nature genetics

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[1,10], 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[1,11] & Peter K Gregersen[2,11]

The c-Src tyrosine kinase, Csk, physically interacts with the intracellular phosphatase Lyp (encoded by *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 the intronic polymorphism rs34933034 (odds ratio (OR) = 1.32; $P = 1.04 \times 10^{-9}$). The risk allele at this SNP is associated with increased *CSK* expression and augments inhibitory phosphorylation of Lyn. In carriers of the risk allele, there is increased B-cell receptor (BCR)-mediated activation of mature B cells, as well as higher concentrations of plasma immunoglobulin M (IgM), relative to individuals with the non-risk haplotype. Moreover, the fraction of transitional B cells is doubled in the cord blood of carriers of the risk allele, due to an expansion of late transitional cells in a stage targeted by selection mechanisms. This suggests that the Lyp-Csk complex increases susceptibility to lupus at multiple maturation and activation points in 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[2]. Several genome-wide associations in systemic lupus have been described in molecules that participate in the BCR signaling pathway[3,4]. 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 (SFKs). Activation of SFKs is regulated by the Lyp-Csk complex. Lyp (the product of the *PTPN22* gene) destabilizes the kinase domains of SFKs through tyrosine dephosphorylation, while Csk phosphorylates C-terminal tyrosines in SFKs, leading to a closed, inactive conformation[5]. The Lyp variant R620W disrupts the interaction between Lyp and the SH3 domain of Csk and has been strongly associated with many autoimmune disorders[6]. Consequently, we searched for evidence of association with SLE at the *CSK* locus in a previous GWAS of SLE in subjects of European ancestry[3]. We observed that several SNPs that tag a *CSK* haplotype (designated here the B haplotype) had a nominal association with SLE in this previously published data set, with associations well below genome-wide levels of statistical significance (OR ~1.15; uncorrected *P* values = 0.004–0.007; 1,311 subjects with SLE and 3,340 controls)[3]. We therefore sequenced the *CSK* exons and five highly conserved intronic regions in a discovery cohort of 24 SLE-affected subjects of European ancestry who are homozygous for the B haplotype (**Fig. 1e**). Of the 4 polymorphisms present in the discovery data set (**Supplementary Table 1**), the minor allele rs34933034[A] of an intronic variant was present in 34 of the 48 chromosomes bearing the B haplotype and was therefore of particular interest.

To assess the relationship of the variant rs34933034[A] allele to the B haplotype and explore its association with risk for SLE, we genotyped 3,769 subjects with SLE and 3,404 controls of European ancestry derived from 11 cohorts and analyzed the data in 3 groups (**Table 1**, Online Methods and **Supplementary Table 2**). The results provide

[1]Center for Autoimmune and Musculoskeletal Disorders, The Feinstein Institute for Medical Research, North Shore–Long Island Jewish (LIJ), Manhasset, New York, USA. [2]Robert S Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, North Shore–LIJ, Manhasset, New York, USA. [3]Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California, San Francisco, San Francisco, California, USA. [4]Department of Medical and Molecular Genetics, Division of Genetics and Molecular Medicine, School of Medicine, King's College London, London, UK. [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, California, USA. [8]Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA. [9]Arthritis and Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA. [10]Present address: Division of Clinical Immunology, McGill University Health Centre, Montreal, Quebec, Canada. [11]These authors contributed equally to this work. Correspondence should be addressed to B.D. (bdiamond@nshs.edu) or P.K.G. (pgregers@nshs.edu).

Received 11 July; accepted 11 September; published online 7 October 2012; doi:10.1038/ng.2439

