HLA-C Level Is Regulated by a Polymorphic Oct1 Binding Site in the *HLA-C* Promoter Region

Nicolas Vince,[1,2] Hongchuan Li,[1] Veron Ramsuran,[1,2] Vivek Naranbhai,[2,3,4] Fuh-Mei Duh,[1] Benjamin P. Fairfax,[3] Bahara Saleh,[1] Julian C. Knight,[3] Stephen K. Anderson,[1] and Mary Carrington[1,2,*]

Differential HLA-C levels influence several human diseases, but the mechanisms responsible are incompletely characterized. Using a validated prediction algorithm, we imputed HLA-C cell surface levels in 228 individuals from the 1000 Genomes dataset. We tested 68,726 SNPs within the *MHC* for association with HLA-C level. The *HLA-C* promoter region variant, rs2395471, 800 bp upstream of the transcription start site, gave the most significant association with *HLA-C* levels ($p = 4.2 \times 10^{-66}$). This imputed expression quantitative trait locus, termed impeQTL, was also shown to associate with *HLA-C* expression in a genome-wide association study of 273 donors in which *HLA-C* mRNA expression levels were determined by quantitative PCR (qPCR) ($p = 1.8 \times 10^{-20}$) and in two cohorts where HLA-C cell surface levels were determined directly by flow cytometry ($n = 369$ combined, $p < 10^{-15}$). rs2395471 is located in an Oct1 transcription factor consensus binding site motif where the A allele is predicted to have higher affinity for Oct1 than the G allele. Mobility shift electrophoresis demonstrated that Oct1 binds to both alleles in vitro, but decreased *HLA-C* promoter activity was observed in a luciferase reporter assay for rs2395471_G relative to rs2395471_A on a fixed promoter background. The rs2395471 variant accounts for up to 36% of the explained variation of HLA-C level. These data strengthen our understanding of HLA-C transcriptional regulation and provide a basis for understanding the potential consequences of manipulating HLA-C levels therapeutically.

Variation in the *MHC* associates with an extensive number of human diseases and traits, with about 500 (30%) reported in the human genome-wide association study (GWAS) catalog,[1] particularly that occurring within the *HLA* class I and II genes. Extreme polymorphism of the *HLA* loci[2] along with their central importance in both the acquired and innate immune response accounts for this over-representation relative to the rest of the genome. HLA class I and II molecules bind and present an extensive array of antigenic peptides to cytotoxic T lymphocytes and CD4 T cells, respectively, in order to initiate the acquired immune response.[3,4] The class I molecules also serve as ligands for killer cell immunoglobulin-like receptors (KIR) expressed on natural killer cells.[5] Relative to *HLA-A* (MIM: 142800) and *-B* (MIM: 142830), *HLA-C* (MIM: 142840) exhibits limited diversity, lower cell surface level,[6,7] and a more widely distributed role as ligands for KIR. *HLA-C* alleles are associated with many disease traits,[1] primarily with regard to autoimmune diseases.[8,9] High HLA-C level associates with better HIV control (MIM: 609423), but also with increased risk of developing Crohn disease (MIM: 266600).[10] Hence, understanding the mechanisms that determine HLA-C level could provide important insights into the management of complex human disease.

An insertion/deletion polymorphism in the 3′ UTR of *HLA-C* determines the binding and inhibition of *HLA-C* expression by the microRNA miR-148a (*MIR148A* [MIM: 613786]), contributing to differential HLA-C levels.[11,12] However, this polymorphism accounts for only 9% of the explained variation in HLA-C level, indicating that additional mechanisms participate in determining allele-specific expression levels at this locus. In the current study, we identified variants within and near *HLA-C* that significantly associate with imputed HLA-C levels among individuals from the 1000 Genomes dataset (1KG).[10,13,14] The most significant variant was then tested for its association with both RNA expression and cell surface HLA-C levels measured directly, validating the imputation approach, and the mechanism underlying this association was shown to involve the transcription factor (TF) Oct1.

HLA-C levels vary in an allele-specific manner over a range of 7-fold in a pattern that is consistent between African and European Americans and highly reproducible across study groups.[10] Based on the level value characteristic for each given HLA-C allotype as determined previously (Table S1), we imputed HLA-C levels for 228 European 1KG individuals who have previously been typed for *HLA-C*.[14] We restricted our analysis to Europeans with homogeneous ancestry background (Figure S1), which included 52 CEU (Utah residents with Northern and Western European ancestry), 87 GBR (British in England and Scotland), and 89 TSI (Toscani in Italy).

