

Extended exome sequencing identifies *BACH2* as a novel major risk locus for Addison's disease

■ D. Eriksson^{1,2,*}, M. Bianchi^{3,*}, N. Landegren^{1,4}, J. Nordin³, F. Dalin^{1,4}, A. Mathioudaki³, G. N. Eriksson⁵, L. Hultin-Rosenberg³, J. Dahlqvist³, H. Zetterqvist^{3,4}, Å. Karlsson³, Å. Hallgren^{1,4}, F. H. G. Farias³, E. Murén³, K. M. Ahlgren⁴, A. Lobell⁴, G. Andersson⁶, K. Tandre⁴, S. R. Dahlqvist⁷, P. Söderkvist⁸, L. Rönnblom⁴, A.-L. Hulting⁵, J. Wahlberg⁹, O. Ekwall^{10,11}, P. Dahlqvist⁷, J. R. S. Meadows³, S. Bensing^{2,5}, K. Lindblad-Toh^{3,12}, O. Kämpe^{1,2,4,†} & G. R. Pielberg^{3,†}

From the ¹Department of Medicine (Solna), Center for Molecular Medicine, Karolinska Institutet; ²Department of Endocrinology, Metabolism and Diabetes Karolinska University Hospital, Stockholm; ³Science for Life Laboratory, Department of Medical Biochemistry and Microbiology; ⁴Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala; ⁵Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm; ⁶Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala; ⁷Department of Public Health and Clinical Medicine, Umeå University, Umeå; ⁸Department of Clinical and Experimental Medicine; ⁹Department of Endocrinology, Department of Medical and Health Sciences, Department of Clinical and Experimental Medicine, Linköping University, Linköping; ¹⁰Department of Pediatrics, Institute of Clinical Sciences, Sahlgrenska Academy; ¹¹Department of Rheumatology and Inflammation Research, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; and ¹²Broad Institute of MIT and Harvard, Cambridge, MA, USA

Abstract. Eriksson D, Bianchi M, Landegren N, Nordin J, Dalin F, Mathioudaki A, Eriksson GN, Hultin-Rosenberg L, Dahlqvist J, Zetterqvist H, Karlsson Å, Hallgren Å, Farias FHG, Murén E, Ahlgren KM, Lobell A, Andersson G, Tandre K, Dahlqvist SR, Söderkvist P, Rönnblom L, Hulting A-L, Wahlberg J, Ekwall O, Dahlqvist P, Meadows JRS, Bensing S, Lindblad-Toh K, Kämpe O, Pielberg GR (Karolinska Institutet; Metabolism and Diabetes Karolinska University Hospital, Stockholm; Uppsala University; Swedish University of Agricultural Sciences, Uppsala; Umeå University, Umeå; Linköping University, Linköping; University of Gothenburg, Gothenburg; Sweden; Broad Institute of MIT and Harvard, Cambridge, MA, USA). Extended exome sequencing identifies *BACH2* as a novel major risk locus for Addison's disease. *J Intern Med* 2016; **280**: 595–608.

Background. Autoimmune disease is one of the leading causes of morbidity and mortality worldwide. In Addison's disease, the adrenal glands are targeted by destructive autoimmunity. Despite being the most common cause of primary adrenal failure, little is known about its aetiology.

Methods. To understand the genetic background of Addison's disease, we utilized the extensively characterized patients of the Swedish Addison Registry. We developed an extended exome capture array comprising a selected set of 1853 genes and their potential regulatory elements, for the purpose of sequencing 479 patients with Addison's disease and 1394 controls.

Results. We identified *BACH2* (rs62408233-A, OR = 2.01 (1.71–2.37), $P = 1.66 \times 10^{-15}$, MAF 0.46/0.29 in cases/controls) as a novel gene associated with Addison's disease development. We also confirmed the previously known associations with the HLA complex.

Conclusion. Whilst *BACH2* has been previously reported to associate with organ-specific autoimmune diseases co-inherited with Addison's disease, we have identified *BACH2* as a major risk locus in Addison's disease, independent of concomitant autoimmune diseases. Our results may enable future research towards preventive disease treatment.

Keywords: Addison Disease, *BACH2* protein, genetic, genetic association studies, genetic predisposition to disease, human, polymorphism.

Introduction

Autoimmune destruction of the adrenal cortex is the predominating cause of primary adrenal

insufficiency, and it is referred to as autoimmune Addison's disease (AAD) [1]. The ensuing loss of adrenal hormones is ultimately lethal if cortisol is not replaced [2]. The prevalence of AAD amongst Caucasians is 87–144 per million [3–9], and for the last decades, an increasing incidence of AAD has been reported [3, 6, 7, 10]. Typically, the disease

*These authors contributed equally to this work.

†These authors jointly supervised this work.

onset occurs in the fourth or fifth decade of life, although individuals of all ages can be affected [8], with a preponderance for women (60%–65%) [8, 11].

Despite its rarity, AAD displays high concordance rates in twins and aggregation in families, suggesting a strong genetic contribution [12–18]. A recent twin concordance study (Skov J. *et al.*, submitted) reports a heritability higher than that of, for instance, type 1 diabetes mellitus (0.69–0.88) [19–23] and Graves' disease (0.79) [24].

The success of genetic case–control disease studies, including autoimmune diseases, depends to a large extent on the phenotypic homogeneity of the case group. Diagnosis of AAD can be confirmed with a serological marker, namely 21-hydroxylase autoantibodies in patient sera [2, 25]. 21-hydroxylase autoantibodies are present in about 86% of patients presenting with idiopathic primary adrenal insufficiency, with a specificity of about 95% in experienced labs [8, 26]. Therefore, patients with AAD are likely to constitute one of the most homogenous autoimmune case groups available for genetic studies.

The Swedish Addison Registry (SAR) is one of the world's largest AAD biobanks. At the commencement of this study, the registry included 700 cases, representing approximately half of the estimated patients with AAD currently living in Sweden [3]. The majority of these patients have concomitant tissue-specific autoimmune diseases such as type 1 diabetes, autoimmune thyroid disease or pernicious anaemia. The high rate of comorbidity makes it possible to study the shared genetic basis of autoreactivity.

Similar to most other autoimmune diseases, AAD has a complex inheritance. Apart from the well-characterized association with the HLA complex [27], little is known about the genetic variants that contribute to disease development [27]. Until now, candidate gene studies have linked genes implicated in co-occurring autoimmune diseases to AAD [27, 28]. The rarity of AAD has however, made extensive unbiased genomewide association studies unfeasible.

Here, we performed exome sequencing of a comprehensive selection of 1853 genes, extended to include their promoters and other potential

regulatory elements, as well as their splice sites. Using a well-defined AAD phenotype, based on all available diagnostic modalities ranging from clinical signs to autoantibody markers, we identified *BACH2* as a novel major risk locus in AAD.

Materials and methods

Study subjects

All participants in this study were recruited in Sweden. In total, 1501 controls were retrieved from directed sampling of healthy individuals ($n = 653$), or from blood donors ($n = 848$) [29], not known to suffer from AAD. All the 700 AAD samples included were obtained from the Swedish Addison Registry (SAR). SAR incorporates detailed clinical information and serological data compiled by the responsible physician. All cases fulfilled diagnostic criteria of primary adrenal insufficiency, that is low morning serum cortisol and elevated plasma adrenocorticotropic hormone (ACTH) [2, 30]. In equivocal cases, insufficient response to corticotropin stimulation confirmed the diagnosis. Blood and serum samples were stored at -80°C until use.