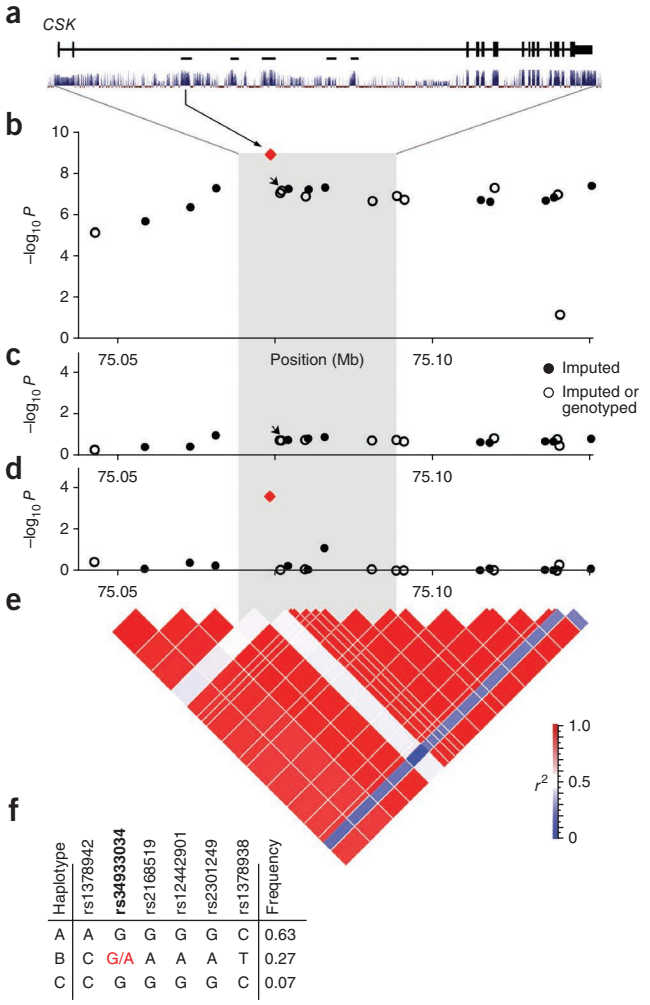

Figure 1 Genetic structure of the *CSK* locus. (**a**) *CSK* spans 14 exons (vertical boxes). The plot shows the sequence homology among mammals, with highly conserved regions shown as peaks (UCSC Genome Browser). The horizontal bars highlight the conserved areas that were selected for sequencing. The arrow from the first conserved region that was sequenced shows the position of the rs34933034 variant. (**b**) Association of SNP markers with SLE. A region of ~80 kb surrounding the *CSK* locus is shown. The gray area highlights the position of the *CSK* gene. The markers shown were either imputed or genotyped in some samples and were imputed in others. SNPs were imputed using 1000 Genomes Project data. The arrow indicates rs8033381, one of many SNPs that tag the B haplotype, and rs34933034 is represented as a diamond. (**c**) After conditioning on rs34933034, all association signals within the *CSK* gene are eliminated. (**d**) After conditioning on rs8033381, there was still an association signal for 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 value for rs34933034 with this block (r^2 ~0.5). (**f**) The major haplotypes for the *CSK* locus and their frequencies in European controls are shown.

convincing evidence for association of the variant rs34933034[A] allele with SLE (allelic OR = 1.32; $P = 1.04 \times 10^{-9}$). Linkage disequilibrium analysis showed 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 the other SNPs that define this haplotype (**Fig. 1**). Two SNPs that are present in the *CSK* B haplotype have recently been associated with several autoimmune diseases[7,8]. Therefore, we imputed SNPs, using 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 association with the rs34933034[A] allele. For example, the rs8033381 marker tagging the common B haplotype shows an association ($P = 9 \times 10^{-8}$) that is eliminated after conditioning on rs34933034 ($P = 0.19$) (**Fig. 1**, arrowhead). In contrast, conditioning on rs8033381 shows an association at rs34933034 ($P = 2.49 \times 10^{-4}$). Notably, we did not observe evidence of genetic interaction between rs34933034 and rs2476601 (encoding the PTPN22 R620W variant).

The Encyclopedia of DNA Elements (ENCODE) database reports the presence of DNase-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[9]. This was confirmed by evaluation of the pattern of *CSK* expression in different subsets of peripheral blood mononuclear cells (PBMCs) from healthy individuals, including the transitional, naive and memory subsets of B cells (all subpopulations were recovered in at least five subjects), as well as $CD4^+$ and $CD8^+$ T cells and monocytes in three subjects (**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 had recently emerged from the bone marrow had higher *CSK* expression than mature naive B cells, which in turn had higher expression than memory B cells. We sorted naive B cells from 29 healthy donors with either genotype at rs34933034, all of whom were homozygous for the B haplotype. Higher levels of *CSK* transcripts in naive B cells were significantly associated with the presence of the rs34933034[A] allele (**Fig. 2c**).

