HHS Public Access

Author manuscript

*Nat Genet*. Author manuscript; available in PMC 2013 July 18.

Published in final edited form as:

*Nat Genet*. 2012 November ; 44(11): 1227–1230. doi:10.1038/ng.2439.

# CSK regulatory polymorphism is associated with systemic lupus erythematosus and influences B cell signaling and activation

**Nataly Manjarrez-Orduño**[1,2], **Emiliano Marasco**[1], **Sharon A. Chung**[3], **Matthew S. Katz**[2], **Jenna F. Kiridly**[2], **Kim R. Simpfendorfer**[2], **Jan Freudenberg**[2], **David H. Ballard**[2], **Emil Nashi**[1,*], **Thomas J. Hopkins**[2], **Deborah S Cunninghame Graham**[4], **Annette T. Lee**[2], **Marieke J. H. Coenen**[5], **Barbara Franke**[5], **Dorine W. Swinkels**[6], **Robert R. Graham**[7], **Robert P. Kimberly**[8], **Patrick M. Gaffney**[9], **Timothy J. Vyse**[4], **Timothy W. Behrens**[7], **Lindsey A. Criswell**[3], **Betty Diamond**[1,§], and **Peter K. Gregersen**[2,§]

[1]Center for Autoimmune and Musculoskeletal Disorders, The Feinstein Institute for Medical Research, North Shore-LIJ, Manhasset, NY. 11030 [2]Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, North Shore-LIJ, Manhasset, NY. 11030 [3]Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California, San Francisco, CA. 94117 [4]Department of Medical and Molecular Genetics, Division of Genetics and Molecular Medicine, School of Medicine, King's College London, London SE1 9RT, United Kingdom [5]Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands [6]Department of Laboratory Medicine, Laboratory of Genetic Endocrine and Metabolic Diseases, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands [7]Deparment of Human Genetics, Genentech, South San Francisco, CA. 94080 [8]Department of Medicine, University of Alabama at Birmingham, AL 35294 [9]Arthritis and Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK. 73104

## Abstract

C-src tyrosine kinase, Csk, physically interacts with the intracellular phosphatase Lyp (*PTPN22*) and can modify the activation state of downstream Src kinases, such as Lyn, in lymphocytes. We identified an association of Csk with systemic lupus erythematosus (SLE) and refined its location to an intronic polymorphism rs34933034 (OR 1.32, $p = 1.04 \times 10^{-9}$). The risk allele is associated with increased *CSK* expression and augments inhibitory phosphorylation of Lyn. In carriers of the risk allele, B cell receptor (BCR)-mediated activation of mature B cells, as well as plasma IgM,



§Jointly directed this work.

*Current address: Division of Clinical Immunology, McGill University Health Centre. Montreal, Canada

**Contributions**

N.M.O, B.D. and P.K.G. designed the study. N.M.O., S.A.C., D.S.G.C., J.F. D.H.B., T.V., L.A.C., and A.L. performed genetic analysis. N.M.O., E.M., J.F.K., M.S.K., K.R.S., T.J.H. performed experiments. E.N. gave the initial insight into Csk. M.J.H.C., B.F., D.W.S., R.G., R.P.K., T.J.V., T.W.B., P.M.G., L.A.C. provided samples. N.M.O., B.D. and P.K.G. analyzed and interpreted the data and prepared the manuscript.

**Competing financial interests**

T.W.B. and R.R.G. are full-time employees of Genentech.

are increased. Moreover, the fraction of transitional B cells is doubled in the cord blood of carriers of the risk allele compared to non-risk haplotypes due to an expansion of the late transitional cells, a stage targeted by selection mechanisms. This suggests that the Lyp-Csk complex increases susceptibility to lupus at multiple maturation and activation points of B cells.

Genome-wide association studies (GWAS) have identified hundreds of common risk variants that implicate multiple signaling pathways in the development of autoimmune diseases[1]. Many of these risk variants may act through lineage- or maturation-specific mechanisms that depend on threshold effects on signaling responses[1]. Several genome-wide associations in systemic lupus have been described in molecules that participate in the BCR signaling pathway[2, 3]. Yet, there is scant and often conflicting information on how these variants modify normal lymphocyte signaling and predispose to autoimmunity.

In lymphocytes, one of the early consequences of engagement of either the T cell or B cell antigen receptor is activation of members of the Src family of tyrosine kinases (SFK). Activation of SFK is regulated by the Lyp-Csk complex. Lyp (the product of the *PTPN22* gene) destabilizes the kinase domain of SFK through tyrosine dephosphorylation, while Csk phosphorylates a C-terminal tyrosine in SFK, leading to a closed, inactive conformation[4]. The Lyp variant R620W disrupts the interaction between Lyp and the SH3 domain of Csk and has been strongly associated with many autoimmune disorders[5]. Consequently, we searched for evidence of association with SLE at the *CSK* locus in a prior GWAS of SLE in subjects of European ancestry[2]. We observed that several single-nucleotide polymorphisms (SNPs) that tag a *CSK* haplotype “B” have a nominal association with SLE in this previously published dataset, well below genome-wide levels of statistical significance (OR ~ 1.15, uncorrected p-values 0.004–0.007, 1311 cases and 3340 controls)[2]. We therefore sequenced the *CSK* exons and five highly conserved intronic regions in a discovery cohort of 24 SLE-affected subjects of European ethnicity, homozygous for the “B” haplotype (Figure 1e). Of the four polymorphisms present in the discovery dataset (Supplementary Table 1), the minor allele (A) of the intronic variant rs34933034 (G > A) was present in 34 of the 48 chromosomes bearing the “B” haplotype, and was therefore of particular interest.

