Exomechip-based rare variant association study in restless legs syndrome

Erik Tilch[1,2¶], Barbara Schormair[1,2¶*], Chen Zhao[1], Birgit Högl[3], Ambra Stefani[3], Klaus Berger[4], Claudia Trenkwalder[5,6], Cornelius G. Bachmann[7,#a], Magdolna Hornyak[8,9,#b], Ingo Fietze[10], Martina Müller-Nurasyid[11,12,13,14], Annette Peters[15,16,17], Stefan Herms[18,19,20], Markus M. Nöthen[20,21], Bertram Müller-Myhsok[22,23,24], Konrad Oexle[1&], Juliane Winkelmann[1,23,2&*]

1 Institute of Neurogenomics, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany.

2 Institute of Human Genetics, Technische Universität München, Munich, Germany.

3 Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria.

4 Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany.

5 Clinic for Neurosurgery, University Medical Centre, Georg August University Göttingen, Göttingen, Germany.

6 Paracelsus-Elena Hospital, Center of Parkinsonism and Movement Disorders, Kassel, Germany.

7 Department of Neurology, Paracelsus Klinik, Osnabrueck, Germany.

8 Department of Neurology, University of Ulm, Ulm, Germany.

9 Neuropsychiatry Centre Erding/München, Erding, Germany.

10 Charité - Universitätsmedizin Berlin, Center of Sleep Medicine, Berlin, Germany.

11 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.

12 The Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Faculty of Medicine, Ludwig-Maximilians-Universität (LMU) Munich, Munich, Germany.

13 Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, Mainz, Germany.

14 Department of Internal Medicine I (Cardiology), Hospital of the Ludwig-Maximilians-University (LMU) Munich, Munich, Germany.

15 Institute of Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany.

16 German Centre for Cardiovascular Research (DZHK), partner site Munich Heart Alliance, Munich, Germany.

17 German Centre for Diabetes Research (DZD), Neuherberg, Germany.

18 Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland.

19 Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland.

20 Department of Genomics, Life & Brain Research Center, University of Bonn, Bonn, Germany.

21 Institute of Human Genetics, University of Bonn School of Medicine & University Hospital Bonn, Bonn, Germany.

22 Max Planck Institute of Psychiatry, Munich, Germany.

23 Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.

24 Institute of Translational Medicine, University of Liverpool, Liverpool, UK.

[#a]Current Address: Somnodiagnostics, Martinistrasse 63-65, Osnabrück, Germany and Department of Neurology, University Medicine Göttingen, Germany.

[#b]Current Address: Private practice of Magdolna Hornyak, Erding, Germany

* Corresponding author

barbara.schormair@helmholtz-muenchen.de (BS), juliane.winkelmann@tum.de (JW)

[¶]These authors contributed equally to this work.

[&]These authors also contributed equally to this work.

Abstract

Restless legs syndrome (RLS) is a common sleep-related movement disorder in populations of European descent and disease risk is strongly influenced by genetic factors. Common variants have been assessed extensively in several genome-wide association studies, but the contribution of rarer genetic variation has not been investigated at this scale. We therefore genotyped a case-control set of 9,246 individuals for mainly rare and low frequency exonic variants using the Illumina ExomeChip. However, standard single variant and gene-level association tests were negative. This does not preclude a role of rare variants in RLS, but is likely due to the small sample size and the limited selection of rare genetic variation captured on the array. Therefore, exome or whole genome sequencing should be performed rather than increasing the sample size of ExomeChip studies in order to identify rare risk variants for RLS.

Keywords: Restless legs syndrome, RLS, genome-wide association study, rare variant association study

Introduction

Restless legs syndrome (RLS) is one of the most common sleep-related movement disorders in populations of European descent with a prevalence of up to 10%. Existing treatment options are limited due to their side effects and preventive measures are not available yet. Since RLS is a complex disease with a substantial genetic contribution to disease susceptibility, identification of the underlying genetic risk factors is important for progress in therapy and prevention.[1]

In complex traits, variants across the entire allele frequency spectrum can contribute to disease risk, ranging from rare (minor allele frequency (MAF) < 1%) to low-frequency (1% ≤ MAF < 5%) to common (MAF ≥ 5%).[2] For RLS, common variants have been assayed comprehensively in several genome-wide association studies (GWAS), which have identified 23 risk variants to date.[3,4] In contrast, rare and low-frequency variants have not been studied extensively so far. Only a few individual families have been screened in linkage and in exome sequencing studies, but no causal variants were detected (reviewed in [5]). We have performed a targeted resequencing of 84 genes in GWAS risk loci in a case-control sample and prioritized 14 as candidates based on an enrichment of rare and low-frequency variants in them.[6] Based on the observation of some similarities to monogenic diseases such as large families with Mendelian inheritance patterns, early-onset RLS, and pediatric RLS, rare variants with larger effects, especially coding variants, could contribute to RLS.

Whole genome sequencing would provide the most comprehensive data, but it is still rather expensive for larger samples sizes. The ExomeChip was designed as a transitional solution to allow less costly genotyping of rare variants. Besides a backbone of common variants, this array includes a selection of rare and low-frequency coding variants across the genome identified in sequencing studies of other traits (https://genome.sph.umich.edu/wiki/Exome_Chip_Design). These included type 2 diabetes, cardiovascular traits, and depression, i.e. traits where RLS is a common comorbidity.[1] We hypothesized that a substantial proportion of the variants would also be present in RLS cases and testable in an association study. Therefore, we performed an association study in 3,678

RLS cases and 5,407 population-based controls of European ancestry using the ExomeChip to identify rare and low-frequency risk variants for RLS.