Next, we examined whether increased *CSK* expression affects B-cell function. It has been reported that, in resting mouse T cells, Csk-mediated C-terminal phosphorylation of SFKs maintains them in an inactive conformation[10]. In B cells, Lyn is the most abundant SFK; therefore, we hypothesized that greater amounts of Csk would increase basal phosphorylation at the C-terminal tyrosine of Lyn in B cells. We analyzed Tyr508 phosphorylation by flow cytometry in resting naive B cells ($CD20^+CD27^-$) from 27 healthy adults (**Fig. 3a**).

Table 1 Frequencies of the rs34933034 *CSK* risk allele in cases and controls of European ancestry

Analytic group[a]	*N*		Genotype frequency						MAF (%)[b]		OR[c] (95% CI)	*P* (allelic)[d]	*P* (additive)[e]	*P* (additive)[f]
			AA		AG		GG							
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls				
1	1,378	919	0.04	0.02	0.28	0.23	0.68	0.75	18.17	13.54	1.42 (1.20–1.67)	3.2×10^{-5}	4.3×10^{-5}	9.2×10^{-5}
2	1,368	1,228	0.03	0.03	0.29	0.24	0.68	0.73	17.50	14.78	1.22 (1.05–1.42)	7.8×10^{-3}	8.6×10^{-3}	9.8×10^{-3}
3	1,023	1,257	0.02	0.02	0.31	0.24	0.67	0.74	17.74	13.76	1.35 (1.15–1.59)	2.3×10^{-4}	1.9×10^{-4}	4.1×10^{-4}
Total	3,769	3,404	0.03	0.02	0.29	0.24	0.68	0.74	17.81	14.07	1.32 (1.20–1.44)	1.0×10^{-9}	1.35×10^{-9}	3.35×10^{-8}

[a]A total of 12 data sets were used in 3 stages of genotyping and testing. Details of the data sets used are provided in **Supplementary Table 2**. [b]MAF, minor allele frequency. [c]OR calculated for an allelic model. The samples were matched for European ancestry. [d]χ^2 *P* value. [e]Armitage *P* value. [f]After correction for European substructure by principal components in each group, *P* values were calculated using the sample size–based meta-analysis strategy[18].

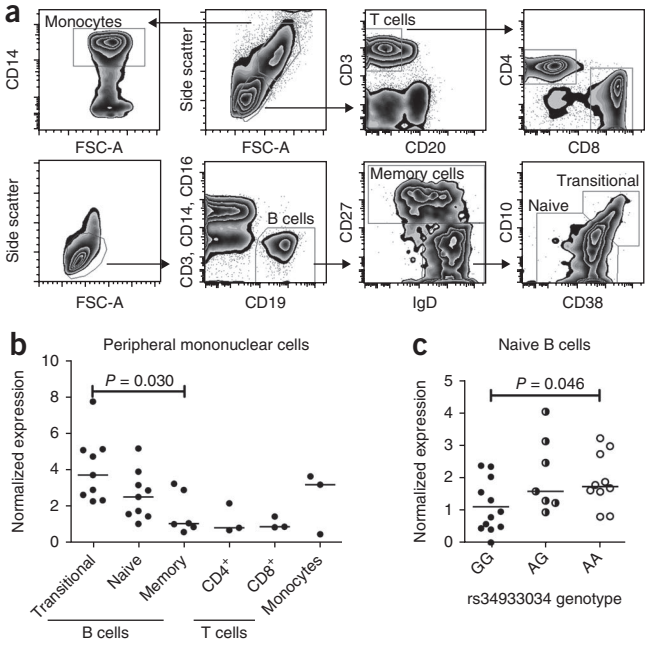

Figure 2 *CSK* expression varies in lymphocyte subsets and is associated with *CSK* genotype. (**a**) Gating strategy employed to distinguish monocytes and T cells (top) and B-cell subpopulations (bottom). FSC-A, forward scatter. (**b**) *CSK* expression decreases in peripheral B cells as they mature 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 naive B cells, the rs34933034[A] risk allele 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 quantitative PCR (qPCR) with cDNA synthesized from the RNA of isolated subpopulations of peripheral mononuclear cells from non-genotyped blood donors in **b** or haplotype-matched Genotype and Phenotype (GaP) registry subjects in **c**. The horizontal bar marks the median of the values.

