

ARTICLE

Dilution of candidates: the case of iron-related genes in restless legs syndrome

Konrad Oexle^{*1}, Barbara Schormair², Janina S Ried³, Darina Czamara⁴, Katharina Heim², Birgit Frauscher⁵, Birgit Högl⁵, Claudia Trenkwalder^{6,7}, G Martin Fiedler⁸, Joachim Thiery⁸, Peter Lichtner², Holger Prokisch², Michael Specht⁴, Bertram Müller-Myhsok⁴, Angela Döring^{3,9}, Christian Gieger³, Annette Peters⁹, H-Erich Wichmann^{3,10}, Thomas Meitinger^{1,2} and Juliane Winkelmann^{1,2,11}

Restless legs syndrome (RLS) is a common multifactorial disease. Some genetic risk factors have been identified. RLS susceptibility also has been related to iron. We therefore asked whether known iron-related genes are candidates for association with RLS and, vice versa, whether known RLS-associated loci influence iron parameters in serum. RLS/control samples ($n=954/1814$ in the discovery step, $735/736$ in replication 1, and $736/735$ in replication 2) were tested for association with SNPs located within 4 Mb intervals surrounding each gene from a list of 111 iron-related genes using a discovery threshold of $P=5 \times 10^{-4}$. Two population cohorts (KORA F3 and F4 with together $n=3447$) were tested for association of six known RLS loci with iron, ferritin, transferrin, transferrin-saturation, and soluble transferrin receptor. Results were negative. None of the candidate SNPs at the iron-related gene loci was confirmed significantly. An intronic SNP, rs2576036, of *KATNAL2* at 18q21.1 was significant in the first ($P=0.00085$) but not in the second replication step (joint nominal P -value = 0.044). Especially, rs1800652 (C282Y) in the *HFE* gene did not associate with RLS. Moreover, SNPs at the known RLS loci did not significantly affect serum iron parameters in the KORA cohorts. In conclusion, the correlation between RLS and iron parameters in serum may be weaker than assumed. Moreover, in a general power analysis, we show that genetic effects are diluted if they are transmitted via an intermediate trait to an end-phenotype. Sample size formulas are provided for small effect sizes.

European Journal of Human Genetics (2013) **21**, 410–414; doi:10.1038/ejhg.2012.193; published online 29 August 2012

Keywords: restless legs syndrome; iron parameters; *MEIS1* haplotype; power calculation; linear regression; logistic regression

INTRODUCTION

Restless legs syndrome (RLS)¹ is a sensory-motor disorder characterized by an urge to move and unpleasant sensations in the lower limbs at rest. In Caucasians, RLS is present in up to 10% of the population.^{2,3} Pregnancy, uremia, celiac disease, and iron deficiency are considered to be risk factors.⁴ RLS also has a considerable heritability and is associated with multiple genetic risk factors. Genome-wide association studies^{5–10} identified variants in six loci encompassing the genes *MEIS1*, *BTBD9*, *MAP2K5/SKOR1*, *PTPRD*, and *TOX3/non-coding RNA*.

Aiming for the identification of further genetic risk factors of RLS, we addressed the apparent relation of RLS susceptibility to iron metabolism and iron availability¹¹ that has been explained by cerebral iron being crucial in the etiology of RLS.¹² We therefore tested the SNPs at the loci of a candidate set of genes known to be involved in iron metabolism for association with RLS. Vice versa, we asked whether RLS genes are involved in iron metabolism because the RLS-associated SNP rs3923809 at the *BTBD9* locus has previously been reported to be associated with the serum level of ferritin in a study on RLS subjects and their relatives.⁶

We also performed a general power consideration (ie, sample size calculation) concerning genetic effects on an end-phenotype (eg, a disease) that depend entirely on transmission via an intermediate phenotype (eg, a metabolic parameter).

MATERIALS AND METHODS

Informed consent, written in the respective language, was obtained from each participant. The work has been approved by the institutional review boards of the contributing centers. The primary review boards were located in Munich, Bayerische Ärztekammer und Technische Universität München.

RLS samples and controls

RLS cases were of German or Austrian origin (954 in the discovery step, 735 in replication step 1, and 736 in replication step 2). Diagnosis was based on the diagnostic criteria of the International RLS Study Group¹³ as assessed in a personal interview conducted by an RLS expert. Patients with probably secondary RLS in case of uremia, dialysis, or anemia because of iron deficiency were excluded. The presence of secondary RLS was determined by clinical interview, physical and neurological examination, blood chemistry, and nerve conduction studies whenever deemed clinically necessary.

