A distant trophoblast-specific enhancer controls HLA-G expression at the maternal–fetal interface

Leonardo M. R. Ferreira[a,b,1], Torsten B. Meissner[a], Tarjei S. Mikkelsen[a,c,d], William Mallard[a,c], Charles W. O'Donnell[d], Tamara Tilburgs[a], Hannah A. B. Gomes[a], Raymond Camahort[a], Richard I. Sherwood[e], David K. Gifford[f], John L. Rinn[a,b,c,d,g], Chad A. Cowan[a,c,d], and Jack L. Strominger[a,1]

[a]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138; [b]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138; [c]The Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142; [d]Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138; [e]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115; [f]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; and [g]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA 02156

Contributed by Jack L. Strominger, February 23, 2016 (sent for review February 5, 2016; reviewed by Koichi S. Kobayashi, Peter Parham, and Peter J. van den Elsen)

HLA-G, a nonclassical HLA molecule uniquely expressed in the placenta, is a central component of fetus-induced immune tolerance during pregnancy. The tissue-specific expression of HLA-G, however, remains poorly understood. Here, systematic interrogation of the *HLA-G* locus using massively parallel reporter assay (MPRA) uncovered a previously unidentified *cis*-regulatory element 12 kb upstream of *HLA-G* with enhancer activity, *Enhancer L*. Strikingly, clustered regularly-interspaced short palindromic repeats (CRISPR)/Cas9-mediated deletion of this enhancer resulted in ablation of HLA-G expression in JEG3 cells and in primary human trophoblasts isolated from placenta. RNA-seq analysis demonstrated that *Enhancer L* specifically controls HLA-G expression. Moreover, DNase-seq and chromatin conformation capture (3C) defined *Enhancer L* as a cell type-specific enhancer that loops into the *HLA-G* promoter. Interestingly, MPRA-based saturation mutagenesis of *Enhancer L* identified motifs for transcription factors of the CEBP and GATA families essential for placentation. These factors associate with *Enhancer L* and regulate HLA-G expression. Our findings identify long-range chromatin looping mediated by core trophoblast transcription factors as the mechanism controlling tissue-specific HLA-G expression at the maternal–fetal interface. More broadly, these results establish the combination of MPRA and CRISPR/Cas9 deletion as a powerful strategy to investigate human immune gene regulation.

human immune gene regulation | pregnancy | immune tolerance | MPRA | CRISPR/Cas9

During pregnancy, a semiallogeneic fetus expressing paternally derived antigens is nurtured for months without suffering rejection by the maternal immune system (1). This state of immune tolerance is established at a precise anatomical location, the placenta, a transient organ consisting of fetal trophoblasts and a specialized uterine mucosa, the decidua. During implantation, HLA-G$^+$ extravillous trophoblasts (EVTs) invade the maternal tissue, defining the boundary between mother and fetus (2).

HLA-G, a nonclassical nonpolymorphic major histocompatibility complex (MHC) class I molecule, is uniquely expressed by EVTs (3, 4), where it plays a central role in inducing immune tolerance. Several inhibitory receptors present on natural killer (NK) cells, the most abundant immune cell type at the maternal–fetal interface, and on myeloid cells, have been shown to bind to HLA-G (5–7). An HLA-G cycle between decidual NK cells and EVTs provides for both NK cell tolerance and antiviral immunity (8–10). Importantly, HLA-G is sufficient to inhibit NK cell cytotoxicity (11) and required to protect trophoblasts against NK cell-induced lysis (12). Several pregnancy-related disorders, including miscarriage, recurrent fetal loss, and preeclampsia, have been associated with polymorphisms resulting in reduced *HLA-G* expression levels (13, 14). Intriguingly, HLA-G expression has also been detected in tumor lesions, where it may facilitate immune evasion (15, 16). However, despite substantial effort, the mechanism by which the EVT-specific expression of HLA-G is obtained has remained elusive for more than two decades (13, 17, 18).

Tissue-specific gene expression is primarily regulated at the level of transcription by distant *cis*-regulatory elements—enhancers (19, 20). Traditionally, enhancer discovery has relied on examining features predictive of enhancer activity, such as chromatin accessibility, DNA and chromatin covalent modifications, and sequence conservation between species (21). This approach has been successfully used to gain important insights into immune gene regulation, such as the discovery of enhancers controlling the expression of murine Foxp3, a transcription factor governing the commitment and stability of regulatory T cells (22). However, substantial differences in regulatory sequences between species limit the ability to derive conclusions from model organisms regarding human gene regulation. In particular, the MHC locus differs significantly between mouse and humans (23), and *HLA-G* lacks a clear ortholog in mice.

In this study, we used an unbiased high-throughput approach, massively parallel reporter assay (MPRA) (24), to interrogate a

Significance

Successful pregnancy poses an immunological paradox, as the mother's immune system does not reject a fetus, even though it is a partially foreign tissue. Fetal extravillous trophoblasts (EVTs) deeply invade the uterus and interact with maternal immune cells without facing rejection. The nonclassical major histocompatibility complex (MHC) molecule HLA-G is essential for immune tolerance induction in pregnancy, yet the mechanism by which EVTs uniquely express HLA-G remains unknown. Using high-throughput *cis*-regulatory element dissection and genome editing tools, we discovered a remote enhancer essential for HLA-G expression in human EVTs, describing the basis for its selective expression at the maternal–fetal interface. These findings provide insight into immune tolerance induction during pregnancy and may yield new therapeutic targets for pregnancy-related disorders.

Author contributions: L.M.R.F., T.B.M., R.C., J.L.R., C.A.C., and J.L.S. designed research; L.M.R.F., T.B.M., T.S.M., W.M., C.W.O., H.A.B.G., R.C., and R.I.S. performed research; L.M.R.F., T.S.M., W.M., C.W.O., R.I.S., and D.K.G. contributed new reagents/analytic tools; L.M.R.F., T.B.M., T.T., H.A.B.G., R.C., D.K.G., J.L.R., C.A.C., and J.L.S. analyzed data; and L.M.R.F., T.B.M., and J.L.S. wrote the paper.

Reviewers: K.S.K., College of Medicine, Texas A&M Health Science Center; P.P., Stanford University School of Medicine; and P.J.v.d.E., Leiden University Medical Center.

Conflict of interest statement: C.A.C. is a founder and scientific advisor of CRISPR Therapeutics. J.L.S. is a consultant for King Abdulaziz University (Jeddah, Saudi Arabia).

Data deposition: RNA-seq data are available in the NCBI Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE79779).

[1]To whom correspondence may be addressed. Email: leonardoferreira@fas.harvard.edu or jlstrom@fas.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1602886113/-/DCSupplemental.

27-kb region spanning the HLA-G locus for functional activation of transcription. Our results uncover a private enhancer, which controls the tissue-specific expression of HLA-G at the maternal–fetal interface, and provide a relevant methodology to dissect human immune gene regulation without prior sequence knowledge.

Results

Identification of a Trophoblast-Specific Enhancer 12 kb Upstream of HLA-G. To systematically interrogate the *HLA-G* locus for active *cis*-regulatory elements, we set up a MPRA screen (24). For this purpose, 12,000 partially overlapping 121-bp-long elements (tiles) spanning 27 kb of the *HLA-G* locus were synthesized, coupled to unique DNA tags, and cloned into plasmids containing an invariant promoter and a firefly luciferase reporter gene. For greater confidence, two different promoters were used in parallel libraries, a strong promoter (SV40P) and a minimal TATA box synthetic promoter (minP). The resulting libraries were cotransfected into JEG3 cells, an HLA-G$^+$ choriocarcinoma cell line commonly used to model EVTs (25). To measure the relative enhancer activity of each tested element, we performed high-throughput sequencing and quantified the relative abundance of each element's tag reads in mRNA isolated from the transfected cells and in the pooled libraries. Enhancer activity was calculated as the median (cDNA count divided by the DNA count) of tags representing a tile, divided by the median ratio for all tags in a library. Nominal candidates were defined as any tile where enhancer activity measurements were >1 and P values were <0.05 for both biological replicates of each library transfection.

Our unbiased MPRA screen yielded several enhancer candidates upstream of *HLA-G* (Fig. 1*A*). The four most confident hits, indicated in Fig. 1 *A* and *B*, were then carried on for further analysis using classical luciferase reporter gene assays. The most confident candidate, located 12 kb upstream of the *HLA-G* gene, was the only tile with enhancer activity greater than 2 with both promoters tested, displaying the highest enhancer activity with minP (8.4) and second highest enhancer activity with SV40P (12.4) overall. This region specifically enhanced firefly luciferase activity upstream of the minimal promoter by 20-fold in HLA-G$^+$ JEG3 cells (Fig. 1*C*). We named this previously unidentified putative regulatory element *Enhancer L*, for being a long-range enhancer discovered with our unbiased enhancer screen. Importantly, *Enhancer L* was not active in HEK293T cells, an HLA-G–negative control cell line (Fig. 1*D*). Moreover, this cell type-specific activity pattern was maintained even when *Enhancer L* was cloned in an inverted orientation (Fig. 1*D*), a classical hallmark of an enhancer element (19). Of note, candidate numbers 3 and 4 from our MPRA screen, located near or even partially overlapping with *Enhancer L*, respectively, displayed negligible activity in JEG3 cells (Fig. 1*C*). Altogether, these observations suggest that *Enhancer L* corresponds to a narrowly defined regulatory region in the *HLA-G* locus that may confer tissue-specific HLA-G expression to trophoblasts.

