nature genetics

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions

Juliane Winkelmann[1–3], Barbara Schormair[1,3], Peter Lichtner[1,3], Stephan Ripke[2], Lan Xiong[4], Shapour Jalilzadeh[1,3], Stephany Fulda[2], Benno Pütz[2], Gertrud Eckstein[1,3], Stephanie Hauk[1,3], Claudia Trenkwalder[5], Alexander Zimprich[6], Karin Stiasny-Kolster[7], Wolfgang Oertel[7], Cornelius G Bachmann[8], Walter Paulus[8], Ines Peglau[9], Ilonka Eisensehr[10], Jacques Montplaisir[11,12], Gustavo Turecki[13], Guy Rouleau[4], Christian Gieger[14], Thomas Illig[14], H-Erich Wichmann[14,15], Florian Holsboer[2], Bertram Müller-Myhsok[2,16] & Thomas Meitinger[1,3,16]

Restless legs syndrome (RLS) is a frequent neurological disorder characterized by an imperative urge to move the legs during night, unpleasant sensation in the lower limbs, disturbed sleep and increased cardiovascular morbidity. In a genome-wide association study we found highly significant associations between RLS and intronic variants in the homeobox gene *MEIS1*, the *BTBD9* gene encoding a BTB(POZ) domain as well as variants in a third locus containing the genes encoding mitogen-activated protein kinase MAP2K5 and the transcription factor LBXCOR1 on chromosomes 2p, 6p and 15q, respectively. Two independent replications confirmed these association signals. Each genetic variant was associated with a more than 50% increase in risk for RLS, with the combined allelic variants conferring more than half of the risk. MEIS1 has been implicated in limb development, raising the possibility that RLS has components of a developmental disorder.

Nightwalkers, as individuals with RLS call themselves, are forced to move their legs during periods of rest especially in the evening and night to relieve uncomfortable or painful sensations in the deep calf[1]. This diurnal variation leads to impaired sleep onset, and the periodic leg movements during sleep in the majority of patients contribute to sleep disruption and a reduced quality of life as a major consequence[2]. There are recognized secondary forms of RLS such as in iron deficiency, pregnancy and end-stage renal disease and associated morbidity such as increased cardiovascular risk[2,3]. RLS is one of the most common neurological disorders, with an age-dependent prevalence of up to 10% in the elderly in North America and Europe[2]. Dopaminergic agents originally developed for Parkinson's disease have been used to treat RLS, with an unknown mode of action[2]. Neurophysiological, pharmacological and neuroimaging studies suggest that the characteristic symptoms originate in the central nervous system, yet the underlying neurobiology remains obscure[4]. A family history of RLS is present in more than 50% of affected individuals, and similar figures have been reported for heritability in twin studies[5,6]. Linkage analysis uncovered five loci based on recessive (RLS1) or dominant inheritance (RLS2–RLS5), but so far

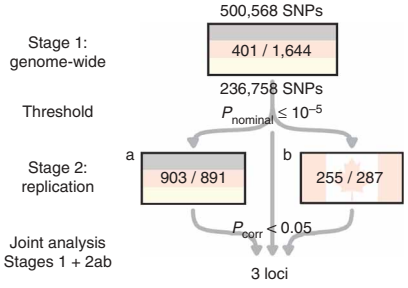

Figure 1 Study overview. Numbers refer to cases and controls and SNPs genotyped and analyzed. The 13 most significant SNPs together with neighboring SNPs were replicated in a German ('a') and a Canadian ('b') case/control sample. Three loci were confirmed in both stage 2 samples of the study.

[1]Institute of Human Genetics, GSF National Research Center of Environment and Health, D-85764 Neuherberg, Munich, Germany. [2]Max Planck Institute of Psychiatry, D-80804 Munich, Germany. [3]Technical University, Institute of Human Genetics, D-81675 Munich, Germany. [4]Laboratoire d'étude des maladies du cerveau, Centre de recherche du CHUM, Hôpital Notre-Dame, Université de Montréal, Montréal, Québec H2L 4M1, Canada. [5]Paracelsus-Elena-Hospital, 34128 Kassel, Germany. [6]Neurological Department, Medical University of Vienna, 1090 Vienna, Austria. [7]Philipps University Marburg, Department of Neurology, 35039 Marburg, Germany. [8]University of Göttingen, Department of Clinical Neurophysiology, 37070 Göttingen, Germany. [9]Neurologische Praxis, 10969 Berlin, Germany. [10]Neurologische Praxis Sendlingerstrasse, 80331 Munich, Germany. [11]Centre d'étude du sommeil, Hôpital du Sacré-Cœur de Montréal, Montréal, Québec H4J 1C5, Canada. [12]Centre de recherche en sciences neurologiques, Université de Montréal, Montréal, Québec H4J 1C5, Canada. [13]Departments of Psychiatry and Human Genetics, McGill University, Douglas Hospital, Montreal, Quebec H4H 1R3, Canada. [14]Institute of Epidemiology, GSF National Research Center for Environment and Health, 85764 Neuherberg, Munich, Germany. [15]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, 81377 Munich, Germany, [16]These authors contributed equally to this work. Correspondence should be addressed to J.W. (janew@mpipsykl.mpg.de) or T.M. (meitinger@gsf.de).