Materials and Methods

Genotyping and quality control

The 3,804 RLS cases were of Austrian/German descent. They have been diagnosed and recruited based on the IRLSSG criteria.[3] Population-based controls were from a Bavarian region in the south (2,921, KORA[7]) and the Ruhr in the north of Germany (2,476, HNR[8]) (Supplementary Table 1). Study participants provided informed written consent and the study was approved by the institutional review board. Individuals were genotyped with the Illumina HumanExome BeadChip 12v1_A ("ExomeChip") according to manufacturer's protocol at two facilities in Bonn and Munich. We used Illumina software (GenomeStudio V2011.1V, "Genotyping" module 1.9.4, "Illumina Genome Viewer" module 1.9.0) and the CHARGE cluster file 1.0 for initial 238,876 markers [9].

Quality control (QC) was done with PLINK v1.07 and R (Supplementary Table 2). Initial QC at the genotyping facilities was performed separately for HNR controls, KORA controls, and RLS cases. Markers/individuals were filtered on the call rate of at least 98%. A QC by sex-check was applied. The pre-cleaned datasets were then merged and again filtered for call rates of at least 98%. Only autosomal variants were used in further analyses. A heterozygosity check was applied for the individuals. To estimate family relatedness, we calculated pairwise proportion of alleles shared by IBD (identity-by-descent) with pruned markers [10] as present on array version 1.1. Then, we removed markers with HWE violations (p < 1E-07), individuals with many large IBD proportions (n(IBD > 0.09375) ≥ 10), two individuals with missing age data, and one of each pair of duplicated individuals by keeping the best genotyped one. In an update of marker QC, we removed duplicated and genotyping error-prone markers based on an internal black list (n = 717). We determined a set of pruned markers[10] suitable for calculations of genetic similarities or principal components (PCs) in later analyses.

Single variant association tests

Association testing was performed for all variants which passed the quality control and which were not monomorphic in our dataset. A genome-wide mixed model association scan was done using an additive model including age and sex as covariates in FaST-LMM v2.06 (without genotype normalization, without proximal contamination within 2 Mb)[11] and GMMAT [12] 0.7 (with a centered GRM from GEMMA 0.94.1 [13]). Results were annotated using CADD v1.3 https://cadd.gs.washington.edu/info), VEP (https://www.ensembl.org/info/docs/tools/vep/index.html) and the CHARGE annotation file v5 [9].

Gene-level association tests

For gene-level association tests, polygenic residuals were obtained using GRAMMAR[14] in GenABEL v.1.8-0 (https://CRAN.R-project.org/package=GenABEL) with an IBS (identity-by-state) matrix and covariates (age, sex). Applying these residuals as the phenotype, we ran four different association test, i.e. SKAT v0.93[15], BRV ("Burden of Rare Variant Test")[16], and combination tests (Fisher's and minimum-p method[17]), on genes with at least two markers which both had to have a MAF of less than 5% in cases or in controls. Gene definitions for the ExomeChip were based on the Illumina annotation v1 files (ftp://ftp.ncbi.nlm.nih.gov/geo/platforms/GPL18nnn/GPL18544/suppl/). We accounted for missing genotypes as described in Wu et al. and Auer et al.[15,16]. Markers were weighted uniformly or based on the variants' normalized raw CADD scores. In all four tests, the significance was determined empirically by at least 10,000 permutations. To enable a correction for multiple testing, the number of hypotheses was determined which could adjust the vector of the smallest random p values (one for each gene, from 10,000 phenotype permutations) to a uniform distribution by Sidak's approach (= number of independent null hypotheses).

Pathway analysis

For pathway analysis, variants were annotated to genes based on the ExomeChip marker annotation for HUGO gene symbols as provided by Illumina. Pathways were defined

by Reactome v7.1 annotations which were accessed via the GSEA Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb).[18,19] We ran the pathway analysis in the spirit of gene-level testing using the optimal sequence kernel association test SKAT-O (v1.3.2.1, no variant weights, imputation by "best guess") including a maximum sample of unrelated individuals (IBD ≤ 0.09375), variants in stringent HWE ($p > 1 \times 10^{-4}$) and with MAF ≤ 0.05 in cases or controls, and covariates (age, sex, two PCs).[15,20,21]

Results

Single variant associations

A total of 234 individuals and 7,010 variants were removed due to failing quality control. Exclusion of 94,855 monomorphic variants resulted in9,012 individuals and 137,011 polymorphic autosomal variants available for association analysis (Supplementary Table 2). No novel variants were significantly associated with RLS after correction for multiple testing, but associations in known GWAS loci were confirmed by common and low-frequency/rare variants in LD with the published GWAS lead SNPs (table 1). One independent low-frequency missense variant reached candidate status (MAF = 0.02, rs34377632, *RASGRP4*, $p \leq 1 \times 10^{-4}$ in both FaST-LMM and GMMAT score test). P values were slightly inflated for rare variants (GMMAT [12] score test, MAF ≤ 0.01, $\lambda_{GC} = 1.19$, Supplementary Figures 1 and 2).