¹Institute of Human Genetics, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany; ²Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; ³Institute of Genetic Epidemiology I, Helmholtz Zentrum München, Neuherberg, Germany; ⁴Max Planck Institute of Psychiatry, Munich, Germany; ⁵Department of Neurology, Universität Innsbruck, Innsbruck, Austria; ⁶Department of Neurology, Paracelsus-Elena-Klinik, Kassel, Germany; ⁷Department of Neurology, Universität Göttingen, Göttingen, Germany; ⁸Institute of Laboratory Medicine, Universitätsklinikum Leipzig, Leipzig, Germany; ⁹Institute of Genetic Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany; ¹⁰Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany; ¹¹Department of Neurology, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany

*Correspondence: Dr K Oexle, Institut für Humangenetik, Klinikum rechts der Isar, Technische Universität München, Trogerstraße 32, 81675 Munich, Germany. Tel: +49 89 4140 6386; Fax: +49 89 4140 6382, E-mail: oexle@humangenetik.med.tum.de

Received 17 April 2012; revised 9 July 2012; accepted 19 July 2012; published online 29 August 2012

Controls were recruited from the KORA S3/F3 and F4 surveys of the Cooperative Health Research in the South-East German region of Augsburg. KORA procedures and samples have been described before.¹⁴ For the discovery phase, we included 1814 subjects. For the replication steps 1 and 2, we included 736 and 735 subjects, respectively.

Iron-related serum parameters in the KORA cohorts

In the KORA surveys F3 ($n=1638$) and KORA F4 ($n=1809$), iron-related serum parameters (ferritin, transferrin, and soluble transferrin receptor) were determined. Serum parameters were measured by standard laboratory methods, that is, by electrochemiluminescence immunoassay (Roche Diagnostics, Mannheim, Germany) for ferritin, Tina-quant immunoturbidometry (Roche) for soluble transferrin receptor, colorimetry (Roche) for iron, and immunonephelometry (Siemens Healthcare Diagnostics, Eschborn, Germany) for transferrin. For more details, see Oexle *et al.*¹⁵

Genotyping and statistical analysis

Genome-wide genotyping was performed on Affymetrix Human SNP Arrays using 5.0 arrays for RLS cases, and 500K or 6.0 arrays for KORA subjects with all RLS controls being genotyped on 6.0 arrays. Calling, quality control, and imputing procedures have been described previously.^{10,15} To identify and correct for population stratification, multidimensional scaling (MDS; identifying 18 outliers) and Genomic Control analyses were performed. For association testing, logistic regression as implemented in PLINK 1.07 (<http://pngu.mgh.harvard.edu/~purcell/plink/>)¹⁶ was applied after standard filtering, that is, 98% calling for both SNPs and individuals, $HWE > 0.00001$, and minor allele frequency (MAF) $> 5\%$ (except for the hemochromatosis-causing SNP rs1800562 of the *HFE* gene, which was included in the RLS discovery sample in spite of a $MAF < 5\%$).

After filtering, there remained 301 495 SNPs in 922 cases and 1526 controls for the discovery step of the RLS case-control analysis. In this step, we applied age, sex, and the first four axes of variation resulting from MDS analysis as covariates. We performed a candidate-based analysis focusing on SNPs within 4 Mb intervals (2 Mb in each direction) surrounding each of 111 genes (see Supplementary Table 1) known to be involved in iron metabolism^{15,17,18} or in neurodegeneration with brain iron accumulation (NBIA), NBIA1–NBIA3 (*PANK2*, *PLA2G6*, and *FTL*). From that set of SNPs, we selected those with a nominal P -value $< 5 \times 10^{-4}$ for replication. SNPs of known RLS genes were ignored. Genotyping for replication (24 SNPs including technical replicates) was performed on the MassARRAY system using MALDI-TOF mass spectrometry with the iPLEX Gold chemistry (Sequenom Inc., San Diego, CA, USA). Automated genotype calling was done with SpectroTYPER 3.4. Genotype clustering was visually checked by an experienced evaluator. Except one (rs8029116), all SNPs in replication step 1 were genotyped successfully. Replication step 2, intended to provide finemapping of genes in the vicinity of rs2576036 that had appeared to replicate in replication step 1, was performed in the same way and comprised 33 SNPs of whom 32 were genotyped successfully. SNPs for finemapping were selected using the tag SNP selection algorithm “Tagger” as implemented in Haploview 4.1 (Broad Institute, Cambridge, MA, USA). In both replication steps, age and sex were used as covariates.