Imputed HLA-C levels were tested as a continuous variable for association with 68,726 SNPs within the *MHC* using linear regression in order to identify *cis*-acting variants that may cause (or mark) differential level of HLA-C. The peak association was centered in the *HLA-C* promoter region (Figure 1A), and correction for population structure did not alter the results.[17] The top signal identified was rs2395471 ($p = 4.2 \times 10^{-66}$), which is 800 bp upstream of the transcription start site (Figure 1B), and

[1]Cancer and Inflammation Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA; [2]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA 02139, USA; [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; [4]Center for the AIDS Programme of Research in South Africa, University of KwaZuluNatal, Durban 4091, South Africa
*Correspondence: carringm@mail.nih.gov
http://dx.doi.org/10.1016/j.ajhg.2016.09.023.

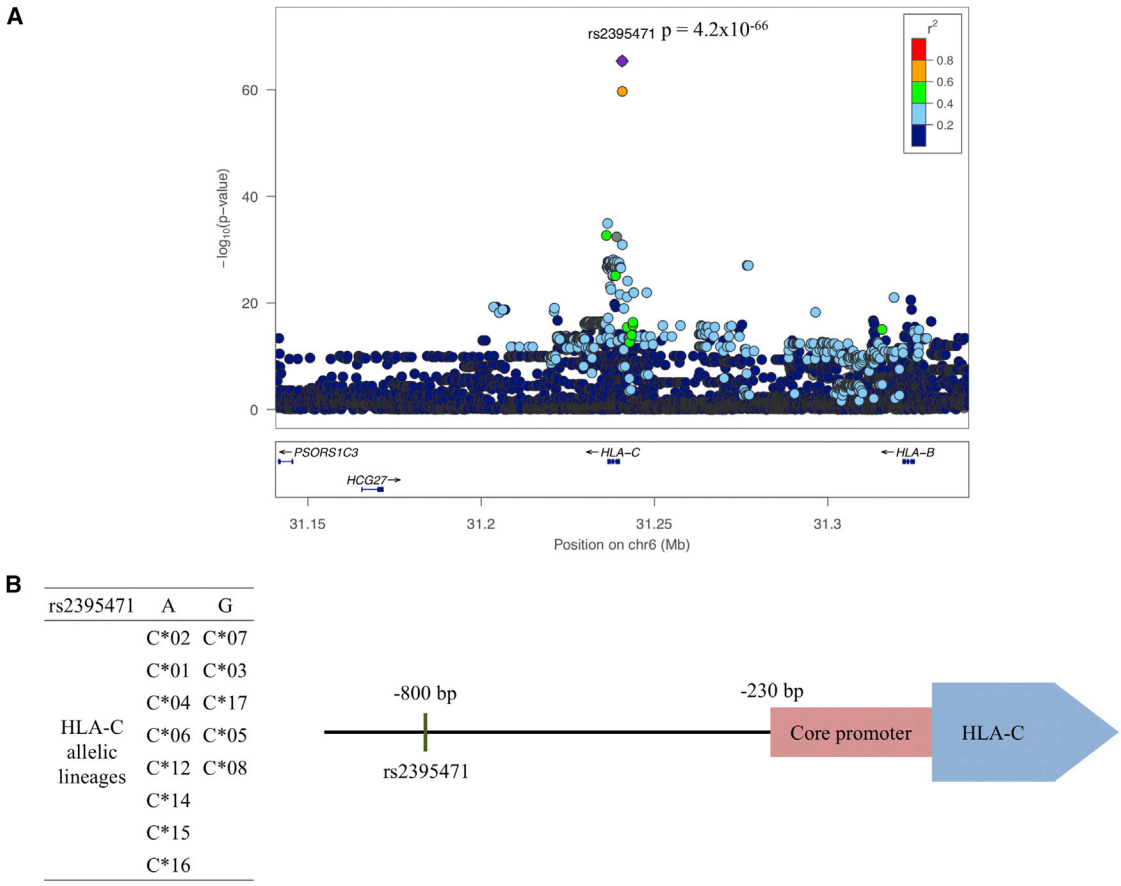

Figure 1. HLA-C Level Is Associated with rs2395471 in the Promoter Region of *HLA-C*

(A) Association between SNPs located in a 200 kb window around *HLA-C* and imputed HLA-C level. The Manhattan plot was created with LocusZoom.[15] Each dot represents a SNP for which the color depicts the linkage disequilibrium (LD) score (r^2) computed with PLINK[16] in the European 1KG dataset (n = 228). A complete list of LD scores between rs2395471 and the surrounding 1 Mb SNPs is presented in Table S2. Logistic regression tests for association between the 68,726 SNPs in the *HLA* locus and HLA-C level was performed in R (version 3.2.4) using the imputed HLA-C level values as a continuous variable.

