**Supporting Information**

Table S1. Relative pairs in Bangladeshi cohort

|  |  |  |
| --- | --- | --- |
| **Relationship type** | **# Relative pairsa** | **Proportion IBD** |
| Parent-offspring | 210 | 0.5 |
| Siblings | 583 | 0.4-0.5 |
| Avuncular/grandparents | 605 | ~0.25 |
| Cousins | 945 | ~0.125 |
| Distant relatives | 5122 | 0.05-0.125 |
|  | | |

a Relative pairs present in our data consisting of ~12.8M pairs

Table S2. Baseline characteristics of Bangladeshi cohorts by type of telomere length measurement: qPCR and Luminex-based method (QGP), 2000-2009 (n=5,075)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **qPCR** | **Luminex method** | |
|  |  | HEALS | BEST | ACE |
| **n** |  | (2,203) | (1,825) | (1,047) |
| **Sex** | Male | 1,060 (48.1%) | 996 (54.6%) | 396 (37.8%) |
|  | Female | 1,143 (51.9%) | 829 (45.4%) | 651 (62.2%) |
|  |  |  |  |  |
| **Age** | Years | 38.3 (10.6) | 43.7 (10.6) | 36.8 (10.7) |
|  |  |  |  |  |
| **BMI** | <18.5 | 863 (39.6%) | 698 (38.3%) | 431 (41.4%) |
|  | 18.5-24.9 | 1,173 (53.8%) | 963 (52.8%) | 549 (52.7%) |
|  | >25 | 145 (6.7%) | 162 (8.9%) | 62 (5.9%) |
|  |  |  |  |  |
| **Smoking** | Never | 1,298 (58.9%) | 1,110 (60.8%) | 735 (70.2%) |
|  | Former | 176 (8%) | 198 (10.9%) | 58 (5.5%) |
|  | Current | 729 (33.1%) | 517 (28.3%) | 254 (24.3%) |
|  |  |  |  |  |
| **Telomere length** | T/S ratio or TQIa | 0.765 (0.17) | 0.672 (0.12) | 0.679 (0.12) |
| **Adjusted TL** | Standardized residualsb | 0.02 (0.99) | 0.00 (0.96) | 0.00 (1.06) |

a These are the raw T/S (for qPCR) and TQI (for luminex-method) measures, prior to mixed effects modeling.

b Standardized residuals from mixed-effects models used in GWA analysis as TL phenotype. Table S3. Minor allele frequency of rs2297439 for the 26 populations of the 1000 Genomes Project

|  |  |  |
| --- | --- | --- |
| **Populationa** | **Super populationb** | **MAF** |
| CHB | EAS | 0.03 |
| JPT | EAS | 0.05 |
| CHS | EAS | 0.02 |
| CDX | EAS | 0.05 |
| KHV | EAS | 0.03 |
| CEU | EUR | 0.06 |
| TSI | EUR | 0.09 |
| FIN | EUR | 0.07 |
| GBR | EUR | 0.08 |
| IBS | EUR | 0.13 |
| YRI | AFR | 0.07 |
| LWK | AFR | 0.02 |
| GWD | AFR | 0.06 |
| MSL | AFR | 0.04 |
| ESN | AFR | 0.02 |
| ASW | AFR | 0.06 |
| ACB | AFR | 0.03 |
| MXL | AMR | 0.06 |
| PUR | AMR | 0.12 |
| CLM | AMR | 0.06 |
| PEL | AMR | 0.03 |
| GIH | SAS | 0.39 |
| PJL | SAS | 0.23 |
| BEB | SAS | 0.24 |
| STU | SAS | 0.24 |
| ITU | SAS | 0.32 |

a Names of the 26 populations can be found at <http://www.internationalgenome.org/faq/which-populations-are-part-your-study/>, BEB is Bengali from Bangladesh

b The 26 populations have been divided into the following super populations: African (AFR), Admixed American (AMR), East Asian (EAS), European (EUR), and South Asian (SAS).

Table S4. Characteristics of proxy variants in LD with rs2297439 based on 1000 Genomes BEB population