To assess the relationship of the A allele of the rs34933034 variant to the *B* haplotype and explore its association with risk for SLE, we genotyped 3769 SLE cases and 3404 controls of European ancestry derived from eleven cohorts and analyzed the data in three groups (Methods, Table 1 and Supplementary Table 2). The results provide convincing evidence for an association between SLE and the A allele of the variant rs34933034: allelic O.R. 1.32, *p*-value $1.04 \times 10^{-9}$. Linkage disequilibrium analysis shows that the rs34933034 A allele nearly exclusively exists on the B haplotype ($D' = 0.95$), and has a relatively low overall correlation ($r^2 < 0.5$) with any of the other SNPs that define this haplotype, as summarized in Figure 1. Two SNPs that are present in the *CSK* B haplotype have recently been associated with several autoimmune diseases[6, 7]. Consequently, we imputed SNPs with data from the 1000 Genomes project (Supplementary Table 3) and performed conditional analysis for rs34933034, showing that the associations seen for variants that tag the B haplotype are secondary to the rs34933034 A allele. For example, the rs8033381 marker tagging the common B haplotype shows an association (P-value: $9 \times 10^{-8}$) that is eliminated

after conditioning on rs34933034 (P-value: 0.19), (Figure 1, arrowhead). In contrast, conditioning on rs8033381 shows an association in rs34933034 (P-value: $2.49 \times 10^{-4}$). Of note, we did not observe evidence of genetic interaction between rs34933034 and rs2476601 (*PTPN22*, R620W).

The ENCODE data reports the presence of DNA sensitive sites around rs34933034, suggesting that this variant lies within an intronic regulatory region of *CSK*. Likewise, it has been reported that *CSK* expression is highest in cells of the immune system, particularly in B cell subsets[8]. This was confirmed by evaluation of the pattern of *CSK* expression in different subsets of peripheral blood mononuclear cells (PBMC) of healthy individuals, including the transitional, naïve and memory subsets of B cells (all subpopulations were recovered in at least 5 subjects), as well as CD4+ and CD8+ T cells and monocytes (n = 3) (Fig 2a). As previously reported, *CSK* expression was highest in B cells (Fig 2b), and was inversely correlated with B cell maturity (n = 5, p = 0.03). Thus, transitional B cells that have recently emerged from the bone marrow have higher *CSK* expression than mature naïve B cells, which in turn have higher expression than memory B cells. We sorted naïve B cells from 29 healthy donors of either rs3493034 genotype, all of whom were homozygous for the B haplotype. As can be seen in Figure 2c higher levels of *CSK* transcripts in naïve B cells are significantly associated with the rs3493034 A allele.

Next, we examined whether increased *CSK* expression affects B cell function. It has been reported that in resting murine T cells, Csk-mediated C-terminal phosphorylation maintains SFKs in an inactive conformation[9]. In B cells, Lyn is the most abundant SFK; therefore, we hypothesized that elevated Csk would increase basal phosphorylation at the C-terminal tyrosine of Lyn in B cells. We analyzed Tyr508 phosphorylation by flow cytometry in resting naïve B cells (CD20+, CD27−) from 27 healthy adults. As shown in Figure 3b and Supplementary Figure 1, naïve B cells from subjects carrying the *CSK* risk allele exhibit higher Tyr508 phosphorylation of Lyn than non-risk allele carriers.

Interestingly, and in contrast to the activating role of Lck in T cells, Lyn has been reported to mediate negative regulation of BCR signaling[10] Therefore, we measured calcium mobilization in naïve B cells of 11 donors homozygous for the B haplotype after BCR cross-linking in B cells. As expected, naïve B cells homozygous for the A allele exhibit enhanced calcium mobilization triggered by BCR cross-linking (Figure 3c and 3d; no differences in signaling capacity related to *CSK* genotype were observed upon stimulation with ionomycin – data not shown). This result is consistent with reports that Lyn has negative regulatory effects on B cell activation[10]. Moreover, healthy subjects who carry the *CSK* risk allele also have more IgM in plasma (n = 44, Figure 3e), compared to subjects with the non-risk allele, consistent with enhanced activation of mature B cells.

Given that *CSK* expression is highest in transitional B cells, we investigated how the risk allele would affect early B cell differentiation. To study B cell maturation in the most pristine situation, we analyzed B cell subpopulations in 27 samples of umbilical cord blood and found that newborn subjects homozygous for the *CSK* risk allele have double the frequency of transitional cells ($CD38^{hi}$, $CD10^{hi}$) compared to levels observed in the cord blood of non-risk subjects (Figure 4a and 4b). Further dissection of this B cell compartment

shows that individuals with the risk allele have more “late” transitional cells than their non-risk counterparts; these cells are also characterized by high surface expression of CD21, and have acquired surface IgD while still retaining a high expression of CD38 and CD10 (Figure 4c, and Supplementary Figure 2).

Our results clearly support the hypothesis that *CSK* expression levels can modify normal B cell biology. We have demonstrated that *CSK* expression is highest at the earliest stages of B cell maturation (transitional B cells), and that normal individuals who carry the risk allele exhibit expansion of the late transitional B cell population. This might reflect an impairment of a tolerance checkpoint or an enhanced positive selection of late transitional B cells. Either model leads to a prediction that there would be an increased number of autoreactive B cells in the peripheral repertoire of healthy carriers of the *CSK* risk allele, consistent with an increased likelihood of autoantibody production.

The Lyp-Csk complex acts by setting thresholds for BCR signaling, but the specific effects may vary according to the B cell developmental stage and the particular SFKs that are functional at that stage. In B cells that carry the *PTPN22* risk allele, both reduced[11] and augmented[12, 13] responses to BCR cross-linking have been reported, as well as higher auto-reactivity in the early B cell repertoire of healthy subjects carrying the risk allele[14]. The results reported here show that levels of Csk differ at different stages of development as the signaling apparatus undergoes maturation. We note that the expression and activity of Lyp (*PTPN22*) in human transitional B cells have not yet been studied in detail. Our results clearly indicate that any investigation of *PTPN22* effects on signaling thresholds in transitional B cells must account for the *CSK* genotype, as we have done here for *PTPN22*. Moreover, two SFKs, *LYN* and *BLK*, have been associated with risk for SLE[2, 15]. Research from our group suggests that the *BLK* risk allele also predominantly influences expression differences in the early stages of B cell development[16], although the functional consequences of differences in Blk expression have not yet been determined. This emphasizes the need for further exploration of the role of signaling molecules in immune cells at different stages of development in order to understand the mechanisms that lead to autoimmune disease and thereby inform the rational development of targeted approaches to therapy.