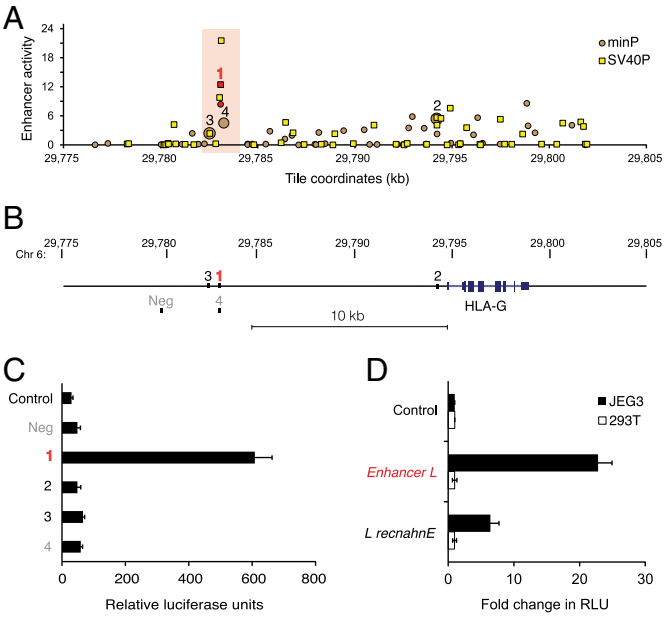

Fig. 1. *Enhancer L* is a trophoblast-specific enhancer upstream of *HLA-G*. (*A*) Massively parallel reporter assay (MPRA) covering the *HLA-G* locus. Enhancer activity of tiles upstream of the minP (circles) and SV40P (squares) promoters, calculated as the median count of any tags representing a tile, divided by the median ratio for all tags in the library, plotted against genomic coordinates (genome build hg19). Only tiles with $P < 0.05$ for both biological replicates are shown. Top-ranked tiles are numbered in decreasing order of confidence. The most confident hit (1) is in red type, and the region surrounding it is highlighted with a red box. (*B*) Schematic representing the location of the most confident hits from the MPRA relative to *HLA-G*, together with a negative control region (Neg). (*C*) *Enhancer L*, marked in red, was found to be active in JEG3 cells (HLA-G$^+$), as determined by luciferase reporter gene activity in combination with the minP promoter. Control, empty vector; Neg, negative control region. (*D*) *Enhancer L* remains active specifically in JEG3 cells when its direction is inverted. Control, empty vector; "*L recnahnE*," inverted *Enhancer L*; RLU, relative luciferase units. Error bars represent SEM of three independent experiments.

***Enhancer L* Is Essential for HLA-G Expression in JEG3 Cells.** Next, we sought to investigate whether *Enhancer L* modulates endogenous HLA-G expression. To directly target *Enhancer L* in JEG3 cells, we used a clustered regularly-interspaced short palindromic repeats (CRISPR)/Cas9 dual-guide approach (26, 27) by targeting two guide RNAs (gRNAs) to sites flanking *Enhancer L* (Fig. 2*A*). We used a *Streptococcus pyogenes* Cas9 linked via a self-cleaving 2A peptide to a green fluorescent protein (GFP) to facilitate identification of Cas9-expressing cells. GFP$^+$ cells were sorted and plated at clonal density and the emerging single-cell–derived colonies were transferred 10 d postplating into 96-well plates. PCR analysis of CRISPR/Cas9 targeted single-cell–derived clones was used to identify homozygous *Enhancer L* KO clones (Fig. 2*B*). We observed a clonal targeting efficiency of 29.5%, with homozygous deletions occurring at a frequency of 8.7%. Four independent *Enhancer L*-null clones and three WT clones were selected for further characterization. As expected, Sanger sequencing demonstrated excision of the DNA between the predicted Cas9 cleavage sites (three bases 5′ of the PAM sequence), with three out of four clones having the same exact deletion of 154 bp (Fig. 2*C*).

Strikingly, deletion of *Enhancer L* resulted in complete ablation of HLA-G expression, as determined by flow cytometry (Fig. 2*D*) and quantitative real-time PCR (qRT-PCR) (Fig. 2*E*). Surveying the whole genome for chromatin accessibility using genome-wide DNase-seq revealed that *Enhancer L* is located within a DNase I hypersensitivity site (DHS) in JEG3 cells (Fig. 2*C*), supporting the hypothesis that *Enhancer L* is indeed an active regulatory element in its endogenous chromatin context.

Deletion of *Enhancer L* in JEG3 Cells Uniquely Ablates HLA-G Expression. Following our observation that *Enhancer L* is required for HLA-G expression, we then asked whether *Enhancer L* acts specifically on HLA-G. Previous studies have identified enhancers that affect multiple genes spanning regions of hundreds of kilobases (28, 29). To investigate whether *Enhancer L* also regulates other genes in the HLA locus or elsewhere on chromosome 6, we sequenced polyA$^+$ mRNA from three *Enhancer L* KO JEG3 clones, as well as three WT clones and two independent samples of the parental JEG3 cell line as controls. RNA-seq confirmed that *HLA-G* is completely ablated across all KO clones (Fig. S1*A*), and that it is the only such gene within 2 Mb of *Enhancer L* (Fig. S1*B*), suggesting that *HLA-G* is the only direct *cis* target of *Enhancer L*. Looking beyond chromosome 6, transcriptome-wide analysis revealed statistically

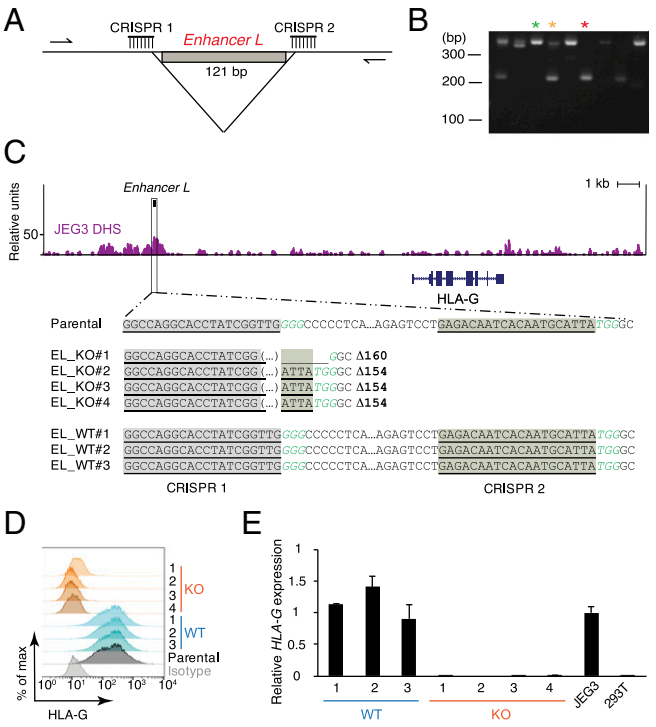

Fig. 2. *Enhancer L* is required for HLA-G expression in the JEG3 trophoblast cell line. (*A*) Dual-CRISPR guide strategy to delete *Enhancer L*. Arrows represent the primers used for PCR screening. (*B*) PCR screening of CRISPR/Cas9-targeted JEG3 single-cell–derived clones. Green*, wild type; yellow*, heterozygote; red*, null clone. (*C*) Sanger sequencing of four independent homozygous *Enhancer L* KO clones and three independent WT clones resulting from CRISPR/Cas9 targeting of *Enhancer L* (black box) in JEG3 cells using a dual-CRISPR guide RNA approach. Binding sites for the gRNAs targeting *Enhancer L* are underlined and shaded. PAM motifs are italicized in green type. *Enhancer L* is part of a DNase I hypersensitive site (DHS) in JEG3 cells, as determined by genome-wide DNase-seq. EL, *Enhancer L*. (*D*) Combined FACS histogram demonstrating complete ablation of HLA-G surface expression in *Enhancer L* KO JEG3 clones. (*E*) *HLA-G* transcript levels of *Enhancer L* KO clones, with JEG3 cells and HEK293T cells as controls. Gene expression normalized to *GAPDH* expression. Error bars represent SEM of replicates of a representative experiment ($n = 2$).

significant differences in the expression of 321 genes using Cuffdiff [false-discovery rate (FDR) < 0.05]. To rule out the possibility that these changes were caused by CRISPR/Cas9-induced off-target effects, we performed in silico off-target analyses of our *Enhancer L* gRNAs using the CRISPR design tool at crispr.mit.edu (30). The top 50 predicted off-target sites yielded maximum scores of 3.3 for gRNA 1, and 0.9 for gRNA 2 (out of 100), suggesting that the observed global changes in gene expression are not likely to be a result of off-target cleavage at these sites. Gene set enrichment analysis (GSEA) of the most differentially expressed genes revealed statistically significant enrichment (FDR < 0.05) for six gene sets, all of which are related to steroid hormone biosynthesis and G-protein–coupled receptor signaling, processes expected to play a role in trophoblast physiology. Pairwise comparison of all three experimental groups (WT, KO, parental JEG3), however, revealed that, despite the observed transcriptome-wide changes in gene expression, *HLA-G* was by far the most down-regulated gene upon *Enhancer L* deletion at the whole-transcriptome level (Fig. S1*C*), indicating that *Enhancer L* uniquely modulates HLA-G expression.

***Enhancer L* Is Required for HLA-G Expression in Primary EVTs.** To confirm the role of *Enhancer L* in primary human trophoblasts, we obtained villi from first-trimester human placental tissue and purified HLA-G$^+$ EVTs by flow cytometry (31). Cas9-2A-GFP and gRNAs targeting *Enhancer L* were successfully codelivered into primary EVTs using lentiviral particles, as assessed by GFP expression (Fig. 3*A*). As expected, *Enhancer L* deletion resulted in a significant decrease in *HLA-G* mRNA levels [74.12 ± 13.61% (SEM); $n = 3$] (Fig. 3*B*).

Loss of HLA-G surface expression as a result of lentiviral CRISPR/Cas9-mediated ablation of *Enhancer L* was first evaluated in JEG3 cells, which divide rapidly in culture. We observed complete loss of HLA-G surface expression 1 wk posttransduction in a large percentage of transduced cells [61.9 ± 1.93% (SEM); $n = 3$] (Fig. S2*A*). Detecting changes in HLA-G surface expression in primary EVTs, however, is hampered by the unusually long half-life of HLA-G protein on the cell membrane (32), and the fact that primary EVTs can only be cultured ex vivo for a short period (<5 d). Despite these technical limitations, we were able to detect a significant reduction in HLA-G surface expression 5 d after targeting *Enhancer L* in primary EVTs [60.71 ± 10.68% (SEM); $n = 3$] (Fig. 3 *C* and *D*, and Fig. S2*B*). Successful genomic deletion of *Enhancer L* was confirmed by PCR sequencing (Fig. S2 *C* and *D*). Our results demonstrate that *Enhancer L* is indeed necessary for HLA-G expression in primary human EVTs.