To strictly define the case group, we excluded cases where other causes of adrenal failure were suspected: cases with fragmentary clinical information, adrenoleukodystrophy/adrenomyeloneuropathy, autoimmune polyendocrine syndrome type 1 (APS1), adrenal failure secondary to corticosteroid treatment, congenital adrenal hyperplasia (CAH), adrenal metastases, isolated ACTH deficiency, hypogonadotropic hypogonadism with adrenal insufficiency due to *DAX1* mutations or tuberculosis and cases that had undergone adrenalectomy. Furthermore, AAD cases with autoantibodies against interferon- α , interferon- ω or interleukin-22, and cases lacking 21-hydroxylase autoantibodies were also excluded (Figure S1 in the supplementary information). AAD cases with anticytokine antibodies were excluded because of potentially undiagnosed APS1. To prioritize specificity over sensitivity, we required positive results from two 21-hydroxylase assays, performed at two different laboratories.

All study subjects gave their informed consent. The study was performed in accordance with the Declaration of Helsinki and was approved by the local ethics committees (2008/296-31/2, M177-07, 2012/52-31, 2009/013, 08-109M § 137/08, 2010/182-32).

Genomic DNA samples

Genomic DNA was isolated from whole blood EDTA using QIAmp Blood Midi Kit (51185; Qiagen, Venlo, Netherlands) following the manufacturer's protocol, or using a LGC Genomics extraction service. The concentration of the DNA was measured using NanoDrop ND-1000 Spectrophotometer or Qubit 2.0 Fluorometer (ThermoFisher Waltham, MA, US) with Qubit dsDNA HS Assay Kit (Q32851, ThermoFisher). The rough assessment of high molecular weight DNA was performed by electrophoretic separation of 150 ng of genomic DNA on 1% agarose gel.

Sequence capture array

A custom SeqCap EZ Choice XL Library (06266517001; Roche NimbleGen, Basel, Switzerland) was designed targeting 1853 genes. Genes were selected based on their known role in basic immune functions, inflammation and autoimmune disease. A systematic inclusion of genes associated with immunological or autoimmune disease, for instance type 1 diabetes and coeliac disease, was made based on candidate loci identified in genome-wide association studies. By searching the pathway databases KEGG (www.genome.jp/kegg) and Ingenuity® (QIAGEN, www.qiagen.com/ingenuity) with keywords such as autoimmunity, cytokines and T cells, additional genes involved in immune pathways were retrieved and selected. Genes causing monogenic adrenal disease were also targeted. The gene intervals for all alternative transcripts were retrieved from human assembly NCBI36/hg18 using UCSC Genome Table Browser (<https://genome.ucsc.edu/cgi-bin/hgTables>) known gene and microRNA tracks. Specific regions of interest were the following: 5' and 3' UTRs, coding exons, potential promoter regions (± 2 kb from transcription start sites of all alternative transcripts) and splice sites (20 bp of intronic sequences adjacent to exons). In addition, we targeted a selection of potential regulatory elements from the extended gene intervals (± 100 kb of the gene). These elements were defined based on 29 mammals conservation, and the elements with SiPhy LOD-score higher than 7 were included [31]. The total size of the targeted regions of interest was >36 Mb, whereas the designed array probes covered 89.6% of it (97.7% including the 100-bp offset).

Sequence capture experiment

The sequencing library was prepared by ultrasonication of up to 1 μ g of high molecular weight DNA

to 400-bp fragments (Covaris E220). 3' and 5' overhangs were repaired using T4 DNA polymerase (EP0062; Thermo Fisher Scientific), Klenow fragment (EP0052; Thermo Fisher Scientific) and T4 polynucleotide kinase (EK0032; Thermo Fisher Scientific). Adenylation with Klenow fragment exo⁻ (EP0422; Thermo Fisher Scientific) and ligation of NEXTflex-96™ DNA barcode adapters (514106; Bio Scientific Austin, TX, US) using T4 DNA Ligase (EL0011; Thermo Fisher Scientific) were performed according to the manufacturer's protocol. Samples with a final fragment library concentration of less than 130 ng μ L⁻¹ ($n = 118$ control samples) were amplified 4–7 cycles using Phusion High-Fidelity PCR Master Mix (F531S; Thermo Fisher Scientific) prior to hybridization. Pools of eight samples were hybridized according to Roche NimbleGen protocol and sequenced with 100-bp paired-end reads using Illumina HiSeq 2500 version 3 or version 4 chemistry.

Bioinformatic analysis

Raw reads from individual samples were mapped to the hg19 human reference genome with the Burrows-Wheeler aligner 0.7.4 [32], and duplicate reads were marked by Picard tools 1.92 (<http://broadinstitute.github.io/picard>). GATK 2.7.2 was applied for realignment around indels [33–35]. Alignment quality was evaluated using Samtools' flagstat [36] and Picard tools CalculateHsMetrics. Samples with mean target coverage less than 10 \times were excluded. Variants were called on individual samples by GATK 3.2.2 HaplotypeCaller in gVCF mode, according to the procedures recommended for single samples all-sites calling [34, 35]. Genotyping was made jointly for the complete case group with GATK 3.2.2 GenotypeGVCFs, and thereafter, only single nucleotide polymorphisms (SNPs) were considered in further analyses. Genotype filters were applied with VCFtools [37] requiring a minimum read depth of 8 and a minimum genotype Phred quality score of 20 [38]. SNPs were assigned probability scores using the GATK 3.2.2 VariantRecalibrator and filtered at tranche level 99.0. Furthermore, heterozygous calls were kept if their variant allele balance across all samples was between 0.2 and 0.8. Positions deviating from Hardy-Weinberg equilibrium ($P < 10^{-6}$) and monomorphic sites were excluded. SNPs were annotated with identifiers from dbSNP 137 [39] and had their impact predicted with SnpEff 4.1 [40]. Annotations for noncoding variants were retrieved using HaploReg v4.1 [41] and RegulomeDB [42].

Furthermore, we used the IGV [43] and UCSC genome browser [44] for sequence alignments and functional investigation, respectively.

Sample ancestry and relatedness estimation

We determined the study subjects' ancestry based on off-target genetic differences using the LASER software, University of Michigan [45] (Figure S2). We set the dimensions of the reference space (k) to 4, and we used 20 principal components (K) for calculation and projection of the study samples after 10 iterations (r). Study subjects were superimposed on the Human Genome Diversity Projects reference panel and predicted non-Europeans were excluded if they did not fall within the European reference coordinates defined by Wang *et al.* [45]. Intersample relatedness coefficients were inferred pairwise with VCFtools [37] that implements the algorithm by Manichaikul *et al.* [46]. The coefficients representing first-degree relatedness ($\Phi_{ij} > 0.177$) were used to identify closely related pairs. The sample with the lowest mean target coverage in case–case and control–control pairs was excluded, as well as both samples in case–control pairs. Furthermore, seeming twins, not born on the same day, were discarded.

Quality control

Samples with discordance between reported sex and sex inferred from sequencing data were excluded. Furthermore, samples with a rate of missing data, heterozygosity ratio, or transition–transversion ratio deviating more than five standard deviations from the mean, were discarded. Finally, samples with singleton counts, calculated on variants called in the majority of samples (0.95 call rate), deviating by five standard deviations or more, were removed. With this adjustment, we were able to pinpoint both low number of singletons revealing duplicates or contamination and high singleton counts indicating low-quality DNA or suggesting foreign ancestry (Figures S3–S5).