Received 11 April; accepted 12 June; published online 18 July 2007; doi:10.1038/ng2099

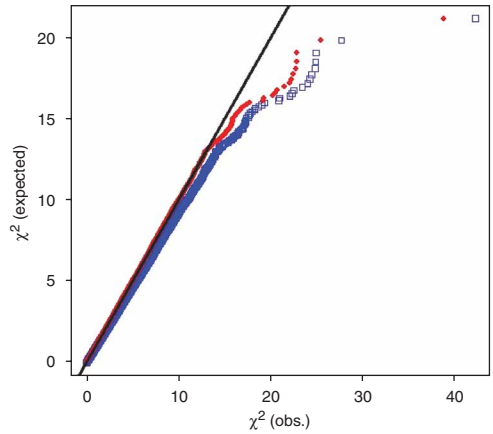

Figure 2 Extent of population stratification. The distribution of expected (under the null hypothesis) versus observed χ^2 values (all *P* values obtained in the analysis of sample 1, using Armitage trend test with age and sex as covariates) before (blue) and after (red) correction by division with λ. Adherence to the diagonal indicates lack of inflation of the statistic. As can be seen in the uncorrected plot, there is evidence for a systematic deviation toward higher-than-expected values. After the correction, there is near-perfect adherence to the diagonal for most of the values obtained, indicating that the correction performed well.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

no causally related sequence variants have been identified[5–7]. With SNP arrays becoming a mature technology, we conducted a genome-wide association study (GWAS), typing 500,568 SNPs in individuals with RLS and in a large control cohort from the general population.

Genome-wide association

The study design involved an exploratory stage (stage 1) followed by replication in two further case-control samples (stages 2a and 2b) (**Fig. 1**). In stage 1, we performed a GWAS, typing cases and controls on a single platform with the Affymetrix 500K Array Set. To enrich for risk alleles and minimize phenotypic heterogeneity, we selected subjects with familial RLS ($n = 401$). Controls were selected randomly from a population-based cohort ($n = 1{,}644$, from the KORA-S3/F3 survey, described previously)[8]. For statistical analysis, we selected SNPs by including only high-quality genotypes to reduce the number of false-positive signals (**Supplementary Table 1** online). A total of 236,758 SNPs passed all quality control filters (mean call rate = 99.48%). The effect of population stratification was negligible (inflation factor $\lambda = 1.09$ via genomic control)[9] (**Fig. 2**). Eigenvalue-based analysis showed only minimal population substructure (**Fig. 2**). An Armitage trend test uncovered four SNPs with *P* values $< 10^{-6}$ (**Fig. 3** and **Supplementary Table 2** online). After correcting for multiple testing, we identified a single SNP within *MEIS1* that reached genome-wide significance (rs2300478, $P_{corrected} < 0.0002$).

Replication of genome-wide findings

For stage 2 replication, 13 SNPs passed our inclusion criteria based on *P* value, location within a linkage peak and visual inspection of clustering data. We selected these and 15 neighboring SNPs for replication. They mapped to six discernible regions. Of these 28 SNPs, 25 were successfully genotyped in stage 2a and 24 in stage 2b (**Supplementary Table 3** online). Individuals in 2a ($n = 903$, familial or sporadic RLS) had been recruited separately using the sampling design of stage 1. Control subjects were selected from KORA-S4 ($n = 891$).

In stage 2a, we found nominally significant evidence for association in five regions, of which three withstood correction for multiple testing (**Fig. 4** and **Supplementary Table 4** online). The first region was on 2p, located in a 32-kb linkage disequilibrium (LD) block containing exon 9 of *MEIS1*. Here, two of three SNPs showed significant association ($P < 10^{-11}$). *MEIS1* is a member of a family of highly conserved *TALE* homeobox genes. Heterodimers of MEIS1 with PBX and HOX proteins augment the affinity and specificity of DNA binding by HOX proteins[10]. MEIS1 has been found to be overexpressed in acute myeloid leukemia[10], and studies in *Xenopus laevis* have shown involvement in neural crest development[11]. In addition, there are several potential links to RLS: during embryonic development, MEIS1 is essential for proximo-distal limb formation[12], and children with restless legs syndrome are often described as having growing pains[13]. MEIS1 is part of a Hox transcriptional regulatory network that specifies spinal motor neuron pool identity and connectivity[14]. Notably, spinal hyperexcitability is an established component in the genesis of periodic leg movements found in individuals with RLS[15]. Specific functions of MEIS1 in postembryonic tissues still remain to be established. The protein is known to be expressed in the adult mouse brain in cerebellar granule cells, the forebrain and, notably, in dopaminergic neurons of the substantia nigra[16].

The second region with significant association was on chromosome 6p, within a 113-kb LD block in intron 5 of the *BTBD9* gene. All five SNPs tested were significant, four of these with *P* values $< 10^{-5}$. Little is known about BTBD9 other than that it belongs to the BTB(POZ) proteins. BTB stands for *broad complex*, *tramtrack* and *bric à brac*, genes that in *Drosophila melanogaster* are required for embryonic