***Enhancer L* Is a Distant Regulatory Element That Loops into the HLA-G Proximal Promoter.** Next, we aimed to characterize the mechanism by which *Enhancer L* activates HLA-G expression at a distance. The current model of long-range gene regulation postulates that remote *cis*-regulatory elements come into close proximity to the promoters of the genes they regulate via chromatin looping (33). To test for the involvement of looping in *Enhancer L–HLA-G* promoter long-range communication, we carried out chromatin conformation capture (3C) assays in JEG3 and HLA-G–negative HEK293T cells (Fig. S3 *A* and *C*, and *Materials and Methods*). We detected a looping interaction between *Enhancer L* and the classical promoter of *HLA-G* specifically in JEG3 cells (Fig. S3*B*), confirming the nature of the resulting hybrid DNA molecule consisting of *Enhancer L* and the proximal promoter by sequencing (Fig. S3*D*). Of note, this looping interaction was absent in

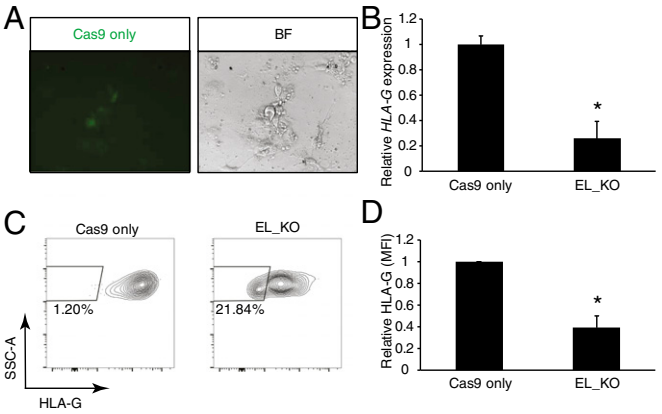

Fig. 3. *Enhancer L* is necessary for HLA-G expression in primary extravillous trophoblasts (EVTs). (*A*) Transduction of first-trimester HLA-G$^+$ EVTs with lentiviral Cas9 and *Enhancer L* gRNAs, assessed based on GFP expression. BF, bright-field (40× magnification). (*B*) Reduction of *HLA-G* expression at the mRNA level following *Enhancer L* deletion. Bars represent average ± SEM of three independent experiments. Gene expression normalized to *GAPDH* expression. *$P < 0.05$, paired Student's *t* test. EL, *Enhancer L*. (*C*) *Enhancer L* deletion leads to significant reduction in HLA-G surface expression in primary EVTs, as assessed by FACS. One representative experiment is shown ($n = 3$). (*D*) Significant reduction in HLA-G surface expression upon *Enhancer L* deletion in primary EVTs [mean fluorescence intensity (MFI)]. Bars represent average ± SEM of three independent experiments. *$P < 0.05$, paired Student's *t* test.

HEK293T cells (Fig. S3*B*), in agreement with the lack of *Enhancer L* activity in these cells (Fig. 1*D*).

MPRA-Based Scanning Mutagenesis Reveals Motifs Controlling *Enhancer L* Activity. Having established *Enhancer L* as a bona fide enhancer upstream of *HLA-G*, we sought to identify the transcriptional regulators that mediate its action. To our surprise, truncation of *Enhancer L* invariably led to loss of enhancer activity in firefly luciferase reporter gene assays, suggesting multiple active motifs spread across its length (Fig. S4*A*). To fine map the active regulatory motifs responsible for *Enhancer L* activity, we carried out an MPRA-based scanning mutagenesis at the single-base pair resolution (24). In brief, we generated a total of 12,000 *Enhancer L* variants, representing all possible single substitutions, as well as small insertions or deletions at all positions. To reduce experimental noise, each variant was coupled to 16 tags on average, for a total of 200,000 distinct variant–tag combinations. As before, this complex library was cotransfected into JEG3 cells, followed by RNA harvesting and sequencing analysis. This fine mapping of *Enhancer L* led to the identification of five putative regulatory motifs, consistent across both promoters tested (SV40P and minP) (Fig. 4*A* and Fig. S4*B*). Reporter gene assays with truncated versions of *Enhancer L* lacking each one of these motifs (M1 through M5) showed that each one of them is essential for optimal *Enhancer L* activity in JEG3 cells (Fig. S4*C*). Subsequent in silico analysis using the TRANSFAC database (34) predicted binding of CEBP and GATA family transcription factors within these five motifs (Fig. 4*A* and Fig. S4*B*).

CEBP and GATA Factors Regulate Trophoblast-Specific HLA-G Expression. Motif sequence analysis alone does not allow discrimination between different members of transcription factor families. We reasoned that the transcription factors controlling HLA-G expression via *Enhancer L* must be highly expressed specifically in HLA-G$^+$ trophoblasts. Microarray analysis of primary cells isolated from human placental tissue, and JEG3 cells (31), revealed that *CEBPA*, *CEBPB*, *GATA2*, and *GATA3* are the most highly expressed genes within their respective transcription factor families (Fig. 4*B*). Our whole-transcriptome RNA-seq analysis (Fig. 4*C*) confirmed high expression levels of *CEBPB*, *GATA2*, and *GATA3* in JEG3 cells. In addition, a survey of publicly available gene expression profiles (*BioGPS*) revealed that these three transcription factors are highly coexpressed in human placenta, and also more restricted in expression to this tissue than any other CEBP or GATA transcription factor family member. Importantly, CEBPβ, GATA2, and GATA3 have been implicated in murine placental development and trophoblast-specific gene regulation (35–37), making them strong candidates for transcriptional regulators of HLA-G expression in human trophoblasts.

To test our prediction, we sought to determine whether CEBPβ, GATA2, and GATA3 bind to *Enhancer L*. Indeed, chromatin immunoprecipitation (ChIP) using validated ChIP-grade antibodies, followed by qPCR analysis (ChIP-qPCR), revealed a 40-fold enrichment for CEBPβ on *Enhancer L* (Fig. 4*D*). Similarly, a significant enrichment for GATA2 and GATA3 (fivefold) was detected on *Enhancer L* (Fig. 4*E*), indicating that, in JEG3 cells, endogenous CEBPβ, GATA2, and GATA3 associate with *Enhancer L*. In addition, all three factors were found to bind to the proximal promoter of *HLA-G* (Fig. 4 *D* and *E*), providing further evidence for the existence of a chromatin loop between *Enhancer L* and the core promoter of *HLA-G*, possibly established by GATA2 and GATA3 (38, 39). Of note, Pol II associated with both *Enhancer L* and the *HLA-G* core promoter (Fig. S4*D*), suggesting that active transcription is involved in the formation of this long-range chromatin loop. Consistent with a role in *HLA-G* transcriptional activation, transient overexpression of CEBPβ, GATA2, and GATA3 individually in JEG3 cells led to an up to eightfold increase in *HLA-G* expression, indicating that these three factors are transcriptional activators of *HLA-G* expression (Fig. 4*F*). Taken together, our data support a model where CEBPβ and GATA2/3

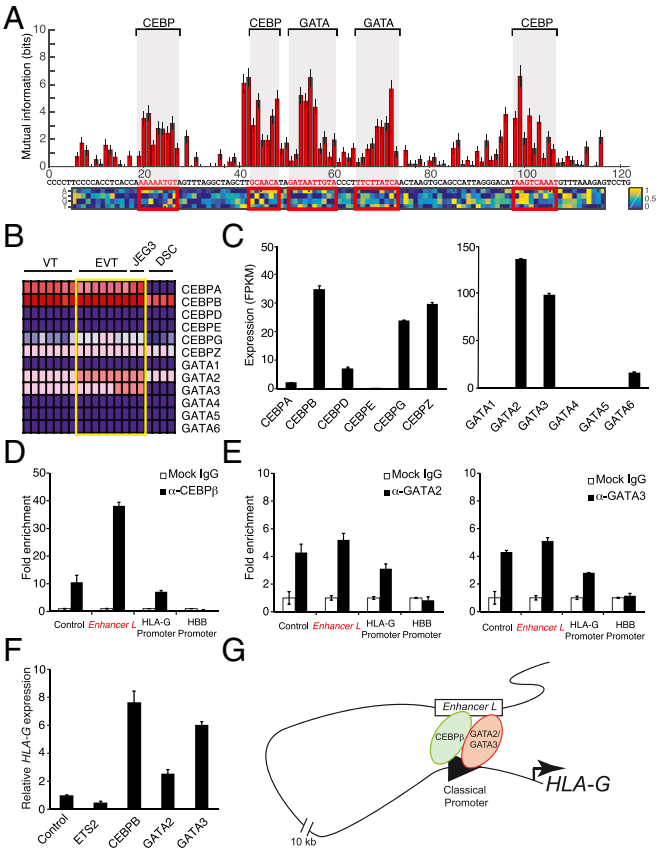

Fig. 4. Trophoblast CEBP and GATA factors regulate HLA-G expression. (*A*) Identification of five putative regulatory motifs required for *Enhancer L* activity using MPRA-based scanning mutagenesis in combination with an SV40 promoter. Red bars indicate a significant change from original *Enhancer L* activity (Mann–Whitney *U* test, 5% FDR); blue bars, not significant. The matrix represents the estimated additive contribution of each nucleotide to *Enhancer L* activity. Transcription factor binding site prediction was performed using the TRANSFAC database. (*B*) Expression levels of genes belonging to the two transcription factor families predicted to bind to *Enhancer L*, CEBP and GATA, according to published microarray data. The heat map was generated using *GenePattern*, with dark blue representing lowest expression, and dark red, highest expression. DSC, decidual stromal cells. (*C*) CEBP and GATA gene expression levels in JEG3 cells, as determined by whole-transcriptome RNA-seq. FPKM, fragments per kilobase of exon per million fragments mapped. (*D* and *E*) CEBPβ, the most highly expressed CEBP transcription factor in JEG3 cells (*D*), and GATA2 and GATA3, the most abundant GATA factors in JEG3 cells (*E*), associate with *Enhancer L* and with the *HLA-G* classical promoter, as assessed by ChIP-qPCR ($n = 2$). Control, positive control region predicted to be bound by the respective transcription factor according to *ENCODE* data; *HBB* promoter, negative control. (*F*) Ectopic expression of CEBPβ, GATA2, or GATA3 up-regulate HLA-G expression in JEG3 cells, as measured by qPCR. The transcription factor ETS2 was used as a negative control. Control, empty vector. Error bars represent SEM of replicates of a representative experiment ($n = 2$). (*G*) Proposed model of trophoblast-specific HLA-G transcriptional regulation by CEBPβ, GATA2, and GATA3 via *Enhancer L*.

mediate long-range chromatin interactions between *Enhancer L* and the classical promoter of *HLA-G* (Fig. 4*G*), driving HLA-G expression specifically in EVTs at the maternal–fetal interface.