(B) A list of *HLA-C* lineages that carry either the A or G at the rs2395471 SNP and a map of the *HLA-C* promoter region are shown.

only one other neighboring SNP, rs2249741, showed a similar level of significance ($p = 2.0 \times 10^{-60}$). The A versus G frequency at rs2395471 was fairly evenly distributed across *HLA-C* alleles (Figure 1B). We term these variants "imputed expression quantitative trait loci" (impeQTL) to distinguish them from those associating with expression levels of genes that were measured directly. impeQTL can only be used to identify candidates of expression modifiers in *cis* of the gene when its expression or protein level is imputed.

Genotyping of rs2395471 in two independent cohorts where HLA-C levels were measured directly by flow cytometry confirmed the association between this SNP and cell surface levels of HLA-C on CD3-positive cells: $p = 9.3 \times 10^{-16}$ in a cohort of 195 African Americans for whom levels were determined previously[10] (Figure 2A) and $p = 2.9 \times 10^{-17}$ in a cohort of 174 European Americans (Figure 2B). The explained variation based on the R^2 indicated that rs2395471 accounts for 28% and 36% of the HLA-C level variation in the African American and European American, respectively.[18] The rs2395471_A allele, which marks high level of HLA-C and is the ancestral allele, has a global frequency of 53% in the 1KG dataset ranging from 46% in East Asia to 62% in South Asia.

The association between rs2395471 and *HLA-C* expression was independently corroborated by a genome-wide expression quantitative trait loci (eQTL) study of *HLA-C* expression measured by qPCR in peripheral blood mononuclear cells (PBMCs) from 273 donors of European descent from Great Britain.[19] A locus containing two variants spaced 20 bp apart and in moderate linkage disequilibrium ($r^2 = 0.6$), namely rs2249741 (effect allele frequency [EAF] 48%, $p = 1.8 \times 10^{-24}$) and rs2395471 (EAF 36%, $p = 1.8 \times 10^{-20}$), was the most significantly associated locus genome wide, demonstrating that this locus probably represents the strongest eQTL for *HLA-C* expression across the genome (Figure S2A). We note that the differences in allele frequency may account for the relative difference in degree of statistical significance, but the effect sizes are comparable (rs2395471: beta = 0.72; SE = 0.074 versus rs2249741: beta = 0.74; SE = 0.071). The mean expression level of individual *HLA-C* alleles in

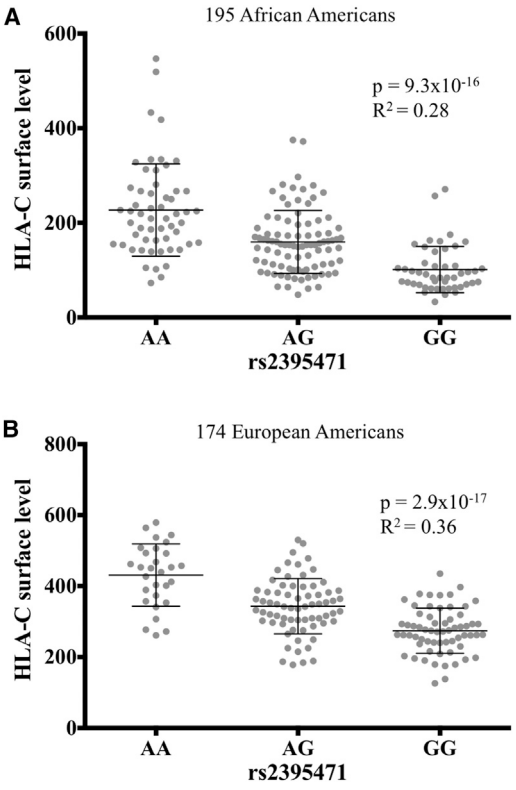

Figure 2. Association between rs2395471 and HLA-C Cell Surface Level

Cell surface level was measured directly in two independent cohorts of (A) 195 African Americans and (B) 174 European Americans. HLA-C cell surface level was measured on $CD3^+$ cells from peripheral blood by flow cytometry. For each plot, the p value from linear regression and R^2 (the level of variation explained by the SNP) are provided. The rs2395471 genotypes were determined by Sanger sequencing. The mean with standard deviation is represented. The National Health Institutes office of human subjects research protection approved this study.