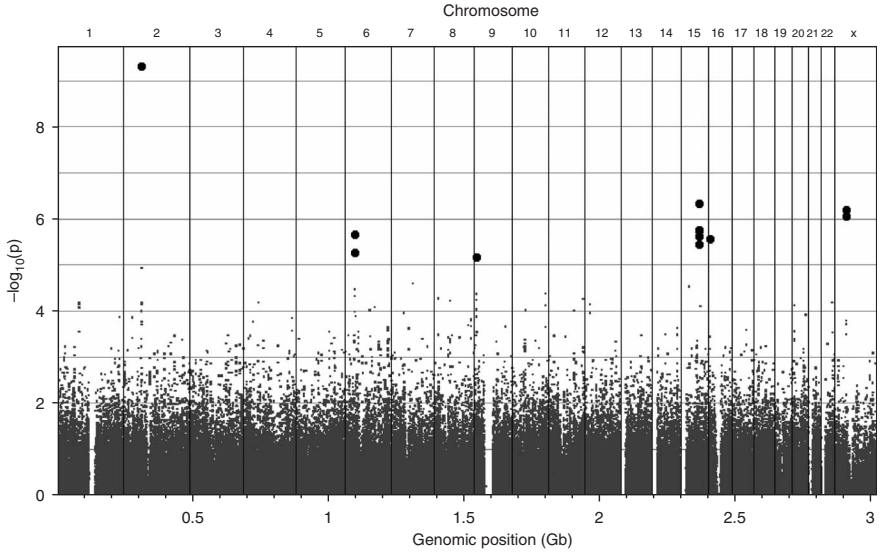

Figure 3 Genome-wide association study for RLS susceptibility loci. The analysis compared 393 successfully genotyped RLS cases with 1,602 population-based KORA controls. The *x*-axis represents genomic position, and the *y*-axis shows $-\log_{10}(P)$. Thirteen SNPs that passed inclusion criteria for the replication study of stage 2 are highlighted in bold. Note that the *P* values of three SNPs on chromosome 15 are very similar, and these SNPs appear as one single dot.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