Discussion

Genome-wide association studies (GWAS) have uncovered an astonishing number of disease-associated noncoding loci (40), posing a challenge to functionally validate and characterize putative regulatory elements. MPRA represents an unbiased high-throughput method for de novo discovery and validation of *cis*-regulatory

regions. In this study, the most confident candidate from our MPRA screen, located 12 kb upstream of *HLA-G,* was found to be active specifically in the HLA-G$^+$ JEG3 choriocarcinoma cell line (Fig. 1), suggesting that it may be involved in tissue-specific *HLA-G* transcriptional regulation. Indeed, CRISPR/Cas9 genome editing revealed that this previously unidentified enhancer, *Enhancer L*, is essential for trophoblast expression of HLA-G (Figs. 2 and 3, and Figs. S1 and S2).

Previous studies established that MHC gene expression is mainly controlled at the level of a conserved proximal promoter. Upon interaction with a transcriptional activator—CIITA for class II and NLRC5 for class I genes—a multiprotein transcription factor complex is assembled, forming the MHC enhanceosome (17, 41, 42). Even though the enhanceosome is essential for basal and induced expression of MHC class I genes, its relevance in trophoblasts is uncertain: EVTs do not express NLRC5 or CIITA (31) and the *HLA-G* proximal promoter harbors several nonfunctional motifs (18), suggesting that tissue-specific HLA-G expression is mediated by a distinct mechanism. Although several studies have described *cis*-regulatory regions involved in *HLA-G* transcriptional regulation (13, 43, 44), the present study is the first (to our knowledge) to report a noncoding sequence, *Enhancer L*, absolutely required for the tissue-specific expression of HLA-G in trophoblasts.

Interestingly, *Enhancer L* is contained within a long terminal repeat (LTR) sequence, LTR7 (45), associated with a human endogenous retroviral element (ERV), ERV1, as indicated in Fig. S5*A*. LTR sequences have been co-opted by mammalian genomes as regulatory elements, especially in the placenta (46). Well-known examples include the placenta-specific promoter of *CYP19* (47) and MER20, regulatory sequences found upstream of progesterone-responsive genes essential for decidualization (48). *Enhancer L* sequence is unique in the human genome and well conserved across apes and Old World monkeys, yet absent in New World monkeys (Fig. S5 *A* and *B*), where HLA-G appears to be a classical MHC molecule (49–51). Intriguingly, the orangutan genome, the only ape genome containing a functional *HLA-G* promoter (X2 and Y *cis*-elements matching those in the *HLA-A* promoter; Fig. S5*C*), does not harbor the *Enhancer L* sequence. In addition, similar to New World monkeys, the orangutan HLA-G ortholog is a polymorphic MHC molecule. Perhaps in orangutans, because they are predominantly monogamous and thus less exposed to allogeneic fetuses (49), *HLA-G* functions as a classical antigen-presenting molecule. The observation that *Enhancer L* is only found in genomes that lack a functional *HLA-G* classical promoter raises the possibility that a retroviral element was co-opted during evolution to function in trophoblast-specific tolerogenic MHC expression.

Previous literature suggests that differential expression of transcription factors plays a role in cell type-specific HLA-G transcription. The identity of such factors, however, has remained elusive (52, 53). In our study, MPRA-based saturation mutagenesis allowed us to fine map the regulatory elements responsible for *Enhancer L* activity, ultimately pointing toward CEBP and GATA factors as candidates for transcriptional activators of HLA-G expression in trophoblasts (Fig. 4). Indeed, ChIP and transient transfection studies revealed that CEBPβ, GATA2, and GATA3 associate with *Enhancer L* (Fig. 4 *D* and *E*) and are positive regulators of *HLA-G* expression (Fig. 4*F*).

3C revealed that *Enhancer L* loops across a 12-kb distance into the classical promoter of *HLA-G* (Fig. S3). Consistent with this long-range chromatin interaction, genome-wide DNase-seq demonstrated that *Enhancer L* is part of a DHS specifically in HLA-G$^+$ JEG3 cells (Fig. 2*C* and Fig. S6). Publicly available ChIP-seq data indicates CTCF binding flanking *Enhancer L* and the *HLA-G* coding sequence (Fig. S6). This CTCF binding pattern suggests the existence of an insulated chromatin domain (54) for *HLA-G* transcriptional regulation, corroborated by our observation that *Enhancer L* deletion does not significantly alter the expression of any gene other than *HLA-G* on chromosome 6 (Fig. S1). Interestingly, a long-range chromatin interaction mediated by the insulator CTCF has been described in the MHC class II locus (55). Our data suggest that the looping interaction between *Enhancer L* and the promoter of *HLA-G* is mediated by GATA2/3, possibly in association with CEBPβ (38, 39, 56).

In conclusion, we have demonstrated that trophoblast HLA-G expression is contingent upon the activity of a remote enhancer, *Enhancer L*. Our data are consistent with a model where CEBPβ and GATA2/3 associate with *Enhancer L*, are recruited to the core promoter of *HLA-G* via chromatin looping, and up-regulate HLA-G expression (Fig. 4*G*). These findings establish chromatin looping mediated by lineage-specific transcription factors as a mechanism governing tissue-specific immune gene expression at the maternal–fetal interface. Future studies further dissecting the transcriptional regulation of *HLA-G* will not only shed light on immune privilege during pregnancy, but may also enable us to specifically control HLA-G expression to induce tolerance in transplantation therapies.

Materials and Methods

All of the human tissue used for this research was deidentified, discarded clinical material. The Committee on the Use of Human Subjects [the Harvard institutional review board (IRB)] determined that this use of all of this human material is exempt from the requirements of IRB review.

Cell Culture. JEG3 and HEK293T cells were cultured in RPMI-1640 medium (Gibco) supplemented with 10% (vol/vol) FBS, Glutamax, and penicillin–streptomycin. Transfections were carried out using FuGENE 6 (Promega) according to the manufacturer's instructions and analyzed 48 h posttransfection.

Flow Cytometry. Cells were harvested, blocked in 4% (vol/vol) FBS for 30 min, stained with HLA-G PE (clone MEMG/9; Abcam) in 1% FBS for 1 h, washed thrice, and resuspended in 1% FBS. Cells were acquired using either a FACSCalibur or an LSR-II instrument (BD Biosciences) and analyzed with FlowJo (Tree Star) software.

qRT-PCR Analysis. Total RNA was isolated using TRIzol (Life Technologies), according to manufacturer's instructions. A total of 1,500 ng of RNA was used for cDNA synthesis with the qScript cDNA SuperMix (Quanta Biosciences). A total of 30 ng of cDNA was used per qRT-PCR, performed using SYBR Green (Life Technologies) on a ViiA7 system real-time PCR system (Life Technologies). Target gene expression levels were normalized to *GAPDH*. Primer pairs used are listed in Table S1.

Details of molecular biology, MPRA, luciferase reporter gene assay, genome-wide DNase-seq, 3C, CRISPR/Cas9 genome editing, transcriptome-wide RNA-seq, first-trimester primary EVT isolation and transduction, and ChIP (ChIP-qPCR) experiments are given in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We are indebted to Xiaolan Zhang and Alexander Melnikov for help with the MPRA, and Chiara Gerhardinger for help with RNA-seq library preparation. L.M.R.F. is supported by the Molecules, Cells and Organisms PhD training grant (Department of Molecular and Cellular Biology, Harvard University) and a doctoral fellowship from the Portuguese Foundation for Science and Technology. This project was supported by National Human Genome Research Institute Grant R01HG006785 (to T.S.M.); Grant 1K01DK101684-01, Human Frontier Science Program (to R.I.S.); National Heart, Lung, and Blood Institute (NHLBI) Grant U01HL100408, NHLBI Grant U01HL10744, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Grant R01DK097768, and NIDDK Grant R01DK072041 (to C.A.C.); and NIH/National Institute of Allergy and Infectious Diseases Grant AI053330 (to J.L.S.).

1. Medawar PB (1953) Some immunological and endocrinological problems raised by the evolution of viviparity in vertebrates. *Symp Soc Exp Biol* 44:320–338.
2. Moffett A, Loke C (2006) Immunology of placentation in eutherian mammals. *Nat Rev Immunol* 6(8):584–594.
3. Ellis SA, Sargent IL, Redman CW, McMichael AJ (1986) Evidence for a novel HLA antigen found on human extravillous trophoblast and a choriocarcinoma cell line. *Immunology* 59(4):595–601.
4. Kovats S, et al. (1990) A class I antigen, HLA-G, expressed in human trophoblasts. *Science* 248(4952):220–223.