Naive B cells from subjects carrying the *CSK* risk allele showed higher Lyn phosphorylation at Tyr508 than B cells from carriers of the non-risk allele (**Fig. 3b** and **Supplementary Fig. 1**). Notably and in contrast to the activating role of Lck in T cells, Lyn has been reported to mediate negative regulation of BCR signaling[11]. Therefore, we measured calcium mobilization in naive B cells from 11 donors homozygous for the B haplotype after BCR cross-linking. As expected, naive B cells homozygous for the rs34933034[A] allele showed enhanced calcium mobilization triggered by BCR cross-linking (**Fig. 3c**,**d**) (no differences in calcium mobilization 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[11]. Moreover, healthy subjects ($n = 44$) who carried the *CSK* risk allele also had higher plasma concentrations of IgM (**Fig. 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 at *CSK* would affect early B-cell differentiation. To study B-cell maturation in the most pristine situation possible, we analyzed B-cell subpopulations in 27 samples of umbilical cord blood and found that newborn subjects homozygous for the *CSK* risk allele had double the frequency of transitional ($CD38^{hi}CD10^{hi}$) cells than were observed in subjects homozygous for the non-risk allele (**Fig. 4a**,**b**). Further dissection of this B-cell compartment showed that individuals with the risk allele had more 'late' transitional cells than their counterparts with the non-risk allele; these cells were also characterized by high surface expression of CD21 and had acquired surface immunoglobulin D (IgD), while still retaining high expression of CD38 and CD10 (**Fig. 4c** and **Supplementary Fig. 2**).

Our results clearly support the hypothesis that *CSK* expression levels can modify normal B-cell biology. We have shown that *CSK* expression is highest at the earliest stages of B-cell maturation (in transitional B cells) and that normal individuals who carry the risk allele have expansion of the late transitional B-cell population. This expansion might reflect impairment of a tolerance checkpoint or enhanced positive selection of late transitional B cells. Either model leads to the prediction that there would be increased numbers of autoreactive B cells in the peripheral repertoire of healthy carriers of the *CSK* risk allele, consistent with 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[12] and augmented[13,14] responses to BCR cross-linking have been reported, and higher autoreactivity in the early B-cell repertoire of healthy subjects carrying the risk allele has been observed[15]. The results reported