this eQTL study (Figure S2B) was highly correlated with that published previously[12] ($p = 0.001$, $R = 0.95$), underscoring the reproducibility of the expression level measurements.

Given the location of the rs2395471 and rs2249741 variants in the promoter region of *HLA-C*, the possibility that one or both may alter TF binding was considered using the AliBaba online prediction tool.[20] A potential TF binding site for POU, a family of TFs containing well-conserved homeodomains,[21] was predicted to overlap with the rs2395471 variant. No TF binding site was predicted for rs2249741. The presence of a TF binding site within a sequence containing the rs2395471_A/G variant was confirmed by an electrophoretic mobility shift assay (EMSA) using rs2395471_A versus _G oligonucleotides and HeLa cell nuclear extracts (Figure 3A). POU-specific antibodies showed that Oct1 (*POU2F1* [MIM: 164175]), but not Oct2 (*POU2F2* [MIM: 164176]) or Oct3/4 (*POU5F1* [MIM: 164177]), can bind to the promoter region containing rs2395471 (Figure 3A). Further, the intensity of Oct1 binding was lower in the presence of oligonucleotide containing the G allele relative to that containing the A allele (Figures 3A and 3B). These experiments were reproduced using Jurkat cell line nuclear extract (Figure S3). Nuclear extracts of PBMCs from two healthy donors also demonstrated Oct1 binding within the region containing the rs2395471 variant (Figure 3C).

Given the limitations in terms of sensitivity and in vitro application, we proceeded to design a luciferase reporter assay to evaluate whether the rs2395471 alleles differentially impact *HLA-C* promoter activity. The promoter regions of two high expression *HLA-C* alleles that carry the rs2395471_A allele, *C*0102* and *C*0401*, were cloned into a pGL3 plasmid and alternates containing G at this position were generated by site-directed mutagenesis. In addition, the lower expression alleles *C*03:04* and *C*08:02*, both of which carry rs2395471_G, were used to generate a pGL3 construct together with an alternate that contains A at this position. Luciferase activity was assayed 48 hr after transfection of HeLa cells, revealing a significant decrease in promoter activity upon the single base pair change from A to G in a *C*01:02*, as well as a *C*04:01* background (Figure 4). These results strongly suggest that rs2395471 has a direct effect on *HLA-C* expression by impacting the binding affinity of the Oct1 TF. The promoter activities of the wild-type *C*03:04* and *C*08:02* alleles (i.e., rs2395471_G) were similar to the activity of *C*01:02* or *C*04:01* after the A residue in the Oct1 site was changed to G. However, replacement of the G residue with A in the Oct1 binding site of *C*03:04* and *C*08:02* did not increase the promoter activity of either allele (Figure 4), indicating that negative regulatory factors can dominate over the enhancer activity of the Oct1 binding site. Of note, wild-type *C*08:02* promoter activity was lower than that of *C*04:01* (Figure 4), in spite of their virtually identical levels observed previously on $CD3^+$ cells.[10] This discrepancy may be due to escape of *C*08:02* from miR-148a downregulation due to a deletion polymorphism in the 3′ UTR of this allele,[11] whereas *C*04:01* binds and is inhibited by miR-148a. These data demonstrate the complexity of *HLA-C* regulation, and this complexity further illustrates the remarkable observation that the Oct1 binding site variant has the most significant association with cell surface levels across *HLA-C* alleles overall.

Here we have shown that imputation of HLA-C level data allowed the identification of the *cis* variant rs2395471, located in the *HLA-C* promoter region, that contributes to determining allele-specific *HLA-C* mRNA expression and HLA-C cell surface levels. The association involving this impeQTL was confirmed in three independent cohorts (cumulative $n = 694$) with HLA-C expression and cell surface levels measured directly by either qPCR or flow cytometry. Importantly, this locus has the most significant effect on *HLA-C* expression levels genome wide based on qPCR data in which GWAS data were also available, indicating that it is likely the leading regulator of HLA-C levels.