For association analysis of top RLS-associated SNPs^{5,7–10} with iron-related serum parameters, we selected rs12469063 and rs2300478 in *MEIS1*, rs9357271 in *BTBD9*, rs1975197 in *PTPRD*, rs12593813 in *MAP2K5/SKOR1*, rs6747972 in an intergenic region on chromosome 2p14, and rs3104767 in *TOX3/BC034767*. Moreover, we included rs3923809 in *BTBD9* that has been reported to be associated with ferritin⁶ and rs2576036 in *KATNAL2* that we initially suspected to be associated with RLS in this study (see Results section). Imputed SNPs were used if selected SNPs were not genotyped directly. Imputation was performed with Impute¹⁹ for KORA F3 and F4 separately using hapmap 2 as reference. The genetic association of selected SNPs was tested with a linear regression on $\log_{10}$ -transformed iron traits with age and sex as covariates. The results of the two cohorts F3 and F4 were combined by meta-analysis using a fixed-effect model analogous to Oexle *et al.*¹⁵ Calculations were done using PLINK¹⁶ v1.07 and METAL (<http://www.sph.umich.edu/csg/abecasis/Metal/index.html>).

RESULTS

For the present investigation, we used the same discovery sample as in Winkelmann *et al.*¹⁰ that is, 954 RLS cases and 1814 controls (before filtering, see above), but relaxed the cut-off level for replication from 1×10^{-4} to 5×10^{-4} and focused on a set of genes that have been related to iron metabolism or to NBIA. Besides the three NBIA genes (see OMIM database), this set comprised 35 genes listed by the HealthIron¹⁷ consortium and another 72 genes that have been discussed in recent reviews on iron metabolism.^{15,18} The complete list is given in Supplementary Table 1. Of the SNPs within the 4 Mb-intervals surrounding these genes by 2 Mb in both directions, 18 had P -values below the cut-off level (ignoring the SNPs in high LD with a SNP in one of the already known RLS genes). The set of top hits did not contain rs1800562 ($P > 0.1$). This SNP was specifically included in the discovery step in spite of its MAF being $< 5\%$ because, of all known genetic polymorphisms, it explains the largest fraction of the variance of the body iron storage indicator ferritin.^{15,20} (In the KORA cohorts, it explained 0.5% (KORA F3) to 0.8% (KORA F4) of the variance of the age- and sex-adjusted $\log_{10}$ (ferritin) values).

Including technical replicates, 24 SNPs were then tested for replication in a sample of 736 RLS cases and 736 controls (Supplementary Table 2). Genotyping of one SNP failed (rs8029116 on chr15: 72 610 147 bp) but a nearby replicate (rs11072496) was genotyped successfully. For rs2576036, an intronic SNP of *KATNAL2* on chromosome 18q21.1 (chr18: 42.85 Mb), a significant association with the RLS phenotype was detected in the first replication step ($P = 0.00085$, logistic regression using age and sex as covariates).

In order to further evaluate the possible effect of rs2576036 on expression, we checked an in-house whole blood transcriptome database for association with transcript levels of neighboring genes. Neither the expression of *KATNAL2* nor of any other transcripts was significantly associated with rs2576036. The selection of rs2576036 for this study resulted from its being located within the intervals surrounding the iron-related candidate genes *SMAD2* at chr18: 43.65 Mb and *SMAD7* at chr18: 44.71 Mb. However, neither the expression of *SMAD2* nor the expression of *SMAD7* was associated with rs2576036.

To confirm and to finemap the seeming association of the rs2576036 locus with RLS, we run a second replication analysis on 736 German RLS cases and 735 German controls, specifically addressing *KATNAL2*, its neighbor *PIAS2*, and *CORL2* (*SKOR2*, *FUSSEL18*). *CORL2* was included because it is located in the same region but was not represented in the expression database. For this replication step 2, 33 SNPs were selected, of which 32 were genotyped successfully. None of these SNPs resulted in a significant association signal. SNP rs2576036, which was significant in replication step 1, now showed a P -value of 0.78. The joint analysis for rs2576036 of steps 1 and 2 yielded a P -value of 0.044 which, after Bonferroni correction for 18 loci in replication step 1, also was not significant.