After removing all low-quality samples, variants with more than 20% missing data across all subjects were excluded as well as subjects with more than 35% missing data (Figure S6). We set these filter thresholds for call rates to reach final variant and individual call rates $\geq 95\%$. By limiting the number of variants and individuals discarded, whilst keeping adequate average call rates, we aimed at increasing power in both aggregate tests

and single-variant association analyses. Lastly, variants were pruned out if their call rate in cases versus controls differed more than ten percentage points (Figure S6).

Association analyses – common variants

Logistic regression was applied for association analyses of common single variants with a minor allele frequency above 1% ($MAF > 0.01$) [47]. The logistic regression analysis was performed with an additive model, which is the most common model for unknown modes of inheritance [48, 49]. Principal component analysis (PCA) was compared with multidimensional scaling (MDS) in correcting population stratification in our data set, with a method adapted from Wang *et al.* [50] (Figures S7–S9, Table S1). In our data set, MDS performed better in decreasing the inflation factor (Figure S10) and the three first MDS-generated dimensions were included as regression covariates to account for population substructure. Alternative covariates such as sex and amplification status were also tested consecutively. Lambda values were calculated with and without the top-associated loci, and P values were adjusted by means of genomic control [51] to account for cryptic relatedness inflating the test statistic (Tables S2–S3). The MDS and single-variant association analysis were performed in PLINK [52]. PCA was calculated using EIGENSOFT [53]. LD structures were measured (r^2) and visualized with Haploview [54].

Rare variant gene-based test

To gain statistical power to associate rare variants with disease, variants have to be evaluated collectively in predefined groups. We aggregated genetic variants by the genes in which they were located and calculated the risk associated with each gene and intergenic space using the SKAT statistic implemented in the SCORE-Seq software [55]. Aggregates including less than two variants were disregarded. Using gene boundaries defined by SnpEff-based annotation, comprised of Ensembl GRCh37 genes and intergenic regions, we tested the effect of rare variants ($MAF \leq 0.01$) [47] on AAD susceptibility. We made two analyses with rare variants. First, we tested the whole set of rare variants, and in a second analysis, a subset comprised of only moderate- and high-impact variants. In both tests, we included three covariates of MDS dimensions but no genomic control [56]. We used a Bonferroni-corrected threshold

(3.3×10^{-6} , $\alpha = 0.05$ corrected for the total number of aggregates in the two analyses; 15 089) to evaluate the statistical significance of our findings.

In silico HLA typing

HLA types were inferred from *de novo* assembly of sequencing reads using MOSAIK aligner and the ATHLATES software [57, 58]. HLA reference definitions were updated to the IMGT/HLA release 3.22.0 with sequences aligned 10 October 2015. Alleles at loci A, B and C of class I, and DPA, DPB, DQA, DQB and DRB of class II were deduced. Haplotype frequencies were compared both per haplotype, and per haplotypic pair. *P* values were computed from χ^2 -test in comparison with the most equally distributed haplotype. The odds ratio 95% confidence intervals were calculated according to Szumilas *et al.* [59].

Results

Our study initially included 700 primary adrenal insufficiency case samples and 1501 control samples. We set our focus on characterizing the genetic background of complex inherited autoimmune Addison's disease (AAD). Therefore, we stringently excluded 173 cases where other causes of primary adrenal failure were suspected, including all the 119 cases that were negative for 21-hydroxylase autoantibodies (Materials and methods), thereby ensuring a case group with homogeneous disease pathoaetiology. We also excluded 25 case and 28 control samples estimated to be either of non-European ancestry or highly related. After the application of all quality control measures, 479 case and 1394 control samples remained for further analyses. Of the 479 cases, 298 (62.2%) were female, and the median age of onset was 34 years (range 3–78). Figures S1–S5 depict in detail the sample quality control process. The average target coverage of the remaining individuals was $37.2 \times$ (SD 12.9).

The final data set included 103 120 common SNPs (MAF > 1%) and 304 338 rare SNPs (MAF \leq 1%). The mean variant call rate (96%) and the mean individual call rate (95%) indicated a data set of high quality.

BACH2 is a novel AAD susceptibility locus

We conducted a single-SNP association analysis comparing common variants with MAF higher than

1%, across all study participants. To account for multiple testing, we used a well-established statistical significance threshold for genomewide association studies, $P = 5 \times 10^{-8}$ [48], which is conservative for our extended exome study. MDS plots revealed population substructures, separating study subjects sampled in northern versus central/southern Sweden (Figure S7). The overdispersion of the test statistic was partially reduced by including the first three MDS-generated dimensions as covariates, and residual inflation was corrected with genomic control to account for cryptic relatedness ($\lambda_{GC} = 1.02$) (Tables S3, Figures S11–S12).

The single-variant association test identified two peaks of association on chromosome 6 (Fig. 1). In consecutive analyses, using no or alternative covariates, the peaks remained statistically significant. With sequential conditioning of the lead SNPs, we identified three statistically significant independent signals (Table 1). We identified the gene *BACH2* (*BTB domain and CNC homolog 2*) as a novel AAD risk locus (rs62408233-A, OR = 2.01 (1.71–2.37), $P = 1.66 \times 10^{-15}$, MAF 0.46/0.29 in cases/controls). Minor allele frequencies from the 1000 Genomes Project [60] and call rates for the associated SNPs in *BACH2* are presented in Table S4. Furthermore, we found two signals of association in the HLA region (rs41315836-G, OR = 0.20 (0.13–0.29), $P = 5.73 \times 10^{-16}$; rs17221059-A, OR = 2.29 (1.84–2.84), $P = 9.59 \times 10^{-13}$). A comprehensive summary of top-associated SNPs is presented in Table 1 and conditional tests in Figure S13.

Separating AAD from concomitant autoimmune diseases

BACH2 had previously been reported to have an association with other tissue-specific autoimmune diseases such as type 1 diabetes [61], autoimmune thyroid disease [62], coeliac disease [63], inflammatory bowel disease [64] and vitiligo [65]. Because AAD cases have a high degree of autoimmune comorbidity, the potential risk of confounding in an assorted AAD case group is substantial. It was also known that *BACH2* had been associated with the presence of thyroid peroxidase (TPO) autoantibodies [62]. Because the Swedish Addison Registry (SAR) includes detailed clinical information and SAR samples have been assayed for TPO autoantibodies, we were able to identify cases diagnosed with any, present or previous, autoimmune diseases besides AAD, and furthermore, to identify TPO autoantibody-positive cases [66].