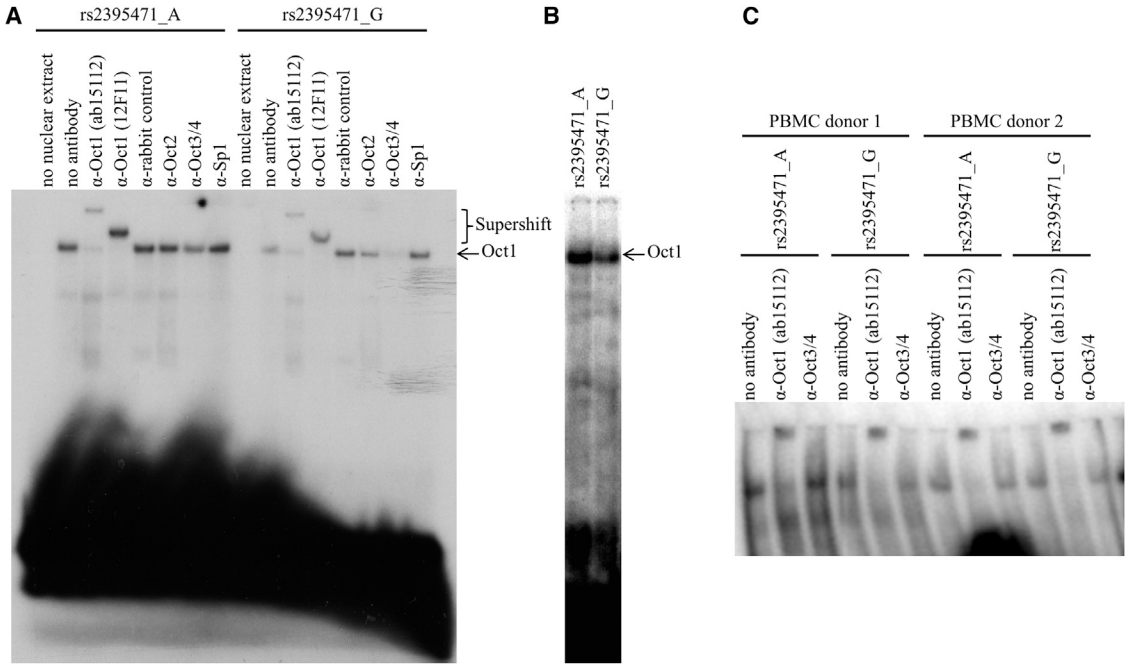

Figure 3. The Oct1 Transcription Factor Binds to the Genomic Region Containing rs2395471
Electrophoretic mobility shift assay (EMSA) was performed using nuclear protein extract from HeLa cell line (A and B) and PBMCs (C). Two oligonucleotides were designed, one containing the rs2395471_G allele and one containing the rs2395471_A allele (detailed methods can be found in Li et al.,[22] primers are available upon request).
(A) The presence of a shift (arrow) indicates that the genomic region containing rs2395471 binds a protein from the HeLa nuclear extract. A supershift is observed when anti-Oct1 antibodies (2 different clones) are added, but not anti-Oct2, anti-Oct3/4, or anti-Sp1 antibodies, indicating that Oct1 is a transcription factor that binds to the rs2395471 genomic region.
(B) HeLa nuclear extract binds to oligonucleotides containing rs2395471_A more strongly than it does to rs2395471_G.
(C) Nuclear proteins derived from PBMCs of two donors were extracted and incubated with the oligonucleotides and antibodies to anti-Oct1 and anti-Oct3/4. A supershift was detected only in the presence of the anti-Oct1 antibody, confirming the specificity of this TF binding site for Oct1.

