index [40] and SRS data retrieval system [41], respectively. As our study aimed at the analysis of the genetic variability, we omitted autoantigens whose corresponding gene could not be unambiguously identified from sequence annotation. 'Other human genes' were human genes other than those for autoantigens. Control genes were human genes other than those for autoantigens, with at least three GO nodes annotated with high-quality evidence (evidence tags 'traceable author statement' or 'inferred from direct assay'). The disease-associated genes were human genes in the OMIM database with confirmed status, other than those for autoantigens.

Statistical analysis of SNP counts

Various classes of SNP [coding SNP (synonymous and non-synonymous), coding non-synonymous SNP, SNP in introns, UTR and splice sites] were counted per gene and averaged for

all genes in a group. To account for differences in gene length and complexity, SNP counts for each gene were also divided by the number of splice sites (SNP at splice sites), the number of coding base pairs (coding and non-synonymous SNP), or the number of intron and UTR base pairs (SNP in introns and UTR, respectively). For genes with multiple transcripts, lengths of functional regions and numbers of splice sites were averaged over all variants. To estimate the significance of an observed difference of means between two groups of genes, a Monte Carlo permutation test was used: All genes in the first group (n_1 genes, mean number of SNP m_1) and in the second group (n_2 genes, mean number of SNP m_2) were pooled and randomly split into two new groups of n_1 and n_2 genes (mean numbers of SNP m_1' and m_2'). The difference of means was calculated for the two resampled groups ($\Delta m_{\text{resampled}} = m_1' - m_2'$), and the so-generated difference of means was compared with the observed real difference of means between group one and group two ($\Delta m_{\text{observed}} = m_1 - m_2$). This procedure was repeated 10,000 times, and the p value was calculated as the number of resampled differences $\Delta m_{\text{resampled}}$ equal to or larger than the real difference $\Delta m_{\text{observed}}$ plus one, divided by the total number of resampled differences plus one.

Acknowledgements: This work was funded by a governmental grant from CTI, the Swiss Innovation Promotion Agency, project no. 6369.1 in the context of a collaboration with ZLB Behring, Bern, Switzerland.