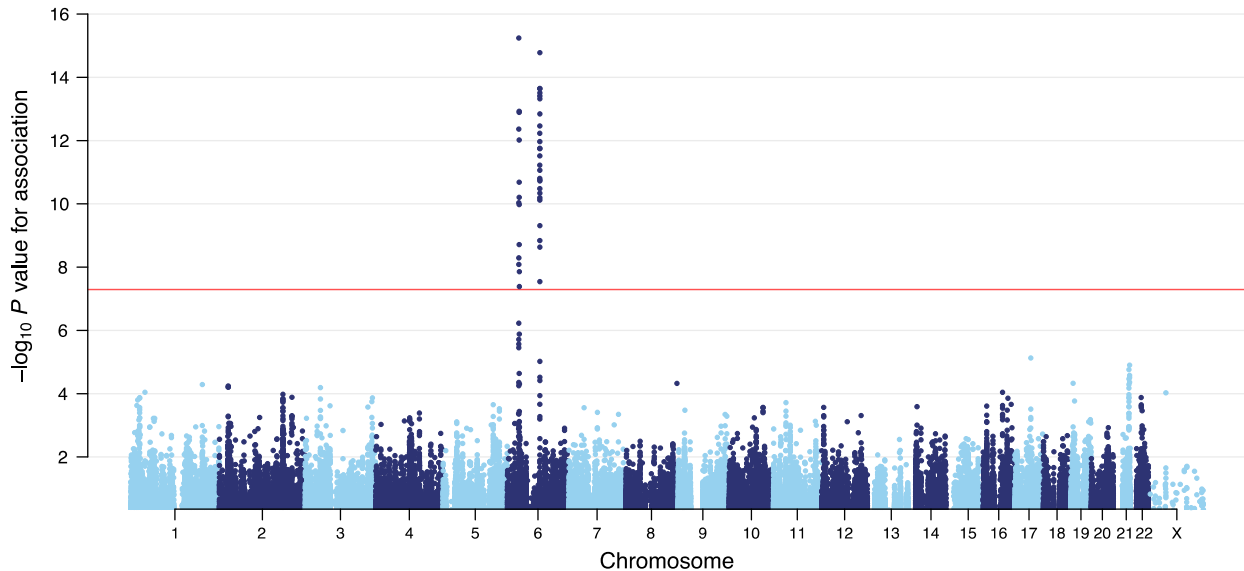


Fig. 1 Association between single-nucleotide polymorphisms (SNPs) and autoimmune Addison's disease (AAD), exhibited as P values and genomic location. The association between SNPs and AAD, as calculated by an additive logistic regression model, is plotted as the negative log of the P value against chromosomal location. Three MDS generated dimensions are included as covariates and P values are corrected with genomic control. The horizontal red line denotes the statistical significance level of 5×10^{-8} . Bands in shades of blue depict the chromosome boundaries.

Table 1 Lead variants independently associated with autoimmune Addison's disease in single-variant association analysis

Lead variant	Chromosome	Position	Locus	Minor/Maj. allele	Minor allele frequency		Association	
					Cases	Controls	OR (95% CI) ^a	P value ^b
rs41315836	6	31 975 151	HLA	G/T	0.04	0.17	0.20 (0.13–0.29)	5.73×10^{-16}
rs17221059	6	33 048 348	HLA	A/G	0.20	0.09	2.29 (1.84–2.84)	9.59×10^{-13}
rs62408233	6	90 976 609	BACH2	A/G	0.46	0.29	2.01 (1.71–2.37)	1.66×10^{-15}

^aThe odds ratio (OR) represents the effect of the minor allele on the odds of observing AAD, under an additive logistic regression model, such that an OR greater than one implies that the minor allele increases odds. ^bThe presented P values are calculated with population substructure covariates and corrected with genomic control.

Recalculating the logistic regression using only the cases with isolated AAD and without TPO autoantibodies (119 cases), the signal in *BACH2* remained statistically significant ($P = 3.9 \times 10^{-8}$, Figure S14). TPO autoantibody-positive and autoantibody-negative cases had the same mean age, lowering the possibility that negative cases had a genetic risk that was not yet penetrant (Figure S15, Table S5).

From *BACH2* variants to disease

From the total of 103 120 common variants, 40 exceeded our stringent statistical significance

level ($P < 5 \times 10^{-8}$), 26 of which were located in *BACH2*. Conditioning on the top SNP in *BACH2* deflated all association signals at this locus, underlining the linkage between associated markers in the region (Figure S13). The associated haplotype block (HSA6:90,815,190–91,029,762) covers the 5' noncoding exons and introns of *BACH2* (Fig. 2). In general, it is complex to decipher which single genetic variants have an effect on disease development. The presence of linkage disequilibrium and the difficulty of detecting the effects of single variants confound the dissection of which genes have an important impact on disease predisposition. Therefore, we