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SNP Identifier** | **Position** | **Major/Minor allele** | **MAF (BEB)** | **MAF (EUR)a** | **Distanceb** | ***r2*** | **Regulome DBc** | **Imputation score** |
| rs77552606 | [62314054](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62313804-62314304&snp142=pack&hgFind.matches=rs75031349) | (A/G) | 0.24 | 0.07 | 2,926 | 1 | 2b | 0.82 |
| rs114703330 | [62309392](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62309142-62309642&snp142=pack&hgFind.matches=rs114703330) | (T/C) | 0.24 | 0.07 | 1,736 | 1 | 5 | 0.84 |
| rs75031349 | [62311129](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62310879-62311379&snp142=pack&hgFind.matches=rs77552606) | (T/C) | 0.24 | 0.07 | 1 | 1 | 5 | 0.83 |
| rs78430218 | [62313878](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62313628-62314128&snp142=pack&hgFind.matches=rs78430218) | (C/T) | 0.23 | 0.07 | 2,750 | 0.94 | 4 | 0.77 |
| rs16983884 | [62320674](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62320424-62320924&snp142=pack&hgFind.matches=rs16983884) | (C/T) | 0.24 | 0.07 | 9,546 | 0.91 | 2b | 0.80 |
| rs3787099 | [62307517](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62307267-62307767&snp142=pack&hgFind.matches=rs3787099) | (A/G) | 0.24 | 0.07 | 3,611 | 0.91 | 5 | 0.82 |
| rs79981941 | [62307938](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62307688-62308188&snp142=pack&hgFind.matches=rs79981941) | (C/A) | 0.24 | 0.07 | 3,190 | 0.91 | 6 | 0.82 |
| rs373458837 | [62304449](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62304199-62304699&snp142=pack&hgFind.matches=rs373458837) | (A/-) | 0.25 | 0.10 | 6,679 | 0.85 | -- | -- |
| rs76636901 | [62282821](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62282571-62283071&snp142=pack&hgFind.matches=rs76636901) | (C/T) | 0.23 | 0.07 | 28,307 | 0.82 | 4 | 0.78 |
| rs143190905 | [62291767](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62291517-62292017&snp142=pack&hgFind.matches=rs143190905) | (G/T) | 0.23 | 0.07 | 19,361 | 0.82 | 7 | 0.82 |
| rs73137724 | [62308517](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62308267-62308767&snp142=pack&hgFind.matches=rs73137724) | (T/C) | 0.30 | 0.15 | 2,611 | 0.62 | 5 | 0.60 |
| rs56057703 | [62363512](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr20%3A62363262-62363762&snp142=pack&hgFind.matches=rs56057703) | (C/T) | 0.28 | 0.17 | 52,384 | 0.48 | 5 | 0.74 |

a Minor allele frequencies for North and Western European populations (CEU, FIN, and GBR) b Distance from rs2297439 in base-pairs c Higher RegulomeDB scores represent increasing probability of being a regulatory variant (http://www.regulomedb.org/help)

Table S5. Attempted replication of suggestive association signals in ENGAGE Consortium data

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SNP Identifier** | **Chr** | **Closest Gene** | **ENGAGE data** | | | | **Bangladesh data** | | |
| **MAF** | **βb** | **P** | **N** | **MAF** | **βb** | **P** |
| rs1151814 | 1 | *PARP1* | 0.23 | -0.02 | 0.09 | 29,538 | 0.10 | -0.16 | 2.0x10-6 |
| rs9357354 | 6 | *NCR2* | 0.31 | -0.002 | 0.82 | 26,705 | 0.17 | 0.13 | 1.4x10-6 |
| rs75283006a | 8 | *RBPMS-AS1* | -- | -- | -- | -- | 0.09 | 0.15 | 1.7x10-5 |
| rs9537514 | 13 | *PRR20* | 0.25 | -0.01 | 0.12 | 37,406 | 0.21 | 0.11 | 4.4x10-6 |
| rs28790308a | 15 | *ONECUT1* | -- | -- | -- | -- | 0.25 | -0.10 | 3.6x10-6 |

a r75283006 and rs28790308 were not imputed in the ENGAGE Consortium data

b β corresponds to the association of minor allele with TL, major allele is the reference

Table S6. Observed associations for genetic variants reported to be associated with TL in prior GWA studies

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SNP Identifier** | **Chr** | **Gene** | **Major/Minor allele** | **Bangladesh Data** | | | | **ENGAGE data from Codd, et al.** | | |
| **MAF** | **βa** | **SE** | ***P*** | **MAF** | ***P*** | **Same directionality** |
| rs11125529 | 2 | *ACYP2* | C/A | 0.10 | 0.06 | 0.03 | 0.06 | 0.14 | 4.5x10-8 | Yes |
| rs10936599 | 3 | *TERC* | C/T | 0.24 | -0.13 | 0.02 | 2.8x10-8 | 0.25 | 2.5x10-31 | Yes |
| rs7675998 | 4 | *NAF1* | G/A | 0.12 | -0.008 | 0.03 | 0.79 | 0.22 | 4.3x10-16 | Yes |
| rs2736100 | 5 | *TERT* | C/A | 0.44 | -0.09 | 0.02 | 8.2x10-6 | 0.49 | 4.4x10-19 | Yes |
| rs9420907 | 10 | *OBFC1* | A/C | 0.10 | 0.05 | 0.03 | 0.13 | 0.13 | 6.9x10-11 | Yes |
| rs2535913 | 14 | *DACF4* | C/T | 0.18 | -0.02 | 0.03 | 0.37 | 0.34 | 6.4x10-10 | Yes |
| rs3027234 | 17 | *CTC1* | T/C | 0.07 | 0.003 | 0.04 | 0.95 | 0.21 | 2.3x10-8 | Yes |
| rs412658 | 19 | *ZNF676* | C/T | 0.41 | 0.03 | 0.02 | 0.10 | 0.35 | 9.8x10-9 | Yes |
| rs8105767 | 19 | *ZNF208* | A/G | 0.37 | 0.06 | 0.02 | 0.003 | 0.29 | 1.1x10-9 | Yes |
| rs755017 | 20 | *RTEL1* | A/G | 0.22 | 0.002 | 0.02 | 0.92 | 0.13 | 6.7x10-9 | Yes |

a β corresponds to the association of minor allele with TL, major allele is the reference.