Having tested whether iron-related genes influence the genesis of RLS, we then asked whether RLS-associated genes influence iron-related parameters. We selected seven top hits from the previously reported RLS loci as well as rs3923809 in *BTBD9* that has been reported to be associated with ferritin⁶ and rs2576036 in *KATNAL2* that we initially suspected to be associated with RLS (see Materials and Methods section). In a set of altogether 3447 KORA individuals, none of these SNPs were associated with serum iron or any of the iron-related parameters in serum (ferritin, transferrin, transferring saturation, and soluble transferrin receptor; see Supplementary Table 3 for association results). In view of a recent report of Catoire *et al.*²¹ we further tested whether the risk haplotype of the RLS gene

MEIS1 (G alleles of rs12469063 and rs2300478) is associated with any serum iron parameter. This test also gave a negative result with no *P*-value being smaller than 0.45 (KORA F3)/0.27 (KORA F4).

DISCUSSION

Three possible causes may contribute to the failure of our candidate approach to detect RLS-associated genes among a set of iron-related genes. First, the association between serum iron parameters and RLS may be weaker than usually assumed. Second, our study may be biased. Third, our study may lack power because of a dilution of the genetic effect by the transmission via an intermediate trait. In the following, we discuss all three. The third is presented in terms of a general power analysis.

(1) Iron deficiency has been considered to have a causal role in RLS ever since the first modern description of RLS in the middle of the last century.¹² Iron substitution is a common therapeutic approach to RLS. Several association studies on RLS and serum iron parameters have been performed. In a retrospective study on 18 cases and 18 matched controls, O'Keefe *et al*¹¹ described a significant association to serum ferritin, an indicator of the level of body iron storage. Other retrospective studies with sample sizes between 27 and 302 RLS patients identified associations of serum ferritin to RLS severity and/or the need for therapeutic augmentation.^{22–25} Recently, low serum ferritin was described as a significant predictor of RLS in 301 hospital patients older 50 years of whom 55 had RLS.²⁶ On the other hand, cross-sectional studies on 365, 701, and 714 individuals from German,² Tyrolean,³ and Korean²⁷ population cohorts with 36, 74, and 59 RLS cases, respectively, did not show an association to serum ferritin. The same was true for most other serum iron parameters except for the soluble transferrin receptor in the study that included 74 RLS patients. Although the results of these well-designed studies do not exclude the possibility that iron, especially¹² cerebral iron, is involved in the pathophysiology of RLS (at least in a subgroup of patients), the association between peripheral iron parameters and RLS may be weaker than assumed, thus impeding the power of our approach.

(2) The set of iron-related candidate genes that we selected from the literature is biased by the current state of knowledge. It cannot be excluded that future insights in iron physiology will identify genes that have a stronger effect on the pathogenesis of RLS. Moreover, in the discovery step of the association we only considered polymorphisms with MAFs >5% (except for the *HFE* missense mutation C282Y). This filtering is reasonable in association studies because for small values the MAF is inversely proportional to the power (necessary sample size) of a study (see Appendix). However, it is possible that rare variants of iron-related genes with MAF <5% but strong effect contribute to the genesis of RLS. Detection of such variants will be difficult and, besides next-generation sequencing, necessitate specific study design.²⁸ A further possible bias of our study resides in the fact that our sampling scheme for the RLS GWAS (on which we based the discovery step of this study) excluded cases that had anemia because of iron deficiency. Although this exclusion criterion only affected cases with severe iron deficiency (which already caused anemia), our study would possibly have been more powerful without this criterion.