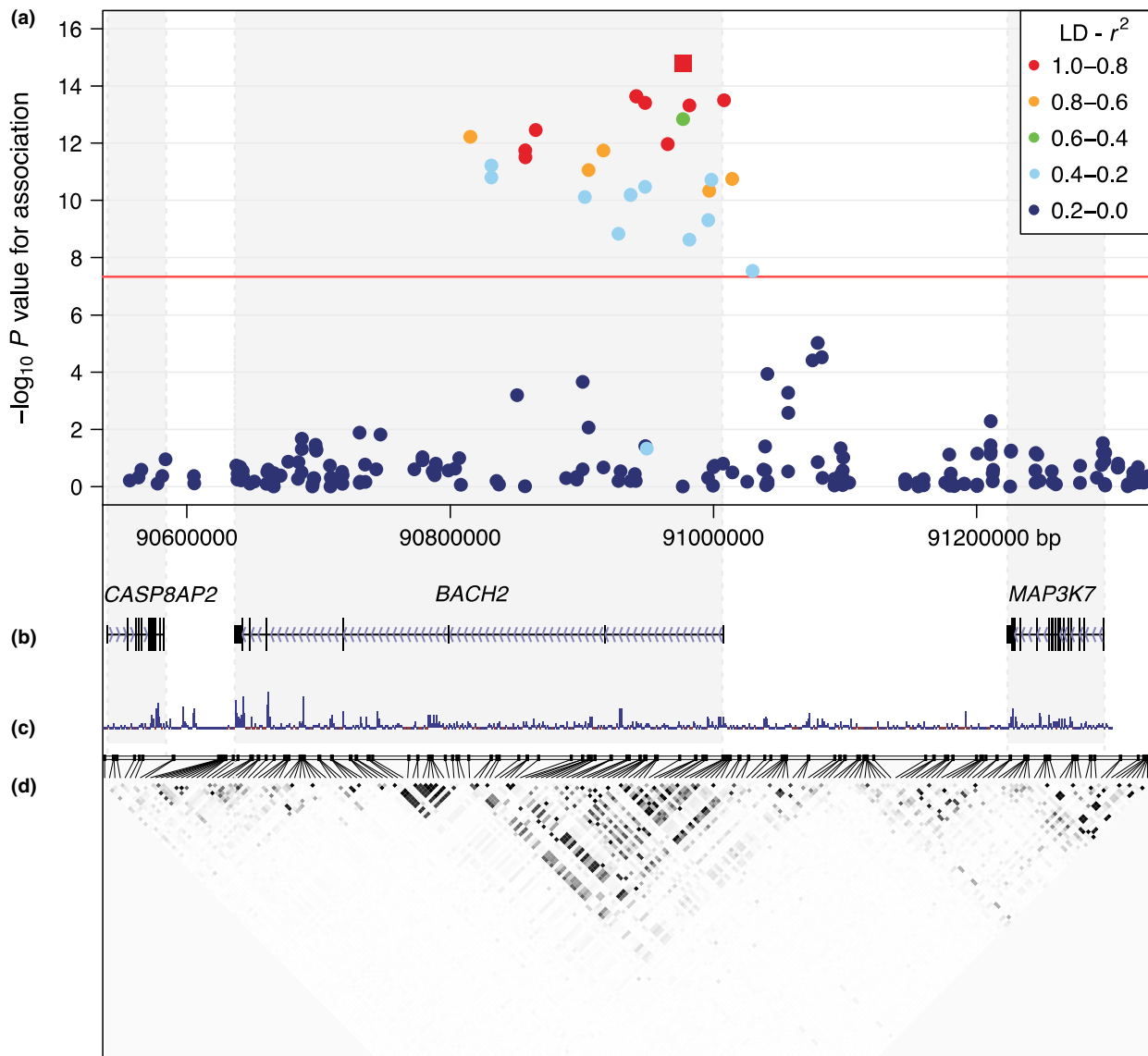


Fig. 2 Association between single-nucleotide polymorphisms (SNPs) and autoimmune Addison's disease (AAD), the location of nearby genes, conservation, and the surrounding linkage disequilibrium (LD) structure, in a region of chromosome 6. Shown is a region around BACH2 and its neighboring genes on chromosome 6q15. In panel (a), the association between SNPs and AAD, as calculated by an additive logistic regression model, is plotted as the negative log of the P value against chromosomal position. The horizontal red line denotes the level considered statistically significant, 5×10^{-8} . SNPs are colored according to their level of genotype correlation (r^2) with the most significant SNP (red square, rs62408233), which has a P value of 1.7×10^{-15} . Panel (b) depicts the genomic locations of the genes present in the region, and (c) a track with basewise conservation between 100 vertebrates, both from the genome browser at UCSC, assembly hg19. Arrows indicate the direction of transcription (right to left for BACH2). The main BACH2 transcript, shown here, has three 5' non-coding exons. Panel (d) illustrates the linkage disequilibrium (LD), as measured by pairwise correlation (r^2) of SNPs, and plotted by Haploview. Darker shading marks higher genotype correlation and stronger LD.

compiled a table with available epigenetic data from both HaploReg v4.1 [41] and RegulomeDB [42] to highlight relative functional importance of

associated SNPs (Table S6). Of the 26 associated SNPs in this region, all except two overlap enhancer histone marks in blood (B and T

lymphocytes) and four are associated with *BACH2* eQTLs.

HLA typing confirmed known AAD risk alleles

Using ATHLATES for HLA typing from sequencing reads [57], we recapitulate previously established Scandinavian associations in the HLA class II region (Table 2). For instance, we could confirm the risk associated with the haplotypes DQB1*02:01-DQA1*05:01-DRB1*03:01, corresponding to the serotype DR3-DQ2, and DQB1*03:02-DQA1*03:01-DRB1*04:04 corresponding to the serotype DR4-DQ8 [8, 67]. We could also confirm the protective effect of DQB1*06:03-DQA1*01:03-DRB1*13:01. Moreover, we could replicate the association with the risk allele DRB1*04:03, previously identified in Russian patients with AAD [68]. This allelic association was also present on the haplotype DQB1*03:02-DQA1*03:01-DRB1*04:03.

By further analysing haplotypes in the pairs in which they occur, we could confirm that homozygosity for risk haplotypes is linked to an increased susceptibility to AAD ($6.4 < OR < 8.8$) (Table 3). The pair of haplotypes associated with the highest risk was the compound heterozygous DQB1*02:01-DQA1*05:01-DRB1*03:01 and DQB1*03:02-DQA1*03:01-DRB1*04:04 ($OR = 18$ (7.2–46), $P = 4.9 \times 10^{-12}$), also concordant with current knowledge [8, 67, 69].

Rare variant contribution to AAD

We included all detected variants with MAF lower than or equal to 1% in a gene-based aggregate test. First, we performed association analysis with the whole set of variants. Thereafter, we selected only high- and moderate-impact variants as enriching for harmful alleles could increase statistical power in aggregate tests. The inflation factors (λ) were 1.07 and 1.02, respectively (Table S3). Neither with the inclusion of all variants nor with only high- or moderate-impact variants did any gene exceed the statistical significance threshold ($P = 3.3 \times 10^{-6}$) (Figure S16). Hence, no gene aggregate, powered with only rare variants, was associated with AAD in this study.

Discussion

This is the first large-scale study aiming at identifying the genetic background of autoimmune Addison's disease (AAD). For this purpose, we developed an array to capture a comprehensive selection of potential causative genes and surrounding noncoding regions. Our Swedish collection of AAD samples is one of the largest available for investigation, and the detailed phenotypic characterization of all cases underpins the foundation for this study.

We identified a novel risk locus, *BACH2*, associated with AAD and confirmed a well-established

Table 2 Most frequent HLA class II haplotypes and their associations with autoimmune Addison's disease

DQB DQA DRB	Cases <i>n</i> (%) ^a	Controls <i>n</i> (%) ^a	Odds ratio (95% CI)	<i>P</i> value ^b
02:01 05:01 03:01	358 (37.4)	341 (12.2)	2.7 (1.8–4.1)	$3.7 \times 10^{-6***}$
03:02 03:01 04:04	174 (18.2)	149 (5.3)	3 (1.9–4.6)	$1.9 \times 10^{-6***}$
06:02 01:02 15:01	93 (9.7)	430 (15.4)	0.55 (0.35–0.86)	$1.2 \times 10^{-2*}$
03:02 03:01 04:01	66 (6.9)	172 (6.2)	0.98 (0.6–1.6)	1.0
04:02 04:01 08:01	35 (3.7)	89 (3.2)	1 (ref)	
03:02 03:01 04:03	18 (1.9)	12 (0.4)	3.8 (1.7–8.7)	$2.1 \times 10^{-3**}$
06:03 01:03 13:01	15 (1.6)	230 (8.2)	0.17 (0.086–0.32)	$1.2 \times 10^{-8***}$
Other haplotypes	119 (20.8)	1365 (49.0)		
Total	958 (100.2)	2788 (99.9)		

^aThe table is restricted to include haplotypes with at least 10 cases and 10 controls. The remaining haplotypes are summed in 'Other haplotypes'. ^b*P* values from the χ^2 distribution are calculated with the counts from the most equally distributed HLA haplotype as reference (ref). **P* value smaller than 0.05, ***P* value smaller than 0.01, ****P* value smaller than 0.001.

Table 3 Most frequent pairs of HLA class II haplotypes and their associations with autoimmune Addison's disease