**Fig S1.** Layout of quantitative PCR plate (design 1) consisting of triplicates of 14 participant DNA samples and six replicates of a reference DNA sample, each measured for telomere (light gray cells) and single copy gene (dark gray cells) for a total of 96 samples per plate.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| A | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| B | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 |
| C | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 6 | 6 |
| D | 7 | 7 | 7 | 7 | 7 | 7 | 8 | 8 | 8 | 8 | 8 | 8 |
| E | 9 | 9 | 9 | 9 | 9 | 9 | 10 | 10 | 10 | 10 | 10 | 10 |
| F | 11 | 11 | 11 | 11 | 11 | 11 | 12 | 12 | 12 | 12 | 12 | 12 |
| G | 13 | 13 | 13 | 13 | 13 | 13 | 14 | 14 | 14 | 14 | 14 | 14 |
| H | ref | ref | ref | ref | ref | ref | ref | ref | ref | ref | ref | ref |

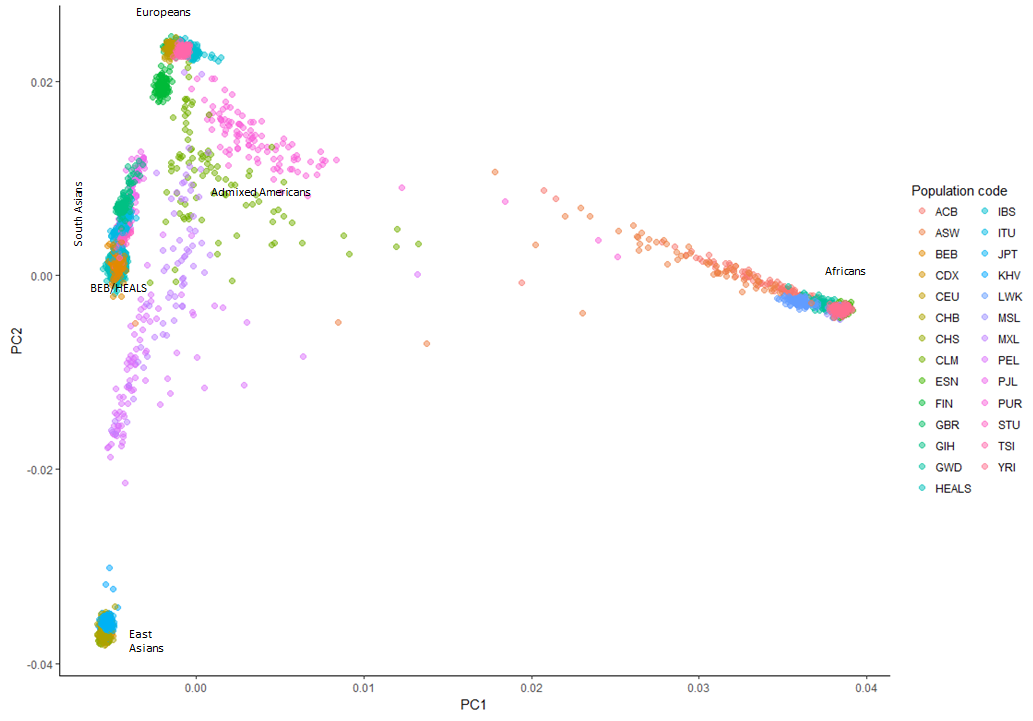
**Fig S2.** Layout of quantitative PCR plate (design 2) consisting of triplicates of 31 participant DNA samples and triplicates of a reference DNA sample, each measured for either telomere or single copy gene at the same position on paired plates. Light gray represents the first plate consisting of telomere primers only, dark gray represents the second plate consisting of single copy gene primers only.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| A | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 4 |
| B | 5 | 5 | 5 | 6 | 6 | 6 | 7 | 7 | 7 | 8 | 8 | 8 |
| C | 9 | 9 | 9 | 10 | 10 | 10 | 11 | 11 | 11 | 12 | 12 | 12 |
| D | 13 | 13 | 13 | ref | ref | ref | 14 | 14 | 14 | 15 | 15 | 15 |
| E | 16 | 16 | 16 | 17 | 17 | 17 | 18 | 18 | 18 | 19 | 19 | 19 |
| F | 20 | 20 | 20 | 21 | 21 | 21 | 22 | 22 | 22 | 23 | 23 | 23 |
| G | 24 | 24 | 24 | 25 | 25 | 25 | 26 | 26 | 26 | 27 | 27 | 27 |
| H | 28 | 28 | 28 | 29 | 29 | 29 | 30 | 30 | 30 | 31 | 31 | 31 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| A | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 4 |
| B | 5 | 5 | 5 | 6 | 6 | 6 | 7 | 7 | 7 | 8 | 8 | 8 |
| C | 9 | 9 | 9 | 10 | 10 | 10 | 11 | 11 | 11 | 12 | 12 | 12 |
| D | 13 | 13 | 13 | ref | ref | ref | 14 | 14 | 14 | 15 | 15 | 15 |
| E | 16 | 16 | 16 | 17 | 17 | 17 | 18 | 18 | 18 | 19 | 19 | 19 |
| F | 20 | 20 | 20 | 21 | 21 | 21 | 22 | 22 | 22 | 23 | 23 | 23 |
| G | 24 | 24 | 24 | 25 | 25 | 25 | 26 | 26 | 26 | 27 | 27 | 27 |
| H | 28 | 28 | 28 | 29 | 29 | 29 | 30 | 30 | 30 | 31 | 31 | 31 |