Table 1: Top SNPs in single variant association analysis with p values $\leqslant$ 1E-04 in both FaST-LMM and GMMAT score test.

dbSNP ID	Chr	Pos	Maj/min	MAF	*P*	*OR*	Consequence	Gene	CADD
rs2300478§	2	66,781,453	A/C	0.29	1.4E-35	1.60 [1.49; 1.73]	i, nc, nmd	*MEIS1*	5.067
rs3923809§	6	38,440,970	A/G	0.27	5.2E-28	1.55 [1.43; 1.67]	i, nmd	*BTBD9*	8.722
rs1026732§	15	68,095,085	G/A	0.29	2.1E-19	1.42 [1.32; 1.53]	i, nc	*MAP2K5*	0.871
rs12593813	15	68,036,852	G/A	0.30	5.0E-19	1.41 [1.31; 1.52]	i, nc, u	*MAP2K5*	14.560
rs4489954§	15	68,072,075	C/A	0.28	3.3E-18	1.41 [1.31; 1.53]	i, nc, r	*MAP2K5*	5.544
rs9357271§	6	38,365,873	A/G	0.21	1.8E-16	1.43 [1.31; 1.56]	i, nmd	*BTBD9*	2.729
rs9296249§	6	38,365,841	A/G	0.20	2.6E-16	1.43 [1.31; 1.56]	i, nmd	*BTBD9*	0.266
rs3104767§	16	52,624,738	C/A	0.39	4.2E-16	1.34 [1.25; 1.44]	i, nc	*CASC16*	0.634
rs3112612§	16	52,635,164	G/A	0.39	7.6E-16	1.34 [1.25; 1.44]	i, nc	*CASC16*	2.951
rs11897119§	2	66,772,000	A/G	0.36	1.8E-15	1.34 [1.25; 1.44]	i, nc, nmd	*MEIS1*	0.983
rs2241423§	15	68,086,838	G/A	0.20	2.9E-13	1.38 [1.27; 1.50]	i, nc	*MAP2K5*	3.451
rs2814899§	6	38,389,301	A/G	0.50	1.1E-12	1.28 [1.20; 1.38]	i, nmd	*BTBD9*	2.695
rs1975197§	9	8,846,955	G/A	0.17	1.9E-09	1.31 [1.20; 1.44]	i	*PTPRD*	1.784
rs6747972§	2	68,070,225	G/A	0.47	2.4E-07	1.20 [1.12; 1.29]	ig	-	4.560
rs7819412	8	11,045,161	A/G	0.48	3.4E-06	1.17 [1.10; 1.25]	i, nmd, nc	*XKR6*	2.013
rs7173826§	15	67,528,374	A/C	0.31	1.3E-05	1.18 [1.10; 1.27]	m, d, ncx	*AAGAB*,	23.900
rs7824557	8	11,104,111	A/G	0.39	3.4E-05	1.16 [1.08; 1.24]	d	*LINC00529*	3.101
rs11215690	11	115,685,475	G/A	0.45	5.4E-05	1.15 [1.08; 1.24]	i, nc	*LINC02698*	5.775
rs34377632	19	38,901,633	G/A	0.02	5.4E-05	1.58 [1.26; 1.98]	d, m, ncx, nmd	*RASGRP4*	24.300
rs4256842§	1	107,163,979	A/G	0.30	5.9E-05	1.16 [1.08; 1.25]	ig	-	6.015
rs2293583§	7	88,423,881	A/G	0.09	6.0E-05	1.27 [1.13; 1.43]	m, i	*C7orf62*, *ZNF804B*	0.001
rs4793601§	17	46,791,801	A/C	0.46	6.4E-05	1.15 [1.08; 1.24]	d	*COX6B1P2*	1.480
rs2031577	10	4,050,003	G/A	0.39	8.4E-05	1.15 [1.07; 1.24]	ig	-	2.866

Chr, chromosome; Pos, hg19 position; Maj/min, major/minor allele (risk allele underscored); MAF, minor allele frequency in the dataset; P, p value from GMMAT score test; OR, odds ratio with 95% confidence interval from GMMAT Wald test effect sizes; Consequence, variant effect (d, downstream gene; i, intronic; ig, intergenic; m, missense; nc, non-coding transcript; ncx, non-coding transcript exon; nmd, NMD transcript; r, regulatory; u, upstream gene); CADD, CADD Phred score; §, locus significant in [3].

Gene-level associations

14,147 genes were tested for gene-level association (109,593 rare markers, ≥ 2/gene). We applied four types of association tests (SKAT, BRV, Fisher's method, and minimum-p method) and two variant weighting schemes in order to allow different underlying genetic models since the true model is unknown. Of all conducted BRVs, 11,833 were independent. No association p value remained significant after Bonferroni correction for four methods, two variant weighting schemes, and the number of independent tests (4 x 2 x 11,833 tests). There were 12 subthreshold candidate genes for RLS which had a p value ≤ 1/11,833 in any one of the applied tests (Table 2).

Table 2: Top association *p* values for gene level tests with different variant weighting and association test methods.