## Methods

### Research subjects and specimens

DNA from cohorts of SLE-affected subjects was obtained from the UCSF Lupus Genetics Project[18], the Multiple Autoimmune Diseases Genetics Consortium (MADGC)[19], the University of Minnesota[20], the UK SLE Study[21], ABCoN[2], the University of Alabama[3] and OMRF[3] as previously reported. Control DNA samples were taken from the New York Cancer Project collection[22, 23], the Nijmegen Biomedical study (NBS, a population-based cohort of self-reported, randomly selected inhabitants of Nijmegen, the Netherlands)[24] and the Genotype and Phenotype Registry at The Feinstein Institute for Medical Research (FIMR) and at the University of Alabama (www.gapregistry.org). All subjects gave written informed consent, and Institutional Review Boards reviewed the protocols at their host institutions. SLE diagnosis was established according to ACR guidelines[2, 3, 18–20]. A

detailed explanation of the number of samples per dataset is shown in Supplementary Table 2.

All of the genotype-matched blood samples of healthy controls were obtained from volunteers belonging to the GaP registry; carriers of the *PTPN22* risk allele (rs2476601, T) were excluded from the functional studies. Cord blood samples deemed not suitable for banking and leukocyte units were obtained, de-identified, from the Long Island Blood Bank. The FIMR IRB reviewed and approved all of the protocols, consent was waived for de-identified cord blood and leukocyte units.

## Genotyping

*CSK* sequencing for the 24 SLE samples of the discovery cohort was performed by Polymorphic DNA Technologies (Alameda, California). Genotyping of datasets was performed sequentially in three analytic groups. European ancestry was determined by the ancestry informative markers within each analytic group. Supplementary Table 2 describes each of the three groups of datasets. The *CSK* genotyping for groups I and II was performed by pyrosequencing (Supplementary table 4). Genotyping of *CSK* rs34933034 for the third analytic group was performed by qPCR with the TaqMan assay C__60143137_10 in a ViiA7 machine (Applied Biosciences) in samples from the NBS and OMRF.

SNPs informative for continental origin and European population diversity were obtained from genome-wide typing for each of the cohorts, and used to limit the study to subjects of European ancestry. Matching of cases and controls by principal component analysis[25], as well as association analysis for both allelic and additive genetic models was performed using SNP and Variation Suite (Golden Helix). The meta-analysis p-value for the additive genetic model with correction for principal component analysis was performed using the sample size based analytical strategy reported previously[26].

## Imputation and conditional analysis

For each group of datasets, we obtained genotype information on markers situated 300 Kbp either side of *CSK*. We used 1000 Genomes project data (Phase 1, version 3, March 2012) as reference to impute variants across a fragment of 120 Kbp (chr15:75,024,425–75,145,539, hg19), with *CSK* in the center. Imputation was performed using IMPUTE2.2[17]. Only imputed polymorphisms with probabilities above 90% were used in subsequent analyses, and variants that were not called in more than 90% of the samples per dataset were not used (Supplementary Table 3). Conditional analysis of the variants that passed quality control was performed with SNPtest2[17], using an additive model of association.

## Sample processing, flow cytometry and cell sorting

Mononuclear cells from peripheral blood were recovered by layering over Ficoll-Paque (GE HealthCare). For cell sorting, the cells were stained for ten minutes with an antibody mix containing one of the following antibody cocktails: CD8 (FITC, 551347), CD20 (PE, 555623), CD4 (PerCP, 340671), CD3 (APC, 340440) and CD14 (Pacific Blue, 558121), all from BDBiosciences, for sorting of monocytes and T cell subpopulations (3 samples). For B cells: anti IgD (FITC, 555778), Pacific blue-conjugated CD3 (558124), CD14, CD16

(558122), CD19 (APC-Cy7, 557791) all from BDBiosciences; CD27 (PE, MHCD2704) and CD38 (PE-TR, MHCD3817) from Invitrogen; and CD10 (PE-Cy7, 312214) from Biolegend. Cells were sorted in a FACSAria instrument (BDBiosciences) and after exclusion of doublets by forward and side scatter area and width parameters, gates were set for monocytes ($CD3^-$, $CD20^-$, $CD14^+$), $CD4^+$ and $CD8^+$ cells (after a $CD20^-$ and $CD3^+$ gate). B cells were defined as $CD19^+$, $CD3/14/16^-$ and subsequently gated for B cell subpopulations: memory ($CD19^+$, $CD27^+$) and preimmune ($IgD^+$, $CD27^-$). To divide the preimmune B cell populations, the upper limit of CD10 fluorescence in T cells (negative) was used to define the lower limits of CD10 positivity for transitional B cells ($CD38^{hi}$, $CD10^{hi}$) and naïve cells ($CD38^{dim}$, $CD10^{low}$). Thirteen of those cord blood samples were stained with CD21 PE to confirm the gates for early and late transitional B cells as defined previously[27]. Statistical analyses of B cell subpopulations were performed using the Kruskal-Wallis one-way analysis of variance.

## Expression analysis

*CSK* expression analysis of peripheral blood cells was performed on sorted cells. RNA was extracted using a Micro RNeasy Isolation Kit (Qiagen). cDNA was synthesized by linear retrotranscription with iScript (Bio-Rad). cDNA was subsequently used for qPCR of *CSK* (Hs01062585_m1, Applied Biosystems) and *POLR2a* (Hs01108291_m1, Applied Biosystems) in a Lightcycler 480 II (Roche). Normalized expression was calculated according to the modified Livak method ΔCt: $2^{(\mathrm{Ct}\ PolR2a\ -\ \mathrm{Ct}\ CSK)}$ [28]. All of the expression assays were performed in duplicate. The analysis of *CSK* expression across B cell subpopulations was performed in a Kruskal-Wallis test; the analysis of *CSK* expression across genotypes was performed using a one way ANOVA.