DQB DQA DRB / DQB DQA DRB	Cases <i>n</i> (%) ^a	Controls <i>n</i> (%) ^a	Odds ratio (95% CI)	<i>P</i> value ^b
02:01 05:01 03:01 / 03:02 03:01 04:04	108 (22.5)	20 (1.4)	18 (7.2–46)	$4.9 \times 10^{-12***}$
02:01 05:01 03:01 homozygous	32 (6.7)	17 (1.2)	6.4 (2.4–17)	$3.0 \times 10^{-4***}$
02:01 05:01 03:01 / 03:01 05:05 11:01	17 (3.5)	7 (0.5)	8.2 (2.5–27)	$6.9 \times 10^{-4***}$
02:01 05:01 03:01 / 03:01 05:05 12:01	13 (2.7)	5 (0.4)	8.8 (2.4–32)	$1.5 \times 10^{-3**}$
03:02 03:01 04:04 homozygous	13 (2.7)	5 (0.4)	8.8 (2.4–32)	$1.5 \times 10^{-3**}$
02:01 05:01 03:01 / 03:01 03:03 04:01	16 (3.3)	14 (1.0)	3.9 (1.3–11)	$2.3 \times 10^{-2*}$
02:01 05:01 03:01 / 03:02 03:01 04:01	28 (5.8)	31 (2.2)	3.0 (1.2–7.8)	$3.1 \times 10^{-2*}$
02:01 05:01 03:01 / 06:02 01:02 15:01	38 (7.9)	52 (3.7)	2.5 (1–6)	7.0×10^{-2}
02:01 05:01 03:01 / 04:02 04:01 08:01	12 (2.5)	14 (1.0)	2.9 (0.96–8.7)	0.10
04:02 04:01 08:01 / 06:02 01:02 15:01	11 (2.3)	13 (0.9)	2.9 (0.93–8.8)	0.12
03:01 03:03 04:01 / 03:02 03:01 04:01	6 (1.3)	6 (0.4)	3.4 (0.85–13)	0.16
03:02 03:01 04:01 / 03:02 03:01 04:04	9 (1.9)	12 (0.9)	2.5 (0.79–8.2)	0.2
06:02 01:02 15:01 / 03:01 05:05 12:01	6 (1.3)	9 (0.6)	2.2 (0.61–8.3)	0.37
02:01 05:01 03:01 / 06:03 01:03 13:01	5 (1.0)	33 (2.4)	0.51 (0.15–1.7)	0.44
03:02 03:01 04:04 / 06:03 01:03 13:01	5 (1.0)	8 (0.6)	2.1 (0.54–8.3)	0.47
03:02 03:01 04:04 / 06:02 01:02 15:01	8 (1.7)	27 (2.1)	1 (ref)	
Other haplotype pairs	152 (31.7)	1121 (80.4)		
Total	479 (99.8)	1394 (99.7)		

^aThe table is restricted to include haplotypes with at least 5 cases and 5 controls. The remaining haplotypes are summed in 'Other haplotype pairs'. ^b*P* values from the χ^2 distribution are calculated with the counts from the most equally distributed HLA haplotype as reference (ref). **P* value smaller than 0.05, ***P* value smaller than 0.01, ****P* value smaller than 0.001.

contribution of HLA risk alleles to the risk of developing this disease. Furthermore, we were able to show that neither autoimmune comorbidity nor the presence of TPO autoantibodies was required for the strong association signal in *BACH2*. From the rare variant association analysis, we revealed no additional candidate loci.

BACH2 encodes a basic leucine zipper transcription repressor with important functions in activated B cells. Acting as a heterodimer with Maf proteins, *BACH2* is an important regulator of the immunoglobulin heavy chain transcription, as well as class switching [70, 71]. Although initially linked to B-cell function, *BACH2* has also been found to be crucial for T-cell differentiation. As the most prominent T-cell super-enhancer, a lack of *BACH2* shifts the T-cell balance towards effector cells (CD8+), at the expense of helper and regulatory T cells [72]. In the light of the *BACH2* locus being associated with the development of other autoimmune diseases, and its functional

characterization, *BACH2* emerges as a major gene candidate in organ-specific autoimmunity. Here, we present a 215-kb *BACH2* region associated with AAD. Even though the new association does not directly infer a causal relationship with AAD, it provides direction for further investigations. The 5' noncoding region of *BACH2*, including both upstream and intronic sequence and potentially harbouring regulatory variants, poses a candidate locus for functional studies.

The newly identified risk locus includes 26 previously known common variants. Due to high linkage disequilibrium, it is complicated to dissect the effect of single variants in the 5' noncoding region of *BACH2*. The vast majority of these variants are either located in enhancer/promoter regions in relevant cell types or have been associated with the expression of *BACH2* in large-scale eQTL-mapping studies [73]. Therefore, considering both genetic and epigenetic data, we are currently not able to define a single most important AAD-associated

BACH2 variant. Instead, we postulate that multiple variants may contribute to the risk of developing AAD and could be selected for functional studies.

Using *in silico* typing from sequencing reads, we could confirm the risk conferred by certain HLA haplotype pairs [8, 67–69]. The strongest risk factor is by far the combination of HLA alleles DRB1*03:01 and DRB1*04:04, as previously shown in other populations [8, 67–69]. Despite the fact that we selected only patients with AAD reacting against the same autoantigen, 21-hydroxylase, we detect a great variability at the HLA locus. Interestingly, the majority of AAD cases did not carry any of the high-risk HLA alleles. This suggests that it is not necessary to have complete homogeneity at the HLA locus in order to elicit immune reactivity against 21-hydroxylase, in the context of AAD.

Apart from HLA, other already published loci found to be associated with AAD in candidate gene studies did not exceed our significance threshold [27, 28]. This could be due to several factors. First, a strict significance threshold is needed in our study to avoid false-positive associations when testing multiple hypotheses. Secondly, associated alleles could be population-specific and therefore may not be reproducible in our study population. Lastly, directed candidate gene studies do not have the possibility to account for population substructure derived from genetic data and are therefore more prone to spurious associations.

Even though correction for population stratification is extensively enabled in our study, rare variant association analysis is limited by the current sample size [74]. Because not all genes and not all noncoding regions could be selected in our study, whole-genome sequencing would be a desirable next step. That would also enable the assessment of the potential role of structural variation in the disease development [75]. Functional studies of potentially regulatory variants and replication in other populations should be performed to build on the association with *BACH2*.

The current study adopts the benefits of high-throughput sequencing to tackle the shortcomings of candidate gene studies. By extending the exome with noncoding conserved elements, we gain the opportunity to identify regulatory variants, which

are suggested to be major contributors to complex diseases [76].

Conclusion

We have identified *BACH2* as a major risk locus in AAD. Knowledge gained from this study may aid the future development of preventive treatment.

Author contributions

The study was conceived by KL-T, OK and GRP. The study design and optimization was performed by NL, JN, AM, EM, KMA, AL, GA, LR, JRSM, KL-T, OK and GRP. Acquisition of study material and data was implemented by MB, NL, JN, FD, AM, GNE, LHR, JD, HZ, ÅK, ÅH, FHGF, EM, KT, SRD, PS, LR, JRSM, A-LH, JW, OE, PD, SB, OK and GRP. Analysis and interpretation of data was carried out by DE, MB and GRP. The manuscript was drafted by DE, MB, NL, OK and GRP, with input from all co-authors.

Conflict of interest statement

Dr. Rönnblom reports grants from AstraZeneca, grants from the Swedish Research Council, grants from the KAW Foundation, grants from the Swedish Rheumatism Foundation, grants from the King Gustaf V's 80-year Foundation, during the conduct of the study, grants from UCB, outside the submitted work. Dr. Olle Kämpe reports grants from the Swedish Research Council, the Novo Nordisk Foundation and the Torsten and Ragnar Söderberg's Foundations. He is board member of Olink Biosciences. The remaining authors had no conflict of interests to disclose.

Acknowledgements

We would like to thank all patients and control subjects participating in this study for making this research possible. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala, supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. Computational resources were provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), cluster and storage project b2014026. We would

especially like to thank Dr. Maija-Leena Eloranta at the Dept. of Medical Sciences, Uppsala University, Uppsala, Sweden, for contributing to both control sample collection and capture array design and Dr. Cindy Wong for critical review of the manuscript. We would also like to acknowledge Emma Ivansson, Malin Melin and Maria Wilbe for their technical support, as well as the Medical Biobank of Northern Sweden. Sources of Research Support: The Swedish Research Council, Torsten and Ragnar Söderberg Foundations, the European Union Seventh Framework Programme grant 201167 EurAdrenal fp7 consortium, the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Society for Medical Research, the Swedish Society of Medicine, the NovoNordisk Foundation, Tore Nilson's Foundation for Medical Research, Karolinska Institutet, the Swedish Research Council Formas, the Knut and Alice Wallenberg Foundation, the Marianne and Marcus Wallenberg Foundation, the Swedish Rheumatism Foundation, the King Gustaf V's 80-year Foundation and the Åke Wiberg Foundation. An AstraZeneca Science for Life Laboratory Research Collaboration grant (DISSECT) supported sequencing of a subset of controls.