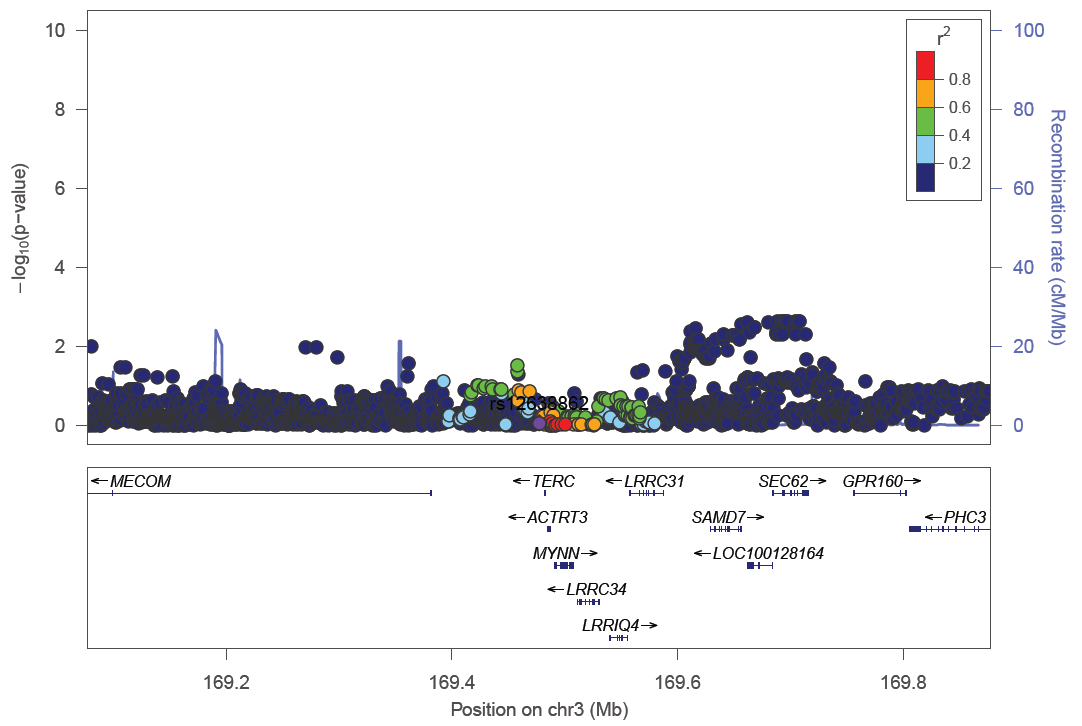
**Fig S3. Generation of an adjusted telomere length phenotype using a mixed-effects regression method.** A total of 5,075 telomere length (TL) measures were collected using two methods (qPCR and Luminex) from three cohorts (HEALS, ACE, and BEST). We ran mixed-effects models (MEM) for each different method and plate design used, and residuals obtained from the MEMs were standardized and used directly as TL phenotypes in the GWA analysis.

**Fig S4. Principal Components Analysis (PCA) using 1000 Genomes reference populations.** This PCA of 111,018 uncorrelated SNPs includes 2,504 1000 Genomes reference samples and 1,524 unrelated participants from our Bangladeshi cohort. The HEALS population code signifies our study participants.