## Cell signaling

Cell signaling assays were performed with PBMC of healthy subjects homozygous for the B haplotype. Cells were washed with PBS and allowed to rest for 1 h at 37°C in RPMI containing 2% FCS. For pLyn, cells were fixed, washed and surface stained with CD20, CD24, CD38, CD27, and CD3. After a permeabilization step with BDPerm, cells were stained with the anti-pLck (Y505) antibody (BDPhosflow, 557879), which cross-reacts with pLyn. Median Fluorescence Intensity (MFI) in naïve B cells was obtained by analysis of the data with FlowJo software. No differences were observed if the data was analyzed solely on the basis of MFI or MFI minus isotype control. The data was normalized to the mean of the MFI per experiment. Given the correction for multiple testing, pairwise comparison was considered significant at the $p < 0.0166$ level.

For calcium flux, after resting for 1 h, the cells were loaded with Indo-1, and subsequently labeled with anti CD2, CD14, CD16, CD20, CD27, CD38 and CD10 as before. Data was collected in a BD LSRII machine as the ratio of $Ca^{2+}$-bound Indo (405 nm) over free Indo (450nm) for one minute. Cells were then stimulated with 10 μg/ml $F(ab')_2$ goat anti-human IgM (Southern Biotech) and data were recorded for a further 5 minutes. All of the samples were subsequently activated with ionomycin as a control. Data was analyzed with FlowJo software to calculate the area under the activation curve for the first 90 seconds and 4 minutes.

SAS version 9.2 was used to perform Hierarchical Linear Mixed Models (HLMM) to analyze the normalized values of the area under the curve. In order to adjust for day-to-day variation in experimental conditions, we "blocked" on day and considered subjects nested within days. Furthermore, the day of the experiment was considered a random effect. A result was considered statistically significant at the $p < 0.05$ level.

### ELISA

96 well plates were pre-coated with 10 μg/mL of anti-human IgM antibody (Southern Biotech) and blocked with PBS (1% BSA). Samples were plated and after one hour of incubation wells were washed and incubated with anti human IgM conjugated to alkaline phosphatase, and later developed with p-nitrophenyl disodium (Sigma) in carbonate buffer. The optical density at 405nm was measured. Samples were run in duplicate and IgM was quantified using an IgM standard curve (Sigma Aldrich).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors thank the volunteers who participated in this study; the GaP registry (www.gapregistry.org), Mss. Keogh, DeFranco, Mason, Metz and the Biorepository at FIMR for recruiting subjects and collecting samples; H. Borrero for technical assistance; the Biostatistics Unit of the FIMR and M. Akerman for assistance. Work supported by: NIH RC2AR059092, The Alliance for Lupus Research, the Kirkland Scholar Award, and NIH/NCRR 5 M01 RR-00079 (L.A.C.). The authors have no conflicting financial interests. Nijmegen Biomedical Study's principal investigators are L.A.L.M. Kiemeney, M. den Heijer, A.L.M. Verbeek, D.W. Swinkels and B. Franke.

## References

1. Cho JH, Gregersen PK. Genomics and the Multifactorial Nature of Human Autoimmune Disease. New England Journal of Medicine. 2011; 365:1612–1623. [PubMed: 22029983]
2. Liston A, Lesage S, Gray DHD, Boyd RL, Goodnow CC. Genetic lesions in T-cell tolerance and thresholds for autoimmunity. Immunological Reviews. 2005; 204:87–101. [PubMed: 15790352]
3. Hom G, et al. Association of Systemic Lupus Erythematosus with C8orf13-BLK and ITGAM-ITGAX. NEJM. 2008; 358:900–909. [PubMed: 18204098]
4. Nath SK, et al. A nonsynonymous functional variant in integrin-[alpha]M (encoded by ITGAM) is associated with systemic lupus erythematosus. Nat Genet. 2008; 40:152–154. [PubMed: 18204448]
5. Levinson NM, Seeliger MA, Cole PA, Kuriyan J. Structural Basis for the Recognition of c-Src by Its Inactivator Csk. Cell. 2008; 134:124–134. [PubMed: 18614016]
6. Gregersen PK, Lee HS, Batliwalla F, Begovich AB. PTPN22: Setting thresholds for autoimmunity. Seminars in Immunology. 2006; 18:214–223. [PubMed: 16731003]
7. Trynka G, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet. 2011; 43:1193–1201. [PubMed: 22057235]
8. Martin JE, et al. Identification of CSK as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up. Human Molecular Genetics. 2012
9. Su AI, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:6062–6067. [PubMed: 15075390]
10. Zikherman J, et al. CD45-Csk Phosphatase-Kinase Titration Uncouples Basal and Inducible T Cell Receptor Signaling during Thymic Development. Immunity. 2010; 32:342–354. [PubMed: 20346773]

11. Hasegawa M, et al. A CD19-dependent signaling pathway regulates autoimmunity in Lyn-deficient mice. J Immunol. 2001; 167:2469–2478. [PubMed: 11509585]
12. Arechiga AF, et al. Cutting Edge: The PTPN22 Allelic Variant Associated with Autoimmunity Impairs B Cell Signaling. J Immunol. 2009; 182:3343–3347. [PubMed: 19265110]
13. Zikherman J, et al. PTPN22 deficiency cooperates with the CD45 E613R allele to break tolerance on a non-autoimmune background. J Immunol. 2009; 182:4093–4106. [PubMed: 19299707]
14. Zhang J, et al. The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness. Nat Genet. 2011; 43:902–907. [PubMed: 21841778]
15. Menard L, et al. The PTPN22 allele encoding an R620W variant interferes with the removal of developing autoreactive B cells in humans. J Clin Inv. 2011; 121
16. Lu R, et al. Genetic associations of LYN with systemic lupus erythematosus. Genes Immun. 2009; 10:397–403. [PubMed: 19369946]
17. Simpfendorfer KR, et al. The autoimmunity-associated BLK haplotype exhibits cis-regulatory effects on mRNA and protein expression that are prominently observed in B cells early in development. Human Molecular Genetics. 2012
18. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007; 39:906–913. [PubMed: 17572673]