Gene	***M***	**No weights**				**CADD weights**			
		BRV	**SKAT**	**Min-p**	**Fisher**	**BRV**	**SKAT**	**Min-p**	**Fisher**
DMPK	08	1.0E-3	2.6E-4	5.1E-4	**1.2E-5***	4.9E-4	3.3E-4	6.6E-4	**9.0E-6***
EYA2	12	4.3E-4	4.7E-3	8.5E-4	**4.0E-5***	3.8E-4	2.1E-3	7.5E-4	**4.0E-5***
NECAP1	02	9.0E-5	6.7E-4	1.7E-4	**8.0E-5***	1.1E-4	9.4E-4	2.1E-4	**7.0E-5***
NENF	03	4.0E-4	2.4E-3	7.9E-4	**5.0E-5***	4.1E-4	2.3E-3	8.1E-4	**8.0E-5***
OLFML2B	15	2.5E-3	8.1E-4	1.6E-3	1.1E-4	1.4E-3	7.4E-4	1.4E-3	**8.0E-5***
OSBP	03	4.7E-3	9.0E-5	1.7E-4	**4.0E-5***	3.2E-3	1.5E-4	2.9E-4	**7.1E-5***
OSGIN1	15	1.5E-3	4.2E-4	8.3E-4	**3.0E-5***	2.2E-3	4.2E-4	8.4E-4	**3.2E-5***
PCDHB5	04	5.5E-4	7.2E-4	1.1E-3	9.0E-5	5.0E-4	7.0E-4	9.9E-4	**4.0E-5***
PDE11A	20	8.7E-3	4.4E-4	8.7E-4	1.2E-4	3.1E-3	3.6E-4	7.1E-4	**5.0E-5***
RASGRP4	11	8.0E-3	7.8E-3	1.5E-2	7.0E-4	9.8E-4	2.3E-4	4.5E-4	**1.5E-5***
TREM1	04	4.6E-4	2.4E-3	9.1E-4	**3.0E-5***	1.6E-4	1.6E-3	3.3E-4	**2.0E-5***
UBL4B	04	8.3E-4	1.2E-3	1.6E-3	**6.0E-5***	6.4E-4	1.2E-3	1.3E-3	**5.0E-5***

M, number of loci of variants in gene level tests; BRV, burden of rare variants test; SKAT, sequence kernel association test; Min-p, minimum-p method; Fisher, Fisher's method; *, p ≤ 8.45E-5 (= 1/11,833); No weights, results of association tests without weighting variants; CADD weights; results of association tests using CADD score for variant weighting.

Pathway enrichment

We also performed pathway enrichment analysis in order to aggregate association signals of rare and low-frequency variants across sets of genes. We tested 1,532 biological pathways annotated in Reactome for an enrichment of low-frequency and rare variants using the SKAT-O test. In line with the results from single variant and gene-based testing, no significant associations remained after adjusting for multiple testing ($p \leq 0.05/1,532$). The lowest p value observed was 4.5×10^{-5} for the pathway "cytokine signaling in immune system" (Supplementary Table 3).

Discussion

Here, we assessed the role of rare and low-frequency variants in RLS using the ExomeChip design, which allows a genome-wide analysis, albeit limited to the preselected set of variants present on the array. Nevertheless, we could test 137,011 variants, of which 76,700 were rare with a MAF < 1% and had not been accessible in previous GWAS. However, only common variants representing the known GWAS loci reached genome-wide significance ($p < 5 \times 10^{-8}$) in the single variant analysis. Testing the combined effect of rare and low-frequency variants on the gene level as well as on the pathway level did not return significant results.

Our negative results do not exclude a role of rare and low-frequency variants in RLS. They rather indicate that the ExomeChip is not suitable to screen RLS cases for rare and low-frequency variants neither in a scientific nor a clinical context. Only a fraction of all rare/low-frequency variants are covered by this technology (e.g. only about 3% of the genetic variants reported in the Exome Aggregation Consortium data are represented [9]), limiting the number of variants that can be tested. This coverage issue carries over to gene-level testing, where power is reduced when relevant variants are missing. Another limiting factor is the sample size of our study. ExomeChip studies of comparable size on other traits also had low yields.[22] Only studies at least one order of magnitude larger than the present study were able to detect significant associations of rare or low-frequency variants.[23] However, our study would have been sufficiently powered to detect rare variants with strong effects, e.g. odds ratios of 3 or

larger for a MAF of 0.5%. This suggests that such risk variants for RLS are not present on the ExomeChip and that further increasing the sample size would still most likely not result in the detection of strong effect variants for RLS with this array.

Therefore, we do not recommend further use of the ExomeChip for genetic analyses in RLS. Ideally, future studies would create, use, and share high-coverage whole genome sequencing (WGS) data. However, the large sample sizes needed for association studies of rare and low-frequency variants are still difficult to obtain. Currently, the most efficient option seems to be genotyping with arrays optimized for GWAS such as the Illumina GSA or Affymetrix Axiom array and creating large datasets within a consortium such as the International EU-RLS-GENE consortium or by performing meta-analyses.[3,4,24] This will enable discovery of additional common risk variants as well as the calculation of polygenic risk scores, two important pillars of therapeutic advances. In addition, imputation of low-frequency and even rare variants in such datasets is continuously improving due to the availability of larger and larger reference panels and better imputation tools.[25-27] Especially imputation of rare variants may also benefit from low-coverage WGS data, which could be available more easily and therefore sooner than high-coverage genomes.[28] We expect GWAS-focused array genotyping and potentially low-coverage WGS to be driving the discovery of genetic risk factors underlying the common, multifactorial form of RLS in the coming years.

Acknowledgements

We would like to thank all study participants, and colleagues and staff at the participating centers for their help with recruitment of patients. We thank the German Restless Legs Syndrome Foundation for continuously supporting our research. The COR (Course of RLS) study was supported by the Swiss RLS Patient Association and with unrestricted grants to the University of Münster by the German Restless Legs Syndrome Foundation, and a consortium formed by Vifor Pharma, UCB (Switzerland & Germany), Roche Pharma, Axxonis

Pharma, Mundipharma Research, Boehringer Ingelheim. The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The Heinz Nixdorf Recall Study was supported by the Heinz Nixdorf Foundation Germany, the German Ministry of Education and Research (BMBF), and the Deutsche Forschungsgemeinschaft (DFG).