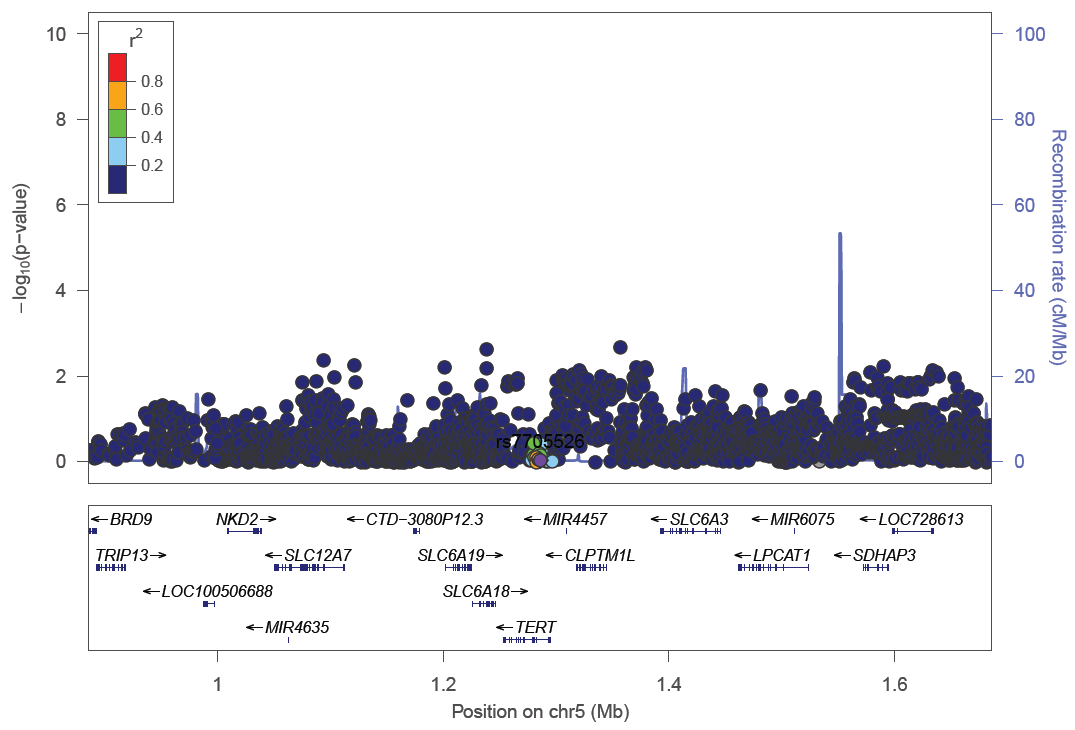


**Fig S5. Regional association plots of conditional analysis of lead SNP in *TERC, TERT, & RTEL1.*** Each conditional analysis is adjusted for the lead SNP in that locus. The results of these analyses suggest there are no clear secondary signals after adjusting for the lead SNP in the significance peak.

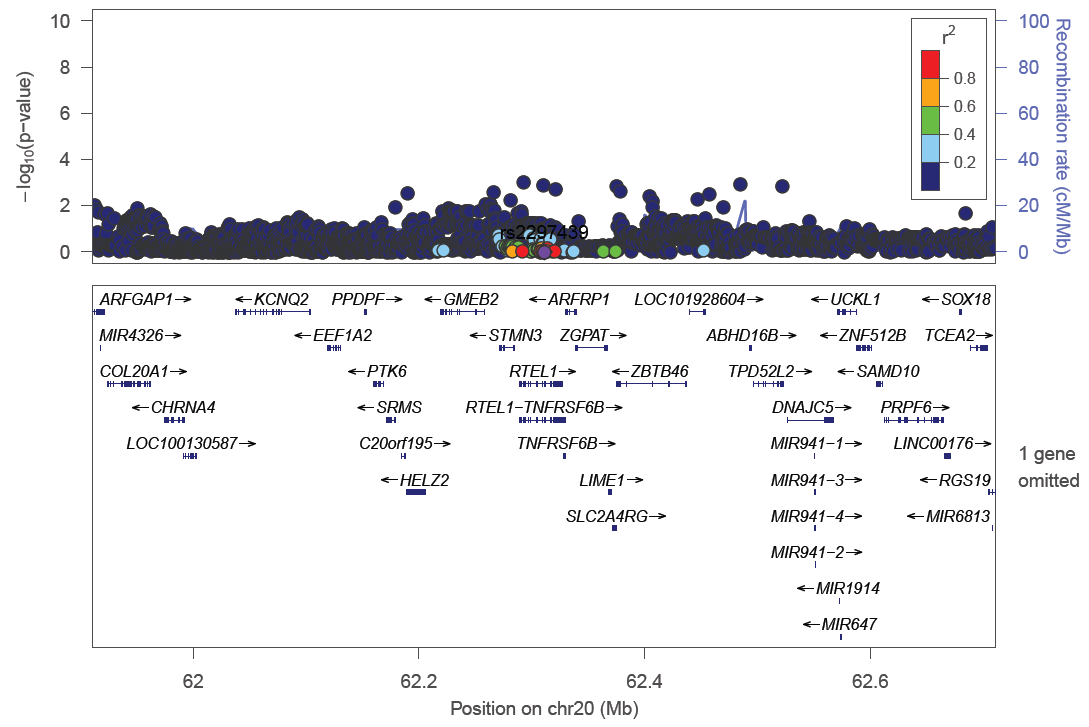
a)



b)

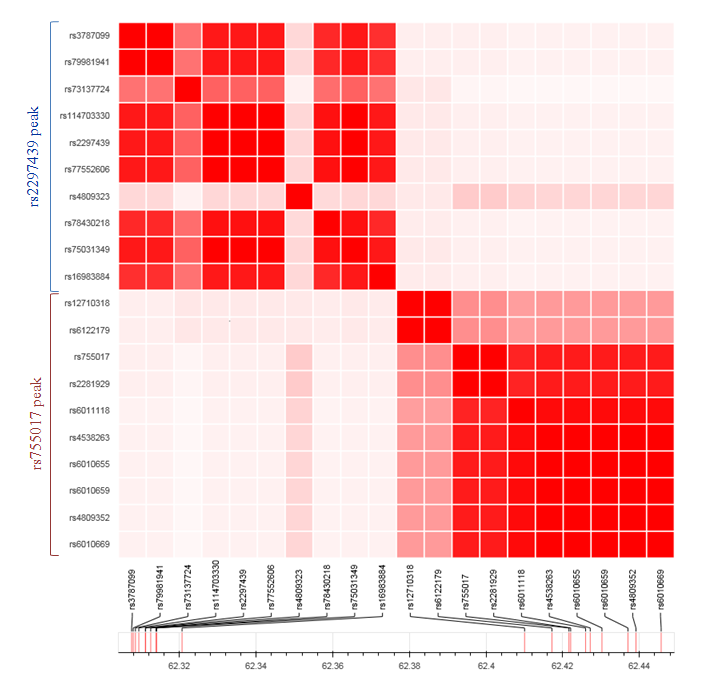


c)