## References for Online methods

19. Thorburn CM, et al. Association of PDCD1 genetic variation with risk and clinical manifestations of systemic lupus erythematosus in a multiethnic cohort. Genes and immunity. 2007; 8:279–287. [PubMed: 17344889]
20. Criswell LA, et al. Analysis of Families in the Multiple Autoimmune Disease Genetics Consortium (MADGC) Collection: the PTPN22 620W Allele Associates with Multiple Autoimmune Phenotypes. Am J Hum Genet. 2005; 76:561–571. [PubMed: 15719322]
21. Graham RR, et al. A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. Nat Genet. 2006; 38:550–555. [PubMed: 16642019]
22. Cunninghame Graham DS, et al. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with Systemic Lupus Erythematosus. PLoS Genet. 2011; 7:e1002341. [PubMed: 22046141]
23. Plenge RM, et al. TRAF1-C5 as a Risk Locus for Rheumatoid Arthritis — A Genomewide Study. NEJM. 2007; 357:1199–1209. [PubMed: 17804836]
24. Gregersen PK, et al. REL, encoding a member of the NF-[kappa]B family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat Genet. 2009; 41:820–823. [PubMed: 19503088]
25. Wetzels JFM, Kiemeney LALM, Swinkels DW, Willems HL, Heijer MD. Age- and gender-specific reference values of estimated GFR in Caucasians: The Nijmegen Biomedical Study. Kidney Int. 2007; 72:632–637. [PubMed: 17568781]
26. Tian C, et al. European population genetic substructure: further definition of ancestry informative markers for distinguishing among diverse European ethnic groups. Mol Med. 2009; 15:371–383. [PubMed: 19707526]
27. Willer CJ, Li Y, Abecasis GAR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]
28. Suryani S, et al. Differential expression of CD21 identifies developmentally and functionally distinct subsets of human transitional B cells. Blood. 2010; 115:519–529. [PubMed: 19965666]
29. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-ΔΔCT Method. Methods. 2001; 25:402–408. [PubMed: 11846609]

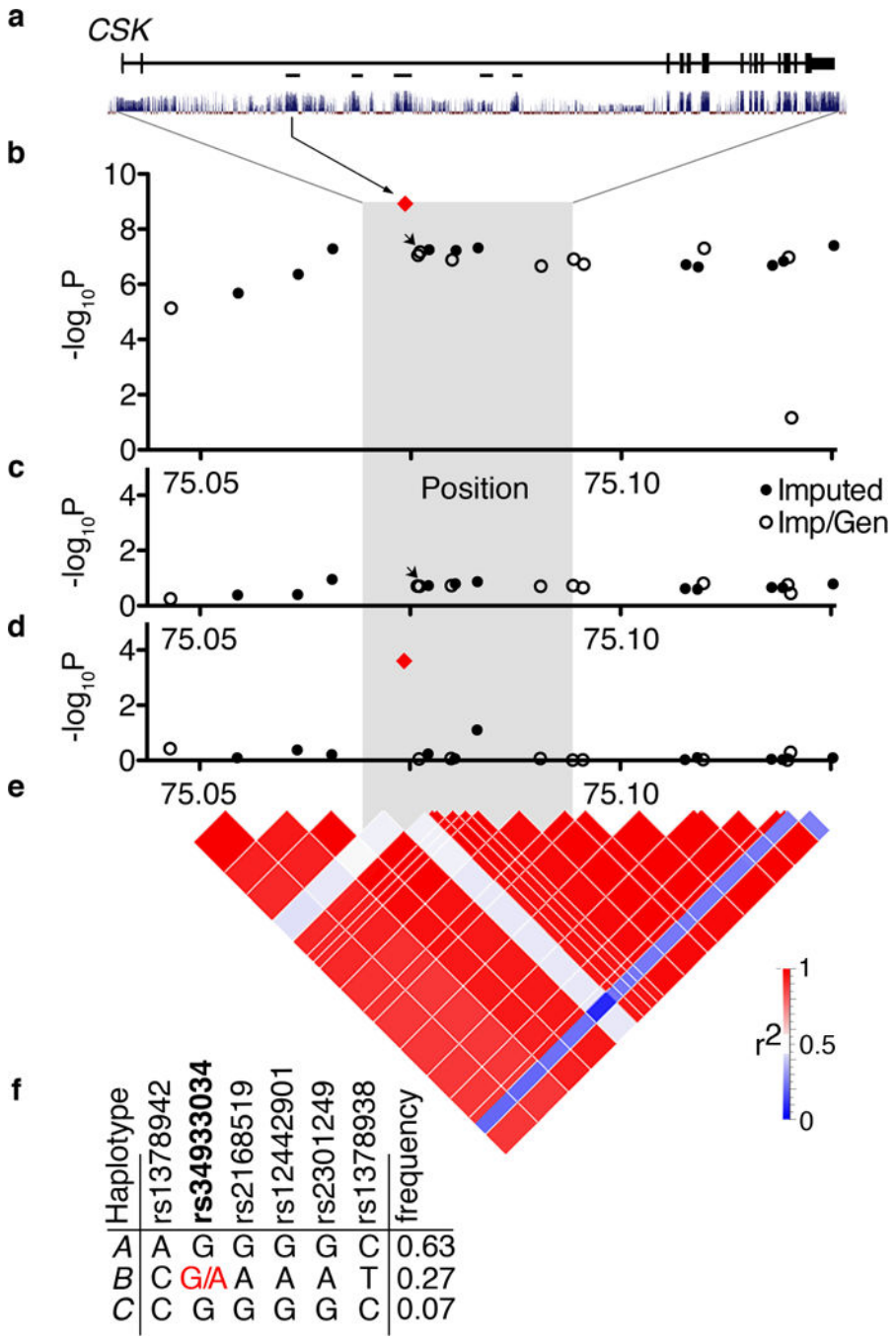


**Figure 1.**

Genetic structure of the *CSK* locus. (a) *CSK* spans 14 exons indicated by the vertical boxes. The plot shows the sequence homology among mammals, with highly conserved regions shown as peaks (taken from genome.ucsc.edu). The horizontal bars highlight the conserved areas that were selected for sequencing. The vertical tick-mark below the first conserved region that was sequenced shows the position of the variant rs34933034 ( ◆). (b) Association of SNP markers with SLE; a region of ~80,000bp surrounding the *CSK* locus is shown. SNPs were imputed using 1000 Genomes data. The gray area highlights the position