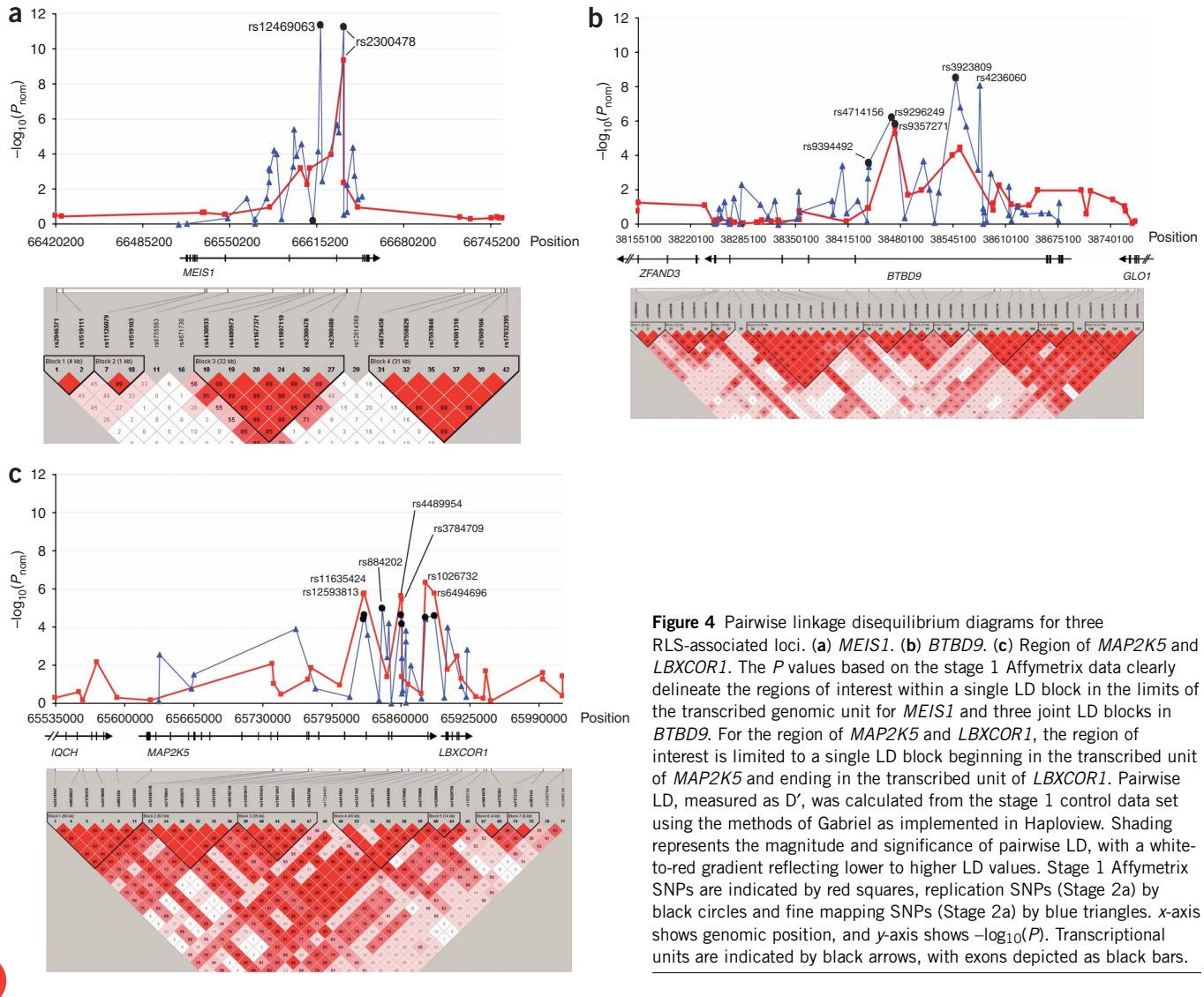

Figure 4 Pairwise linkage disequilibrium diagrams for three RLS-associated loci. (**a**) *MEIS1*. (**b**) *BTBD9*. (**c**) Region of *MAP2K5* and *LBXCOR1*. The *P* values based on the stage 1 Affymetrix data clearly delineate the regions of interest within a single LD block in the limits of the transcribed genomic unit for *MEIS1* and three joint LD blocks in *BTBD9*. For the region of *MAP2K5* and *LBXCOR1*, the region of interest is limited to a single LD block beginning in the transcribed unit of *MAP2K5* and ending in the transcribed unit of *LBXCOR1*. Pairwise LD, measured as D′, was calculated from the stage 1 control data set using the methods of Gabriel as implemented in Haploview. Shading represents the magnitude and significance of pairwise LD, with a white-to-red gradient reflecting lower to higher LD values. Stage 1 Affymetrix SNPs are indicated by red squares, replication SNPs (Stage 2a) by black circles and fine mapping SNPs (Stage 2a) by blue triangles. *x*-axis shows genomic position, and *y*-axis shows $-\log_{10}(P)$. Transcriptional units are indicated by black arrows, with exons depicted as black bars.

development, cell fate determination in the eye, metamorphosis and pattern formation in the limbs[17,18]. Functions of BTB(POZ) proteins include transcription repression, cytoskeleton regulation, tetramerization and gating of ion channels, as well as ubiquitin-dependent protein degradation[17]. The modular nature of this protein and the universal occurrence of the particular domains of BTBD9 make assignment of a specific function difficult at present.

The third region, defined by seven SNPs tested on 15q, showed significant evidence for association, with $P < 10^{-4}$. This region contains a 48-kb LD block overlapping the 3′ end of *MAP2K5*, a member of the mitogen-activated protein kinase family, and the adjacent *LBXCOR1* gene. MAPK pathways are conserved from yeast to human and are activated by a signaling cascade that mediates the transduction of extracellular signals to cytoplasmic nuclear effectors[19]. MAP2K5 is a specific upstream activator of ERK5, and this pathway is activated by oxidative stress, hyperosmolarity and growth factors. In addition, MAP2K5 and ERK5 are abundantly expressed in heart and skeletal muscles, and the MAP2K5/ERK5 MAP kinase cascade is critical at early stages of muscle cell differentiation[19]. The possible link between RLS risk alleles and known biological functions of the MAP2K5-ERK5 pathway is of particular interest, as this pathway is important in neuroprotection of dopaminergic neurons[20]. *LBXCOR1* is annotated as being downstream of *MAP2K5* and acting as a transcriptional corepressor of LBX1. This homeobox gene is critical in the development of sensory pathways in the dorsal horn of the spinal cord that relay pain and touch[21]. Three SNPs within the *PTPRD* gene in the chromosome 9 linkage region (RLS3) and one SNP on chromosome 16 in the A2BP1 gene were nominally significant.

In stage 2b, we genotyped the same SNPs in affected individuals ($n = 255$) and controls ($n = 287$) from a French-Canadian population. Here, we found nominally significant evidence for association in four regions (two SNPs on chromosome 2p, five SNPs on 6p, seven SNPs on 15q and one SNP on 16p, **Supplementary Table 5** online). The same three regions as in stages 1 and 2a remained significant after correction for multiple testing. Odds ratios (ORs) and risk alleles were very similar to those for stage 2a. **Table 1** shows those nine SNPs in the three loci confirmed in all three sample sets and in joint analysis withstanding genome-wide correction for multiple testing.

Table 1 Confirmed association results

dbSNP ID	Chr	Genome position	Gene	MAF (cases)	Risk allele	MAF (controls)
rs2300478	2p	66634957	*MEIS1*	0.367 (G)	G	0.241 (G)
rs9296249	6p	38473819	*BTBD9*	0.162 (C)	T	0.235 (C)
rs9357271	6p	38473851	*BTBD9*	0.165 (C)	T	0.238 (C)
rs12593813	15q	65823906	*MAP2K5*	0.258 (A)	G	0.330 (A)
rs11635424	15q	65824632	*MAP2K5*	0.257 (A)	G	0.330 (A)
rs4489954	15q	65859129	*MAP2K5*	0.239 (T)	G	0.311 (T)
rs3784709	15q	65859329	*MAP2K5*	0.251 (T)	C	0.321 (T)
rs1026732	15q	65882139	*MAP2K5*	0.252 (A)	G	0.327 (A)
rs6494696	15q	65890260	[*MAP2K5*/*LBXCOR1*]	0.253 (C)	G	0.326 (C)

dbSNP ID	OR (95% c.i.)	Stage 1 P_{nom}	Stage 2a P_{nom}	Stage 2b P_{nom}	Stage 1+2a+2b P_{nom}	Stage 1+2a+2b $P^{a}_{corrected}$
rs2300478	1.74 (1.57–1.92)	4.89E–10	5.93E–12	2.19E–03	3.41E–28	8.08E–23
rs9296249	1.67 (1.49–1.89)	2.19E–06	1.61E–06	4.14E–03	3.99E–18	9.44E–13
rs9357271	1.66 (1.48–1.87)	5.48E–06	1.85E–06	2.48E–03	6.31E–18	1.50E–12
rs12593813	1.50 (1.36–1.66)	1.85E–06	4.95E–05	1.57E–02	1.06E–15	2.51E–10
rs11635424	1.51 (1.37–1.67)	1.77E–06	2.54E–05	6.60E–03	3.65E–16	8.64E–11
rs4489954	1.51 (1.36–1.67)	2.44E–06	2.60E–05	1.66E–02	2.68E–15	6.35E–10
rs3784709	1.52 (1.37–1.68)	3.56E–06	7.46E–05	1.79E–03	4.06E–16	9.61E–11
rs1026732	1.53 (1.39–1.70)	4.67E–07	2.78E–05	5.22E–03	6.09E–17	1.44E–11
rs6494696	1.52 (1.38–1.69)	1.79E–06	5.20E–05	5.22E–03	2.00E–16	4.74E–11