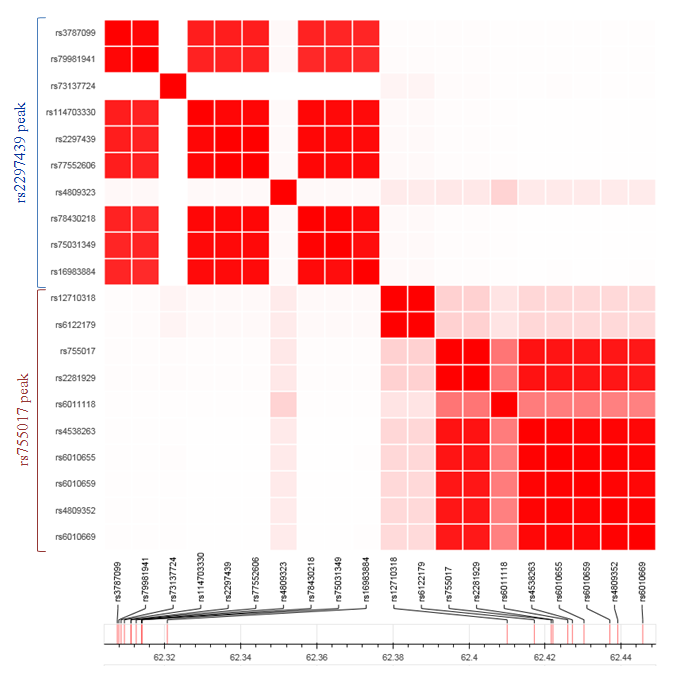


**Fig S6. LD matrix of top 10 SNPs in rs755017 and rs2297439 peaks.** LD patterns for the top ten SNPs in the rs755017 and rs2297439 peaks are shown for (a) Bangladeshi (BEB) and (b) European (EUR) populations, separately. There is high correlation within SNPs of the same peak but no correlation between SNPs of different peaks.

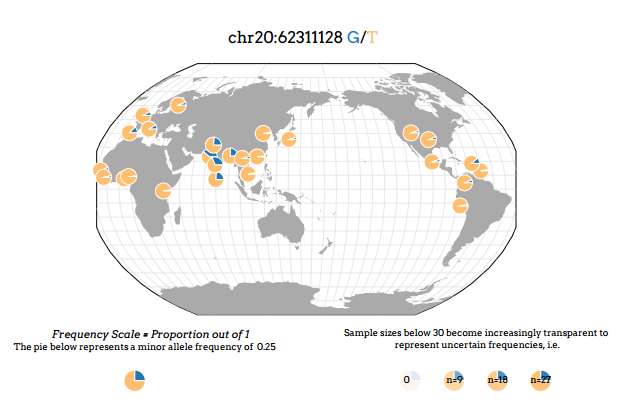
a)

****

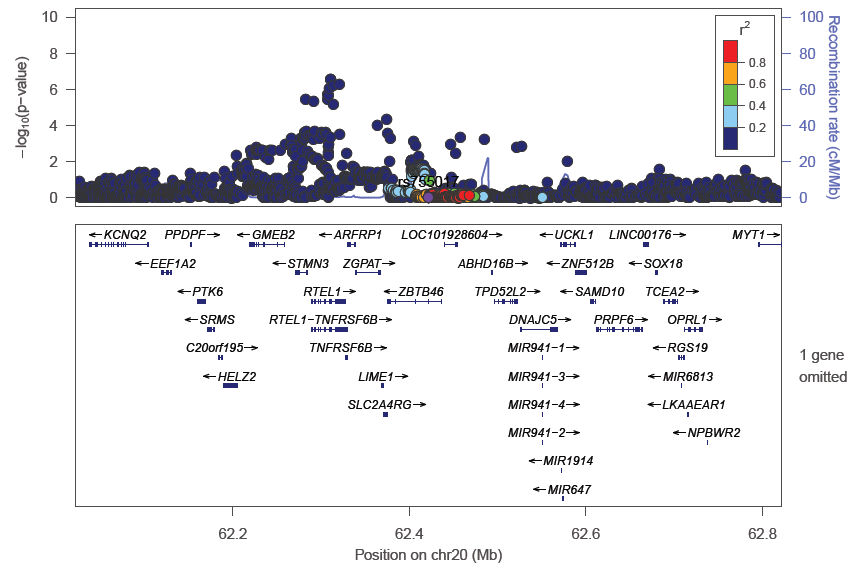
b)



**Fig S7. Worldwide minor allele frequency of rs2297439 in the *RTEL1* locus.** Global minor allele frequency of the SNP rs2297439 for each of the 26 populations determined by the 1000 Genomes Project is represented by a pie. Exact values of each pie are illustrated in S Table 4.