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Manconi M, Garcia-Borreguero D, Schormair B, et al. Restless legs syndrome. *Nat Rev Dis Primers*. Nov 3 2021;7(1):80. doi:10.1038/s41572-021-00311-z
2. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. Oct 8 2009;461(7265):747-53. doi:10.1038/nature08494
3. Schormair B, Zhao C, Bell S, et al. Identification of novel risk loci for restless legs syndrome in genome-wide association studies in individuals of European ancestry: a meta-analysis. *The Lancet Neurology*. 2017/11/01/ 2017;16(11):898-907. doi:https://doi.org/10.1016/S1474-4422(17)30327-7
4. Didriksen M, Nawaz MS, Dowsett J, et al. Large genome-wide association study identifies three novel risk variants for restless legs syndrome. *Commun Biol*. Nov 25 2020;3(1):703. doi:10.1038/s42003-020-01430-1
5. Jimenez-Jimenez FJ, Alonso-Navarro H, Garcia-Martin E, Agundez JAG. Genetics of restless legs syndrome: An update. *Sleep Med Rev*. Jun 2018;39:108-121. doi:10.1016/j.smrv.2017.08.002
6. Tilch E, Schormair B, Zhao C, et al. Identification of Restless Legs Syndrome Genes by Mutational Load Analysis. *Annals of neurology*. Feb 2020;87(2):184-193. doi:10.1002/ana.25658
7. Wichmann HE, Gieger C, Illig T. KORA-gen - resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*. Aug 2005;67 Suppl 1:S26-30. doi:10.1055/s-2005-858226
8. Schmermund A, Mohlenkamp S, Stang A, et al. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *American heart journal*. Aug 2002;144(2):212-8.
9. Grove ML, Yu B, Cochran BJ, et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PloS one*. 2013;8(7):doi: 10.1371/journal.pone.0068095. doi:10.1371/journal.pone.0068095

10. Huyghe JR, Jackson AU, Fogarty MP, et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nature genetics*. Feb 2013;45(2):197-201. doi:10.1038/ng.2507
11. Listgarten J, Lippert C, Kadie CM, Davidson RI, Eskin E, Heckerman D. Improved linear mixed models for genome-wide association studies. *Nature methods*. Jun 2012;9(6):525-6. doi:10.1038/nmeth.2037
12. Chen H, Wang C, Conomos MP, et al. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *American journal of human genetics*. Apr 7 2016;98(4):653-66. doi:10.1016/j.ajhg.2016.02.012
13. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nature genetics*. Jul 2012;44(7):821-4. doi:10.1038/ng.2310
14. Belonogova NM, Svishcheva GR, van Duijn CM, Aulchenko YS, Axenovich TI. Region-based association analysis of human quantitative traits in related individuals. *PloS one*. 2013;8(6):doi: 10.1371/journal.pone.0065395. doi:10.1371/journal.pone.0065395
15. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *American journal of human genetics*. Jul 15 2011;89(1):82-93. doi:10.1016/j.ajhg.2011.05.029
16. Auer PL, Wang G, Leal SM. Testing for rare variant associations in the presence of missing data. *Genet Epidemiol*. Sep 2013;37(6):529-38. doi:10.1002/gepi.21736
17. Derkach A, Lawless JF, Sun L. Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. *Genet Epidemiol*. Jan 2013;37(1):110-21. doi:10.1002/gepi.21689
18. Fabregat A, Sidiropoulos K, Garapati P, et al. The Reactome pathway Knowledgebase. *Nucleic acids research*. Jan 04 2016;44(D1):D481-7. doi:10.1093/nar/gkv1351
19. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. Oct 25 2005;102(43):15545-50. doi:10.1073/pnas.0506580102
20. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics (Oxford, England)*. Sep 2012;13(4):762-75. doi:10.1093/biostatistics/kxs014
21. Richardson TG, Timpson NJ, Campbell C, Gaunt TR. A pathway-centric approach to rare variant association analysis. *European journal of human genetics : EJHG*. Jan 2016;25(1):123-129. doi:10.1038/ejhg.2016.113
22. Dankowski T, Buck D, Andlauer TF, et al. Successful Replication of GWAS Hits for Multiple Sclerosis in 10,000 Germans Using the Exome Array. *Genet Epidemiol*. Dec 2015;39(8):601-8. doi:10.1002/gepi.21933
23. Wessel J, Chu AY, Willems SM, et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nature communications*. 2015;6:doi: 10.1038/ncomms6897. doi:10.1038/ncomms6897
24. Verlouw JAM, Clemens E, de Vries JH, et al. A comparison of genotyping arrays. *European journal of human genetics : EJHG*. Nov 2021;29(11):1611-1624. doi:10.1038/s41431-021-00917-7
25. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nature communications*. Sep 14 2015;6:8111. doi:10.1038/ncomms9111
26. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics*. Oct 2016;48(10):1279-83. doi:10.1038/ng.3643

27. Stanaway IB, Hall TO, Rosenthal EA, et al. The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype. *Genet Epidemiol*. Feb 2019;43(1):63-81. doi:10.1002/gepi.22167
28. Rubinacci S, Ribeiro DM, Hofmeister RJ, Delaneau O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature genetics*. Jan 2021;53(1):120-126. doi:10.1038/s41588-020-00756-0

Supporting information

Supplementary Figure 1: QQ-plots from FaST-LMM single variant association analysis. (a) Variants with MAF ≤ 0.01. (b) Variants with MAF > 0.01, (c) All variants. X-axis, -log10 of expected p values; y-axis, -log10 of observed p values, winsorized at 5E-8

Supplementary Figure 2. QQ-plots from GMMAT score test single variant association analysis. (a) Variants with MAF ≤ 0.01. (b) Variants with MAF > 0.01. (c) All variants. X-axis, -log10 of expected p values; y-axis, -log10 of observed p values, winsorized at 5E-8

Supplementary Table 1: Demographic details of this study.