(3) Dilution of the genetic effect is a third possible reason why none of the iron-related genes was found to be associated with RLS. Consider the constellation delineated in Figure 1a where the influence of a gene on an end-phenotype (eg, disease) entirely depends on the mediation by an intermediate trait (eg, serum parameter), which both are also subject to various other genetic and non-genetic influences. As one can easily show in case of small effect sizes, this constellation

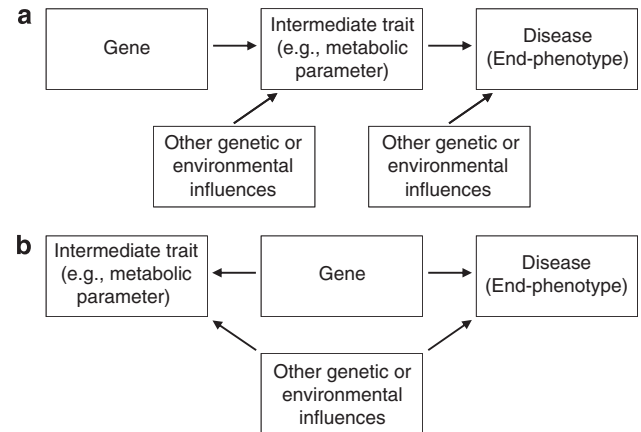


Figure 1 (a) If transmitted by an intermediate trait, the effect of a gene may be diluted by other genetic or environmental influences, which thus impair the power of an association study. (b) If the trait is not truly intermediate and a substantial part of the correlation with the disease results from the pleiotropic effects of the gene, the power of an association study is not impaired in the same way.

implies that the necessary sample size n_{xz} to detect an association between gene (x) and end-phenotype (z) is proportional to product $n_{xy}n_{yz}$ of the sample sizes necessary to detect the associations between gene and intermediate trait (y) and between intermediate trait and end-phenotype. Assume that the intermediate trait y in an individual i is influenced by a genetic effect according to $y_i = a_{xy}x_i + \varepsilon_{xy,i}$ where $x_i \in \{0,1,2\}$ indicates the number of effect alleles, a_{xy} is the effect parameter (assumed to be small, $a_{xy} \ll 1$) and $\varepsilon_{xy,i}$ is a noise parameter with standard normal distribution $\text{No}(0,1)$ that represents a variety of other influences. For simplicity, y is chosen as to have zero mean. As the effect size a_{xy} is small, the variance σ_y^2 is close to 1 and the necessary sample size n_{xy} to detect the genetic influence in a linear regression analysis is proportional to $1/a_{xy}^2$ (see equation (A2) in the Appendix). Second, assume that the influence of y on the occurrence probability $P(z=1|y)$ of the end-phenotype follows a logistic model, $\text{logit}(P(z|y)) = b_0 + b_{yz}y$, where b_0 and $b_{yz} \ll b_0$ are constants again. For a test to successfully detect the association between y and z , a sample size of $n_{yz} \sim 1/b_{yz}^2$ is required (see equation (A3) in the Appendix). Now replace the intermediate trait y by its constituents, that is, $\text{logit}(P(z|x,\varepsilon)) = b_0 + b_{yz}(a_{xy}x + \varepsilon)$, which according to equation (A4) in the Appendix results in $n_{xz} \sim 1/(a_{xy}^2 b_{yz}^2)$ yielding the required proportionality $n_{xz} \sim n_{xy}n_{yz}$ and indicating the dilution phenomenon suggested above.

It has to be considered, however, that the assessment of a candidate gene does not demand the same level of Bonferroni correction as a genome-wide association analysis. Still, this does not entirely compensate for the dilution phenomenon. With $c_{xz}^2 = a_{xy}^2 b_{yz}^2$, equations (A2)–(A4) of the Appendix yield

$$n_{xz} \approx n_{xy}n_{yz} [(Z_\beta + Z_{\alpha_{xz}/2}) / ((Z_\beta + Z_{\alpha_{yz}/2})(Z_\beta + Z_{\alpha_{xy}/2}))]^2$$

where the quantile Z_β represents the required power $1-\beta$ (usually, $Z_\beta = Z_{0.2} = -0.84$) and the $Z_{\alpha/2}$'s are the quantiles of the required significance levels α in two-sided tests with $Z_{\alpha/2}$ necessitating correction for multiple testing, that is, $Z_{\alpha/2} = Z_{0.025} = -1.96$ in a single test, and $Z_{\alpha/2} = Z_{5 \times 10^{-8}} = -5.33$ in a genome-wide test.²⁹ Assuming that the analysis of a candidate gene is a single test and that the association between intermediate trait and disease also was detected in a single test while the GWAS on the intermediate trait to detect the candidate gene required correction for multiple testing, we