References

- Zelissen PM, Bast EJ, Croughs RJ. Associated autoimmunity in Addison's disease. *J Autoimmun* 1995; **8**: 121–30.
- Husebye E, Allolio B, Arlt W *et al.* Consensus statement on the diagnosis, treatment and follow-up of patients with primary adrenal insufficiency. *J Intern Med* 2014; **275**: 104–15.
- Björnsdóttir S, Sundström A, Ludvigsson JF, Blomqvist P, Kämpe O, Bensing S. Drug prescription patterns in patients with Addison's disease: a Swedish population-based cohort study. *J Clin Endocrinol Metab* 2013; **98**: 2009–18.
- Kong MF, Jeffcoate W. Eighty-six cases of Addison's disease. *Clin Endocrinol* 1994; **41**: 757–61.
- Willis AC, Vince FP. The prevalence of Addison's disease in Coventry, UK. *Postgrad Med J* 1997; **73**: 286–8.
- Laureti S, Vecchi L, Santeusano F, Falorni A. Is the prevalence of Addison's disease underestimated? *J Clin Endocrinol Metab* 1999; **84**: 1762.
- Lovas K, Husebye ES. High prevalence and increasing incidence of Addison's disease in western Norway. *Clin Endocrinol* 2002; **56**: 787–91.
- Erichsen MM, Lovas K, Skiningsrud B *et al.* Clinical, immunological, and genetic features of autoimmune primary adrenal insufficiency: observations from a Norwegian registry. *J Clin Endocrinol Metab* 2009; **94**: 4882–90.
- Quinkler M, Dahlqvist P, Husebye ES, Kampe O. A European Emergency Card for adrenal insufficiency can save lives. *Eur J Intern Med* 2015; **26**: 75–6.
- Mason AS, Meade TW, Lee JA, Morris JN. Epidemiological and clinical picture of Addison's disease. *Lancet* 1968; **2**: 744–7.
- Falorni A, Laureti S, De Bellis A *et al.* Italian addison network study: update of diagnostic criteria for the etiological classification of primary adrenal insufficiency. *J Clin Endocrinol Metab* 2004; **89**: 1598–604.
- Fairchild RS, Schimke RN, Abdou NI. Immunoregulation abnormalities in familial Addison's disease. *J Clin Endocrinol Metab* 1980; **51**: 1074–7.
- Heggarty H. Addison's disease in identical twins. *Br Med J* 1968; **1**: 559.
- Hewitt PH. Addison's disease occurring in sisters. *Br Med J* 1957; **2**: 1530–1.
- Russell GA, Coulter JB, Isherwood DM, Diver MJ, Smith DS. Autoimmune Addison's disease and thyrotoxic thyroiditis presenting as encephalopathy in twins. *Arch Dis Child* 1991; **66**: 350–2.
- Simmonds JP, Lister J. Auto-immune Addison's disease in identical twins. *Postgrad Med J* 1978; **54**: 552–4.
- Smith ME, Gough J, Galpin OP. Addison's disease in identical twins. *Br Med J* 1963; **2**: 1316.
- Mitchell AL, Bøe Wolff A, MacArthur K *et al.* Linkage analysis in autoimmune Addison's disease: NFATC1 as a potential novel susceptibility locus. *PLoS One* 2015; **10**: e0123550.
- Condon J, Shaw JE, Luciano M, Kyvik KO, Martin NG, Duffy DL. A study of diabetes mellitus within a large sample of Australian twins. *Twin Res Hum Genet* 2008; **11**: 28–40.
- Kaprio J, Tuomilehto J, Koskenvuo M *et al.* Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* 1992; **35**: 1060–7.
- Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J. Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. *Diabetes* 2003; **52**: 1052–5.
- Kyvik KO, Green A, Beck-Nielsen H. Concordance rates of insulin dependent diabetes mellitus: a population based study of young Danish twins. *BMJ* 1995; **311**: 913–7.
- Mittag F, Büchel F, Saad M *et al.* Use of support vector machines for disease risk prediction in genome-wide association studies: concerns and opportunities. *Hum Mutat* 2012; **33**: 1708–18.
- Brix TH, Kyvik KO, Christensen K, Hegedüs L. Evidence for a major role of heredity in Graves' disease: a population-based study of two danish twin cohorts 1. *J Clin Endocrinol Metab* 2001; **86**: 930–4.
- Winqvist O, Karlsson FA, Kämpe O. 21-Hydroxylase, a major autoantigen in idiopathic Addison's disease. *Lancet* 1992; **339**: 1559–62.
- Falorni A, Bini V, Betterle C *et al.* Determination of 21-hydroxylase autoantibodies: inter-laboratory concordance in the Euradrenal International Serum Exchange Program. *Clin Chem Lab Med* 2015; **53**: 1761–70.
- Mitchell AL, Pearce SH. Autoimmune Addison disease: pathophysiology and genetic complexity. *Nat Rev Endocrinol* 2012; **8**: 306–16.
- Mitchell AL, Macarthur KD, Gan EH *et al.* Association of autoimmune Addison's disease with alleles of STAT4 and GATA3 in European cohorts. *PLoS One* 2014; **9**: e88991.
- Uppsala Bioresource, Uppsala, Sweden [BIORESOURCE].

