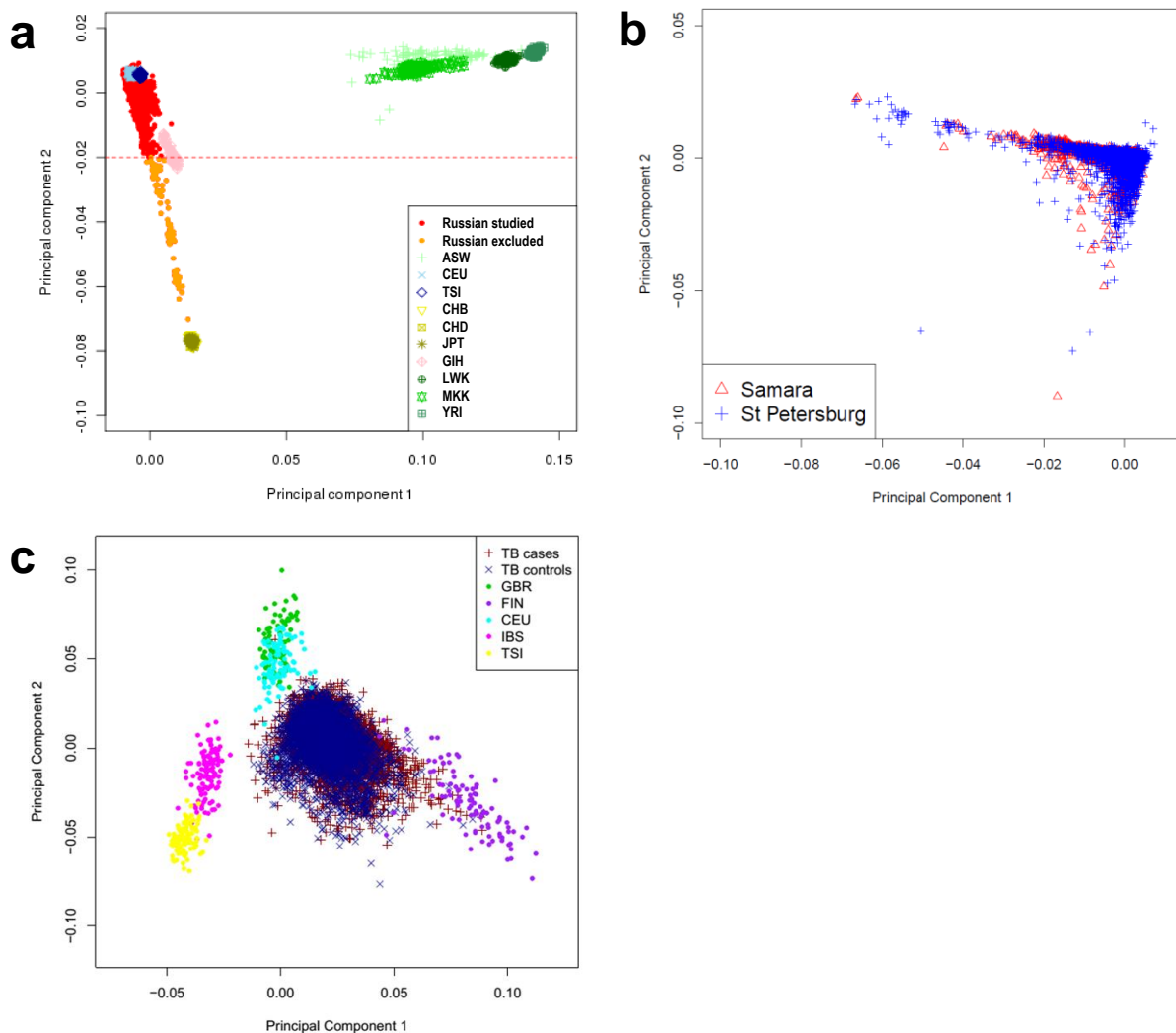


Supplementary Figure 1

**Quality control of the GWAS samples by heterozygosity and missing genotypes rate**

Dashed red lines represent QC thresholds of more than 2% missing genotype data or excess  $\pm 3.5$  standard deviation of heterozygosity rate. Samples circled in red were excluded from the association study.



