Relation between HLA and copy number variation of steroid 21-hydroxylase in a Swedish cohort of patients with autoimmune Addison’s disease

Christian Lundtoft,1,* Daniel Eriksson,2,3,4 Matteo Bianchi,5 Maribel Aranda-Guillén,2 Nils Landegren,2,6 Solbritt Rantapää-Dahlqvist,7 Peter Söderkvist,8 Jennifer R.S. Meadows,5 The DISSECT Consortium,† The ImmunoArray Consortium,† The Swedish Addison Registry Study Group,† Sophie Bensing,9,10 Gerli Rosengren Pielberg,5 Kerstin Lindblad-Toh,5,11 Lars Rönblom,1 and Olle Kämpe2,9,12,13

1Department of Medical Sciences, Uppsala University, Uppsala, Sweden
2Department of Medicine (Solna), Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden
3Department of Clinical Genetics, Uppsala University Hospital, Uppsala, Sweden
4Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden
5Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden
6Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden
7Department of Public Health and Clinical Medicine/Rheumatology, Umeå University, Umeå, Sweden
8Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden
9Department of Endocrinology, Karolinska University Hospital, Stockholm, Sweden
10Department of Medical Sciences, Uppsala University, Linköping, Sweden
11Broad Institute, MIT and Harvard, Cambridge, MA, United States
12Department of Clinical Science, University of Bergen, Bergen, Norway
13K.G. Jebsen Center for Autoimmune Diseases, University of Bergen, Bergen, Norway

*Corresponding author: Uppsala University, DagHammarskjölds Väg 20, 752 37 Uppsala, Sweden. Email: christian.lundtoft@medsci.uu.se

Abstract

Objective: Autoantibodies against the adrenal enzyme 21-hydroxylase is a hallmark manifestation in autoimmune Addison’s disease (AAD). Steroid 21-hydroxylase is encoded by CYP21A2, which is located in the human leucocyte antigen (HLA) region together with the highly similar pseudogene CYP21A1P. A high level of copy number variation is seen for the 2 genes, and therefore, we asked whether genetic variation of the CYP21A2 genes is associated with AAD.

Design: Case-control study on patients with AAD and healthy controls.

Methods: Using next-generation DNA sequencing, we estimated the copy number of CYP21A2 and CYP21A1P, together with HLA alleles, in 479 Swedish patients with AAD and autoantibodies against 21-hydroxylase and in 1393 healthy controls.

Results: With 95% of individuals carrying 2 functional 21-hydroxylase genes, no difference in CYP21A2 copy number was found when comparing patients and controls. In contrast, we discovered a lower copy number of the pseudogene CYP21A1P in patients with autoimmune Addison’s disease and healthy controls. In patients with AAD, the strongest association was found for HLA-DRB1*04:04-DQB1*03:02 haplotype, imposed the greatest risk of AAD.

Conclusions: We identified strong associations between copy number variants in the CYP21 region and risk of AAD, although these associations most likely are due to linkage disequilibrium with disease-associated HLA class II alleles.

Keywords: Addison’s disease, autoimmune disease, autoantibodies, genetics

Significance

The majority of patients with autoimmune Addison’s disease present with autoantibodies against the enzyme steroid 21-hydroxylase that is encoded by CYP21A2. The gene is located in the human leucocyte antigen region together with the similar pseudogene CYP21A1P. Based on a focused analysis of the complex genetic region, we demonstrate extensive copy number variation of CYP21A2 and CYP21A1P in patients with autoimmune Addison’s disease and healthy controls. Although copy numbers differ strongly between patients and controls, HLA risk alleles remain as the strongest risk factor for autoimmune Addison’s disease.

D.E. and M.B. contributed equally.
† Consortia members are listed in Supplementary material.

Received: November 3, 2022. Revised: March 27, 2023. Editorial Decision: June 26, 2023. Accepted: June 26, 2023
© The Author(s) 2023. Published by Oxford University Press on behalf of European Society of Endocrinology. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Introduction

Autoimmune Addison’s disease (AAD) is characterized by an autoimmune destruction of the adrenal glands, resulting in inadequate synthesis of cortisol and aldosterone. The disease has a strong genetic component, as shown in twin studies, and the human leucocyte antigen (HLA) region harbours the main genetic risk factors associated with the disease, including the haplotypes DRB1*03:01-DQB1*02:01 and DRB1*04:DQB1*03:02. In more than 85% of adult patients, the autoimmune aetiology is confirmed by the presence of autoantibodies against the enzyme 21-hydroxylase, which is specifically expressed in the adrenal cortex and is essential for the synthesis of cortisol and aldosterone.