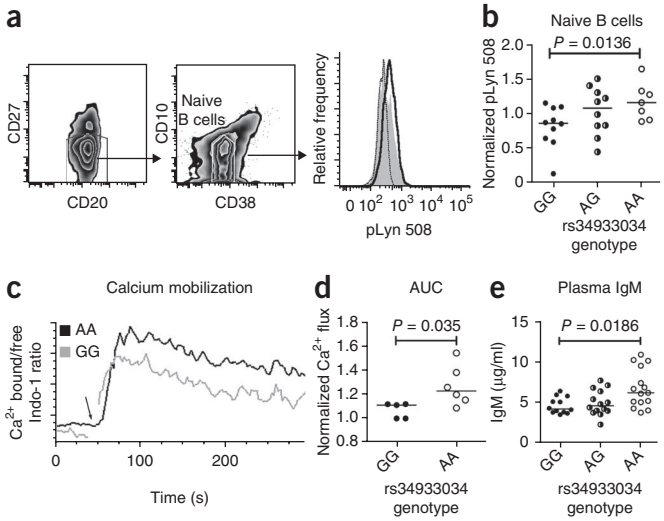

Figure 3 The *CSK* risk allele is associated with increased phosphorylation of Lyn at Tyr508 and enhanced activation of mature B cells. (**a**) Phosphorylated Lyn at Tyr508 (pLyn 508) was measured by flow cytometry (strategy shown at left) in naive B cells from haplotype-matched GaP registry donors homozygous for either allele at rs34933034; the profiles (right) of a subject with the risk allele (black) and a subject with the non-risk allele (gray) are shown. The thin dotted line shows the isotype control. (**b**) Lyn phosphorylation in carriers of the *CSK* risk allele. pLyn 508 levels were normalized to the average signal on each day (27 subjects analyzed; $P = 0.0136$, Mann-Whitney test, see Online Methods for correction after multiple comparison and **Supplementary Fig. 1**; Kruskal-Wallis test $P = 0.0569$). (**c**) A representative experiment analyzing calcium mobilization. Basal calcium levels were measured for 40 s before activation with antibody to IgM $(Fab')_2$. (**d**) Data from 11 different B haplotype donors was normalized to 1 non-risk (GG) subject for comparison on each of 4 separate days of experiments. The area under the curve (AUC) for the first 90 s (similar results at 4 min) is shown (see Online Methods for statistical analysis). (**e**) Plasma concentration of IgM in 42 subjects homozygous for the B haplotype (Kruskal-Wallis test). In all graphs, the horizontal line marks the median.

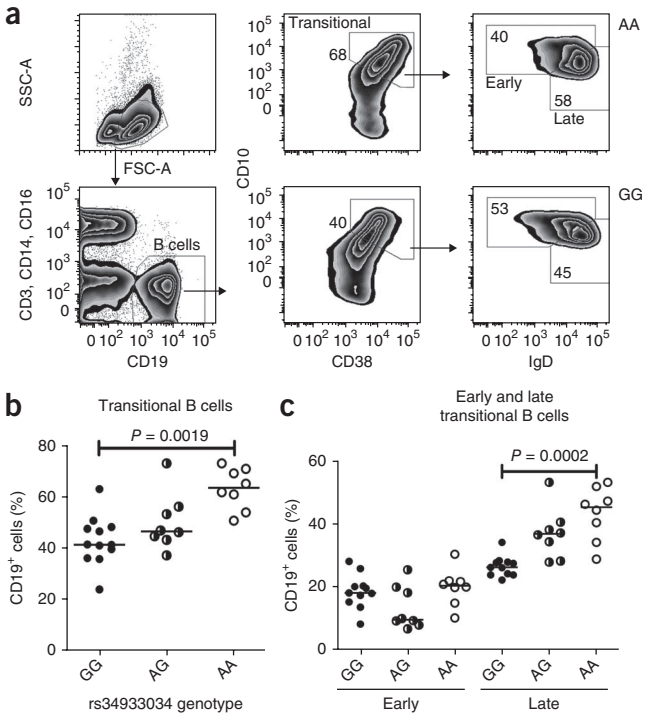

Figure 4 The *CSK* risk allele is associated with expansion of transitional B cells in umbilical cord blood. (**a**) B cells were defined by a $CD19^+$ gate (with CD3, CD14 and CD16 exclusion); transitional cells were defined as $CD38^{hi}CD10^{hi}$. The transitional cells were further divided into early- and late-stage populations by gain of IgD and decreased CD10 in late-stage cells. The panel shows a representative plot for each homozygous genotype. The percentage of cells within each plot is shown. SSC-A, side scatter. (**b**) The percentage of transitional cells in the CD19 compartment is significantly higher in homozygous carriers of the risk allele compared to heterozygotes, with a similar trend in comparison to wild-type homozygous individuals (27 samples analyzed; Kruskal-Wallis test). (**c**) There is expansion of the late transitional compartment in individuals who carry the risk allele (Kruskal-Wallis test *P* values). In all graphs, the horizontal line marks the median.

here show that Csk levels 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 the effects of *PTPN22* 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[3,16]. Research from our group suggests that the *BLK* risk allele also predominantly influences expression differences in the early stages of B-cell development[17], although the functional consequences of differences in Blk expression have not yet been determined. This finding 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 to thereby inform the rational development of targeted approaches to therapy.