SNPs with significant association that were successfully genotyped in all three case-control samples, located in three different genomic regions. Genome positions refer to the human March 2006 (hg18) assembly. [*MAP2K5*/*LBXCOR1*] denotes an intergenic position of the SNP. MAF, minor allele frequency; OR, odds ratio; c.i., confidence interval; P_{nom} = nominal *P* value. MAF refers to stage 2a data only; OR was calculated using combined data from all stages. *P* values for stage 1, 2a and combined analysis were calculated using logistic regression implementing an Armitage trend test and taking sex and age as covariates into account. *P* values in stage 1 and 2a resulting from this regression were further corrected for population stratification by dividing the resulting χ^2 by the inflation factor λ. $^{a}P_{corrected}$ = *P* value corrected for multiple testing using Bonferroni's method, correcting for 236,758 SNPs.

Fine mapping, haplotype and risk analysis

We genotyped tagging SNPs and all known coding and splice-site SNPs for fine mapping in the stage 2a samples. This confirmed the candidate regions defined by the explorative phase of the study (**Fig. 4**). Haplotype analysis for *MEIS1* delineated a haplotype block (rs3890755 to rs12469063). A haplotype completely described by allele A (rs6710341) and allele G (rs12469063) was more strongly associated than each single SNP in this block ($P = 5.87 \times 10^{-20}$, OR = 2.75 [95% confidence interval, 2.23–3.41]). This haplotype was also maximally associated in the Canadian sample ($P = 8.51 \times 10^{-7}$, OR = 2.36 [1.40–3.97], **Fig. 5**). For *BTBD9* and the *MAP2K5* and *LBXCOR1* region haplotype analysis confirmed the results of single-SNP analysis.

In exploratory analysis, we compared the ORs obtained under the allele dosage model to those obtained under the unrestricted model. For *MEIS1* and *BTBD9*, we did not find any significant difference between the models tested (MEIS1: $P = 0.714$; BTBD9: $P = 0.913$), but the allele dosage model was more parsimonious. For the *MAP2K5* and *LBXCOR1* region, the allele dosage model was significantly less likely than the unrestricted model ($P = 0.006$). Estimates pointed to a recessive model. This model was significantly better than the allele dosage model ($P = 0.009$) and not worse than the unrestricted model ($P = 0.395$). There was no difference in effect estimates between samples (**Supplementary Table 6** online).

In the combined German samples, lower limits of the sequential attributable fraction (SAFs)[22,23] were estimated at 0.092, 0.303 and 0.079 for *MEIS1*, *BTBD9* and the *MAP2K5* and *LBXCOR1* region respectively. Corresponding upper limits (equal to the population attributable risk fraction (ARF)) were 0.227, 0.492 and 0.201. In the Canadian sample, the lower limits of the SAFs were 0.075, 0.316 and 0.090, respectively, and we estimated the upper limits at 0.226, 0.550 and 0.258, respectively. We could not identify any statistical interaction between these loci, either in the individual samples or in the combined German or combined German-Canadian samples. Overall, although the single ARF and SAF estimates may be slightly overestimated, they clearly indicate that the three loci account for a large part of the phenotype in the populations studied. We estimated the ARF jointly attributable to the three loci at 68.6% in the German population and 74.2% for the Canadian population.

A comparison of familial versus sporadic cases in the combined stage 1 and 2a data set demonstrated virtually indistinguishable ORs

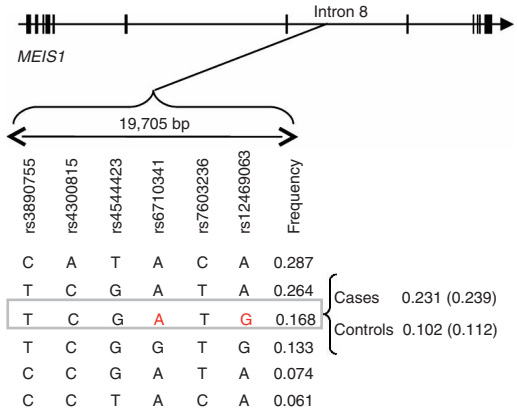

Figure 5 Haplotype structure for *MEIS1*. A haplotype consisting of six SNPs (of which rs6710341 and rs12469063 fully tagged the risk haplotype) is associated with RLS with odds ratios of 2.75 and 2.36 in the stage 2a and 2b samples, respectively. Haplotype frequencies for all haplotypes occurring with these six SNPs are based on cases and controls jointly and are given for cases and controls separately for the risk haplotype. For the Canadian sample, the frequencies are given in brackets and are based on the two tagging SNPs.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

for the regions on 6p and 15q. For the region on 2p, the risk was higher in familial (rs2300478: OR = 1.82 [1.55–2.14]) than in sporadic cases (OR = 1.59 [1.34–1.90]). However, confidence intervals were overlapping with no significant difference in allele distributions ($P = 0.22$, **Supplementary Table 7** online). The familial relative risk figures estimated by the risk to siblings λ_s were 1.13 for *MEIS1*, 1.02 for *BTBD9* and 1.03 for *MAP2K5/LBXCOR1* in the German data set, with almost identical estimates in the Canadian data.