of the *CSK* gene. Many markers that have an intermediate association with SLE tag the B haplotype. The markers shown were either imputed (●), or genotyped in some samples and imputed in others (○). The arrowhead points to rs8033381, one of many that tag the B haplotype, (c) after conditioning on rs34933034, all association signals within the *CSK* gene are eliminated. (d) conditioning on rs8033381 still shows an association signal with rs34933034. (e) An LD heatmap ($r^2$) of the SLE-associated markers in the *CSK* gene reveals a single major haplotype block in subjects of European ancestry. Note the lower $r^2$ for rs3493303 with this block ($r^2 \sim 0.5$). (f) The major haplotypes for the *CSK* locus and their frequencies in European controls are shown.

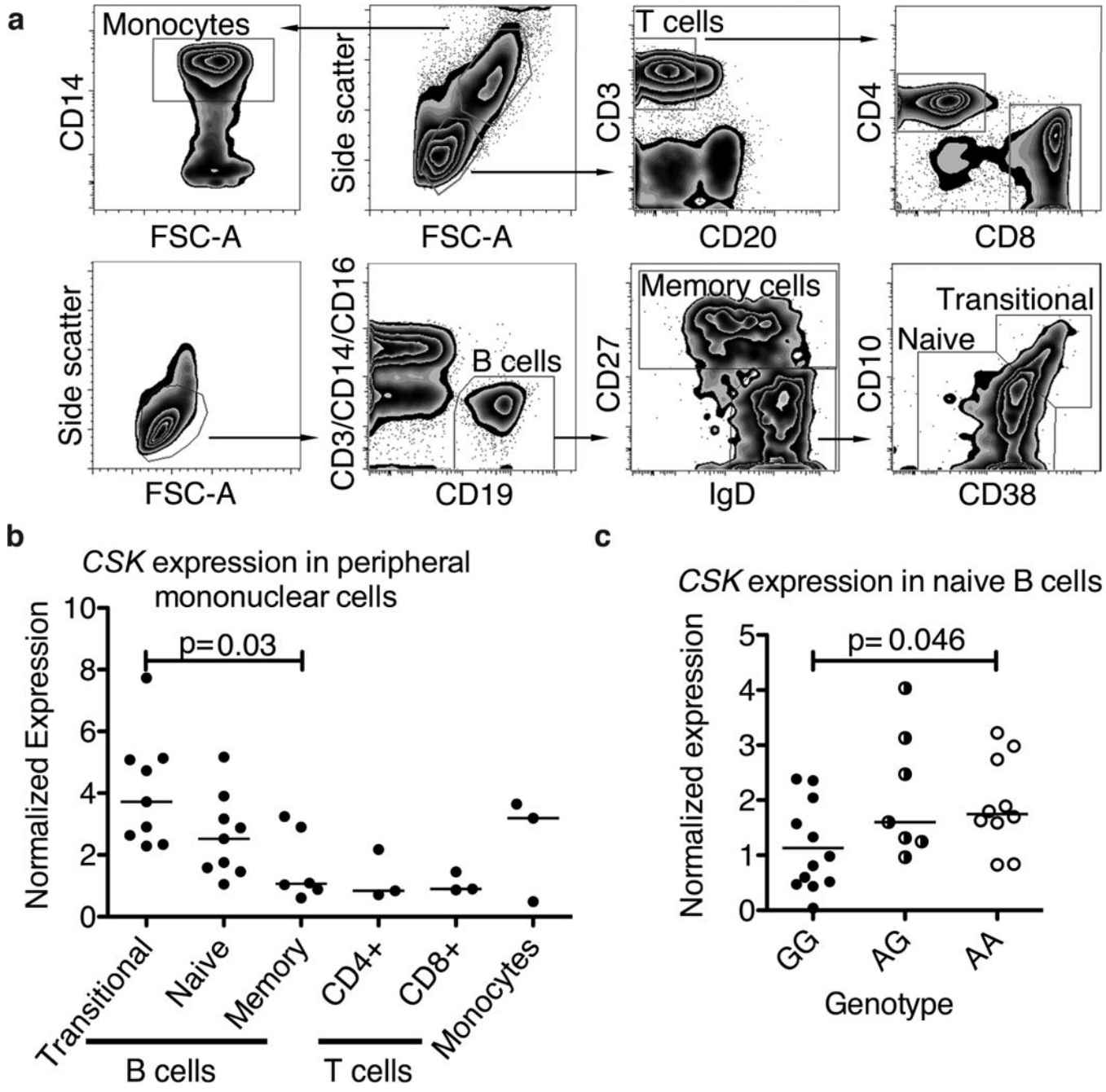


**Figure 2.**

*CSK* expression varies in lymphocyte subsets and is associated with *CSK* genotype. (a) Gating strategy employed to distinguish monocytes and T cells (upper panel) and B cell subpopulations (bottom panel). (b) *CSK* expression decreases in peripheral B cells as they mature from transitional to memory cells (p = 0.030, Kruskal-Wallis test). *CSK* expression in peripheral T cells is low (p = 0.0047, Kruskal-Wallis test between the five cell subpopulations). (c) In naïve B cells, the rs34933034 risk allele (A) is associated with increased *CSK* expression (29 subjects analyzed, Kruskal-Wallis test). Cell subpopulations were isolated from peripheral mononuclear cells by cell sorting as shown in (a). Expression

analysis was performed by qPCR with cDNA synthesized from RNA of isolated subpopulations from mononuclear cells of non-genotyped blood donors (b) or haplotype-matched GaP subjects (c), the horizontal bar marks the median of the values.