References

- Davidson, A. and Diamond, B., Autoimmune diseases. *N. Engl. J. Med.* 2001. **345**: 340–350.
- Plotz, P. H., The autoantibody repertoire: searching for order. *Nat. Rev. Immunol.* 2003. **3**: 73–78.
- Brendel, V., Dohlman, J., Blaisdell, B. E. and Karlin, S., Very long charge runs in systemic lupus erythematosus-associated autoantigens. *Proc. Natl. Acad. Sci. USA* 1991. **88**: 1536–1540.
- Dohlman, J. G., Lupas, A. and Carson, M., Long charge-rich alpha-helices in systemic autoantigens. *Biochem. Biophys. Res. Commun.* 1993. **195**: 686–696.
- Casciola-Rosen, L. A., Anhalt, G. and Rosen, A., Autoantigens targeted in systemic lupus erythematosus are clustered in two populations of surface structures on apoptotic keratinocytes. *J. Exp. Med.* 1994. **179**: 1317–1330.
- Casciola-Rosen, L., Andrade, F., Ulanet, D., Wong, W. B. and Rosen, A., Cleavage by granzyme B is strongly predictive of autoantigen status: implications for initiation of autoimmunity. *J. Exp. Med.* 1999. **190**: 815–826.
- Hanayama, R., Tanaka, M., Miyasaka, K., Aozasa, K., Koike, M., Uchiyama, Y. and Nagata, S., Autoimmune disease and impaired uptake of apoptotic cells in MFG-E8-deficient mice. *Science* 2004. **304**: 1147–1150.
- Leadbetter, E. A., Rifkin, I. R., Hohlbaum, A. M., Beaudette, B. C., Shlomchik, M. J. and Marshak-Rothstein, A., Chromatin-IgG complexes activate B cells by dual engagement of IgM and Toll-like receptors. *Nature* 2002. **416**: 603–607.
- Whitacre, C. C., Sex differences in autoimmune disease. *Nat. Immunol.* 2001. **2**: 777–780.
- Tanriverdi, F., Silveira, L. F., MacColl, G. S. and Bouloux, P. M., The hypothalamic-pituitary-gonadal axis: immune function and autoimmunity. *J. Endocrinol.* 2003. **176**: 293–304.
- Adams, K. M. and Nelson, J. L., Microchimerism: an investigative frontier in autoimmunity and transplantation. *JAMA* 2004. **291**: 1127–1131.
- Lambert, N. and Nelson, J. L., Microchimerism in autoimmune disease: more questions than answers? *Autoimmun. Rev.* 2003. **2**: 133–139.
- Watkins, N. A., Smethurst, P. A., Allen, D., Smith, G. A. and Ouwehand, W. H., Platelet alpha IIb beta 3 recombinant autoantibodies from the B-cell repertoire of a post-transfusion purpura patient. *Br. J. Haematol.* 2002. **116**: 677–685.
- Muller, A. F., Drexhage, H. A. and Berghout, A., Postpartum thyroiditis and autoimmune thyroiditis in women of childbearing age: recent insights and consequences for antenatal and postnatal care. *Endocr. Rev.* 2001. **22**: 605–630.
- Bongu, A., Chang, E. and Ramsey-Goldman, R., Can morbidity and mortality of SLE be improved? *Best Pract. Res. Clin. Rheumatol.* 2002. **16**: 313–332.
- Sundstrom, J. B., Fett, J. D., Carraway, R. D. and Ansari, A. A., Is peripartum cardiomyopathy an organ-specific autoimmune disease? *Autoimmun. Rev.* 2002. **1**: 73–77.
- Dighiero, G. and Rose, N. R., Critical self-epitopes are key to the understanding of self-tolerance and autoimmunity. *Immunol. Today* 1999. **20**: 423–428.
- Robey, I. F., Edmundson, A. B., Schluter, S. F., Yocum, D. E. and Marchalonis, J. J., Specificity mapping of human anti-T cell receptor monoclonal natural antibodies: defining the properties of epitope recognition promiscuity. *FASEB J.* 2002. **16**: 642–652.
- Mason, D., A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol. Today* 1998. **19**: 395–404.
- Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L. et al., A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001. **409**: 928–933.
- Reich, D. E., Gabriel, S. B. and Altshuler, D., Quality and completeness of SNP databases. *Nat. Genet.* 2003. **33**: 457–458.
- Panagiotopoulos, C., Trudeau, J. D. and Tan, R., T-cell epitopes in type 1 diabetes. *Curr. Diab. Rep.* 2004. **4**: 87–94.
- Routsias, J. G., Tzioufas, A. G. and Moutsopoulos, H. M., The clinical value of intracellular autoantigens B-cell epitopes in systemic rheumatic diseases. *Clin. Chim. Acta* 2004. **340**: 1–25.
- Rizou, C., Ioannidis, J. P., Panou-Pomonis, E., Sakarellos-Daitsiotis, M., Sakarellos, C., Moutsopoulos, H. M. and Vlachoyiannopoulos, P. G., B-cell epitope mapping of DNA topoisomerase I defines epitopes strongly associated with pulmonary fibrosis in systemic sclerosis. *Am. J. Respir. Cell Mol. Biol.* 2000. **22**: 344–351.
- Nakken, B., Jonsson, R. and Bolstad, A. I., Polymorphisms of the Ro52 gene associated with anti-Ro 52-kd autoantibodies in patients with primary Sjogren's syndrome. *Arthritis Rheum.* 2001. **44**: 638–646.
- Fairbrother, W. G., Yeh, R. F., Sharp, P. A. and Burge, C. B., Predictive identification of exonic splicing enhancers in human genes. *Science* 2002. **297**: 1007–1013.
- Boydadjiev, S. A. and Jabs, E. W., Online Mendelian Inheritance in Man (OMIM) as a knowledgebase for human developmental disorders. *Clin. Genet.* 2000. **57**: 253–266.
- Brix, T. H., Hansen, P. S., Kyvik, K. O. and Hegedus, L., Aggregation of thyroid autoantibodies in first-degree relatives of patients with autoimmune thyroid disease is mainly due to genes: a twin study. *Clin. Endocrinol. (Oxf.)* 2004. **60**: 329–334.
- Acers, T. E. and Acers-Warn, A., Incidence patterns of immunogenetic diseases in the North American Indians. *J. Okla. State Med. Assoc.* 1994. **87**: 309–314.
- Fontan, P. A., Pancholi, V., Nociari, M. M. and Fischetti, V. A., Antibodies to streptococcal surface enolase react with human alpha-enolase: implications in poststreptococcal sequelae. *J. Infect. Dis.* 2000. **182**: 1712–1721.
- Mor, F., Weinberger, A. and Cohen, I. R., Identification of alpha-tropomyosin as a target self-antigen in Behcet's syndrome. *Eur. J. Immunol.* 2002. **32**: 356–365.
- Bairoch, A. and Apweiler, R., The SWISS-PROT protein sequence data bank and its new supplement TrEMBL. *Nucleic Acids Res.* 1996. **24**: 21–25.

- 33 Stoesser, G., Baker, W., van den Broek, A., Garcia-Pastor, M., Kanz, C., Kulikova, T., Leinonen, R., Lin, Q., Lombard, V., Lopez, R. et al., The EMBL Nucleotide Sequence Database: major new developments. *Nucleic Acids Res.* 2003. **31**: 17–22.
- 34 Benton, D., Recent changes in the GenBank On-line Service. *Nucleic Acids Res.* 1990. **18**: 1517–1520.
- 35 Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. and Sirotkin, K., dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001. **29**: 308–311.
- 36 Birney, E., Andrews, D., Bevan, P., Caccamo, M., Cameron, G., Chen, Y., Clarke, L., Coates, G., Cox, T., Cuff, J. et al., Ensembl 2004. *Nucleic Acids Res.* 2004. **32 Database issue**: D468–D470.
- 37 Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., Fuellen, G., Gilbert, J. G., Korf, I., Lapp, H. et al., The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* 2002. **12**: 1611–1618.
- 38 Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C. et al., The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 2004. **32 Database issue**: D258–D261.
- 39 Camon, E., Magrane, M., Barrell, D., Lee, V., Dimmer, E., Maslen, J., Binns, D., Harte, N., Lopez, R. and Apweiler, R., The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res.* 2004. **32 Database issue**: D262–D266.
- 40 Putnam, N. C., Searching Medline free on the internet using the National Library of Medicine's PubMed. *Clin. Excell. Nurse Pract.* 1998. **2**: 314–316.
- 41 Zdobnov, E. M., Lopez, R., Apweiler, R. and Etzold, T., The EBI SRS server – new features. *Bioinformatics* 2002. **18**: 1149–1150.