- 30 Arlt W, Allolio B. Adrenal insufficiency. *Lancet* 2003; **361**: 1881–93.
- 31 Lindblad-Toh K, Garber M, Zuk O *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 2011; **478**: 476–82.
- 32 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–60.
- 33 McKenna A, Hanna M, Banks E *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–303.
- 34 DePristo MA, Banks E, Poplin R *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491–8.
- 35 Van der Auwera GA, Carneiro MO, Hartl C *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013; **43**: 11 0 1–33.
- 36 Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–9.
- 37 Danecek P, Auton A, Abecasis G *et al.* The variant call format and VCFtools. *Bioinformatics* 2011; **27**: 2156–8.
- 38 Carson AR, Smith EN, Matsui H *et al.* Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC Bioinformatics* 2014; **15**: 125.
- 39 Database of Single Nucleotide Polymorphisms (dbSNP) [Internet]. Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. Available from: Available at: <http://www.ncbi.nlm.nih.gov/SNP/>. Accessed 3 September 2015
- 40 Cingolani P, Platts A, Wang le L *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012; **6**: 80–92.
- 41 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; **40**: D930–4.
- 42 Boyle AP, Hong EL, Hariharan M *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22**: 1790–7.
- 43 Robinson JT, Thorvaldsdottir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011; **29**: 24–6.
- 44 Kent WJ, Sugnet CW, Furey TS *et al.* The human genome browser at UCSC. *Genome Res* 2002; **12**: 996–1006.
- 45 Wang C, Zhan X, Bragg-Gresham J *et al.* Ancestry estimation and control of population stratification for sequence-based association studies. *Nat Genet* 2014; **46**: 409–15.
- 46 Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010; **26**: 2867–73.
- 47 Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001; **17**: 502–10.
- 48 Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA* 2008; **299**: 1335–44.
- 49 Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. *Nat Protoc* 2011; **6**: 121–33.
- 50 Wang D, Sun Y, Stang P, Berlin JA, Wilcox MA, Li Q. Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. *BMC Proc* 2009; **3(Suppl 7)**: S109.
- 51 Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004. Epub 2001/04/21.
- 52 Purcell S, Neale B, Todd-Brown K *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–75.
- 53 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–9.
- 54 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–5.
- 55 Lin D-Y, Tang Z-Z. A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet* 2011; **89**: 354–67.
- 56 Liu Q, Nicolae DL, Chen LS. Marbled inflation from population structure in gene-based association studies with rare variants. *Genet Epidemiol* 2013; **37**: 286–92.
- 57 Liu C, Yang X, Duffy B *et al.* ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res* 2013; **41**: e142.
- 58 Lee W-P, Stromberg MP, Ward A, Stewart C, Garrison EP, Marth GT. MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. *PLoS One* 2014; **9**: e90581.
- 59 Szumilas M. Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry* 2010; **19**: 227–9.
- 60 The Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; **526**: 68–74.
- 61 Grant SF, Qu HQ, Bradfield JP *et al.* Follow-up analysis of genome-wide association data identifies novel loci for type 1 diabetes. *Diabetes* 2009; **58**: 290–5.
- 62 Medici M, Porcu E, Pistis G *et al.* Identification of novel genetic Loci associated with thyroid peroxidase antibodies and clinical thyroid disease. *PLoS Genet* 2014; **10**: e1004123.
- 63 Garner C, Ahn R, Ding YC *et al.* Genome-wide association study of celiac disease in North America confirms FRMD4B as new celiac locus. *PLoS One* 2014; **9**: e101428.
- 64 Jostins L, Ripke S, Weersma RK *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012; **491**: 119–24.
- 65 Jin Y, Birlea SA, Fain PR *et al.* Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat Genet* 2012; **44**: 676–80.
- 66 Dalin F, Eriksson G, Dahlqvist P *et al.* Clinical and immunological characteristics of Autoimmune Addison's disease: a nationwide Swedish multicenter study. *J Clin Endocrinol Metab* 2016. In press.
- 67 Myhre AG, Undlien DE, Lovas K *et al.* Autoimmune adrenocortical failure in Norway autoantibodies and human leukocyte antigen class II associations related to clinical features. *J Clin Endocrinol Metab* 2002; **87**: 618–23.
- 68 Gombos Z, Hermann R, Kiviniemi M *et al.* Analysis of extended human leukocyte antigen haplotype association with Addison's disease in three populations. *Eur J Endocrinol* 2007; **157**: 757–61.

- 69 Skinningsrud B, Lie BA, Lavant E *et al.* Multiple loci in the HLA complex are associated with Addison's disease. *J Clin Endocrinol Metab* 2011; **96**: E1703–8.
- 70 Muto A, Hoshino H, Madisen L *et al.* Identification of *BACH2* as a B-cell-specific partner for small maf proteins that negatively regulate the immunoglobulin heavy chain gene 3' enhancer. *EMBO J* 1998; **17**: 5734–43.
- 71 Muto A, Ochiai K, Kimura Y *et al.* *BACH2* represses plasma cell gene regulatory network in B cells to promote antibody class switch. *EMBO J* 2010; **29**: 4048–61.
- 72 Roychoudhuri R, Hirahara K, Mousavi K *et al.* *BACH2* represses effector programs to stabilize T(reg)-mediated immune homeostasis. *Nature* 2013; **498**: 506–10.
- 73 Westra H-J, Peters MJ, Esko T *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45**: 1238–43.
- 74 Moutsianas L, Agarwala V, Fuchsberger C *et al.* The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. *PLoS Genet* 2015; **11**: e1005165.
- 75 Weischenfeldt J, Symmons O, Spitz F, Korbel JO. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet* 2013; **14**: 125–38.
- 76 Lowe WL Jr, Reddy TE. Genomic approaches for understanding the genetics of complex disease. *Genome Res* 2015; **25**: 1432–41.

Correspondence: Daniel Eriksson, Center for Molecular Medicine, Department of Medicine (Solna), Karolinska Institutet, Stockholm, Sweden.

(e-mail: daniel.eriksson@ki.se).

and

Gerli Rosengren Pielberg, Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala Biomedical Centre, Uppsala University, Uppsala, Sweden.

(email: gerli.pielberg@imbim.uu.se)

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Flowchart of sample quality control and selection.

Figure S2. Ancestry estimation with the LASER software.

Figure S3. Heterozygosity on the X chromosome, plotted to discover samples with discordant sex information.

Figure S4. Per individual quality control metrics calculated across positions passing the filter pipeline, visualized to enable detection of samples with low DNA quality or technical problems.

Figure S5. Per individual quality control metrics calculated across positions passing the filter pipe-

line, visualized to enable detection of samples with low DNA quality or technical problems.

Figure S6. Individual and variant call rate evaluated.

Figure S7. Population substructure visualized with multidimensional scaling (MDS) plots and coloured with case/control status and sampling region respectively.

Figure S8. Population substructure visualized with Eigenstrat-derived eigenvectors from principal component analysis (PCA) and coloured with case/control status and sampling region respectively.

Figure S9. MDS generated dimensions plotted against their most highly correlated Eigenstrat-derived eigenvector.

Figure S10. Quantile-quantile plots from three additive regression models, comparing covariates from Eigenstrat-derived PCA and MDS respectively.

Figure S11. Quantile-quantile plots showing observed versus expected *P* values for association between single-nucleotide polymorphisms (SNPs) and autoimmune Addison's disease (AAD), as calculated by an additive logistic regression model.

Figure S12. Quantile-quantile plot showing observed versus expected *P* values for association between single-nucleotide polymorphisms (SNPs) and autoimmune Addison's disease (AAD), as calculated by an additive logistic regression model.

Figure S13. Genomic location plots showing *P* values for association between single-nucleotide polymorphisms (SNPs) and autoimmune Addison's disease (AAD), as calculated by an additive logistic regression model.

Figure S14. Genomic location plot showing *P* values for association between single-nucleotide polymorphisms (SNPs) and isolated autoimmune Addison's disease (AAD), as calculated by an additive logistic regression model.

Figure S15. Age at sampling for AAD cases, and distribution of ages as compared to Gaussian distribution quantiles.

Figure S16. Quantile-quantile plots showing observed versus expected *P* values for association between rare variants aggregated per gene or intergenic space, and autoimmune Addison's disease (AAD), as calculated by the SKAT test statistic.

Table S1. Pearson correlation between the first eight principal components and the first eight multidimensional scaling (MDS) dimensions. Absolute values larger than 0.5 are bolded for visualization.

Table S2. Estimates of the logistic regression test statistic inflation with overdispersion factors (λ).

Table S3. Effects of genomic control on *P* values.

Table S4. Call rate and call rate difference between cases and controls for associated SNPs in *BACH2*. Minor allele frequencies from the 1000 genomes project are also presented for comparison.

Table S5. Two sample *t*-test compares the age of AAD cases with versus without TPO autoantibodies.

Table S6. SNPs in *BACH2*, chromosome 6q15, associated with autoimmune Addison's disease, and epigenetic annotation from HaploReg v4.1 and RegulomeDB. ■