The increasing medical attention to RLS in recent years is matched by our ignorance about its underlying molecular basis. The genetic heterogeneity of RLS has made linkage studies notoriously difficult and favors association approaches. In agreement with power calculations, an initial genome-wide screen for common variants in 400 cases and 1,600 controls enabled us to detect risk alleles with odds ratios >1.5. Sample size in the replicate was twice as high as in the initial GWAS and provided unequivocal evidence for the signals. The effects were strong enough that a second replication in a small independent sample from Canada also yielded significant signals for all three regions. A particular feature of our study design is the use of a control group from the general population. This provided us with very accurate estimates of the genotype frequencies and it avoided any bias to which a disease-negative population is prone.

The identification of significant signals in genes that have not been considered candidates from previous biological knowledge is a recurring theme in GWASs[24]. The current knowledge about *MEIS1*, *BTBD9*, *MAP2K5* and *LBXCOR1* opens new avenues of RLS research, and the involvement of developmental genes challenges us to rethink our basic concept of this widespread disease.

A major proportion of the risk for RLS is explained by variants in the loci identified. We could not derive any different contributions from any of these loci to familial versus sporadic RLS. The associated variants all convey very low familial relative risk ($\lambda_s < 1.15$ in all cases). The lack of positive results within the known linkage regions does not argue against the validity of the linkage results. The nominally significant signals detected in the RLS3 linkage region might indicate an allelic series of variants conferring weak and strong effects within the same gene.

This study is not exhaustive in identifying genetic factors contributing to RLS, and further investigations will provide a better picture of what constitutes the genetic architecture of the complex phenotype of restless legs syndrome. Future studies should investigate endophenotypes or secondary RLS cases, which might show alternative signal patterns. An interesting question is also whether the loci identified have a role in other dopaminergic disorders such as Parkinson's disease or in other associated disorders such as attention deficit hyperactivity disorder or sleep disorders. Further experimental advances might include features such as higher sample numbers in the exploratory stage, higher SNP density, modification of clustering algorithms[25], inclusion of lower frequency polymorphisms, investigation of copy number changes and use of lower statistical thresholds using *a priori* information.

METHODS

Study population and phenotype assessment. Cases of stages 1 and 2a were of European descent and were diagnosed according to standard criteria[2] in a personal interview. Familial RLS was defined by at least one affected first-degree relative. We excluded subjects with secondary RLS due to uremia, dialysis and iron deficiency.

Controls of stage 1 and 2a were of European descent and from the KORA S3/F3 and S4 surveys, representative of the general population. KORA procedures have been described[8]. For stage 1, we included 1,644 subjects from S3/F3, ages 35–84 years, and for stage 2a, 891 age- and sex-matched subjects from S4. In 2a, 102 affected individuals were outside the age range of KORA and were matched to the next age group.

Affected individuals and controls of stage 2b were of French-Canadian ancestry. Affected individuals ($n = 255$) were diagnosed according to standard criteria[2], and polysomnography was performed in 156 subjects; of those, 82.1% ($n = 128$) showed significant periodic leg movements during sleep. Controls were recruited from the general population (n = 287). Secondary cases were excluded.

Studies were performed according to the declaration of Helsinki and approved by institutional review boards in Germany, Austria, and Canada. Written informed consent was obtained from participants. For demographic data of successfully genotyped samples, see **Supplementary Table 8** online.

Genome-wide assays, SNP genotyping and quality control. Stage 1 genotyping was performed using the Affymetrix 500K Array Set. Genotypes were determined using the BRLMM algorithm with cases and controls undergoing a joint cluster analysis. From 500,568 SNPs, a total of 236,758 were selected for subsequent analyses based on stringent quality control criteria. Exclusion criteria were call rate <98% ($n = 146,297$), minor allele frequency (MAF) <10% ($n = 151,583$), deviations from HWE ($P < 0.00001$, $n = 22,536$) and low number of heterozygotes (<10, $n = 33,122$). 14,069 SNPs were monomorphic. For a detailed breakdown, see **Supplementary Table 1**.

For the 13 SNPs passing the inclusion criteria for genotyping in stages 2, visual inspection of clustering was performed using the Affymetrix SNP Signaling Tool 1.0.0.12. All clusters passed this test. To validate the stage 1 experiment, we genotyped 15 SNPs in 400 samples on another platform (Sequenom MassArray system) with a genotype discordance rate of 0.2%.

Stage 2 and fine-scale mapping were performed using MALDI-TOF mass spectrometry on a Sequenom system (Autoflex HT and SpectroTYPER RT 3.4 analysis software). Assays were designed using AssayDesign 3.1.2.2 with iPLEX Gold chemistry default parameters. **Supplementary Table 9** online lists oligonucleotide sequences of replication and fine mapping.

SNP quality control criteria leading to exclusion were call rate <97%, MAF <10% and $P < 0.001$ for deviations from HWE in controls. This resulted in an exclusion of one SNP (rs2110974) in stage 2a, two SNPs (rs2110974, rs7881785) in stage 2b and 51 SNPs in fine mapping. All coding SNPs were monomorphic. A total of 28 affected individuals and 55 controls in stage 2a and 44 affected individuals and 46 controls in stage 2b were excluded owing to low call rate (<90%) of all SNPs within a single DNA sample.

SNP selection for stage 2. We used the following inclusion criteria: (i) $P < 10^{-6}$ in stage 1 analysis (four SNPs); (ii) $P \leq 10^{-5}$ with two neighboring SNPs (± 100 kb) with $P \leq 10^{-3}$ (eight SNPs); (iii) $P \leq 10^{-4}$ for SNPs within described linkage peaks (one SNP in RLS3). For these 13 SNPs, we chose 15 additional neighboring SNPs based on LD structure for genotyping in the replication samples 2a and 2b (**Supplementary Table 3**).