5. Shiroishi M, et al. (2003) Human inhibitory receptors Ig-like transcript 2 (ILT2) and ILT4 compete with CD8 for MHC class I binding and bind preferentially to HLA-G. *Proc Natl Acad Sci USA* 100(15):8856–8861.
6. Rajagopalan S, Long EO (1999) A human histocompatibility leukocyte antigen (HLA)-G-specific receptor expressed on all natural killer cells. *J Exp Med* 189(7):1093–1100.
7. Koopman LA, et al. (2003) Human decidual natural killer cells are a unique NK cell subset with immunomodulatory potential. *J Exp Med* 198(8):1201–1212.
8. Tilburgs T, Evans JH, Crespo AC, Strominger JL (2015) The HLA-G cycle provides for both NK tolerance and immunity at the maternal–fetal interface. *Proc Natl Acad Sci USA* 112(43):13312–13317.
9. Caumartin J, et al. (2007) Trogocytosis-based generation of suppressive NK cells. *EMBO J* 26(5):1423–1433.
10. Rajagopalan S, Long EO (2012) Cellular senescence induced by CD158d reprograms natural killer cells to promote vascular remodeling. *Proc Natl Acad Sci USA* 109(50): 20596–20601.
11. Pazmany L, et al. (1996) Protection from natural killer cell-mediated lysis by HLA-G expression on target cells. *Science* 274(5288):792–795.
12. Rouas-Freiss N, Gonçalves RM, Menier C, Dausset J, Carosella ED (1997) Direct evidence to support the role of HLA-G in protecting the fetus from maternal uterine natural killer cytolysis. *Proc Natl Acad Sci USA* 94(21):11520–11525.
13. Moreau P, Flajollet S, Carosella ED (2009) Non-classical transcriptional regulation of HLA-G: An update. *J Cell Mol Med* 13(9B):2973–2989.
14. Quach K, Grover SA, Kenigsberg S, Librach CL (2014) A combination of single nucleotide polymorphisms in the 3′untranslated region of HLA-G is associated with pre-eclampsia. *Hum Immunol* 75(12):1163–1170.
15. Rouas-Freiss N, Moreau P, Ferrone S, Carosella ED (2005) HLA-G proteins in cancer: Do they provide tumor cells with an escape mechanism? *Cancer Res* 65(22):10139–10144.
16. Wiendl H, et al. (2002) A functional role of HLA-G expression in human gliomas: An alternative strategy of immune escape. *J Immunol* 168(9):4772–4780.
17. Kobayashi KS, van den Elsen PJ (2012) NLRC5: A key regulator of MHC class I-dependent immune responses. *Nat Rev Immunol* 12(12):813–820.
18. Solier C, et al. (2001) HLA-G unique promoter region: Functional implications. *Immunogenetics* 53(8):617–625.
19. Banerji J, Rusconi S, Schaffner W (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27(2 Pt 1):299–308.
20. Heinz S, Romanoski CE, Benner C, Glass CK (2015) The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* 16(3):144–154.
21. Consortium EP; ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
22. Zheng Y, et al. (2010) Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. *Nature* 463(7282):808–812.
23. Yuhki N, et al. (2003) Comparative genome organization of human, murine, and feline MHC class II region. *Genome Res* 13(6A):1169–1179.
24. Melnikov A, et al. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* 30(3): 271–277.
25. Pattillo RA, Gey GO (1968) The establishment of a cell line of human hormone-synthesizing trophoblastic cells in vitro. *Cancer Res* 28(7):1231–1236.
26. Mandal PK, et al. (2014) Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. *Cell Stem Cell* 15(5):643–652.
27. Meissner TB, Mandal PK, Ferreira LM, Rossi DJ, Cowan CA (2014) Genome editing for human gene therapy. *Methods Enzymol* 546:273–295.
28. Link N, Kurtz P, O'Neal M, Garcia-Hughes G, Abrams JM (2013) A p53 enhancer region regulates target genes through chromatin conformations in cis and in trans. *Genes Dev* 27(22):2433–2438.
29. Melo CA, et al. (2013) eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol Cell* 49(3):524–535.
30. Hsu PD, et al. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31(9):827–832.
31. Tilburgs T, et al. (2015) Human HLA-G+ extravillous trophoblasts: Immune-activating cells that interact with decidual leukocytes. *Proc Natl Acad Sci USA* 112(23):7219–7224.
32. Davis DM, et al. (1997) Impaired spontaneous endocytosis of HLA-G. *Eur J Immunol* 27(10):2714–2719.
33. Sexton T, Cavalli G (2015) The role of chromosome domains in shaping the functional genome. *Cell* 160(6):1049–1059.
34. Matys V, et al. (2006) TRANSFAC and its module TRANSCompel: Transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 34(Database issue):D108–D110.
35. Bégay V, Smink J, Leutz A (2004) Essential requirement of CCAAT/enhancer binding proteins in embryogenesis. *Mol Cell Biol* 24(22):9744–9751.
36. Ma GT, Linzer DI (2000) GATA-2 restricts prolactin-like protein A expression to secondary trophoblast giant cells in the mouse. *Biol Reprod* 63(2):570–574.
37. Cheng YH, Handwerger S (2005) A placenta-specific enhancer of the human syncytin gene. *Biol Reprod* 73(3):500–509.
38. Deng W, et al. (2012) Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149(6):1233–1244.
39. Chen Y, et al. (2012) DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. *Cell Rep* 2(5):1197–1206.
40. Maurano MT, et al. (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337(6099):1190–1195.
41. Meissner TB, et al. (2010) NLR family member NLRC5 is a transcriptional regulator of MHC class I genes. *Proc Natl Acad Sci USA* 107(31):13794–13799.
42. Steimle V, Siegrist CA, Mottet A, Lisowska-Grospierre B, Mach B (1994) Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. *Science* 265(5168):106–109.
43. Ikeno M, et al. (2012) LINE1 family member is negative regulator of HLA-G expression. *Nucleic Acids Res* 40(21):10742–10752.
44. Gobin SJ, Biesta P, de Steenwinkel JE, Datema G, van den Elsen PJ (2002) HLA-G transactivation by cAMP-response element-binding protein (CREB). An alternative transactivation pathway to the conserved major histocompatibility complex (MHC) class I regulatory routes. *J Biol Chem* 277(42):39525–39531.
45. Kelley D, Rinn J (2012) Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol* 13(11):R107.
46. Chuong EB, Rumi MA, Soares MJ, Baker JC (2013) Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* 45(3):325–329.
47. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet* 19(10):530–536.
48. Lynch VJ, Leclerc RD, May G, Wagner GP (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* 43(11):1154–1159.
49. Arnaiz-Villena A, et al. (1999) Evolution of MHC-G in primates: A different kind of molecule for each group of species. *J Reprod Immunol* 43(2):111–125.
50. Slukvin II, Lunn DP, Watkins DI, Golos TG (2000) Placental expression of the non-classical MHC class I molecule Mamu-AG at implantation in the rhesus monkey. *Proc Natl Acad Sci USA* 97(16):9104–9109.
51. Adams EJ, Parham P (2001) Species-specific evolution of MHC class I genes in the higher primates. *Immunol Rev* 183:41–64.
52. Moreau P, et al. (1997) HLA-G gene transcriptional regulation in trophoblasts and blood cells: Differential binding of nuclear factors to a regulatory element located 1.1 kb from exon 1. *Hum Immunol* 52(1):41–46.
53. Moreau P, et al. (1998) Specific binding of nuclear factors to the HLA-G gene promoter correlates with a lack of HLA-G transcripts in first trimester human fetal liver. *Hum Immunol* 59(12):751–757.
54. Dowen JM, et al. (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159(2):374–387.
55. Majumder P, Gomez JA, Chadwick BP, Boss JM (2008) The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J Exp Med* 205(4):785–798.
56. Tong Q, Tsai J, Tan G, Dalgin G, Hotamisligil GS (2005) Interaction between GATA and the C/EBP family of transcription factors is critical in GATA-mediated suppression of adipocyte differentiation. *Mol Cell Biol* 25(2):706–715.
57. Mali P, et al. (2013) RNA-guided human genome engineering via Cas9. *Science* 339(6121):823–826.
58. Ding Q, et al. (2013) Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* 12(4):393–394.
59. Kabadi AM, Ousterout DG, Hilton IB, Gersbach CA (2014) Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res* 42(19): e147.
60. Sanjana NE, Shalem O, Zhang F (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 11(8):783–784.
61. Sherwood RI, et al. (2014) Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat Biotechnol* 32(2):171–178.
62. Hesselberth JR, et al. (2009) Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods* 6(4):283–289.
63. John S, et al. (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* 43(3):264–268.
64. Gavrilov A, et al. (2009) Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. *Methods Mol Biol* 567:171–188.
65. Trapnell C, et al. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* 31(1):46–53.
66. Sun L, et al. (2013) Long noncoding RNAs regulate adipogenesis. *Proc Natl Acad Sci USA* 110(9):3387–3392.
67. Kim D, et al. (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.
68. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.
69. Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102(43):15545–15550.
70. Mootha VK, et al. (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34(3): 267–273.

Supporting Information

Ferreira et al. 10.1073/pnas.1602886113

SI Materials and Methods

Molecular Biology. For CRISPR/Cas9 genome editing in JEG3 cells, a human codon-optimized *Streptococcus pyogenes* Cas9 gene with a C-terminal nuclear localization signal (57) subcloned into a CAG expression plasmid upstream of a 2A-GFP (58) was used. The guide RNAs (gRNAs) were cloned into a separate plasmid containing the human U6 polymerase III promoter (57) using BbsI restriction sites. For lentiviral delivery of CRISPR/Cas9 to primary extravillous trophoblasts (EVTs), Cas9 was instead expressed from a human UbC promoter and upstream of a T2A-GFP (59). gRNAs were subcloned into the lentiGuide-puro vector (60). Lentiviral production was carried out in HEK293T cells using psPAX2 and VSV-G as packaging plasmids, as described (59). Tested transcription factor genes were amplified from JEG3 cDNA and directionally cloned into a plasmid containing a CMV promoter upstream of an IRES-GFP.

Massively Parallel Reporter Assay. First, 12,000 oligonucleotides tiling the *HLA-G* locus (27 kb) coupled to distinguishing tags were generated using microarray-based DNA synthesis. The 121-bp-long tiles and tags are separated by two common restriction sites. The oligonucleotides were then PCR amplified from universal primer sites and directionally cloned into a pGL4 plasmid backbone (Promega) using Gibson assembly. An invariant promoter–firefly luciferase segment containing either a minimal TATA box weak (minP) or strong (SV40P) promoter was then inserted between the tiles and tags by double digestion and directional ligation. The resulting reporter plasmid pools were cotransfected into JEG3 cells using FuGENE 6 (Promega). Two biological replicate massively parallel reporter assay (MPRA) experiments were performed. The relative enhancer activities of the different tiles were inferred by sequencing and counting their corresponding tags from the cellular mRNA and the transfected plasmid pool, as described in ref. 24. Nominal hits were defined as any tile where both enhancer activity measurements were >1 and with P values of <0.05 in both replicates. Those that agreed between SV40P and minP promoter datasets were considered the most confident hits. For the second MPRA experiment, a single-hit scanning mutagenesis (24), 12,000 *Enhancer L* variants were generated, including all possible single substitutions, multiple series of consecutive substitutions, and small insertions at all positions. Each variant was linked to an average of 16 tags each. The remainder of the workflow was as described above.

Luciferase Reporter Gene Assays. Individual candidate regions were amplified from JEG3 genomic DNA and directionally cloned into a pGL4 plasmid (Promega) containing either the minP or the SV40P promoter and firefly luciferase. JEG3 and HEK293T cells were transfected in 24-well plates using FuGENE 6 (Promega) with the individual firefly luciferase constructs and *Renilla* luciferase at a 10:1 ratio. Forty-eight hours posttransfection, firefly luciferase activity was measured using a Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions and normalized to *Renilla* luciferase to control for cell number and transfection efficiency.