URLs. The GaP registry, http://www.gapregistry.org/; UCSC Genome Browser, http://genome.ucsc.edu/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank the volunteers who participated in this study; the Genotype and Phenotype (GaP) registry; M. Keogh, M. DeFranco, C. Mason, C. Metz and the Biorepository at the Feinstein Institute for Medical Research (FIMR) for recruiting subjects and collecting samples; H. Borrero for technical assistance; and the Biostatistics Unit of the FIMR and M. Akerman for assistance. This work was supported by US National Institutes of Health (NIH) grant RC2AR059092; The Alliance for Lupus Research, a Kirkland Scholar Award and NIH/National Center for Research Resources (NCRR) grant 5 M01 RR-00079 (L.A.C.); P01 AR49084 and UL1 TR000165 (R.P.K.); and NIH/National Center for Advancing Translational Sciences grant KL2TR000143 and an American College of Rheumatology Physician Scientist Development Award (S.A.C.). The principal investigators of the Nijmegen Biomedical Study are L.A.L.M. Kiemeney, M. den Heijer, A.L.M. Verbeek, D.W. Swinkels and B. Franke. We thank the Lupus Family Registry and Repository (LFRR) investigators, J.B. Harley, K.L. Moser Sivils, M.H. Weisman and D.J. Wallace, funded by N01AR62277 (K.L.M.S. and J.B.H.).

AUTHOR CONTRIBUTIONS

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

The authors declare competing financial interests: details are available in the online version of the paper.

Published online at http://www.nature.com/doifinder/10.1038/ng.2439.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Cho, J.H. & Gregersen, P.K. Genomics and the multifactorial nature of human autoimmune disease. *N. Engl. J. Med.* **365**, 1612–1623 (2011).
2. Liston, A., Lesage, S., Gray, D.H.D., Boyd, R.L. & Goodnow, C.C. Genetic lesions in T-cell tolerance and thresholds for autoimmunity. *Immunol. Rev.* **204**, 87–101 (2005).
3. Hom, G. *et al.* Association of systemic lupus erythematosus with *C8orf13-BLK* and *ITGAM-ITGAX*. *N. Engl. J. Med.* **358**, 900–909 (2008).
4. Nath, S.K. *et al.* A nonsynonymous functional variant in integrin-α_M (encoded by *ITGAM*) is associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 152–154 (2008).
5. Levinson, N.M., Seeliger, M.A., Cole, P.A. & Kuriyan, J. Structural basis for the recognition of c-Src by its inactivator Csk. *Cell* **134**, 124–134 (2008).
6. Gregersen, P.K., Lee, H.-S., Batliwalla, F. & Begovich, A.B. PTPN22: setting thresholds for autoimmunity. *Semin. Immunol.* **18**, 214–223 (2006).
7. Trynka, G. *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.* **43**, 1193–1201 (2011).
8. Martin, J.-E. *et al.* Identification of *CSK* as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up. *Hum. Mol. Genet.* **21**, 2825–2835 (2012).
9. Su, A.I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **101**, 6062–6067 (2004).
10. Zikherman, J. *et al.* CD45-Csk phosphatase-kinase titration uncouples basal and inducible T cell receptor signaling during thymic development. *Immunity* **32**, 342–354 (2010).
11. Hasegawa, M. *et al.* A CD19-dependent signaling pathway regulates autoimmunity in Lyn-deficient mice. *J. Immunol.* **167**, 2469–2478 (2001).
12. Arechiga, A.F. *et al.* Cutting edge: the PTPN22 allelic variant associated with autoimmunity impairs B cell signaling. *J. Immunol.* **182**, 3343–3347 (2009).
13. Zikherman, J. *et al.* *PTPN22* deficiency cooperates with the CD45 E613R allele to break tolerance on a non-autoimmune background. *J. Immunol.* **182**, 4093–4106 (2009).
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.* **43**, 902–907 (2011).
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. Invest.* **121**, 3635–3644 (2011).
16. Lu, R. *et al.* Genetic associations of *LYN* with systemic lupus erythematosus. *Genes Immun.* **10**, 397–403 (2009).
17. Simpfendorfer, K.R. *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. *Hum. Mol. Genet.* **21**, 3918–3925 (2012).
18. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