Steroid 21-hydroxylase is encoded by the gene CYP21A2 that is located between the HLA class I and class II regions on chromosome 6. CYP21A2 is part of the RP/STK-C4-CYP21-TNX (RCCX) gene module that harbours extensive copy number variation in the general population, generally ranging between 1 and 4 copies on each chromosome (Figure 1A). In addition to the functional 21-hydroxylase gene, a non-functional pseudogene CYP21A1P with ∼98% sequence similarity is located in the same region.

In addition to the CYP21 genes, the genes C4A and C4B encoding complement C4 are also located in the RCCX module (Figure 1A). Previous studies have shown a strong link between DRB1*03:01 and low copy number of C4A in populations of European descent. Although a few studies exist, a systematic evaluation of the CYP21 locus in relation to AAD has been restrained by the high level of copy number variation in combination with the high sequence similarity between functional and non-functional genes.

Here, we performed a focused investigation of genetic variation in the CYP21 region among AAD patients with 21-hydroxylase autoantibodies in comparison with healthy controls. Furthermore, HLA alleles were analysed in order to evaluate the relative contribution of CYP21 copy number variation and HLA haplotypes to the association with AAD.

Methods

Study participants

The 479 AAD patients, all with autoantibodies against 21-hydroxylase, and 1394 healthy controls have been described in detail previously. Study subjects had been recruited in Sweden, and patients fulfilled the diagnostic criteria for primary adrenal insufficiency, whereas patients with other suspected causes of adrenal failure or anti-cytokine autoantibodies had been excluded. In addition, genetic quality control had been performed to exclude samples of low genotyping quality and population outliers.

DNA sequencing and analysis of copy number and HLA

Sequence capture was performed using a custom-array targeting exons and regulatory regions of 1853 genes selected for their role in regulation of the immune system or in autoimmune diseases. DNA sequencing was performed with paired-end reads (100 bp) using an Illumina HiSeq 2500 system, followed by mapping to genome reference consortium human build 37 (GRCh37), genotyping using genome analysis toolkit (GATK) HaplotypeCaller and genetic quality control, as described previously.

A method for analysis of copy number variation of complement C4 described earlier was extended to include copy number variation of the neighbouring CYP21 genes that are also part of the RCCX module. The total copy number of CYP21 and complement C4 was analysed using GATK GermlineCNVCaller as described previously, with few modifications detailed in Supplementary material. Copy numbers of the C4A/C4B and CYP21A2/CYP21A1P paralogues were determined using the relative read depth of paralogue-specific variants. Genetic variants in the RCCX module were analysed using GATK HaplotypeCaller while accounting for the total copy number of C4/CYP21. The method for CYP21 copy number variation was tested on whole exome sequencing data from 90 samples in the 1000 Genomes Project and compared with CYP21 copy number estimates generated from whole genome sequencing data from the same individuals using GenomESTRIP.

As described in Supplementary material, a match in the total copy number of CYP21 was found for 97.8% of the 90 samples, while an agreement in CYP21A1P and CYP21A2 copy number calls was seen for 95.6% of the samples.

HLA alleles of the genes HLA-A, -B, -C, -DPB1, -DQB1, and -DRB1 were typed from sequencing reads at 2-field (ie, 4-digit) resolution using the tool xHLA as described previously.

Statistical analysis

R version 4.0.4 was used for statistical analyses, and the models and covariates included are described in the text or in the figure legends. Two-tailed P-values < 0.05 were considered significant, while a conservative Bonferroni-adjusted GWAS threshold at P < 5 × 10^{-8} was used for variant analysis.

Results

Copy number of CYP21A2 and the pseudogene CYP21A1P

We determined the copy numbers of CYP21A2, together with the pseudogene CYP21A1P, in 479 AAD patients with 21-hydroxylase autoantibodies and 1393 healthy controls. In order to understand the relation to copy number variation in the RCCX modules (Figure 1A), we also determined the copy numbers of complement C4A and C4B in parallel.

An analysis of CYP21A2 revealed that 95% of both patients and controls carried 2 functional 21-hydroxylase genes, while few individuals had 1, 3, or even 4 CYP21A2 genes (Figure 1B, left plot). However, no difference in copy number was observed between cases and controls (P_{CYP21A2} = 0.05). In contrast, a low copy number of the pseudogene CYP21A1P was more common among AAD patients when compared with controls (P_{CYP21A1P} = 5 × 10^{-4}; odds ratio (OR) = 3.44 [95% CI: 2.90-4.10] for each decrease in copy number) (Figure 1B, right plot).