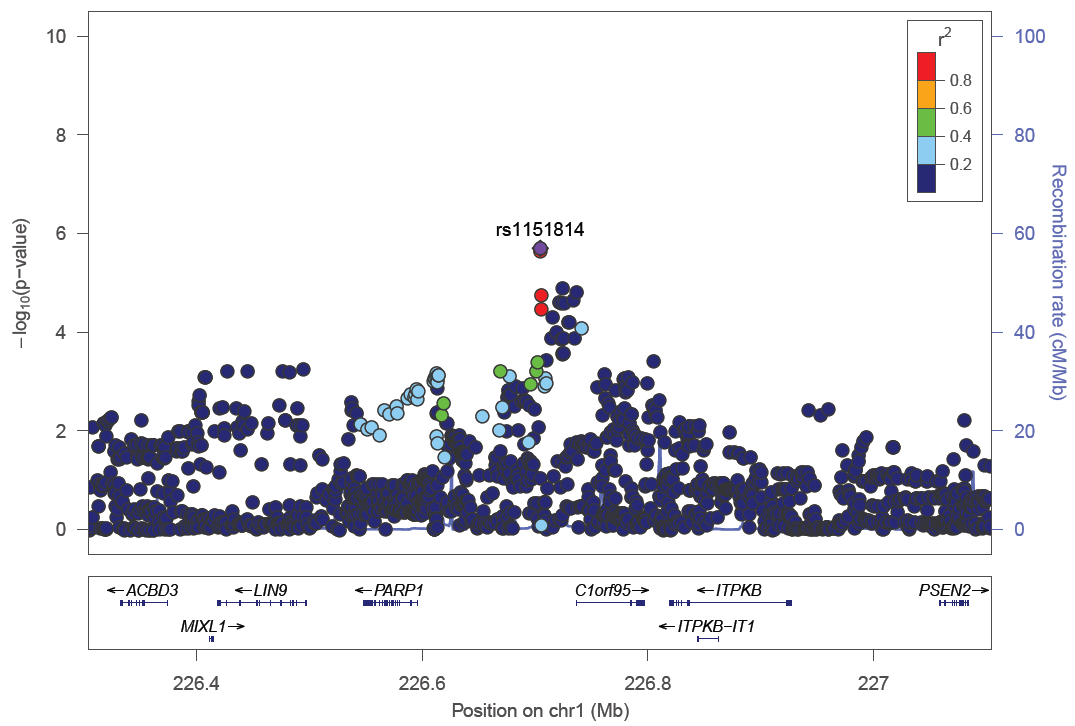


**Fig S8. Regional association plot for the RTEL1 region highlighting previously reported SNP rs755017 and SNPs in LD with rs755071 based on LD in the South Asian (SAS) 1KG reference population.** The –log10(P value) for each SNP is plotted against the base-pair position along each chromosome (Mb). The SNP rs755017 is represented in purple, SNPs are color coded by level of linkage disequilibrium (r2) to the rs755017, and blue peaks represent recombination rates (cM/Mb).

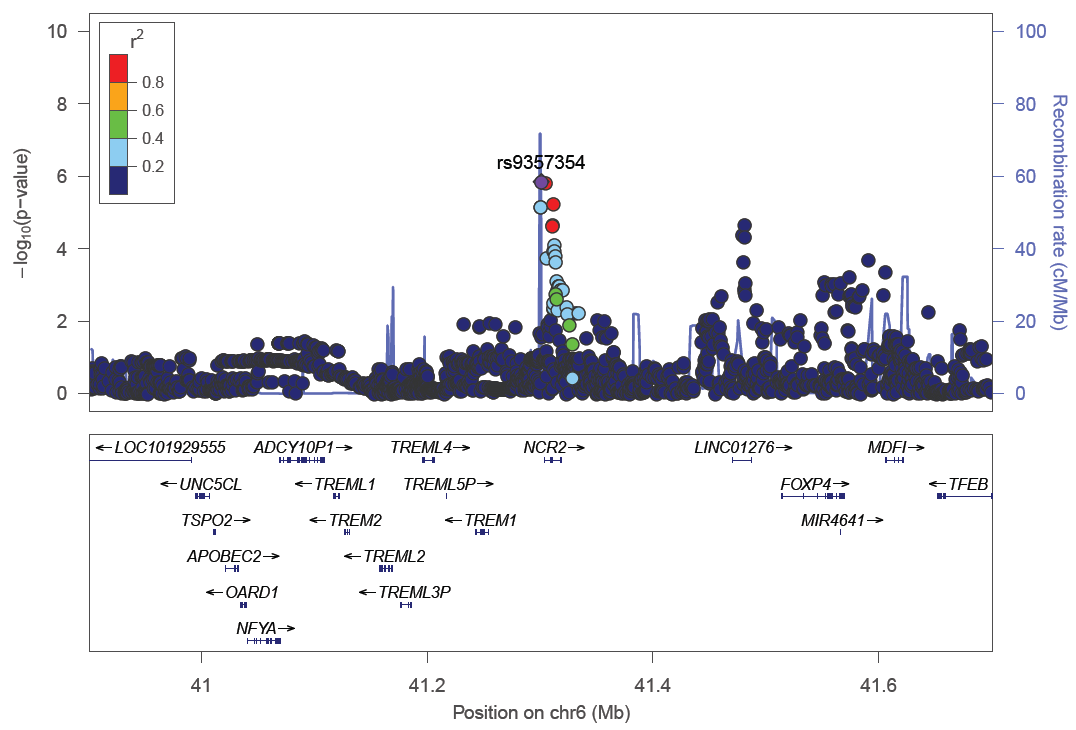


**Fig S9. Regional association plots of novel SNPs (a-d).** These signals show a suggestive association at a threshold *P* of 1x10-5. The –log10(P value) for each SNP is plotted against the base-pair position along each chromosome (Mb). For each locus, the lead SNP is represented in purple, SNPs are color coded by level of linkage disequilibrium (r2) to the lead SNP, blue peaks represent recombination rates (cM/Mb).