POU2F1 (Oct1) is ubiquitously expressed across cell types[23–25] and recognizes the octamer DNA element ATGCAAAT and variations of it.[26] Its activity results in exceptionally diverse biological outcomes. Consistent with its ubiquitous expression, Oct1 is implicated in basal transcription regulation,[24,27,28] and it also appears to have a crucial role during embryogenesis, as Oct1 knockout mice are not viable.[29,30] Increased activity of Oct1 has been implicated in tumorigenesis, particularly epithelial tumors such as gastric (MIM: 613659) and breast (MIM: 114480) cancers,[31–35] and importantly in the context of HLA-C levels, Oct1 is linked to immune regulation of B cells, macrophages, T cells, and NK cells[31] by targeting production of cytokines (e.g., IL2),[36,37] pro-inflammatory mediators (e.g., NOS2),[38] and immunoglobulins.[39] Oct1 also has a role in signal response by serving as an adaptor for other TFs such as NF-κB.[28,40] An NF-κB binding site is predicted 633 bp downstream of the Oct1 binding site in the HLA-C core promoter region, raising the possibility that Oct1 may have a dual function in controlling *HLA-C* expression levels by regulating its basal expression and enhancing NF-κB-mediated effects on *HLA-C* expression levels upon cell activation.

Imputing HLA-C levels has led us to identify a locus, rs2395471, that accounts for up to 36% of the explained variation in HLA-C level. This SNP is not in significant LD with the insertion/deletion polymorphism in the 3′ UTR of *HLA-C* ($r^2 = 0.25$), which was previously shown to account in part for differential HLA-C levels as well.[11] The rs2395471 variant in combination with the *HLA-C* 3′ UTR variant explains up to 40% of the observed variability in measured HLA-C levels in European Americans. The approach described herein could be extended to other imputed expression or protein level data (e.g., additional *HLA* genes) in order to further our understanding of human gene regulation and their impact on disease.

Supplemental Data

Supplemental Data include three figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.09.023.

Acknowledgments

This project has been funded in whole or in part with federal funds from the Frederick National Laboratory for Cancer Research, under contract no. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement

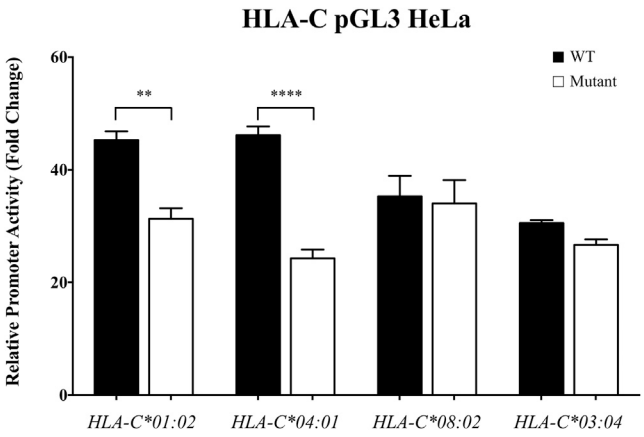

Figure 4. Impact of the rs2395471 on the *HLA-C* Promoter Activity Estimated by Luciferase Assay

The *HLA-C* promoters of two alleles carrying rs2395471_A, *C*01:02* and *C*04:01*, and two alleles carrying rs2395471_G, *C*03:04* and *C*08:02*, were cloned in a pGL3 plasmid. We mutated this position by site-directed mutagenesis to obtain new plasmids with a G nucleotide at the rs2395471 position for *C*01:02* and *C*04:01* and an A at this position for *C*03:04* and *C*08:02* (detailed methods can be found in Li et al.,[22] primers are available upon request). Both *HLA-C*01:02* and *C*04:01* promoters in which rs2395471_A was mutated to rs2395471_G show a significant decrease in promoter activity. Both *HLA-C*03:04* and *C*08:02* promoters in which rs2395471_G was mutated to rs2395471_A show no significant difference in promoter activity. Two-way ANOVA. $^{**}p < 0.01$, $^{****}p < 0.0001$. WT: wild-type. The mean of four replicated experiments with standard error is represented.

by the U.S. Government. This research was supported in part by the Intramural Research Program of the NIH, Frederick National Lab, Center for Cancer Research. J.C.K. is supported by European Research Council funding (grant agreement no. 281824) and the NIHR Oxford Biomedical Research Centre and Wellcome Trust (core facilities grant 090532/Z/09/Z). The Oxford cohort was supported by the Wellcome Trust (grants 074318 [J.C.K.], 088891 [B.P.F.], and 090532/Z/09/Z [core facilities Wellcome Trust Centre for Human Genetics including High-Throughput Genomics Group]), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no. 281824 (J.C.K.), the Medical Research Council (98082 [J.C.K]), and the NIHR Oxford Biomedical Research Centre. V.N. was supported by the Rhodes Trust.