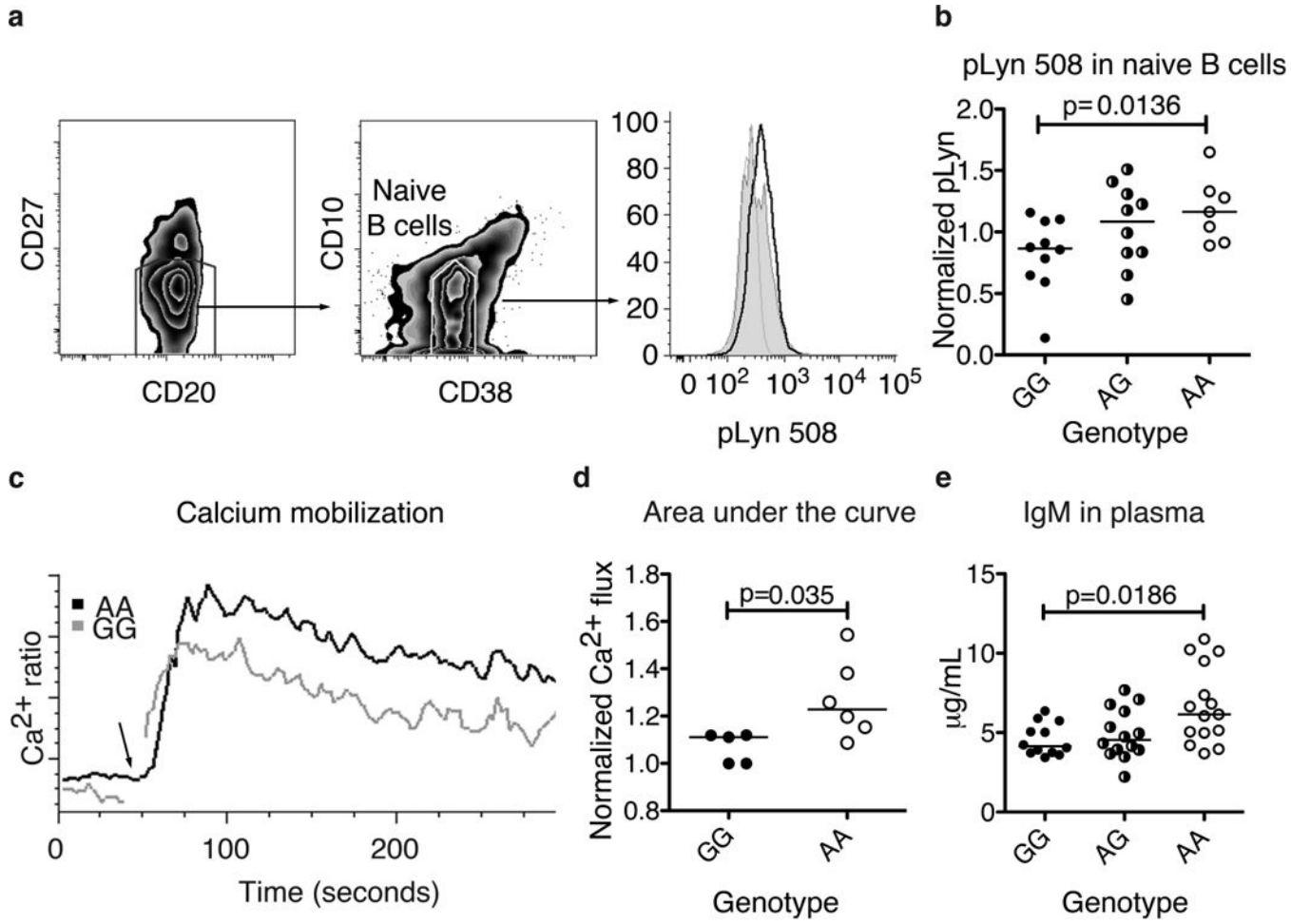


**Figure 3.**

The *CSK* risk allele is associated with increased phospho-Lyn$_{508}$ and enhanced activation of mature B cells. (a) Phosphorylated tyrosine 508 was measured by flow cytometry in naïve B cells of haplotype-matched GaP donors homozygous for either allele of the rs34933034 variant; the profiles of a risk (bold) and non risk (shadowed) subject are shown, the thin line shows the isotype control; (b) pLyn$_{508}$ in carriers of the CSK risk allele. pLyn$_{508}$ was normalized to each day average (n = 27, p = 0.0136, Mann-Whitney test, see methods for correction after multiple comparison, and supplementary Figure 1, Kruskal-Wallis test P = 0.0569). (c) A representative experiment of calcium mobilization. Basal calcium was read for 40 seconds, before activation with anti-IgM (Fab′)$_2$. After at least six minutes of recording, ionomycin was used as a positive control for calcium mobilization and showed no difference between genotypes (data not shown). (d) Data from eleven different “B” haplotype donors was normalized to one non-risk (GG) subject for comparison on each of four separate days of experiments. The area under the curve for the first 90 seconds (similar results at four minutes) is shown, (see methods for statistical analysis). (e) IgM plasma levels in 42 subjects homozygous for the B haplotype, Kruskal-Wallis test. In every graph, the horizontal line marks the median.