ONLINE METHODS

Research subjects and specimens. DNA from cohorts of SLE-affected subjects was obtained from the University of California, San Francisco, Lupus Genetics Project[19], the Multiple Autoimmune Diseases Genetics Consortium (MADGC)[20], the University of Minnesota[21], the UK SLE Study[22], The Autoimmune Biomarkers Collaborative Network (ABCoN)[3], the University of Alabama[4] and the Oklahoma Medical Research Foundation (OMRF)[4] as previously reported. Control DNA samples were taken from the New York Cancer Project collection[23,24], the Nijmegen Biomedical study (NBS; a population-based cohort of self-reported, randomly selected inhabitants of Nijmegen, The Netherlands)[25] and the Genotype and Phenotype Registry at The Feinstein Institute for Medical Research (FIMR) and the University of Alabama. All subjects gave written informed consent, and institutional review boards (IRBs) reviewed the protocols at their host institutions. Diagnosis with SLE was established according to American College of Rheumatology (ACR) guidelines[3,4,19–22]. A detailed explanation of the number of samples per data set is given in **Supplementary Table 2**.

All of the genotype-matched blood samples from 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. Deidentified cord blood samples deemed not suitable for banking and leukocyte units were obtained from the Long Island Blood Bank. The FIMR IRB reviewed and approved all of the protocols; consent was waived for deidentified cord blood and leukocyte units.

Genotyping. *CSK* sequencing for the 24 SLE samples in the discovery cohort was performed by Polymorphic DNA Technologies. Genotyping of the data sets was performed sequentially in three analytic groups. European ancestry was determined by the analysis of ancestry-informative markers within each analytic group. Each of the three groups of data sets are described in **Supplementary Table 2**. *CSK* genotyping for groups 1 and 2 was performed by pyrosequencing (**Supplementary Table 4**). Genotyping of rs34933034 in *CSK* for analytic group 3 of samples from the NBS and OMRF was performed by qPCR with the TaqMan assay C__60143137_10 in a ViiA7 machine (Applied Biosciences).

Data for SNPs informative for continental origin and European population diversity were obtained from genome-wide typing for each of the cohorts and were used to limit the study to subjects of European ancestry. Matching of cases and controls by principal-component analysis[26], as well as association analysis for both allelic and additive genetic models, was performed using the SNP and Variation Suite (Golden Helix). The meta-analysis *P* value for the additive genetic model with correction for principal-component analysis was calculated using the sample size–based analytical strategy reported previously[18].

Imputation and conditional analysis. For each group of data sets, we obtained genotype information on markers situated within 300 kb of either edge of *CSK*. We used 1000 Genomes Project data (Phase 1, version 3, March 2012) as a reference to impute variants across a fragment of 120 kb (chr. 15: 75,024,425–75,145,539, hg19), with *CSK* in the center. Imputation was performed using IMPUTE2.2 (ref. 27). 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 data set were not used (**Supplementary Table 3**). Conditional analysis of the variants that passed quality control was performed with SNPTEST2 (ref. 27) 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 10 min with an antibody mix containing one of the following antibody cocktails: CD8 (FITC, 551347), CD20 (PE, 555623), CD4 (PerCP, 340671), CD3 (APC, 340440) or CD14 (Pacific Blue, 558121), all from BD Biosciences, for sorting of monocytes and T-cell subpopulations (three samples). For B cells, we used antibodies to IgD (FITC, 555778), CD3 (Pacific blue, 558124), CD14 and CD16 (Pacific blue, 558122) and CD19 (APC-Cy7, 557791), all from BD Biosciences; CD27 (PE, MHCD2704) and CD38 (PE-TR, MHCD3817) from Invitrogen; and CD10 (PE-Cy7, 312214) from Biolegend. Cells were sorted on a FACSAria instrument (BD Biosciences), and, after exclusion of doublets by forward and side scatter area and width parameters, gates were set for monocytes ($CD3^-CD20^-CD14^+$) and $CD4^+$ and $CD8^+$ cells (after gating for $CD20^-CD3^+$ cells). B cells were defined as $CD19^+CD3^-CD14^-CD16^-$ and were subsequently gated for B-cell subpopulations, including memory ($CD19^+CD27^+$) and preimmune (IgD^+CD27^-) cells. To divide the preimmune B-cell populations, the upper limit of CD10 fluorescence in T cells (which lack CD10 expression) was used to define the lower limits of CD10 positivity for transitional B cells ($CD38^{hi}CD10^{hi}$) and naive cells ($CD38^{dim}CD10^{low}$). Thirteen of the cord blood samples were stained with PE-conjugated antibody to CD21 to confirm the gates for early and late transitional B cells, as defined previously[28]. Statistical analyses of B-cell subpopulations were performed using the Kruskal-Wallis one-way analysis of variance.