get $n_{xz} \approx n_{xy}n_{yz}/(Z_{0.2} + Z_5 \times 10^{-8})^2 = n_{xy}n_{yz}/38$. Thus, only if the association between the intermediate trait (y) and the disease (z) was strong enough to be detectable with sample size $n_{yz} = 38$, will the association analysis on a candidate gene derived from the GWAS on the intermediate trait have sufficient power with a sample size (n_{xz}) not larger than the sample size (n_{xy}) that was required in the GWAS. For the *HFE*-mutation C282Y (rs1800562), the variant that explains the largest single genetic fraction of the ferritin variance,^{15,20} the KORA cohorts indicated an allele frequency of 4.9% (KORA F3)/4.6% (KORA F4) and an effect size parameter of 0.09 (KORA F3)/0.11 (KORA F4). With the variance of $\log_{10}(\text{ferritin})$ being $(0.41)^2$ in KORA F3 and $(0.42)^2$ in KORA F4, these numbers correspond to a necessary GWAS sample size of about $(-0.84 - 5.33)^2/(0.10^2 \times 2 \times 0.047 \times 0.953/0.42^2) \approx 7500$ (see derivation of equation (A2) in the Appendix with $\sigma_y^2 = 0.42^2 \neq 1$). Thus, taking into account that sample sizes in the range of 36 to 74 failed to confirm the association between RLS and serum ferritin,^{2,3,27} the discovery step in our study with 954 cases and 1814 controls (ie, considerably smaller than 7500) was not powerful enough to detect an influence of *HFE* on RLS if that influence fully depends on mediation by serum ferritin.

We also could not confirm the association between the RLS-associated SNP rs3923809 at the *BTBD9* locus and serum ferritin although our population sample from Southern Germany (KORA F3 and F4 with $n = 1638$ and $n = 1809$, respectively) was considerably larger than the Icelandic sample ($n = 965$ individuals) used by the group that claimed this association.⁶ In fact, none of the other top SNPs at the known RLS loci was associated with any serum iron parameter. Again, this failure may be due to a 'dilution' phenomenon analogous to the one explained above. Recently, Catoire *et al*²¹ reported that in RLS patients the risk haplotype of the RLS gene *MEIS1* is associated with increased thalamic ferritin expression, whereas the expression in another cerebral tissue (pons) or in lymphoblastoid cell lines did not depend on that haplotype. Data on liver expression and data on the general population were not provided but may be desirable in view of the fact that we could not detect an association between this haplotype and serum ferritin in the general population. Of course, our results do not exclude the possibility that *MEIS1* may have a differential influence on ferritin expression in certain cerebral regions.

In summary, the analysis presented here puts some caveat on the expectation that genetic elucidation of intermediate traits will always simplify the genetic dissection of end-phenotypes. Under certain conditions, candidate approaches can be successful, of course. Figure 1b shows a constellation where the seeming intermediate trait is not truly intermediate but is modified by pleiotropic actions of genes that also influence the end-phenotype. If the correlation between the trait and the end-phenotype is largely due to a small number of such genes a candidate approach can be quite powerful.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