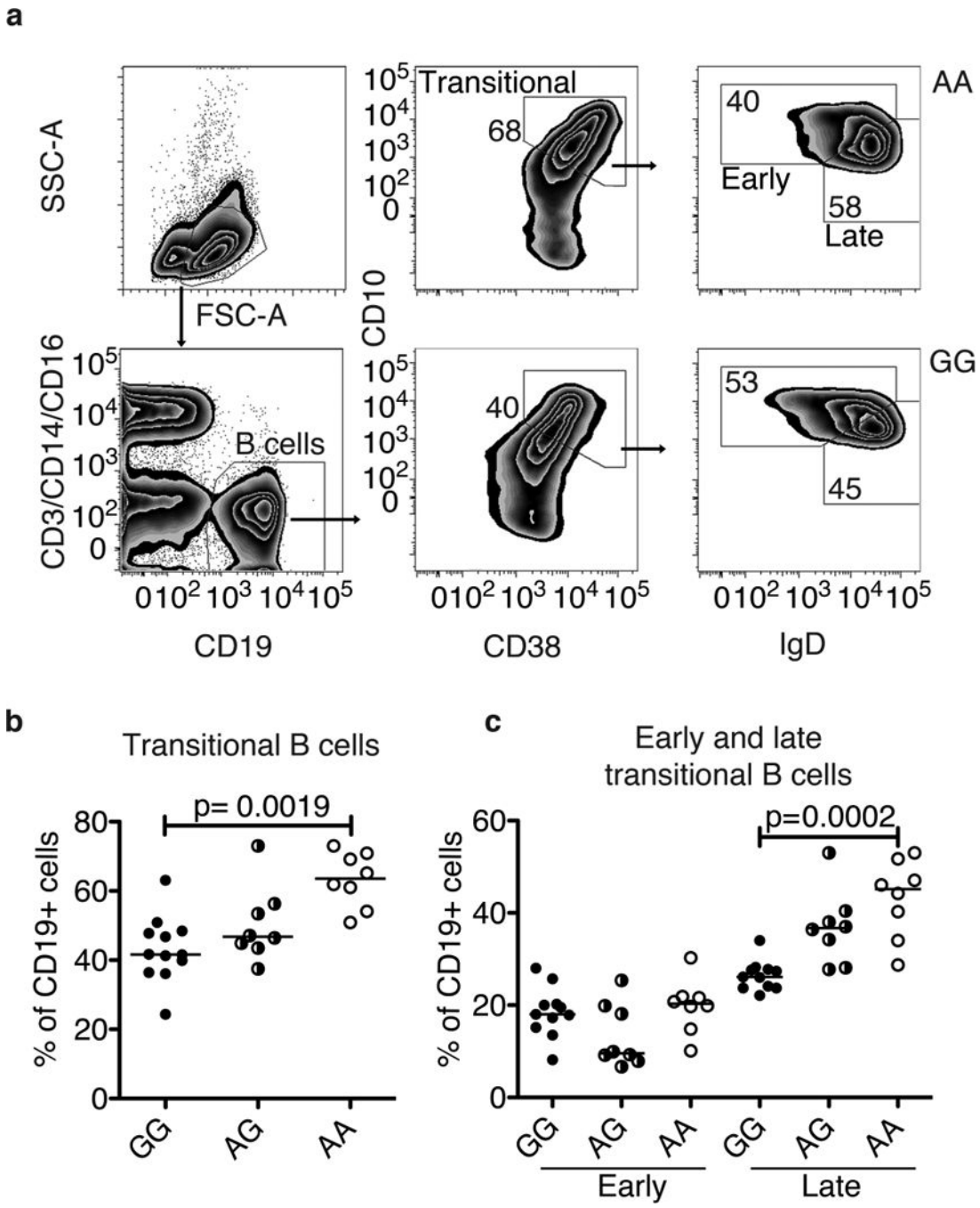


**Figure 4.**

The *CSK* risk allele is associated with expansion of transitional B cells in umbilical cord blood. (a) B cells were defined by a $CD19^{+}$ gate (with CD3/CD14/CD16 exclusion); transitional cells were defined as $CD38^{hi}$, $CD10^{hi}$. The transitional cells were further divided into early and late stage by gain of IgD and decreased CD10 in late stage transitional B cells. The panel shows a representative plot of each homozygous genotype. (b) The percentage of transitional cells in the CD19 compartment is significantly higher in homozygous risk allele carriers compared with heterozygotes, with a similar trend in comparison with wild type

homozygous individuals (n = 27, Kruskal-Wallis test). (c) Among transitional cells, there is an expansion of the late transitional compartment in individuals who carry the risk allele (Kruskal-Wallis test). The horizontal line marks the median.

**Table 1**

Frequencies of the rs34933034 *CSK* risk allele in cases and controls of European ancestry.

| Analytic group[1] | N (Cases / Controls) | Genotype frequency (Cases / Controls) | | | MAF[2], % (Cases/Controls) | OR[3] (95% C.I.) | P value Allelic[4] | P value Additive[5] | P value Additive[6] |
|---|---|---|---|---|---|---|---|---|---|
| | | AA | AG | GG | | | | | |
| I | 1378 / 919 | 0.040 / 0.018 | 0.28 / 0.23 | 0.68 / 0.75 | 18.17 / 13.54 | 1.42 (1.20–1.67) | $3.2 \times 10^{-5}$ | $4.3 \times 10^{-5}$ | $9.2 \times 10^{-5}$ |
| II | 1368 / 1228 | 0.032 / 0.028 | 0.29 / 0.24 | 0.68 / 0.73 | 17.50 / 14.78 | 1.22 (1.05–1.42) | $7.8 \times 10^{-3}$ | $8.6 \times 10^{-3}$ | $9.8 \times 10^{-3}$ |
| III | 1023/1257 | 0.023 / 0.190 | 0.31 / 0.24 | 0.67 / 0.74 | 17.74 / 13.76 | 1.35 (1.15–1.59) | $2.3 \times 10^{-4}$ | $1.9 \times 10^{-4}$ | $4.1 \times 10^{-4}$ |
| Total | 3769/3404 | 0.032 / 0.022 | 0.29 / 0.24 | 0.68 / 0.74 | 17.81 / 14.07 | 1.32 (1.20–1.44) | $1.0 \times 10^{-9}$ | $1.35 \times 10^{-9}$ | $3.35 \times 10^{-8}$ |

[1] A total of 12 datasets were used in three stages of genotyping and testing. Details of the datasets used can be found in Supplementary table 2.

[2] MAF. Minor allele frequency.

[3] O.R. calculated for an allelic model, the samples were matched for European background.

[4] $X^2$ P-value

[5] Armitage P-value

[6] After correction for European substructure by principal components at each group, meta P-value calculated using the sample size based meta-analysis strategy[17].