Genome-wide DNase-seq. DNase I digestion followed by sequencing was performed as previously described (61, 62). In short, 10 million cells were harvested, washed twice with ice-cold PBS, and resuspended in Buffer A containing protease inhibitors and Spermidine (Sigma). Nuclei were extracted using ice-cold 0.05% Nonidet P-40 in Buffer A, centrifuged at 800 × g for 5 min at 37 °C, and gently resuspended in ice-cold PBS. An aliquot was taken to estimate nuclei number and integrity of using a cell counter (Bio-Rad). Intact nuclei were washed twice with ice-cold isotonic buffer and digested with empirically determined limiting concentrations of DNase I (Sigma) for 3 min at 37 °C. Digests were stopped with EDTA, and the samples were incubated with Proteinase K overnight at 55 °C. DNA was phenol/chloroform-extracted and concentrated by ethanol precipitation. Selection of 175- to 400-bp DNA fragments using the E-gel Agarose System (Invitrogen) was performed to select for regions in which DNase I can cut twice (at both ends), enriching for hypersensitive regions (61). Library preparation and sequencing were performed at the MIT BioMicroCenter. Prepared libraries were sequenced on a HiSeq 2000 sequencing system (Illumina) to a depth of 160 million to 230 million reads per sample using paired-end reads with a length of 40 bp. These were aligned to the human genomes (version hg19, canonical chromosomes only) using bwa, version 0.6.2, with default parameters. Quality control tests and regions of DNase hypersensitivity were calculated using the tool Hotspot-SPOT (version 4) (63) with an FDR of 0.01.

Chromatin Conformation Capture. Chromatin conformation capture (3C) assays were carried out essentially as described in refs. 55 and 64. A total of 10^7 cells was resuspended in 9 mL of medium and cross-linked using 2% (vol/vol) formaldehyde for 10 min at room temperature (RT). The cross-linking reaction was quenched with 0.125 M glycine on ice, and the cells were washed twice with cold PBS. Cells were lysed in 5 mL of cell lysis buffer containing protease inhibitors (64) on ice for 10 min. The nuclei were resuspended in NEBuffer DpnII (New England Biolabs) containing 0.3% SDS and incubated in a thermomixer 1 h at 37 °C shaking at 1,400 rpm. Next, 1.8% (vol/vol) Triton X-100 was added to sequester the SDS and the samples were incubated for an additional hour at 37 °C shaking at 1,400 rpm. The cross-linked DNA was digested with 1,000 units of DpnII (New England Biolabs) at 37 °C overnight. DpnII was heat inactivated at 65 °C for 20 min. For ligation of DNA ends, T4 DNA ligase was added and the samples were incubated for 4 h at 16 °C, followed by 30 min at RT. Cross-links were reversed by incubating with Proteinase K (10 mg/mL) at 65 °C overnight. Finally, the DNA was phenol/chloroform-extracted and concentrated by ethanol precipitation. A total of 50 ng of DNA was analyzed by PCR (primers listed on Table S1).

CRISPR/Cas9 Genome Editing. JEG3 cells were transfected with Cas9-2A-GFP and gRNAs targeting *Enhancer L*. GFP^+ cells were sorted 48 h posttransfection and plated at clonal density in 10-cm dishes. Approximately 10 d after plating, single-cell–derived colonies were picked into 96-well plates and cultured for an additional 10 d. For PCR analysis, cells were harvested and genomic DNA extracted using prepGEM Tissue (ZyGEM). Selected WT and KO clones were then expanded and further characterized.

Transcriptome-wide RNA-seq. Total RNA from JEG3 cells was extracted using TRIzol (Life Technologies), according to manufacturer's instructions and then purified by spin column purification (RNeasy mini kit; Qiagen) using a QIAcube system. RNA was quantified using a Nanodrop (Thermo Fisher), and its integrity was assessed on a Bioanalyzer (Agilent) using the RNA 6000 RNA chip. A total of 500 ng of high-quality total RNA (RNA integrity number, ≥8) was used as input for Tru-seq library construction using the TruSeq RNA Sample Preparation Kit

(Illumina), as described in refs. 65 and 66. Library purity, correct fragment size, and concentration were assessed using the Bioanalyzer DNA7500 chip. Libraries free of adapter dimers and with a peak region area (220–500 bp) ≥80% of the total area were individually barcoded, pooled, and sequenced on an Illumina HiSeq 2000 platform. Reads were mapped to the human genome (hg19) using TopHat, version 2.0.14 (67, 68), with the flags: "–no-coverage-search–GTF gencode.v19.annotation.gtf" where gencode.v19.annotation.gtf is the Gencode, version 19, reference transcriptome available at gencodegenes.org. Cufflinks, version 2.2.1 (65), was used to quantify gene expression and assess the significance of differential expression. Briefly, Cuffquant was used to quantify mapped reads against Gencode, version 19, transcripts of at least 200 bp with biotypes: protein_coding, lincRNA, antisense, processed_transript, sense_intronic, sense_overlapping. Cuffdiff was run with default options on the resulting .cxb files. Gene set enrichment analysis was performed with the GSEA, version 2.1.0 (69, 70), from the Broad Institute. Genes with expression greater than 1 fragments per kilobase of exon per million fragments mapped (FPKM) in WT or KO conditions were sorted in decreasing order by the absolute value of their logtwofold change as determined by Cuffdiff. The resulting ranked list was fed to the GSEA preranked tool with default options and a permutation seed of 42. RNA-seq data are available in the NCBI GEO database (www.ncbi.nlm.nih.gov/geo/) under accession number GSE79779.

First-Trimester Primary EVT Isolation and Transduction. Discarded human placental and decidual material (gestational age, 6–12 wk) was obtained from women undergoing elective pregnancy termination at a local reproductive health clinic. All of the human tissue used for this research was deidentified, discarded clinical material. The Committee on the Use of Human Subjects [the Harvard institutional review board (IRB)] determined that this use of all of this human material is exempt from the requirements of IRB review. EVTs were isolated as previously described (31). In total, 50,000–100,000 $CD45^{-}HLA\text{-}G^{+}$ EVTs were plated in 48-well cell culture plates (Costar) precoated with 100 μL of 20 ng/mL fibronectin for 45 min (BD), in Trophoblast Medium, which consisted of DMEM/F12 medium (Gibco) supplemented with 10% (vol/vol) NCS, Glutamax, insulin, transferrin, selenium (100×; Gibco), 5 ng/mL EGF (Peprotech), and 400 units of human gonadotropic hormone (Sigma). Two hours postplating, EVTs were transduced with Cas9-T2A-GFP and *Enhancer L* gRNAs' lentiviral particles pseudotyped with G glycoprotein from vesicular stomatitis virus (VSV-G) in the presence of 8 μg/mL polybrene (hexadimethrine bromide; Sigma). Lentiviral particles were produced using HEK293T cells and were concentrated 20× using Lenti-X Concentrator (Clontech). Transduction was performed three additional times, 12 h apart. Three days after the first transduction, medium was switched to villous stromal cell (VSC)-conditioned medium (RPMI-1640 medium supplemented with 10% (vol/vol) FBS, Glutamax, and penicillin–streptomycin) for an additional 2 d. For analysis by flow cytometry, cells were washed with warm PBS, harvested with trypsin, and resuspended in trophoblast medium for staining with ITGA5 PE (an EVT marker) (31) and HLA-G APC (both Biolegend).

ChIP (ChIP-qPCR). ChIP-qPCR was performed using Dynabeads Protein G (Life Technologies) according to the manufacturer's instructions. JEG3 cells were harvested, resuspended in PBS, and cross-linked using 1% formaldehyde. Glycine was added to stop the cross-linking reaction, and cells were washed twice with ice-cold PBS. Nuclei were isolated using ice-cold Cell Lysis Buffer containing protease inhibitors and PMSF (CalBiochem) and then lysed using ice-cold Nuclei Lysis Buffer containing protease inhibitors and PMSF. Cross-linked chromatin was sheared using a Bioruptor Standard Sonication Device UCD-200 (Diagenode) to 200- to 500-bp fragments, assessed by gel electrophoresis. Samples were then diluted with ChIP Dilution Buffer and incubated with Mock IgG, Pol II, CEBPB, GATA2, or GATA3 antibodies (Santa Cruz Biotechnology) overnight at 4 °C. Dynabeads Protein G were blocked with BSA and glycogen at 4 °C during the same period. The following day, beads were washed and eluted. The eluates were then reverse cross-linked, and DNA was purified using phenol/chloroform extraction. A total of 25 ng of DNA was used for qRT-PCR analysis using SYBR Green (Life Technologies).

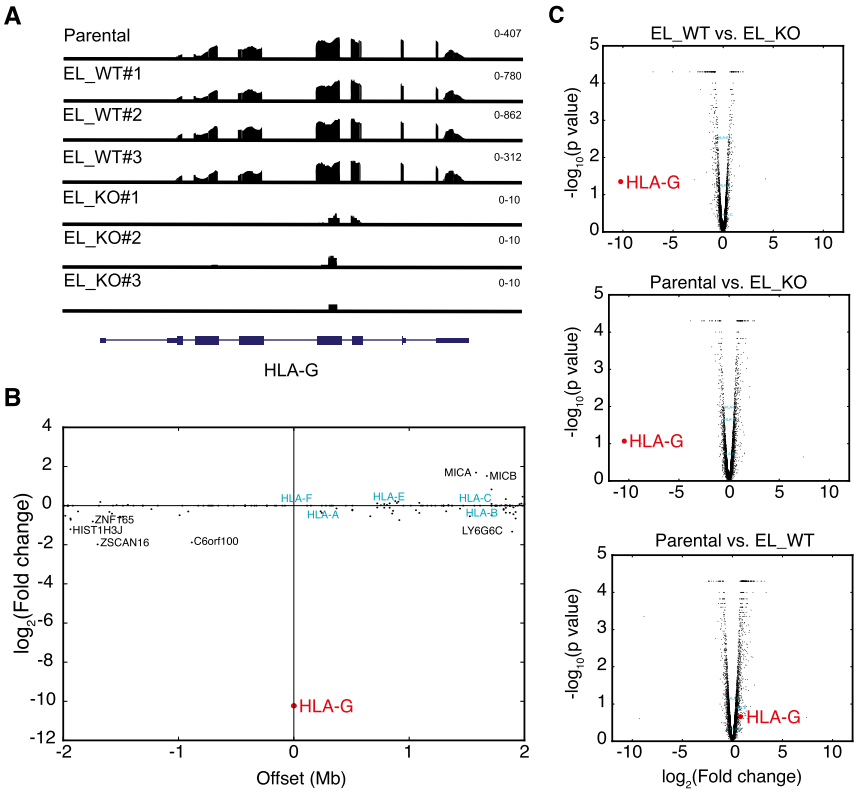

Fig. S1. Deletion of *Enhancer L* specifically abrogates *HLA-G* expression. (*A*) Whole-transcriptome RNA-seq confirming complete loss of *HLA-G* transcript in three independent *Enhancer L* KO JEG3 clones. Scale in fragments per kilobase of exon per million fragments mapped (FPKM) is indicated on the top right corner of each sample. EL, *Enhancer L*. (*B*) *Enhancer L* deletion specifically results in loss of *HLA-G* expression in a radius of 2 Mb centered on *HLA-G*. The fold change in gene expression between combined WT and KO clones is plotted against genomic coordinates, each dot representing a gene. HLA-G is displayed in red (largest fold change), whereas other HLA genes are displayed in blue (no change). (*C*) Transcriptome-wide pairwise comparison between different genotypes, depicted as Volcano plots.

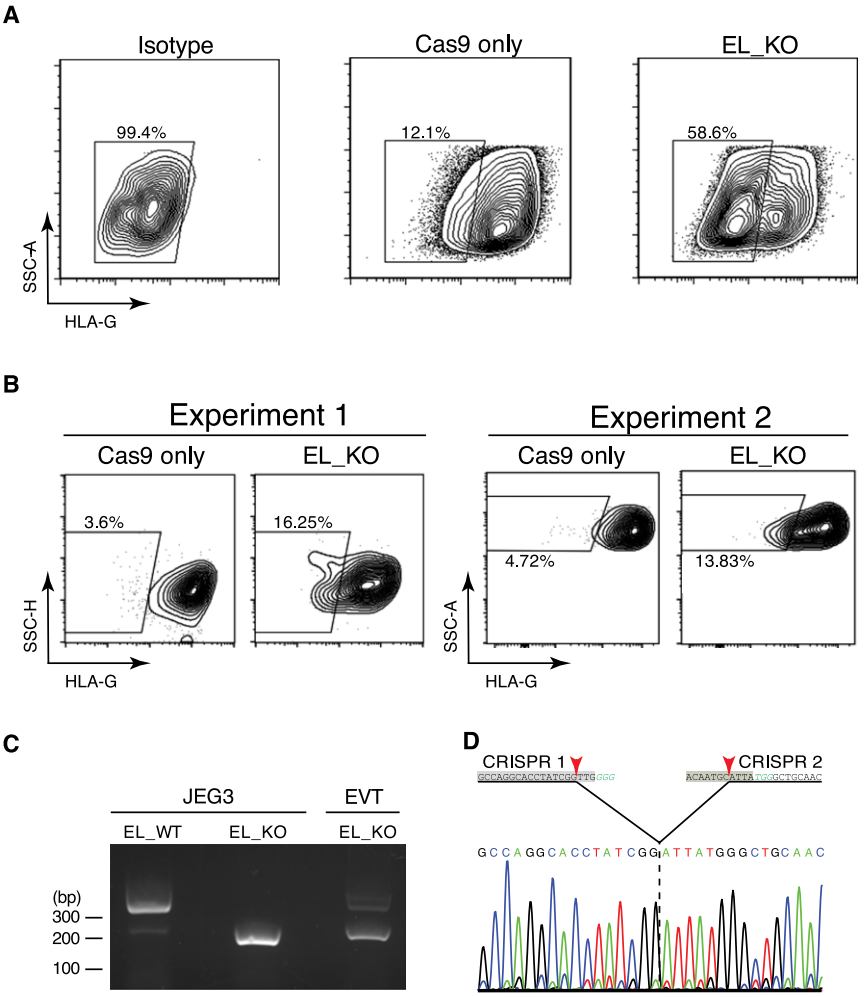

Fig. S2. Lentiviral CRISPR/Cas9 deletion of *Enhancer L* in JEG3 cells and primary EVTs. (*A*) Lentiviral transduction of JEG3 cells with Cas9-2A-GFP and two CRISPR gRNAs targeting *Enhancer L* led to complete loss of HLA-G surface expression ($n = 3$). Cells successfully transduced with lentiviral particles containing the gRNAs were selected with puromycin and analyzed 1 wk posttransduction. E1, *Enhancer L*. (*B*) Two independent experiments showing loss of HLA-G surface expression upon CRISPR/Cas9 targeting of *Enhancer L* in primary EVTs 5 d after the first transduction. EL, *Enhancer L*. (*C*) PCR demonstrating successful *Enhancer L* genomic deletion in EVTs using lentivirally delivered CRISPR/Cas9. An *Enhancer L* WT JEG3 cell clone (EL_WT) and a KO clone (EL_KO) were included as controls. (*D*) Sanger sequencing confirmation of Cas9-mediated genomic deletion of *Enhancer L* in primary EVTs. CRISPR gRNA binding sites are shaded, PAM motifs are italicized in green type, and Cas9 cutting sites are indicated with red arrowheads.

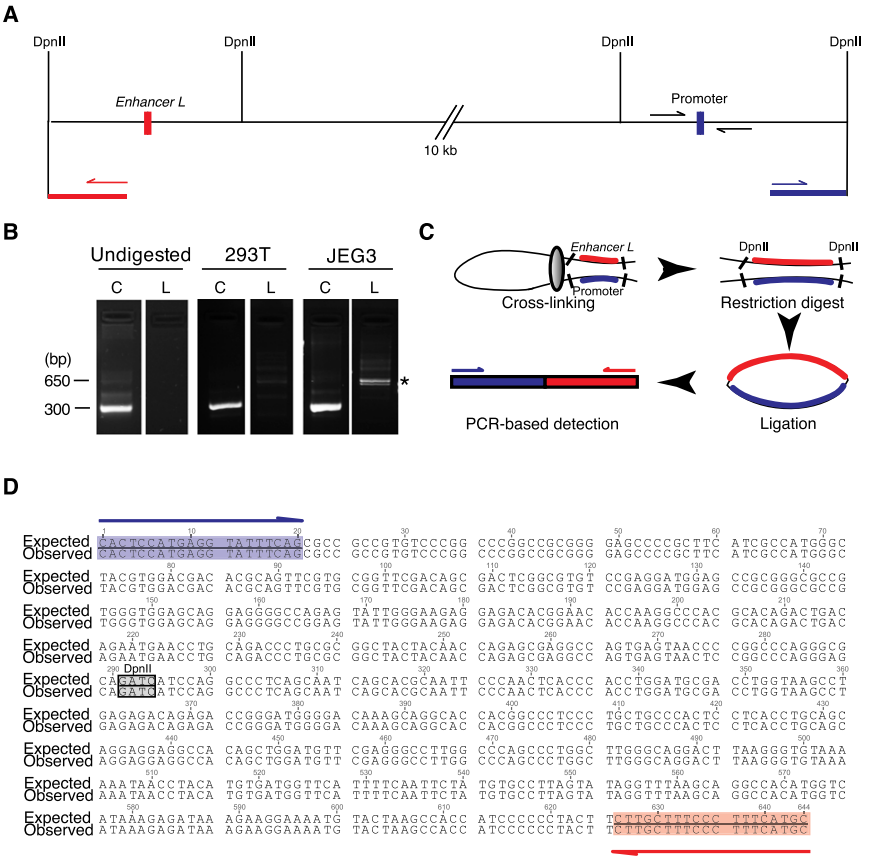

Fig. S3. *Enhancer L* loops into the classical promoter of *HLA-G*. (*A*) Strategy for chromatin conformation capture (3C) analysis of the *HLA-G* locus. DpnII restriction sites flank the two regions of interest: *Enhancer L*, labeled in red, and the *HLA-G* classical promoter, given in blue. Primers in black amplify a product that serves as "loading control," and primers in red and in blue were used to detect loop formation. (*B*) 3C indicates that *Enhancer L* physically interacts with the classical promoter of *HLA-G* in JEG3, but not in HLA-G negative HEK293T cells. C, Loading control PCR product (322 bp); L, *Enhancer L*–classical promoter looping interaction PCR product (640 bp), marked with a star. (*C*) Schematic representing the main steps in 3C: chromatin cross-linking, restriction digest, ligation at low DNA concentrations, and PCR-based detection of looping interactions. (*D*) Sequence confirmation of the physical interaction between *Enhancer L* and the classical promoter of *HLA-G* detected by 3C assays (PCR amplicon marked with a star in Fig. S3*B*). Primer pointing away from the classical promoter is depicted in blue, primer pointing away from *Enhancer L* is in red, and the DpnII restriction site where ligation of the contact regions occurred is enclosed by a gray-shaded box.

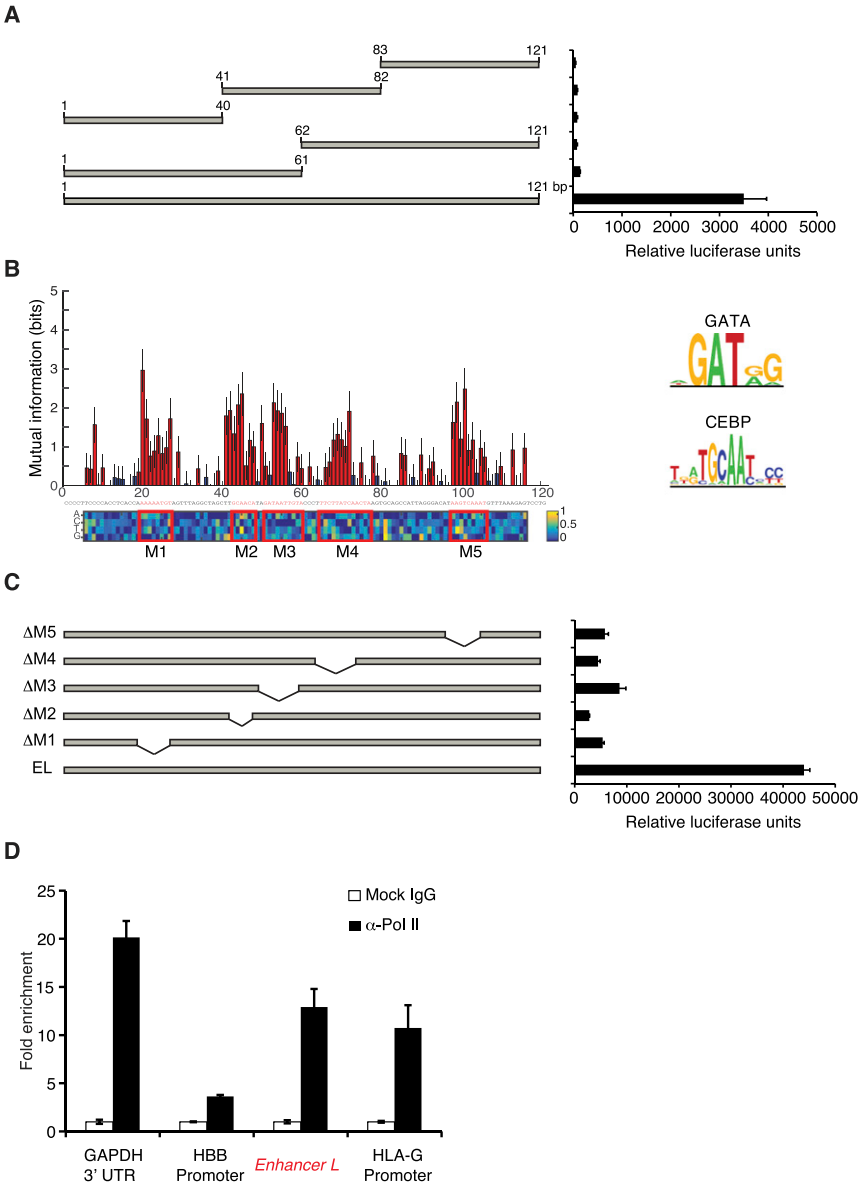

Fig. S4. Dissection and validation of *Enhancer L* regulatory elements. (*A*) Systematic truncation of *Enhancer L* and subsequent assessment of luciferase reporter gene activity indicates that *Enhancer L* requires multiple motifs spread over its length for activity in JEG3 cells (strong SV40 promoter) ($n = 2$). (*B*) Identification of five putative regulatory motifs (M1–M5, in red boxes) involved in *Enhancer L* activity using MPRA-based scanning mutagenesis with a minP promoter. Red indicates a significant change from the original *Enhancer L* sequence [Mann–Whitney *U* test, 5% false-discovery rate (FDR)]; blue means not significant. The matrix represents the estimated additive contribution of each nucleotide to *Enhancer L* activity. CEBP and GATA consensus motifs according to the TRANSFAC database are shown on the *Right*. (*C*) Systematic deletion of the five motifs identified by MPRA (M1–M5, represented in Fig. S4*B*) within *Enhancer L*. Each one was required for maximum *Enhancer L* activity in JEG3 cells, as measured by luciferase reporter gene activity (strong SV40 promoter) ($n = 2$). EL, *Enhancer L*. (*D*) ChIP followed by qPCR (ChIP-qPCR) showing Pol II binding to *Enhancer L* and the *HLA-G* promoter, suggesting the existence of a chromatin loop that may involve active transcription ($n = 2$). *GAPDH* 3′-UTR was used as a positive control and the *HBB* promoter as a negative control for Pol II binding.

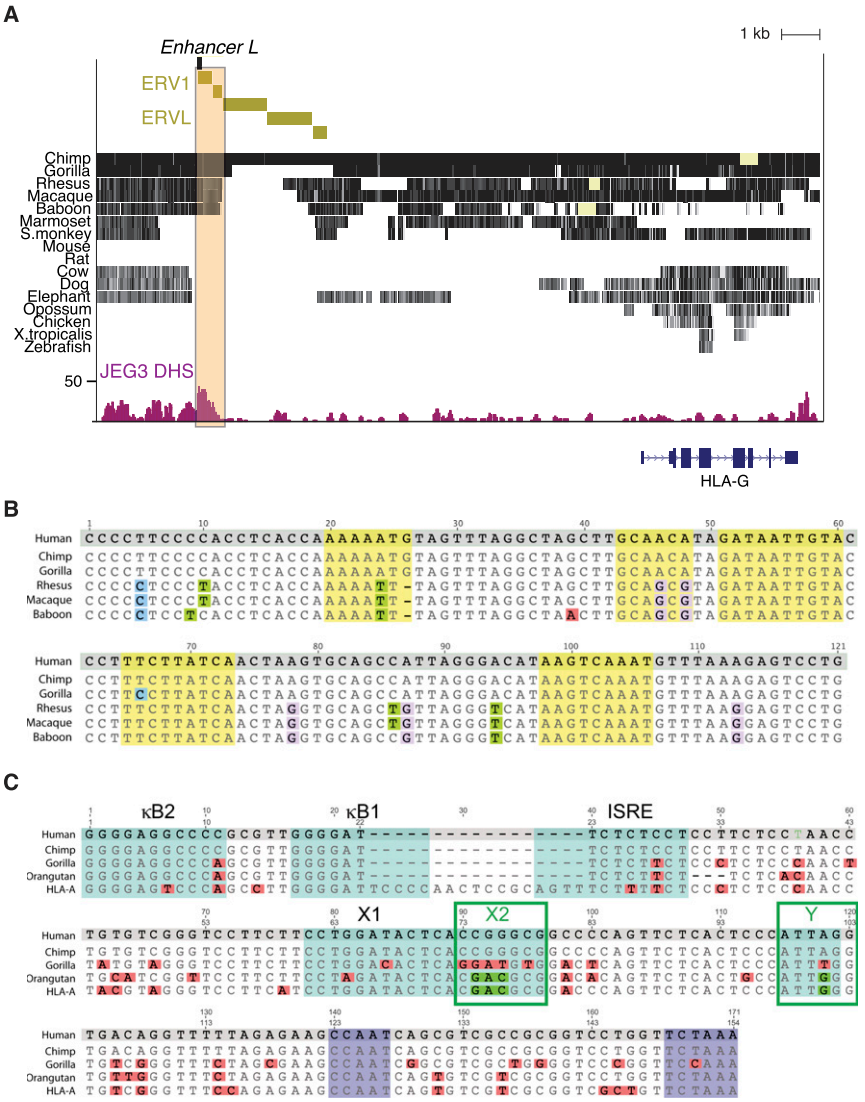

Fig. S5. *Enhancer L* is part of a lineage-specific retrotransposon conserved in apes bearing a nonpolymorphic *HLA-G* gene. (*A*) The region containing *Enhancer L* (beige box) is conserved in apes (with the exception of orangutans) and Old World monkeys, but absent in New World monkeys and other placental mammals. This region is part of a long terminal repeat (LTR) retrotransposon element mostly found in the human genome, ERV1. A different class of retrotransposon elements, ERVL, is also present in the vicinity. (*B*) Alignment of *Enhancer L* sequences across apes and Old World monkeys indicates strong conservation across the analyzed species. *Enhancer L* regulatory motifs are highlighted in yellow and mismatches in blue (C), green (T), pink (A), or purple (G). (*C*) Alignment of the *HLA-G* classical promoter sequences across apes reveals that orangutans, which lack an *Enhancer L* sequence, possess intact X2 and Y motifs (highlighted in green), unlike any other ape. Mismatches are colored pink; kB2, kB1, ISRE, X1, X2, and Y boxes characteristic of an HLA classical promoter are highlighted in blue; the ubiquitous CAAT and TATA boxes necessary for transcription are highlighted in purple.

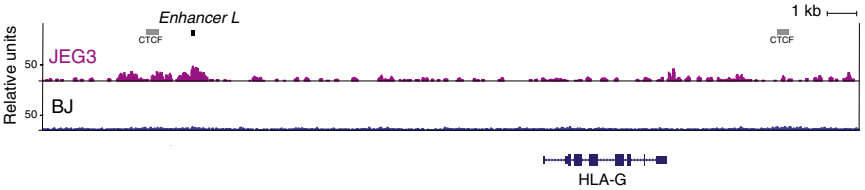

Fig. S6. *Enhancer L* is part of an open chromatin region specifically in JEG3 cells. *Enhancer L* is part of a ~1-kb-long DNase hypersensitive site (DHS) in JEG3 cells that is absent in control BJ fibroblasts (HLA-G^{-}), as revealed by genome-wide DNase-seq. Binding of the insulator CTCF upstream of *Enhancer L* and downstream of *HLA*-G, according to ENCODE ChIP-seq data, indicates potential boundaries of the *HLA-G* regulatory chromatin domain.

Table S1. List of primers used in this study

Primer	Forward/reverse	Sequence
Enhancer L CRISPR gRNAs		
CRISPR 1	Forward	5′-CACCGGCCAGGCACCTATCGGTTG-3′
	Reverse	5′-AAACCAACCGATAGGTGCCTGGCC-3′
CRISPR 2	Forward	5′-CACCGAGACAATCACAATGCATTA-3′
	Reverse	5′-AAACTAATGCATTGTGATTGTCTC-3′
Enhancer L deletion PCR		
	Forward	5′-CATGGTCATAAAGAGATAAAG-3′
	Reverse	5′-CTTACGATCTTCCCGGATGTC-3′
qRT-PCR		
GAPDH	Forward	5′-GAAGGTGAAGGTCGGAGT-3′
	Reverse	5′-GAAGATGGTGATGGGATTTC-3′
HLA-G	Forward	5′-GCTGCCCTGTGTGGGACTGAGTG-3′
	Reverse	5′-GACGGAGACATCCCAGCCCCTTT-3′
3C		
Loading control	Forward	5′-CACAAGAGTAGCGGGGTCAG-3′
	Reverse	5′-GAGCAGCAGGAAGAGGGTTC-3′
Enhancer L-promoter looping	Forward	5′-GCATGAAAGGGAAAGCAAG-3′
	Reverse	5′-CACTCCATGAGGTATTTCAG-3′
ChIP-qPCR		
GAPDH 3′-UTR	Forward	5′-TCGACAGTCAGCCGCATCT-3′
	Reverse	5′-CTAGCCTCCCGGGTTTCTCT-3′
HBB promoter	Forward	5′-CTGGTGGGGTGAATTCTTTGC-3′
	Reverse	5′-AGTCCAAGCTAGGCCCTTTT-3′
CEBP positive control	Forward	5′-AGACTTTGAAGACGATTCAGCA-3′
	Reverse	5′-ACCCCTGATTGCTCAACACT-3′
GATA positive control	Forward	5′-CTCTGGCCGGTCGATGTTATC-3′
	Reverse	5′-GATGGCGGCTGCGATTAAC-3′
Enhancer L	Forward	5′-GGCCAGGCACCTATCGGTTG-3′
	Reverse	5′-TAATGCATTGTGATTGTCTC-3′
HLA-G classical promoter	Forward	5′-GTGGCTCTCAGGGTCTCAGG-3′
	Reverse	5′-CGACGCTGATTGGCTTCTCTA-3′