SNP selection for fine mapping. SNPs in the coding regions and 10 kb of flanking sequences were selected using the Tagger algorithm ($r^2 = 0.8$) implemented in HAPLOVIEW 3.3.2 (ref. 26). In addition, all coding-region SNPs and splice-site SNPs were included. This led to 41 SNPs on chromosome 2p (38 tagging, 1 synonymous and 2 nonsynonymous), 77 SNPs on chromosome 6p (tagging only) and 46 SNPs on chromosome 15q (37 tagging, 1 synonymous, 4 nonsynonymous, 2 splice site, 2 frameshift coding). In total, 164 SNPs were selected, of which 163 were converted into genotyping assays, and 103 with a MAF >10% were analyzed.

Analysis of genetic effects. To test and correct for possible population stratification, we performed an EIGENSOFT[27,28] analysis. We used a random sample of 16,000 SNPs passing the quality criteria for the stage 1 sample and allowed for ten rounds of outlier removal. In the first six rounds, a total of 50 outliers (8 cases and 42 controls) were removed, with none removed in the remaining rounds. To assess stratification, we compared the expected distribution of P values for association versus the expected χ^2 distribution with one degree of freedom[29]. We compared the empirically observed mean of the lower 90% of the distribution of the statistics observed and divided it by its expectation[9]. This led to an inflation factor (λ) of 1.09 (**Fig. 2**).

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

We performed logistic regression analysis coding the number of minor alleles as the dependent variables, thus implementing Armitage's trend test, including age and sex as covariates and allowing for interactions between age, sex and the number of alleles. Odds ratios and confidence interval limits were obtained through logistic regression analysis. The χ^2 values resulting from these analyses (stage 1, 2a and fine mapping) were divided by λ, assuming similar conditions for both German samples.

Haplotype analysis was performed using HAPLOVIEW 3.3.2 (refs. 26,29), with the fraction of strong LD informative comparisons set at 0.9, and using UNPHASED 3.0.8, which allows the incorporation of age and sex as covariates[30]. Haplotype blocks were delineated using the method of Gabriel implemented in HAPLOVIEW[26]. ORs were obtained using logistic regression with age and sex as covariates in the stage 2 samples. χ^2 and P values in 2a were λ corrected. Differences between familial and sporadic cases were tested using Fisher's exact test. Familial attributable risks were calculated using the power calculator described in ref. 22.

Multiple testing. Using WG-PERMER, a program for rapid permutation of genome-wide data, preliminary analysis showed that P values after Westfall-Young and Bonferroni correction, with the number of tests set at the number of SNPs tested ($n = 236{,}758$), were in good agreement. This may reflect stringent criteria for SNPs to enter the analysis, resulting in low average r^2 values between SNPs. To maintain comparability across results, we show Bonferroni-corrected P values for stage 1 and the combined analysis. For stages 2a and 2b, we give Westfall-Young–corrected P values based on 10,000 permutations, as only a few candidate regions were tested with high LD between them, and thus Bonferroni would be conservative.

Power analysis. Power analysis for the combined German sample was performed using the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/). The power of any SNP tested with MAF $\geq$ 0.2 and OR $\geq$ 1.5 (or 1/1.5 or lower) was beyond 90%. The P value used was the P value required for a significant result after Bonferroni correction ($P = 0.05/236{,}758 = 2.112 \times 10^{-7}$).

Testing the mode of inheritance. OR values and likelihoods were obtained using logistic regression analysis with age, sex and samples as covariates in the combined stage 1, 2a and 2b samples. Significance testing between models was done using the likelihood ratio test.

Attributable risk fraction. To quantify the contribution of these loci to RLS, we estimated the population attributable risk fraction (ARF)[22] and the sequential attributable fraction (SAF)[23]. We used the allele dosage model for *MEIS1* and *BTBD9* and the recessive model for *MAP2K5* and *LBXCOR1* and calculated upper and lower limits of SAFs[23]. For each locus, we used the SNP with the lowest P value, aware of the fact that this may lead to slight overestimation of the ARF. The ARF for the three loci combined was calculated by allowing for the possibility of simultaneous exposure to several of the risk genotypes.

Requests for materials: janew@mpipsykl.mpg.de or meitinger@gsf.de.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We are grateful to all patients who participated in this study. The authors also thank T.M. Strom, J. Favor, D. Vogt-Weisenhorn, W. Wurst and I. Tews for discussions and R. Feldmann, J. Golic, K. Junghans, B. Schmick, N. Trapp, M. Petzold, G. Fischer and M. Putz for technical assistance. We acknowledge L. Habersack, H. Rhese and J. Schmidt-Evers from the German RLS patient organization for supporting this study. Part of this work was financed by the National Genome Research Network (NGFN). The KORA study group consists of H.-E. Wichmann (speaker), R. Holle, J. John, T. Illig, C. Meisinger, A. Peters and their co-workers, who are responsible for the design and conduct of the KORA studies. The KORA (Cooperative Research in the Region of Augsburg) research project was initiated and financed by the National Research Centre for Environment and Health (GSF), which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. S.H. was partly supported by a grant from the German RLS patient organization. J.W. was partly supported by a grant form the Bavarian Ministry of Science, Culture and Art. The Canadian part of the study was supported by a Canadian Institutes of Health Research (CIHR) grant to G.R, J.M and G.T.