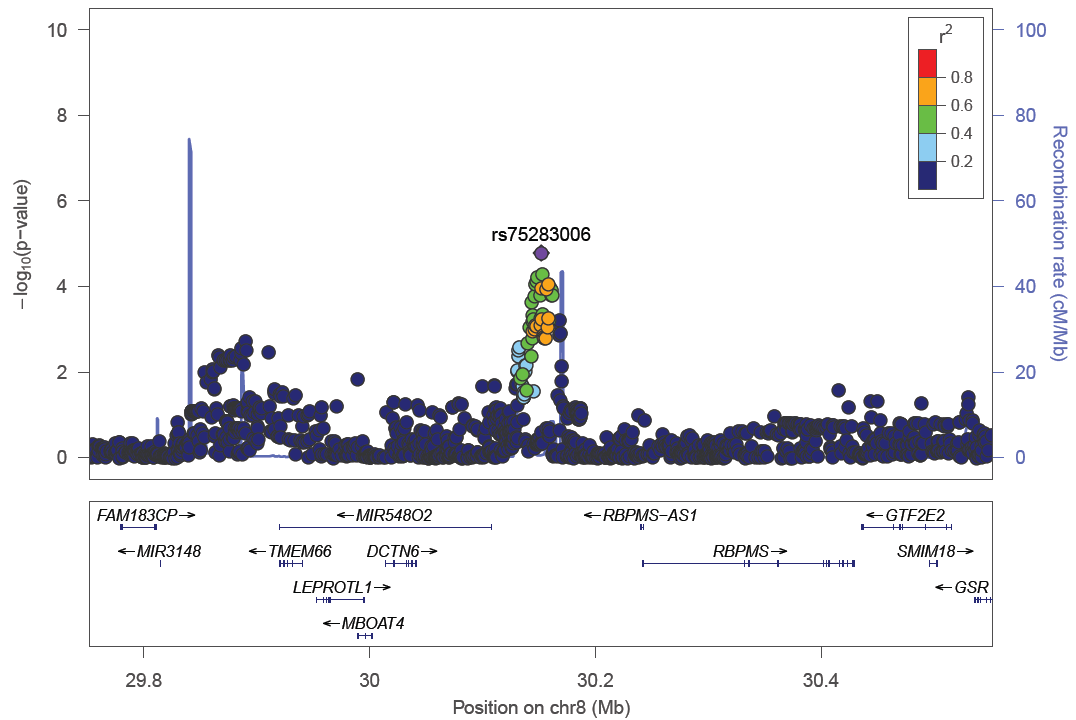
a)



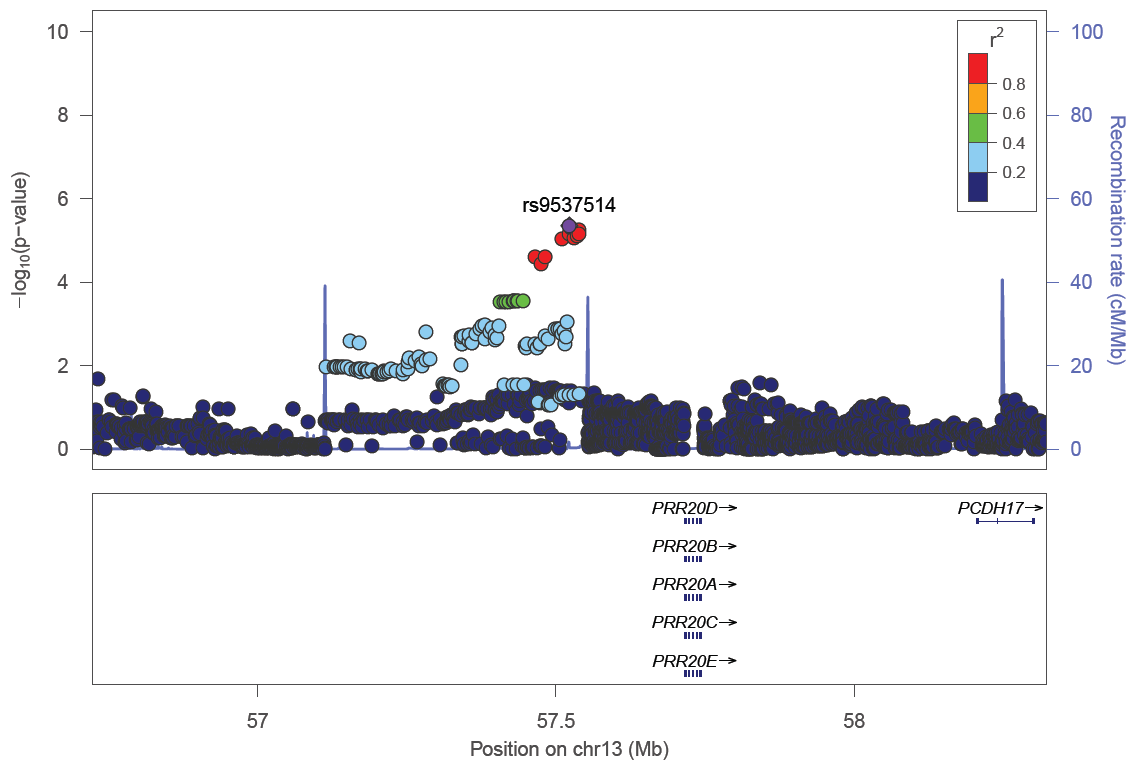
b)



c)



d)



e)