Received: July 15, 2016
Accepted: September 29, 2016
Published: November 3, 2016

Web Resources

1000 Genomes, http://browser.1000genomes.org/index.html
AliBaba, http://www.gene-regulation.com/pub/programs/alibaba2/index.html
EIGENSOFT, https://www.hsph.harvard.edu/alkes-price/software/
GWAS Catalog, http://www.ebi.ac.uk/gwas/
LocusZoom, http://locuszoom.sph.umich.edu/locuszoom/
OMIM, http://www.omim.org/
PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/
R statistical software, http://www.r-project.org/

References

1. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. *42*, D1001–D1006.
2. Robinson, J., Halliwell, J.A., Hayhurst, J.D., Flicek, P., Parham, P., and Marsh, S.G. (2015). The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. *43*, D423–D431.
3. Hansen, T.H., and Bouvier, M. (2009). MHC class I antigen presentation: learning from viral evasion strategies. Nat. Rev. Immunol. *9*, 503–513.
4. Jones, E.Y., Fugger, L., Strominger, J.L., and Siebold, C. (2006). MHC class II proteins and disease: a structural perspective. Nat. Rev. Immunol. *6*, 271–282.
5. Bashirova, A.A., Martin, M.P., McVicar, D.W., and Carrington, M. (2006). The killer immunoglobulin-like receptor gene cluster: tuning the genome for defense. Annu. Rev. Genomics Hum. Genet. *7*, 277–300.
6. Apps, R., Meng, Z., Del Prete, G.Q., Lifson, J.D., Zhou, M., and Carrington, M. (2015). Relative expression levels of the HLA class-I proteins in normal and HIV-infected cells. J. Immunol. *194*, 3594–3600.
7. Parham, P., Lomen, C.E., Lawlor, D.A., Ways, J.P., Holmes, N., Coppin, H.L., Salter, R.D., Wan, A.M., and Ennis, P.D. (1988). Nature of polymorphism in HLA-A, -B, and -C molecules. Proc. Natl. Acad. Sci. USA *85*, 4005–4009.
8. Fellay, J., Shianna, K.V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., et al. (2007). A whole-genome association study of major determinants for host control of HIV-1. Science *317*, 944–947.
9. Wellcome Trust Case Control, C.; Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.
10. Apps, R., Qi, Y., Carlson, J.M., Chen, H., Gao, X., Thomas, R., Yuki, Y., Del Prete, G.Q., Goulder, P., Brumme, Z.L., et al. (2013). Influence of HLA-C expression level on HIV control. Science *340*, 87–91.
11. Kulkarni, S., Savan, R., Qi, Y., Gao, X., Yuki, Y., Bass, S.E., Martin, M.P., Hunt, P., Deeks, S.G., Telenti, A., et al. (2011). Differential microRNA regulation of HLA-C expression and its association with HIV control. Nature *472*, 495–498.
12. Kulkarni, S., Qi, Y., O'hUigin, C., Pereyra, F., Ramsuran, V., McLaren, P., Fellay, J., Nelson, G., Chen, H., Liao, W., et al. (2013). Genetic interplay between HLA-C and MIR148A in HIV control and Crohn disease. Proc. Natl. Acad. Sci. USA *110*, 20705–20710.
13. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
14. Gourraud, P.A., Khankhanian, P., Cereb, N., Yang, S.Y., Feolo, M., Maiers, M., Rioux, J.D., Hauser, S., and Oksenberg, J. (2014). HLA diversity in the 1000 genomes dataset. PLoS ONE *9*, e97282.
15. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.

16. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
17. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.
18. Draper, N.R., and Smith, H. (1998). Applied Regression Analysis (New York: Wiley).
19. Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O., and Knight, J.C. (2012). Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat. Genet. *44*, 502–510.
20. Grabe, N. (2002). AliBaba2: context specific identification of transcription factor binding sites. In Silico Biol. (Gedrukt) *2*, S1–S15.
21. Phillips, K., and Luisi, B. (2000). The virtuoso of versatility: POU proteins that flex to fit. J. Mol. Biol. *302*, 1023–1039.
22. Li, H., Wright, P.W., McCullen, M., and Anderson, S.K. (2016). Characterization of KIR intermediate promoters reveals four promoter types associated with distinct expression patterns of KIR subtypes. Genes Immun. *17*, 66–74.
23. Ferraris, L., Stewart, A.P., Kang, J., DeSimone, A.M., Gemberling, M., Tantin, D., and Fairbrother, W.G. (2011). Combinatorial binding of transcription factors in the pluripotency control regions of the genome. Genome Res. *21*, 1055–1064.
24. Sive, H.L., and Roeder, R.G. (1986). Interaction of a common factor with conserved promoter and enhancer sequences in histone H2B, immunoglobulin, and U2 small nuclear RNA (snRNA) genes. Proc. Natl. Acad. Sci. USA *83*, 6382–6386.
25. Staudt, L.M., Singh, H., Sen, R., Wirth, T., Sharp, P.A., and Baltimore, D. (1986). A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. Nature *323*, 640–643.
26. Reményi, A., Tomilin, A., Pohl, E., Lins, K., Philippsen, A., Reinbold, R., Schöler, H.R., and Wilmanns, M. (2001). Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. Mol. Cell *8*, 569–580.
27. Kim, M.H., and Peterson, D.O. (1995). Stimulation of basal transcription from the mouse mammary tumor virus promoter by Oct proteins. J. Virol. *69*, 4717–4726.
28. Pance, A. (2016). Oct-1, to go or not to go? That is the PolII question. Biochim. Biophys. Acta *1859*, 820–824.
29. Hwang, S.S., Kim, L.K., Lee, G.R., and Flavell, R.A. (2016). Role of OCT-1 and partner proteins in T cell differentiation. Biochim. Biophys. Acta *1859*, 825–831.
30. Wang, V.E.H., Schmidt, T., Chen, J., Sharp, P.A., and Tantin, D. (2004). Embryonic lethality, decreased erythropoiesis, and defective octamer-dependent promoter activation in Oct-1-deficient mice. Mol. Cell. Biol. *24*, 1022–1032.
31. Vázquez-Arreguín, K., and Tantin, D. (2016). The Oct1 transcription factor and epithelial malignancies: Old protein learns new tricks. Biochim. Biophys. Acta *1859*, 792–804.
32. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al.; METABRIC Group (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.
33. Hernández, P., Solé, X., Valls, J., Moreno, V., Capellá, G., Urruticoechea, A., and Pujana, M.A. (2007). Integrative analysis of a cancer somatic mutome. Mol. Cancer *6*, 13.
34. Jeong, S.H., Lee, Y.J., Cho, B.I., Ha, W.S., Choi, S.K., Jung, E.J., Ju, Y.T., Jeong, C.Y., Ko, G.H., Yoo, J., and Hong, S.C. (2014). OCT-1 overexpression is associated with poor prognosis in patients with well-differentiated gastric cancer. Tumour Biol. *35*, 5501–5509.
35. Wang, J., Yang, Y.H., Wang, A.Q., Yao, B., Xie, G., Feng, G., Zhang, Y., Cheng, Z.S., Hui, L., Dai, T.Z., et al. (2010). Immunohistochemical detection of the Raf kinase inhibitor protein in nonneoplastic gastric tissue and gastric cancer tissue. Med. Oncol. *27*, 219–223.
36. Pfeuffer, I., Klein-Hessling, S., Heinfling, A., Chuvpilo, S., Escher, C., Brabletz, T., Hentsch, B., Schwarzenbach, H., Matthias, P., and Serfling, E. (1994). Octamer factors exert a dual effect on the IL-2 and IL-4 promoters. J. Immunol. *153*, 5572–5585.
37. Ullman, K.S., Flanagan, W.M., Edwards, C.A., and Crabtree, G.R. (1991). Activation of early gene expression in T lymphocytes by Oct-1 and an inducible protein, OAP40. Science *254*, 558–562.
38. Kim, Y.M., Ko, C.B., Park, Y.P., Kim, Y.J., and Paik, S.G. (1999). Octamer motif is required for the NF-kappaB-mediated induction of the inducible nitric oxide synthase gene expression in RAW 264.7 macrophages. Mol. Cells *9*, 99–109.
39. Ballard, D.W., and Bothwell, A. (1986). Mutational analysis of the immunoglobulin heavy chain promoter region. Proc. Natl. Acad. Sci. USA *83*, 9626–9630.
40. Xie, Q. (1997). A novel lipopolysaccharide-response element contributes to induction of nitric oxide synthase. J. Biol. Chem. *272*, 14867–14872.