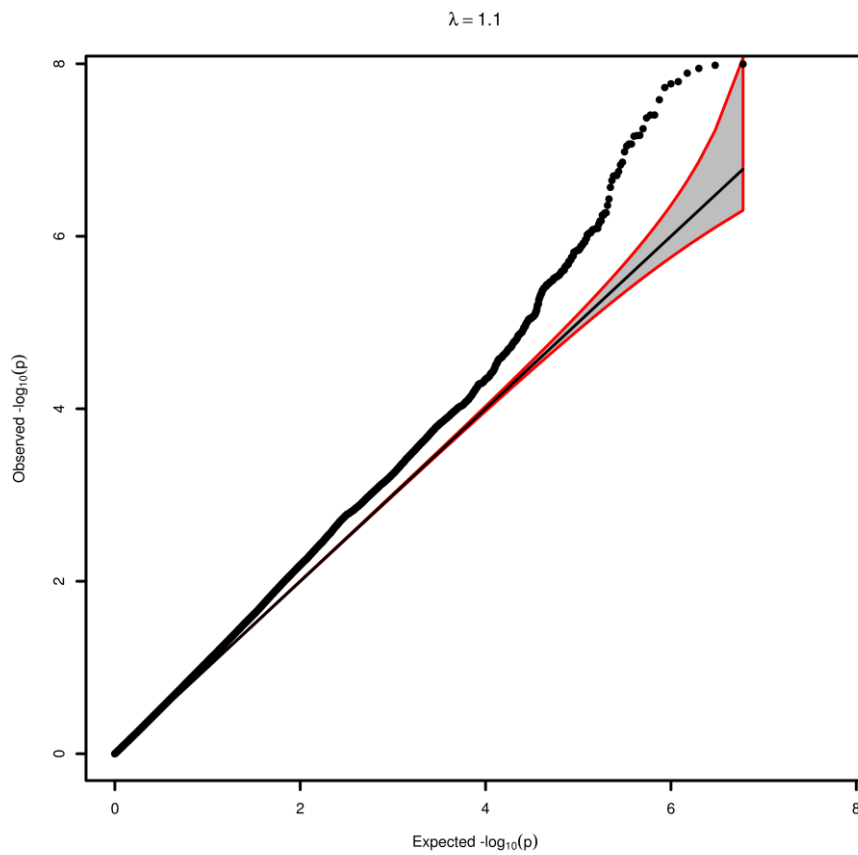
Supplementary Figure 2

### Principal component analysis (PCA) of the Russian GWAS samples

**(a)** First and second principal components. Russian samples are plotted with ten HapMap3 populations (ASW – African ancestry in Southwest USA; CEU – Utah residents of European ancestry; CHB – Han Chinese in Beijing, China; CHD – Chinese in Metropolitan Denver, Colorado; JPT – Japanese in Tokyo, Japan; GIH – Gujarati Indians in Houston, Texas; LWK – Luhya in Webuye, Kenya; MKK – Maasai in Kinyawa, Kenya; TSI – Tuscans in Italy; YRI – Yoruba in Ibadan, Nigeria).

**(b)** All GWAS samples plotted on the first two principal components colored by the city of sample origin.

**(c)** Russian TB cases and controls projected onto the first two principal components using SNP weights precomputed from European samples in the 1000 Genomes Phase III project (CEU — Utah residents with Northern and Western European ancestry; IBS — Iberian populations in Spain; FIN — Finnish in Finland; GBR — British in England and Scotland; and TSI — Toscani in Italy) using SNPweights (Chen, C.Y. *et al.* Improved ancestry inference using weights from external reference panels. *Bioinformatics* **29**, 1399-1406 (2013)).



Supplementary Figure 3

**The quantile - quantile plot of the observed versus the expected  $-\log_{10}P$ -values under the null for 7,614,862 SNPs from the GWAS results after genotype imputation**

The diagonal black line is  $y = x$ , and the grey shapes show 95% confidence interval under the null. The overall distribution has a genomic inflation factor ( $\lambda_{GC}$ ) of 1.10.