- 1 Trenkwalder C, Paulus W: Restless legs syndrome: pathophysiology, clinical presentation and management. *Nat Rev Neurol* 2010; **6**: 337–346.
- 2 Berger K, von Eckardstein A, Trenkwalder C *et al*: Iron metabolism and the risk of restless legs syndrome in an elderly general population - the MEMO-Study. *J Neurol* 2002; **249**: 1195–1199.
- 3 Högl B, Kiechl S, Willeit J *et al*: Restless legs syndrome: a community-based study of prevalence, severity, and risk factors. *Neurology* 2005; **64**: 1920–1924.
- 4 Pantaleo NP, Hening WA, Allen RP, Earley CJ: Pregnancy accounts for most of the gender difference in prevalence of familial RLS. *Sleep Med* 2010; **11**: 310–313.
- 5 Winkelmann J, Schormair B, Lichtner P *et al*: Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat Genet* 2007; **39**: 1000–1006.
- 6 Stefansson H, Rye DB, Hicks A *et al*: A genetic risk factor for periodic limb movements in sleep. *N Engl J Med* 2007; **357**: 639–647.
- 7 Schormair B, Kemlink D, Roeske D *et al*: PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. *Nat Genet* 2008; **40**: 946–948.
- 8 Vilarino-Guell C, Farrer MJ, Lin SC: A genetic risk factor for periodic limb movements in sleep. *N Engl J Med* 2008; **358**: 425–427.
- 9 Kemlink D, Polo O, Frauscher B *et al*: Replication of restless legs syndrome loci in three European populations. *J Med Genet* 2009; **46**: 315–318.
- 10 Winkelmann J, Czamara D, Schormair B *et al*: Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS Genet* 2011; **7**: e1002171.
- 11 O'Keefe ST, Gavin K, Lavan JN: Iron status and restless legs syndrome in the elderly. *Age Ageing* 1994; **23**: 200–203.
- 12 Allen RP, Earley CJ: The role of iron in restless legs syndrome. *Mov Disord* 2007; **22**: S440–S448.
- 13 Allen RP, Picchietti D, Hening WA *et al*: Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. *Sleep Med* 2003; **4**: 101–119.
- 14 Wichmann HE, Gieger C, Illig T: KORA-gen - resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 2005; **67**(Suppl 1): S26–S30.
- 15 Oexle K, Ried JS, Hicks AA *et al*: Novel association to the proprotein convertase PCSK7 gene locus revealed by analysing soluble transferrin receptor (sTfR) levels. *Hum Mol Genet* 2011; **20**: 1042–1047.
- 16 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**: 559–575.
- 17 Constantine CC, Gurrin LC, McLaren CE *et al*: SNP selection for genes of iron metabolism in a study of genetic modifiers of hemochromatosis. *BMC Med Genet* 2008; **9**: 18.
- 18 Hentze MW, Muckenthaler MU, Galy B, Camaschella C: Two to tango: regulation of Mammalian iron metabolism. *Cell* 2010; **142**: 24–38.
- 19 Marchini J, Howie B, Myers S *et al*: A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**: 906–913.
- 20 Whitfield JB, Cullen LM, Jazwinska EC *et al*: Effects of HFE C282Y and H63D polymorphisms and polygenic background on iron stores in a large community sample of twins. *Am J Hum Genet* 2000; **66**: 1246–1258.
- 21 Catoire H, Dion PA, Xiong L *et al*: Restless legs syndrome-associated MEIS1 risk variant influences iron homeostasis. *Ann Neurol* 2011; **70**: 170–175.
- 22 Sun ER, Chen CA, Ho G *et al*: Iron and restless legs syndrome. *Sleep* 1998; **21**: 381–387.
- 23 Frauscher B, Gschliesser V, Brandauer E *et al*: The severity range of restless legs syndrome (RLS) and augmentation in a prospective patient cohort: association with ferritin levels. *Sleep Med* 2009; **10**: 611–615.
- 24 Trenkwalder C, Högl B, Benes H, Kohnen R: Augmentation in restless legs syndrome is associated with low ferritin. *Sleep Med* 2008; **9**: 572–574.
- 25 Curgunlu A, Döventaş A, Karadeniz D *et al*: Prevalence and characteristics of restless legs syndrome (RLS) in the elderly and the relation of serum ferritin levels with disease severity: Hospital-based study from Istanbul, Turkey. *Arch Gerontol Geriatr* 2012; **55**: 73–76.
- 26 Quinn C, Uzbeck M, Saleem I *et al*: Iron status and chronic kidney disease predict restless legs syndrome in an older hospital population. *Sleep Med* 2011; **12**: 295–301.
- 27 Kim KW, Yoon IY, Chung S *et al*: Prevalence, comorbidities and risk factors of restless legs syndrome in the Korean elderly population - results from the Korean Longitudinal Study on Health and Aging. *J Sleep Res* 2010; **19**: 87–92.
- 28 Oexle K: A remark on rare variants. *J Hum Genet* 2010; **55**: 219–226.
- 29 Pe'er I, Yelensky R, Altshuler D, Daly MJ: Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 2008; **32**: 381–385.
- 30 Sokal RR, Rohlf FJ: *Biometry*. New York: WH Freeman and Company, 1995, p 578.
- 31 Hsieh FY, Bloch DA, Larsen MD: A simple method of sample size calculation for linear and logistic regression. *Stat Med* 1998; **17**: 1623–1634.
- 32 Bull SB: Sample size and power determination for a binary outcome and an ordinal exposure when logistic regression analysis is planned. *Am J Epidemiol* 1993; **137**: 676–684.
- 33 Demidenko E: Sample size determination for logistic regression revisited. *Stat Med* 2007; **26**: 3385–3397.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)

APPENDIX

For small effect sizes, the test statistics of linear and logistic regression analyses approximate standard normal distributions. Small effect sizes also imply that the variances under the alternative and the null hypothesis are approximately equal. The formulas in power calculations of necessary sample size n therefore simplify to

$$n \approx (Z_{\beta} + Z_{\alpha/2})^2 / t^2 \quad (\text{A1})$$

where t is the test statistic and $Z_{\alpha/2}$ and Z_{β} are quantiles of a standard normal distribution that correspond to the tolerated error rates α and β , that is, to the required significance and power $(1 - \beta)$, respectively.