Expression analysis. Analysis of *CSK* expression in peripheral blood cells was performed on sorted cells. RNA was extracted using the Micro RNeasy Isolation kit (Qiagen). cDNA was synthesized by linear reverse transcription with iScript (Bio-Rad) and 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 ΔC_T method of $2^{(C_T\,(PolR2A)\,-\,C_T\,(CSK))}$ (ref. 29). All expression assays were performed in duplicate. The analysis of *CSK* expression across B-cell subpopulations was performed by Kruskal-Wallis test; the analysis of *CSK* expression across genotypes was performed using one-way ANOVA.

Cell signaling. Cell signaling assays were performed in PBMCs from 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 (both from Life Technologies). For phosphorylation of Lyn, cells were fixed, washed and surface stained with CD20, CD24, CD38, CD27 and CD3. After a permeabilization step with BD Perm (BD Biosciences), cells were stained with antibody to phosphorylated Lck (Tyr505; BD Phosflow, 557879), which cross-reacts with phosphorylated Lyn. Median fluorescence intensity (MFI) in naive B cells was determined by analysis of the data with FlowJo software. No differences were observed between data analyzed solely on the basis of MFI and data analyzed on the basis of the MFI of phosphorylated Lyn minus the MFI of the isotype control. Data were normalized to the mean MFI for each experiment. With correction for multiple testing, pairwise comparisons were considered significant at $P < 0.0166$.

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

SAS version 9.2 was used to perform hierarchical linear mixed-model (HLMM) analysis of the normalized values for the area under the curve. To adjust for day-to-day variation in experimental conditions, subjects were considered nested within each day. Furthermore, the day of the experiment was considered a random effect. A result was considered statistically significant at $P < 0.05$.

ELISAs. We precoated 96-well plates with 10 μg/ml antibody to human IgM (Southern Biotech) and blocked with PBS (containing 1% BSA). Samples were plated, and, after 1 hour of incubation, wells were washed and incubated with alkaline phosphatase–conjugated antibody to human IgM. Wells were later developed by the addition of p-nitrophenyl disodium (Sigma) in carbonate buffer. Optical density at 405 nm was measured. Samples were run in duplicate, and IgM was quantified using an IgM standard curve (Sigma-Aldrich).

19. Thorburn, C.M. *et al.* Association of *PDCD1* genetic variation with risk and clinical manifestations of systemic lupus erythematosus in a multiethnic cohort. *Genes Immun.* **8**, 279–287 (2007).

20. Criswell, L.A. *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.* **76**, 561–571 (2005).
21. Graham, R.R. *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.* **38**, 550–555 (2006).
22. Cunninghame Graham, D.S. *et al.* Association of *NCF2*, *IKZF1*, *IRF8*, *IFIH1*, and *TYK2* with systemic lupus erythematosus. *PLoS Genet.* **7**, e1002341 (2011).
23. Plenge, R.M. *et al. TRAF1-C5* as a risk locus for rheumatoid arthritis—a genomewide study. *N. Engl. J. Med.* **357**, 1199–1209 (2007).
24. Gregersen, P.K. *et al. REL*, encoding a member of the NF-κB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat. Genet.* **41**, 820–823 (2009).
25. Wetzels, J.F.M., Kiemeney, L.A.L.M., Swinkels, D.W., Willems, H.L. & Heijer, M.d. Age- and gender-specific reference values of estimated GFR in Caucasians: The Nijmegen Biomedical Study. *Kidney Int.* **72**, 632–637 (2007).
26. Tian, C. *et al.* European population genetic substructure: further definition of ancestry informative markers for distinguishing among diverse European ethnic groups. *Mol. Med.* **15**, 371–383 (2009).
27. 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.* **39**, 906–913 (2007).
28. Suryani, S. *et al.* Differential expression of CD21 identifies developmentally and functionally distinct subsets of human transitional B cells. *Blood* **115**, 519–529 (2010).
29. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**, 402–408 (2001).

doi:10.1038/ng.2439