AUTHOR CONTRIBUTIONS

Study design: J.W., P.L., G.R., F.H., B.M.-M., T.M.; recruitment and biobanking of individuals with RLS: J.W., S.H., C.T., A.Z., K.S.-K., W.O., C.B., W.P., I.P., I.E., T.M.; recruitment and biobanking of KORA controls: C.G., T.I., H.-E.W.; recruitment and biobanking of Canadian affected individuals and controls: L.X., J.M., G.T., G.R.; Affymetrix genotyping: B.S., P.L., G.E.; Sequenom genotyping: B.S., P.L., S.J.; supervision of typing of all markers: J.W., P.L.; software development and data processing: S.R.,B.P.; statistical analysis: S.R., B.P., B.M.-M.; clustering of Affymetrix genotypes: S.R., B.M.-M.; manuscript writing: J.W., B.S., S.F., L.X., F.H., B.M.-M., T.M.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Walters, A.S. *et al.* A questionnaire study of 138 patients with restless legs sydrome: the 'night-walkers' survey. *Neurology* **46**, 92–95 (1996).
2. Allen, R.P. *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.* **4**, 101–119 (2003).
3. Winkelman, J.W., Finn, L. & Young, T. Prevalence and correlates of restless legs syndrome symptoms in the Wisconsin Sleep Cohort. *Sleep Med.* **7**, 545–552 (2006).
4. Barrière, G., Cazalets, J.R., Bioulac, B., Tison, F. & Ghorayeb, I. The restless legs syndrome. *Prog. Neurobiol.* **77**, 139–165 (2005).
5. Winkelmann, J. *et al.* Genetics of restless legs syndrome (RLS): state-of-the-art and future directions. *Mov. Disord.,* published online 7 June 2007 (doi:10.1002/mds.21587).
6. Rao, S., Winkelmann, J. & Wang, Q.K. in *Restless Legs Syndrome. Diagnosis and Treatment* (ed. Ondo,W.G.) 111–123 (Informa Healthcare, New York, 2007).
7. Kemlink, D. *et al.* Family-based association study of the restless legs syndrome loci 2 and 3 in a European population. *Mov. Disord.* **22**, 207–212 (2007).
8. Wichmann, H.E., Gieger, C., Illig, T. & MONICA/KORA Study Group KORA-gen-resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67**, S26–S30 (2005).
9. Clayton, D.G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.* **37**, 1243–1246 (2005).
10. Azcoitia, V., Aracil, M., Martinez, A.C. & Torres, M. The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. *Dev. Biol.* **280**, 307–320 (2005).
11. Maeda, R. *et al.* Xmeis1, a protooncogene involved in specifying neural crest cell fate in *Xenopus* embryos. *Oncogene* **20**, 1329–1342 (2001).
12. Mercader, N. *et al.* Conserved regulation of proximodistal limb axis development by Meis/Hth. *Nature* **402**, 425–429 (1999).
13. Rajaram, S.S., Walters, A.S., England, S.J., Mehta, D. & Nizam, F. Some children with growing pains may actually have restless legs syndrome. *Sleep* **27**, 767–773 (2004).
14. Dasen, J.S., Tice, B.C., Brenner-Morton, S. & Jessel, T.M. A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. *Cell* **123**, 477–491 (2005).
15. Bara-Jimenez, W., Aksu, M., Graham, B., Sato, S. & Hallett, M. Periodic limb movements in sleep. State-dependent excitability of the spinal flexor reflex. *Neurology* **54**, 1609–1615 (2000).
16. Allen Institute for Brain Science. Allen Brain Atlas. *http://www.brain-map.org* (2004).
17. Stogios, P.J., Downs, G.S., Jauhal, J.J., Nandra, S.K. & Prive, G.G. Sequence and structural analysis of BTB domain proteins. *Genome Biol.* **6**, R82 (2005).
18. Godt, D., Couderc, J.L., Cramton, S.E. & Laski, F.A. Pattern formation in the limbs of *Drosophila*: bric à brac is expressed in both a gradient and a wave-like pattern and is required for specification and proper segmentation of the tarsus. *Development* **119**, 799–812 (1993).
19. Dinev, D. *et al.* Extracellular signal regulated kinase 5 (ERK5) is required for the differentiation of muscle cells. *EMBO Rep.* **2**, 829–834 (2001).
20. Cavanaugh, J.E., Jaumotte, J.D., Lakoski, J.M. & Zigmond, M.J. Neuroprotective role of ERK1/2 and ERK5 in a dopaminergic cell line under basal conditions and in response to oxidative stress. *J. Neurosci. Res.* **84**, 1367–1375 (2006).
21. Gross, M.K., Dottori, M. & Goulding, M. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* **34**, 535–549 (2002).

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

22. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association. *Nat. Genet.* **38**, 209–213 (2006).
23. Rowe, A.K., Powell, K.E. & Flanders, W.D. Why population attributable fractions can sum to more than one. *Am. J. Prev. Med.* **26**, 243–249 (2004).
24. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
25. Plagnol, V., Cooper, A.D., Todd, J.A. & Clayton, D.G. A method to address differential bias in genotyping in large–scale association studies. *Plos Genetics* **3**, e74 10.1371/journal.pgen.0030074 (2007).
26. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualisation of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
27. Patterson, N., Price, A.L. & Reich, D. Population structure and Eigenanalysis. *Plos Genetics* **2**, e190 (2006).
28. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
29. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
30. Dudbridge, F. UNPHASED user guide. *Technical Report 2006/5*. MRC Biostatistics Unit, Cambridge, UK, (2006).

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics