## SHORT COMMUNICATION

# Independent replication of *STAT3* association with multiple sclerosis risk in a large German case–control sample

**Christina M. Lill · Brit-Maren M. Schjeide · Denis A. Akkad · Paul Blaschke ·
Alexander Winkelmann · Lisa-Ann Gerdes · Sabine Hoffjan · Felix Luessi ·
Thomas Dörner · Shu-Chen Li · Elisabeth Steinhagen-Thiessen ·
Ulman Lindenberger · Andrew Chan · Hans-Peter Hartung · Orhan Aktas ·
Peter Lohse · Tania Kümpfel · Christian Kubisch · Joerg T. Epplen · Uwe K. Zettl ·
Lars Bertram · Frauke Zipp**

**Abstract** Recent genome-wide association studies have implicated the "signal transducer and activator of transcription 3" gene (*STAT3*) as a putative new multiple sclerosis (MS) susceptibility locus. However, independent validation studies are sparse. Therefore, we performed a genetic association study of two *STAT3* polymorphisms (rs744166 and rs2293152) in a large and independent German case–control sample of 5,904 subjects. We

Christina M. Lill and Brit-Maren M. Schjeide are joint first authors.
Lars Bertram and Frauke Zipp are joint last authors.

C. M. Lill · F. Luessi · F. Zipp (✉)
Department of Neurology,
University Medical Center of the Johannes Gutenberg-University Mainz,
Langenbeckstr. 1,
55131 Mainz, Germany
e-mail: zipp@uni-mainz.de

C. M. Lill · B.-M. M. Schjeide · L. Bertram
Department of Vertebrate Genomics,
Max Planck Institute for Molecular Genetics,
Berlin, Germany

D. A. Akkad · S. Hoffjan · J. T. Epplen
Department of Human Genetics,
Ruhr University,
Bochum, Germany

P. Blaschke · A. Winkelmann · U. K. Zettl
Department of Neurology,
University of Rostock,
Rostock, Germany

L.-A. Gerdes · T. Kümpfel
Institute for Clinical Neuroimmunology,
Ludwig Maximilian University,
Munich, Germany

T. Dörner
Department of Medicine, Rheumatology,
and Clinical Immunology,
Charité University Medicine,
Berlin, Germany

S.-C. Li · U. Lindenberger · L. Bertram
Max Planck Institute for Human Development,
Berlin, Germany

E. Steinhagen-Thiessen
Interdisciplinary Metabolic Center, Lipids Clinic,
Charité University Medicine,
Berlin, Germany

A. Chan
Department of Neurology,
Ruhr University,
Bochum, Germany

H.-P. Hartung · O. Aktas
Department of Neurology, Medical Faculty,
Heinrich Heine University,
Düsseldorf, Germany

P. Lohse
Department of Clinical Chemistry,
Ludwig Maximilian University,
Munich, Germany

C. Kubisch
Institute of Human Genetics, University of Ulm,
Ulm, Germany

observed a nominally significant, albeit weak association between rs744166 and MS susceptibility (odds ratio= 1.09, $P$=0.012) in our sample. This study supports the association between *STAT3* and an increase in MS risk. Taking into account the functional role of *STAT3*, our results favour an involvement of $T_h17$ lymphocytes in MS.

**Keywords** STAT3 · Multiple sclerosis · Genetic association · Replication

## Introduction

Genome-wide association studies (GWAS) and GWAS meta-analyses have recently led to the identification of numerous new susceptibility loci in MS [1–3]. *STAT3* (encoding "signal transducer and activator of transcription 3") was initially observed as a putative MS risk gene in a GWAS from Finland. The same group also found evidence for association between *STAT3* and MS risk in other populations of European descent [3]. The strongest, genome-wide significantly associated single nucleotide polymorphism (SNP) in that study was rs744166 (odds ratio (OR)=1.15, $P$=2.8×10$^{-10}$). Recently, association of MS and other SNPs in the *STAT3* locus was reported in a GWAS meta-analysis [4]. Here, the strongest signal was observed with rs2293152 (OR=0.82, $P$=4.1×10$^{-8}$). Interestingly, rs2293152 displays only weak linkage disequilibrium (LD) with the original SNP rs744166 ($r^2$=0.23 in the 1000 Genomes CEU sample). However, since these two studies used partly overlapping datasets, the association has yet to be confirmed in entirely independent samples. A fully independent study assessing the potential association between *STAT3* and MS risk in a Spanish case–control sample did not detect evidence for association [5]. Therefore, additional independent evaluations in large

datasets are needed to assess the potential role of polymorphisms in *STAT3* and MS risk. It is noteworthy that *STAT3*, which encodes a nuclear transcription factor, is also a compelling MS candidate gene based on functional studies. These suggest a role in the differentiation and expansion of $T_h17$ cells and in the development of experimental autoimmune encephalomyelitis, the rodent model of MS, in vivo (e.g., ref. [5, 6]). In this study, we have genotyped rs744166 and rs2293152 in *STAT3* in one of the largest available case–control samples comprising 5,904 Caucasian individuals from Germany who were not included in any of the previous analyses.

## Methods

This sample, collected as part of a German multicenter effort, comprised 2,932 MS cases and 2,972 control subjects (for demographic characteristics see Table 1). All subjects were of self-reported German Caucasian descent. Detailed medical history and/or examination excluded the presence of inflammatory diseases in all controls. All samples were collected with informed written consent and appropriate ethical approval at the respective sites.

SNPs rs744166 and rs2293152 in *STAT3* were genotyped on 384-well microtiter plates using allelic discrimination assays based on TaqMan chemistry (Applied Biosystems) "C___3140282_10" and "C___3140302_1" following the manufacturer's instructions. Each plate contained an approximately equal proportion of cases and controls as well as 5% CEU HapMap control samples that were duplicated between plates. Genotyping and genotype calling were performed blind to the disease status. Power calculations were performed with the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/) assuming a disease prevalence of 0.1%.

**Table 1** Demographic details of the German case–control sample

| Sites (MS cases vs. controls) | $N$ cases/controls | % Female cases/controls | Mean AAE±SD cases/controls | Mean AAO±SD cases |
|---|---|---|---|---|
| Mainz and Berlin (Zipp vs. BASE-II and Dörner) | 715/868 | 68.5/65.5 | 37±9.6/34±14.9 | 28±8.1[a] |
| Rostock (Zettl) | 526/470 | 73.6/69.8 | 42±12.1/52±18.6 | 32±10.7 |
| Munich (Kuempfel/Lohse) | 364/401 | 70.6/49.9 | 44±12.8/42±16.2 | n.g. |
| Bochum/Essen (Epplen) | 1,070/404 | 70.8/43.4 | 42±11.3/43±12.4 | 32±10.2 |
| Duesseldorf/Bonn (Aktas and Hartung vs. Kubisch) | 257/829 | 72.7/74.3 | 39±10.7/44±16.0 | n.g. |
| Total | 2,932/2,972 | 71.0/60.2 | 41±11.4/42±16.9 | – |

Information in parentheses provides details on the source of MS cases and control samples. BASE-II=controls ($n$=760) drawn from the Berlin Aging Study II [9]

$N$ number, *AAE* age at examination, *AAO* age at onset, *SD* standard deviation, *n.g.* data not provided

[a] Data available for 592 cases

All genetic analyses were performed in PLINK v1.07 (http://pngu.mgh.harvard.edu/purcell/plink/). Association statistics were based on an additive transmission model and adjusted for age and sex via logistic regression. Assessment of Hardy–Weinberg equilibrium (HWE) was performed using Pearson's $\chi^2$ as implemented in PLINK. Statistical significance of the association results is expressed as one-tailed $P$ values and 95% confidence intervals. LD between variants in the *STAT3* region was assessed using the SNAP software (www.broadinstitute.org/mpg/snap/) based on CEU data from the 1000 Genomes Pilot 1 Project (http://www.1000genomes.org/). Functional annotation of SNPs was performed using the ENSEMBL database (http://www.ensembl.org/biomart/martview/).

Results and discussion

Overall, genotyping efficiencies were 99.2% and 99.3%, and accuracies were 100% and 99.8% (one inconsistency across 541 HapMap genotypes) for rs2293152 and rs744166, respectively. We had >99% and 97% power to detect association of rs2293152 and rs744166 at one-sided $\alpha = 0.05$ assuming the originally reported ORs (0.82 and 1.15, respectively) as underlying effect sizes. Control genotype frequencies were distributed according to HWE (Table 2). We could confirm a nominally significant association between the G-allele at rs744166 in *STAT3* and MS susceptibility in our dataset (OR, 1.09; 95% confidence interval (CI) 1.01–1.17; $P_{(one\ tailed)} = 0.016$; see Table 2). The second tested SNP, rs2293152, only showed a nonsignificant trend of association with MS risk (OR=0.95; 95% CI 0.88–1.02; $P_{(one\ tailed)} = 0.091$, see Table 2), although the direction of effect was identical to that reported in the GWAS meta-analysis [4]. Logistic regression analysis after adjustment for rs744166 did not reveal a second, independent signal for rs2293152 (data not shown).

Here, we have independently assessed the potential association between MS risk and two common polymorphisms in the *STAT3* gene in 5,904 subjects from Germany. Interestingly, the stronger and statistically more significant effect was observed with polymorphism rs744166, originally implicated by the Finnish GWAS [3], while rs2293152, which showed genome-wide significant evidence for association in the recent GWAS meta-analysis [4] yielded only a weak, statistically nonsignificant effect. In general, the effect sizes estimated in our sample were smaller than those described in the two previous studies (i.e., ORs of 1.15 and 0.82). The most likely explanation for the smaller ORs in our validation sample may be the well-characterized "winner's curse" phenomenon, i.e., an overestimation of the effect size in the GWAS datasets that led to the identification of *STAT3* as an MS risk factor. Actually, the OR of 1.09 for rs744166 in our dataset is well in line with the effect size estimate for *STAT3* (rs9891119: OR=1.10, $P=4.6 \times 10^{-7}$) as well as the median OR (1.11) of all MS susceptibility loci observed in the most recent and largest GWAS [2]. Not surprisingly, rs9891119 is in very strong LD with rs744166 ($r^2 = 0.81$ in the CEU sample of the 1000 Genomes Project), probably tagging the same underlying signal. Whereas in good agreement with the most recent and extensive association data, our results could still be affected by random error or some undetected bias, e.g., population stratification. While the latter possibility could not be directly assessed across the entire dataset owing to the small number of SNPs that have been genotyped in this sample to date, GWAS data are available for 760 of the controls included here with no evidence for admixture (L.B., unpublished data). This is in line with other reports suggesting only a very minor degree of population substructure in the German population (e.g., ref. [7]). Therefore, as all cases and controls were drawn from the same source population, undetected admixture has very likely not affected our results to an appreciable extent.

**Table 2** Association data summary for *STAT3* in the German case–control sample

| SNP | Location (hg18) | Genotypes | Cases | Controls | MAF cases/controls | HWE $P$ value controls | OR (95% CI) | $P$ value |
|---|---|---|---|---|---|---|---|---|
| rs744166 | chr17:40,514,201 | GG | 537 | 500 | 0.433/0.413 | 0.849 | 1.09 (1.01–1.17) | 0.016 |
|  |  | GA | 1,449 | 1,436 |  |  |  |  |
|  |  | AA | 925 | 1,015 |  |  |  |  |
| rs2293152 | chr17:40,481,529 | GG | 497 | 503 | 0.400/0.411 | 0.732 | 0.95 (0.88–1.02) | 0.091 |
|  |  | GC | 1,335 | 1,419 |  |  |  |  |
|  |  | CC | 1,077 | 1,027 |  |  |  |  |

Logistic regression analysis was performed using an additive model with adjustment for age and sex. Statistical significance from the logistic regression analysis is reported as one-sided $P$ values. Hardy–Weinberg equilibrium (HWE) deviation in controls was assessed using Pearson's $\chi 2$ test. Allele names refer to the "+" strand. Location is provided as chromosome ("chr") and base pair position based on build 18 (hg18) of the human genome assembly (NCBI Build 36.2)

*MAF* minor allele frequency, *HWE* Hardy–Weinberg equilibrium, *OR* odds ratio, *CI* confidence interval

Finally, rs744166 and rs2293152 as well as rs9891119 are intronic SNPs that may merely tag the genetic variant(s) functionally responsible for the association of *STAT3* with MS. However, none of these three SNPs is in LD ($r^2 \geq 0.2$ based on the CEU sample of the 1000 Genomes Project) with any potentially functionally relevant variant in the coding region of *STAT3*. Further large fine-mapping studies and functional experiments are warranted to clarify the exact underlying molecular mechanisms. Thus far, our results point towards a potential pathogenic role of *STAT3*, a proinflammatory transcription factor, which is known to be involved in T cell differentiation and activity, particularly in interleukin 17 production and RORγt/Foxp3 balance [8].

## Conclusion

We could confirm the previously reported association between common polymorphisms in *STAT3* and MS risk in this large and independent sample of 5,904 subjects. Further studies are needed to clarify the exact molecular mechanisms underlying this association.

## References

1. De Jager PL et al (2009) Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nat Genet 41:776–782
2. Sawcer S et al (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476:214–219
3. Jakkula E et al (2010) Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. Am J Hum Genet 86:285–291
4. Patsopoulos N, et al. (2011) Genomewide meta-analysis identifies novel multiple sclerosis susceptibility loci. Ann Neurol (in press) Available at "http://onlinelibrary.wiley.com/doi/10.1002/ana.22609/pdf"
5. Cénit MC et al (2010) STAT3 locus in inflammatory bowel disease and multiple sclerosis susceptibility. Genes Immun 11:264–268
6. Liu X, Lee YS, Yu C-R, Egwuagu CE (2008) Loss of STAT3 in CD4+ T cells prevents development of experimental autoimmune diseases. J Immunol 180:6070–6076
7. Simón-Sánchez J et al (2009) Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet 41:1308–1312
8. Yang X-P et al (2011) Opposing regulation of the locus encoding IL-17 through direct, reciprocal actions of STAT3 and STAT5. Nat Immunol 12:247–254
9. Bertram L, et al. (2011) Cohort profile: the Berlin Aging Study II (BASE-II). (under review)