(1) In the case of linear regression analysis, consider a set of independent individuals where each individual i displays a continuous trait y_i that is influenced by an allele of a gene with effect size a_{xy} and frequency q . This allele occurs in $x_i \in \{0, 1, 2\}$ copies and is assumed to be in Hardy–Weinberg equilibrium. In the linear model, y_i is given as $y_i = a_0 + a_{xy}x_i + \varepsilon_i$ where a_0 and a_{xy} are constants and ε_i is a noise parameter. The model is scaled so that ε has a standard normal distribution with zero mean and unit variance. The precise sample size determination^{30,31} yields $n = (Z_{\alpha/2} + Z_{\beta})^2 / C(\rho)^2 + 3$ where $C(\rho) = \frac{1}{2} \ln((1 + \rho)/(1 - \rho))$ is the Fisher transformation of the correlation coefficient $\rho = a_{xy}\sigma_x / \sigma_y$. For small a_{xy} (ie, small ρ), we can approximate $C(\rho)$ by first-order expansion as $C(\rho) \approx \rho$. Moreover, the variance of the noise term being unity and a_{xy} being small, the variance of y is $\sigma_y^2 = a_{xy}^2 \sigma_x^2 + 1 \approx 1$. To calculate ρ we also need σ_x . In Hardy–Weinberg equilibrium with $\mu_x = \sum x_i q^{x_i} (1 - q)^{(2-x_i)} = 2q$ we get $\sigma_x^2 = \sum (x_i - \mu_x)^2 q^{x_i} (1 - q)^{(2-x_i)} = 2q(1 - q)$. For small effect a_{xy}

and, consequently, large sample size ($\gg 3$) we thus arrive at

$$n_{xy} \approx (Z_{\alpha/2} + Z_{\beta})^2 / (2a_{xy}^2 q(1 - q)) \quad (\text{A2})$$

(2) In logistic regression analysis,^{31–33} the association between a binary response (eg, disease present or absent) and an exposure variable y is modeled as $\text{logit}(P(z|y)) = \log(P(1|y)/P(0|y)) = b_0 + b_{yz}y$ where $P(z|y)$, $z \in \{0, 1\}$, is the conditional response probability, and b_0 and b_{yz} are constants. The value of the effect parameter b_{yz} is estimated by maximum likelihood estimation. For a small effect b_{yz} , the variance of the estimator under the alternative hypothesis approximates the variance under the null hypothesis, $[P(1 - P)\sigma_y^2]^{-1}$, where P is the average response probability with $P \approx P(0|0) = e^{b_0} / [1 + e^{b_0}]$. If y has a standard normal distribution with $\sigma_y^2 = 1$, equation (A1) yields the necessary sample size as

$$n_{yz} \approx (Z_{\alpha/2} + Z_{\beta})^2 / [P(1 - P)b_{yz}^2] \quad (\text{A3})$$

The same derivation applies to multiple logistic regression with $\text{logit}(P(z|x, \varepsilon)) = c_0 + c_{xz}x + c_{\varepsilon z}\varepsilon + \dots$, if the effects of the variates x , $\varepsilon, \dots$ are small. Assuming that $x \in \{0, 1, 2\}$ represents a genotype distribution in Hardy–Weinberg equilibrium with allele frequency q (implying $\sigma_x^2 = 2q(1 - q)$, see above), the necessary sample size to estimate the effect size c_{xz} then is

$$n_{xz} \approx (Z_{\alpha/2} + Z_{\beta})^2 / [2q(1 - q)P(1 - P)c_{xz}^2] \quad (\text{A4})$$

where P is the average response probability with $P \approx P(0|x = 0, \varepsilon = 0) = e^{c_0} / [1 + e^{c_0}]$.