To understand the levels of linkage disequilibrium between CYP21 and the neighbouring genes C4A and C4B, we also analysed the copy number of the 2 complement genes. The copy number of C4A generally correlated with the copy number variation and HLA haplotypes to the association with AAD.
number of **CYP21A1P** ($r_{\text{Spearman}} = 0.76$), whereas weaker linkage was seen between the copy numbers of **C4B** and **CYP21A2** ($r_{\text{Spearman}} = 0.13$).

Association analysis showed that a low copy number of **C4A** was associated with AAD to a higher extent than a low copy number of **CYP21A1P** ($P_{\text{C4A}} = 3 \times 10^{-56}$; OR = 3.82 [95% CI: 3.24-4.53] for each decrease in copy number; Figure 1C, right plot). In comparison, the copy number of **C4B** showed a slightly increased risk of AAD ($P_{\text{C4B}} = 2 \times 10^{-10}$; Figure 1C, left plot).

With a stronger association between **C4A** copy number and AAD, we concluded that the copy numbers of **CYP21A2** and **CYP21A1P** were not likely associated with the risk of AAD.

**Genetic variants in the RCCX module**

We continued by evaluating individual nucleotide variants in the RCCX module and their association with AAD. Due to the high-sequence homology of the **C4/CYP21** paralogues (≥98% identical), nucleotide variants in the 2 genes are generally excluded during genotyping, and the high level of copy number variation further challenges the assumption of diploidy in autosomal genes.

By re-mapping sequencing reads to a genomic reference in which RCCX module 1 had been masked while at the same time accounting for the copy number of **C4/CYP21**, we were able to genotype genetic variants in the RCCX module. However, these variants could not be assigned unambiguously to the individual gene paralogues, and instead, we analysed the total number of **C4/CYP21** copies carrying a reference allele and an alternative allele, respectively (Supplementary material).

In total, we identified 317 single-nucleotide variants (SNVs) and insertions/deletions distributed across the RCCX module.

---

**Figure 1.** Copy number variation of the 21-hydroxylase genes in Addison’s disease. (A) Genetic structure of the RCCX modules (**STK19** [formerly **RP**, **C4**, **CYP21**, **TNX**] in the reference genome (GRCh37). A majority of healthy individuals carry 2 RCCX modules on each chromosome, although it may vary between 1 and 4 modules. Pseudogenes have been marked with an asterisk. (B) Copy number of the functional 21-hydroxylase gene **CYP21A2** (left plot) and the pseudogene **CYP21A1P** (right plot) in patients with AAD and controls. (C) Copy number of complement **C4B** (left plot) and complement **C4A** (right plot) in AAD patients and controls. (B, C) Logistic regression of copy numbers from 479 AAD patients and 1393 controls adjusting for sex and population structure (PC1-PC3). The bars indicate the relative proportion of individuals with the indicated copy number. Controls with 5 gene copies are not shown in plots ($n_{\text{CYP21A1P}} = 1$, $n_{\text{C4B}} = 1$, $n_{\text{C4A}} = 5$). **DXO**, decapping and exoribonuclease protein; **STK19**, serine/threonine-protein kinase 19; **TNX**, tenascin X; **AAD**, autoimmune Addison’s disease; **PC**, principal component; **RCCX**, **RP**/**STK**-**C4**-**CYP21**-**TNX**.
(Supplementary material). In addition to the nucleotide variants in exon 26 of C4 that are used in defining the C4A and C4B paralogues,24 we observed 2 variants associated with AAD to a similar extent as the copy number of C4A (Figure 2): a Gly1073Asp missense variant in C4 (chr6:31996297G/A; \( P = 5 \times 10^{-57} \)) and a variant located in the promoter region of CYP21 (chr6:32005267C/T; \( P = 1 \times 10^{-36} \)).25 With multiple genetic variants in the RCCX module showing strong associations with AAD, we continued by evaluating genetic variants in C4/CYP21 in a broader context comprising the entire HLA region.

Analysis of the HLA region

We continued by analyzing the entire HLA region focusing on variants in the RCCX module and HLA alleles. The strongest association was seen for \( DQB1^*02:01 \) (Figure 2): a Gly1073Asp missense variant in C4 (chr6:31996297G/A; \( P = 5 \times 10^{-57} \)) and a variant located in the promoter region of CYP21 (chr6:32005267C/T; \( P = 1 \times 10^{-36} \)).25 With multiple genetic variants in the RCCX module showing strong associations with AAD, we continued by evaluating genetic variants in C4/CYP21 in a broader context comprising the entire HLA region.