Supplementary Table 2: Quality control steps (QC).

Supplementary Table 3: Top pathway enrichment results

Supplementary Information

Exomechip-based rare variant association study in restless legs syndrome

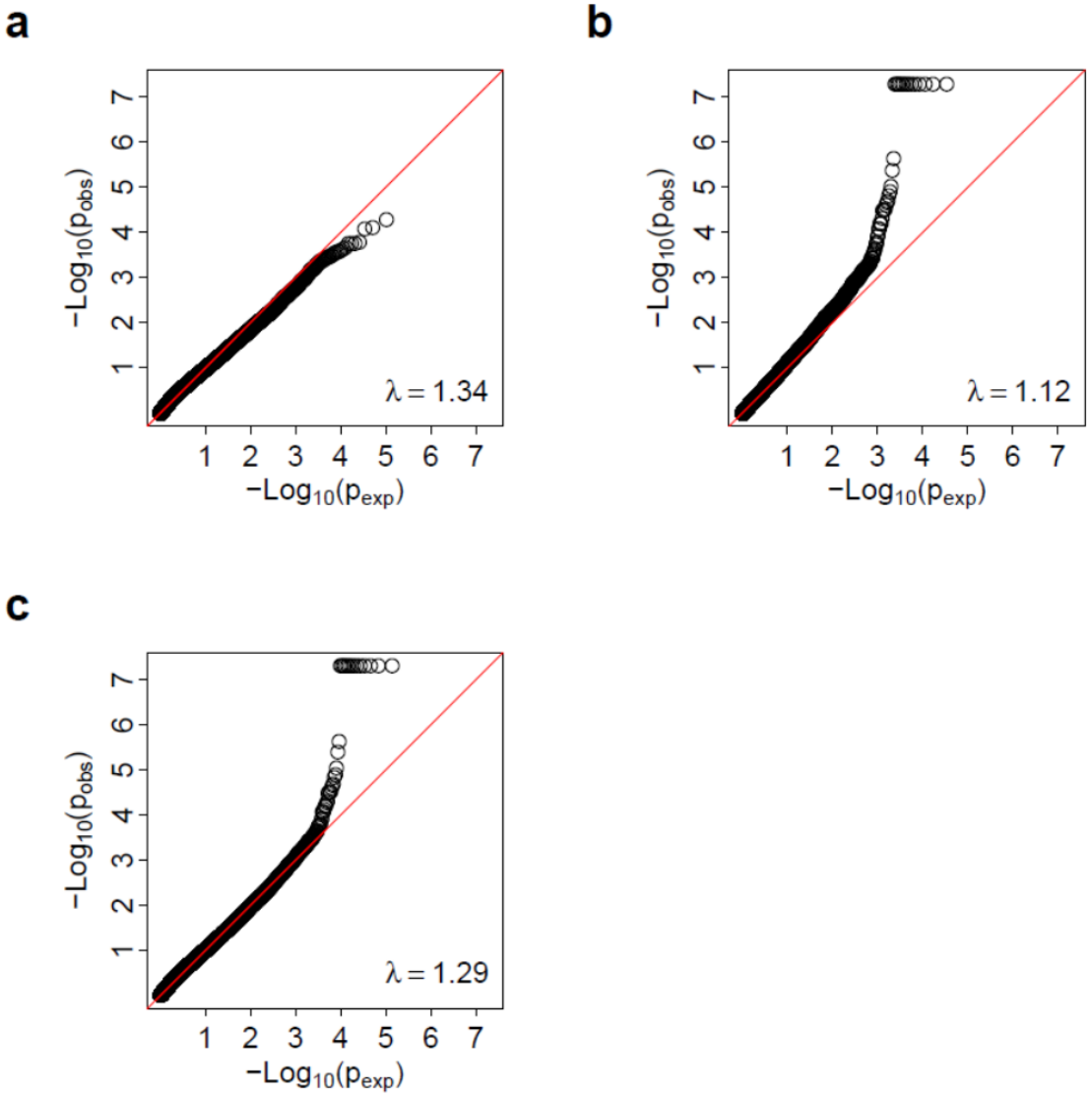
Erik Tilch[1,2¶], Barbara Schormair[1,2¶*], Chen Zhao[1], Birgit Högl[3], Ambra Stefani[3], Klaus Berger[4], Claudia Trenkwalder[5,6], Cornelius G. Bachmann[7,#a], Magdolna Hornyak[8,9,#b], Ingo Fietze[10], Martina Müller-Nurasyid[11,12,13,14], Annette Peters[15,16,17], Stefan Herms[18,19,20], Markus M. Nöthen[21,20], Bertram Müller-Myhsok[22,23,24], Konrad Oexle[1&], Juliane Winkelmann[1,23,2&*]

1 Institute of Neurogenomics, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany.
2 Institute of Human Genetics, Technische Universität München, Munich, Germany.
3 Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria.
4 Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany.
5 Clinic for Neurosurgery, University Medical Centre, Georg August University Göttingen, Göttingen, Germany.
6 Paracelsus-Elena Hospital, Center of Parkinsonism and Movement Disorders, Kassel, Germany.
7 Department of Neurology, Paracelsus Klinik, Osnabrueck, Germany.
8 Department of Neurology, University of Ulm, Ulm, Germany.
9 Neuropsychiatry Centre Erding/München, Erding, Germany.
10, Charité - Universitätsmedizin Berlin, Center of Sleep Medicine, Berlin, Germany.
11 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
12 The Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Faculty of Medicine, Ludwig-Maximilians-Universität (LMU) Munich, Munich, Germany.
13 Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, Mainz, Germany.
14 Department of Internal Medicine I (Cardiology), Hospital of the Ludwig-Maximilians-University (LMU) Munich, Munich, Germany.
15 Institute of Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany.
16 German Centre for Cardiovascular Research (DZHK), partner site Munich Heart Alliance, Munich, Germany.
17 German Centre for Diabetes Research (DZD), Neuherberg, Germany.
18 Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland.
19 Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland.
20 Department of Genomics, Life & Brain Research Center, University of Bonn, Bonn, Germany.
21 Institute of Human Genetics, University of Bonn School of Medicine & University Hospital Bonn, Bonn, Germany.
22 Max Planck Institute of Psychiatry, Munich, Germany.
23 Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.
24 Institute of Translational Medicine, University of Liverpool, Liverpool, UK.
[#a]Current Address: Somnodiagnostics, Martinistrasse 63-65, Osnabrück, Germany and Department of Neurology, University Medicine Göttingen, Germany.
[#b]Current Address: Private practice of Magdolna Hornyak, Erding, Germany

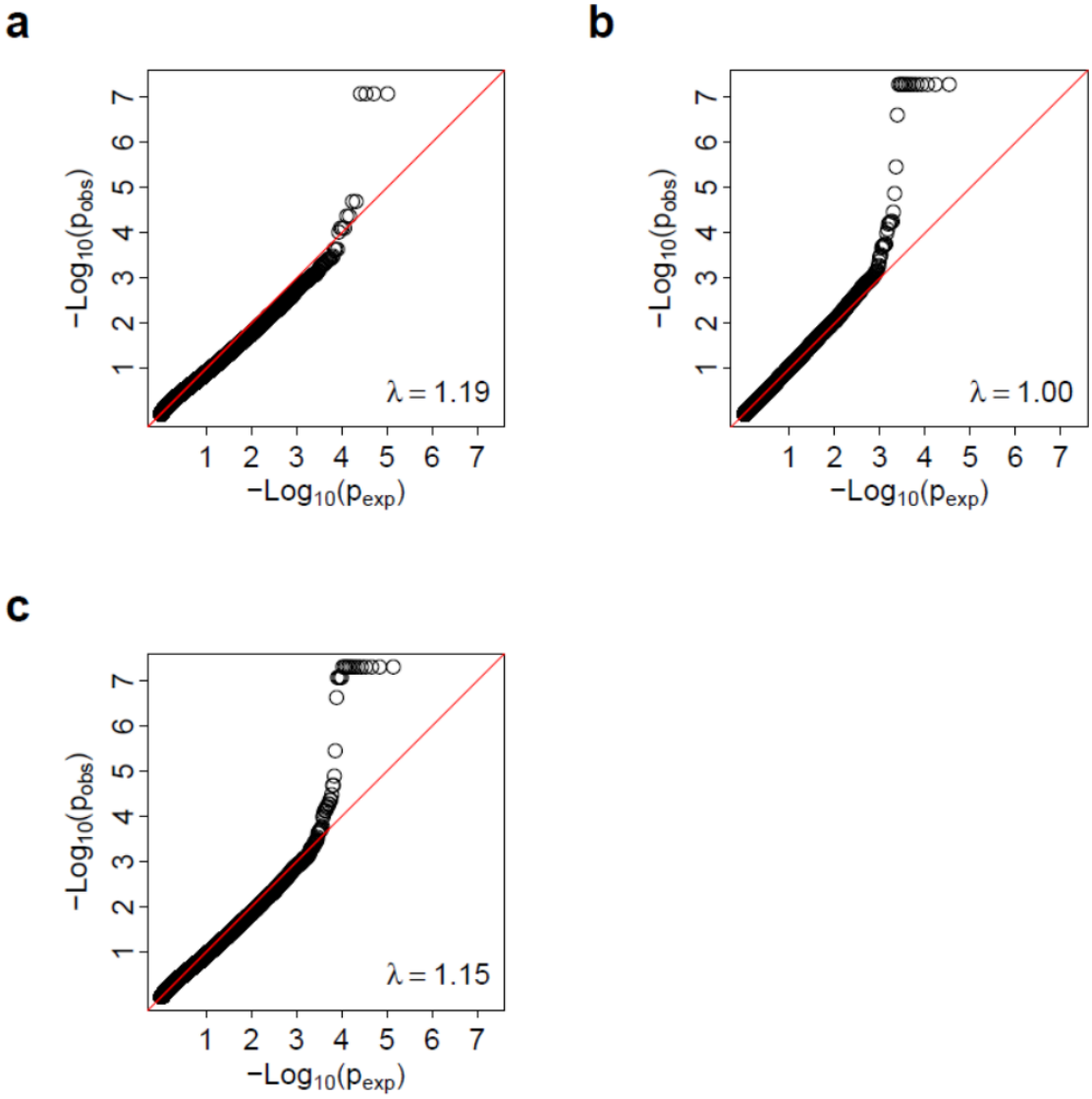
* Corresponding author
barbara.schormair@helmholtz-muenchen.de (BS), juliane.winkelmann@tum.de (JW)

[¶]These authors contributed equally to this work.
[&]These authors also contributed equally to this work.

Supplementary Figure 1: QQ-plots from FaST-LMM single variant association analysis. (a) Variants with MAF ≤ 0.01. (b) Variants with MAF > 0.01, (c) All variants. X-axis, -log10 of expected p values; y-axis, -log10 of observed p values, winsorized at 5E-8



Supplementary Figure 2. QQ-plots from GMMAT score test single variant association analysis. (a) Variants with MAF ≤ 0.01. (b) Variants with MAF > 0.01. (c) All variants. X-axis, -log10 of expected p values; y-axis, -log10 of observed p values, winsorized at 5E-8



Supplementary Table 1: Demographic details of this study.

		Controls from north (HNR)	Controls from south (KORA)	Cases
Ancestry		German	German	German/ Austrian
Sample size	Init	2 486	2 921	3 839
	QC	2 451	2 909	3 652
	Unrel	2 417	2 794	3 599
Median age (95% CI)	QC	60 [47; 73]	49 [27; 72]	64 [33; 84]
	Unrel	60 [47; 73]	49 [27; 72]	64 [34; 84]
Sex (% female)	QC	50.3	51.5	69.2
	Unrel	50.5	51.8	69.1

Init, pre QC stage; QC, post QC stage; Unrel, unrelated individuals of the post QC stage; CI: confidence interval

Supplementary Table 2: Quality control steps (QC).

QC milestone	Previous milestone	QC sub step	*N*(individuals)	*N*(markers)
1) Initial QC		Initial dataset	3 839 RLS 2 921 KORA 2 486 HNR	238 876
		Removed because call rate failed	-15 RLS -0 KORA -15 HNR	-259 RLS -9 KORA -27 HNR
		Remove sex check fails (call rate at 126 ChrY markers greater 50% is male)	-20 RLS -0 KORA -10 HNR	
2) Merge data	1)	Merge data	9 186	238 867
		Remove call rate fails	0	-1 228
3a) Heterozygosity outliers detection part 1	2)	Keep autosomal markers	9 186	232 618
		Keep MAF < 0.01	9 186	198 764
		Remove call rate fails	0	
		Detected heterozygosity fails	24	
3b) Heterozygosity outliers detection part 2	2)	Keep autosomal markers	9 186	232 618
		Keep MAF ≥ 0.01	9 186	33 854
		Remove call rate fails	-13	
		Detected heterozygosity fails	30	
4) Relatedness Calculation	2)	Remove heterozygosity fails	-47	
		Keep BeadChip v1.1 markers		228 263
		Keep MAF >= 0.01		40 244
		Remove HWE fails in controls		-29
		Remove long-range LD		-9 957
		Remove LD		-10 643
5) Merged data filtering, creation of analysis data	2)	Remove heterozygosity check and CR fails	-60	
		Remove HWE fails (controls)		-35
		Remove duplicates, individuals with high IBD counts, individuals with missing age	-114	
		Remove markers from black list		-717
		Final: keep polymorphic autosomal markers	9 012	137 011
6) Markers for similarity estimations	5)	Remove long-range LD		-6 807
		Remove MAF < 0.01		-76 700
		Remove LD		-10 284
		Final	9 012	43 220

Supplementary Table 3: Top pathway enrichment results

Reactome pathway	Individuals with minor alleles	Number of SNPs	P value
CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM	8810	4321	4.48E-05
TRANSPORT_OF_NUCLEOSIDES_AND_FREE_PURINE_AND_PYRIMIDINE_BASES_ ACROSS_THE_PLASMA_MEMBRANE	1903	67	4.83E-05
FCERI_MEDIATED_NF_KB_ACTIVATION	5137	231	0.00052
SLC_TRANSPORTER_DISORDERS	8677	878	0.00096
INNATE_IMMUNE_SYSTEM	8810	6167	0.00158
PLATELET_ACTIVATION_SIGNALING_AND_AGGREGATION	8809	1909	0.00172
TRANSPORT_OF_VITAMINS_NUCLEOSIDES_AND_RELATED_MOLECULES	6462	276	0.00173
ION_HOMEOSTASIS	7339	395	0.00194
REGULATION_OF_SIGNALING_BY_CBL	1590	77	0.00199
MUSCLE_CONTRACTION	8810	2018	0.00210
RESPONSE_TO_ELEVATED_PLATELET_CYTOSOLIC_CA2	8788	1295	0.00276
HEMOSTASIS	8810	4138	0.00304
TAK1_ACTIVATES_NFKB_BY_PHOSPHORYLATION_AND_ACTIVATION_OF_ IKKS_COMPLEX	3599	133	0.00414
CREB1_PHOSPHORYLATION_THROUGH_NMDA_RECEPTOR_ MEDIATED_ACTIVATION_OF_RAS_SIGNALING	4442	139	0.00463
INTERFERON_SIGNALING	8749	1085	0.00481
DEX_H_BOX_HELICASES_ACTIVATE_TYPE_I_IFN_AND_INFLAMMATORY_CYTOKINES _PRODUCTION	688	25	0.00529
NEGATIVE_REGULATION_OF_NMDA_RECEPTOR_MEDIATED_NEURONAL _TRANSMISSION	4839	154	0.00544
STRIATED_MUSCLE_CONTRACTION	8485	785	0.00559
UNBLOCKING_OF_NMDA_RECEPTORS_ GLUTAMATE_BINDING_AND_ACTIVATION	4436	139	0.00578
PYRUVATE_METABOLISM	3046	129	0.00615

Reactome pathway, name of biological pathway as provided in Reactome v7.1; individuals with minor alleles, number of individuals in study who carried minor alleles of the variants tested for the pathway; Number of SNPs, number of variants tested for pathway; P value, nominal p-value from SKAT-O analysis.