Figure 2. Genetic variants in the RCCX module. Logistic regression analysis of C4/CYP21 copy numbers and nucleotide variants in the RCCX module for 479 patients with autoimmune Addison’s disease and 1393 controls. Variants in the RCCX module were analysed by accounting for the total copy number of C4/CYP21, followed by logistic regression using the number of genes carrying the reference allele and the alternative allele, respectively, while adjusting for sex and PC1-PC3. Variants present among ≥5 patients and controls were included in the analysis. Genomic position refers to the RCCX2 module in the GRCh37 reference, and the dashed line indicates GWAS significance threshold (\( P < 5 \times 10^{-8} \)). RCCX, RPSTK-C4-CYP21-I-TNX; PC, principal component.
through linkage among CYP21A1P copy number, C4A copy number, and HLA-DRB1*03:01, which are in high linkage disequilibrium in populations of European descent. Similarly, while several nucleotide variants in the RCCX module were strongly associated with AAD, the strongest association was found for DRB1*03:01-DQB1*02:01, which is in line with a previous study on selected CYP21A2 variants and HLA alleles in 381 AAD patients and 340 healthy controls. As shown previously, the DRB1*03:01-DQB1*02:01 alleles were especially associated with the risk of AAD when present in combination with the DRB1*04:04-DQB1*03:02 alleles.

By accounting for copy number variation while genotyping variants in the RCCX module, we were able to analyze genetic variation in this complex region. Considering the intrinsic limitations of short-read sequencing used in the current study, alternative methods such as long-read sequencing would be better suited for capturing variation of the individual C4/CYP21 paralogues in their complete genetic context. Further, an analysis of patients from other populations with different linkage disequilibrium structures between HLA alleles and variants in the C4/CYP21 region may help elucidating the genetic factors associated with the development of autoantibodies against 21-hydroxylase and AAD.

We note that the method for CYP21 copy number estimation described here will not suffice, for example, in clinical evaluation of congenital adrenal hyperplasia. Instead, we refer to the best practice guidelines for genetic testing of 21-hydroxylase deficiency that comprises techniques such as multiplex ligation–dependent probe amplification and Sanger sequencing. Despite these limitations, the current method is scalable and allows an analysis of a complex genetic region using next-generation sequencing data, which is increasingly being used in the genetic analysis of cohorts of various sizes.

Figure 3. Association analysis of the HLA region. (A) Association of HLA alleles, SNPs, C4/CYP21 copy number, and nucleotide variants in the RCCX module with autoimmune Addison’s disease (n = 479) as compared with controls (n = 1393). HLA alleles for 6 genes are shown (from left: HLA-A, -C, -B, -DRB1, -DQB1, and -DPB1). Genomic position refers to GRCh37, and the dashed line indicates GWAS significance threshold (P < 5 × 10−8). (B) Interaction analysis between HLA-DQB1*02:01 and DRB1*04:04, and the association with autoimmune Addison’s disease compared with controls. The size of the points indicates the number of individuals with each combination of HLA alleles. (A, B) Analysed by logistic regression adjusting for sex and PC1-PC3.

Acknowledgments

DNA sequencing and genotyping were performed at the SNP&SEQ Technology Platform in Uppsala, a part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The computations were enabled by resources in projects sens2017142, sens2017513, sens2018 526, and sens2020577 provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX. Raw sequencing data from SweAS controls were provided following data agreements, from the Uppsala and Linkoping Ankylosing Spondylitis Research Group (Department of Medical
Biochemistry and Microbiology, Uppsala University, Sweden and Department of Biomedical and Clinical Sciences, Linköping University, Sweden). We thank Västerbotten biobank for providing DNA samples from control individuals from Northern Sweden.

Supplementary material

Supplementary material is available at European Journal of Endocrinology online.

Funding

This work was supported by the Swedish Research Council for Medicine and Health, King Gustav V’s 80-year Foundation, the Knut and Wallenberg Foundation, the Novo Nordisk Foundation, and the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet.

Conflict of interest: L.R. has received honorarium for lectures from AstraZeneca and participated in advisory board for UCB. O.K. is a board member of Navinci Diagnostics AB and member of the Scientific Advisory Board for the Leo Foundation Skin Immunology Research Center, University of Copenhagen.

Data availability

Raw data for individual figures are available in the Supplementary material. Genotype data at the individual level are not publicly available since it contains information that could compromise research participant privacy and consent. Scripts for calling CYP21A/C4 copy number in GATK GermlineCNVCaller are available upon request.

References


