Title: Human Genome REWRITE for Off-the-Shelf Stem Cells Reveals an "Epigenetic Ghost"

Authors: Serena F. Generoso[1,†], Sarah Levovitz[1,†], Susanna Jaramillo[1], Minjoo Kim[1], Sumanth Dara[1], Shean Fu Phen[1], Bryan Yi[1], Tomoki Yanagi[1], Thomas L. DesMarais[3], Neta Agmon[3,‡], Megan S. Hogan[3], Leslie A. Mitchell[3], David M. Truong[1,2]*

Affiliations:

[1] Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, NY, 11201 USA.

[2] Department of Pathology, NYU Grossman School of Medicine, New York, NY 10010, USA.

[3] Neochromosome Inc, Long Island City, NY 11101, USA.

‡ Present address: Alagene LTD, Rehovot, Israel.

*Corresponding author. Email: truond01@nyu.edu.

†These authors contributed equally to the work.

Abstract: Human leukocyte antigen (HLA) polymorphism hinders off-the-shelf cell therapies. We developed REWRITE, a modular platform for iterative, scar-minimized genome writing of synthetic constructs >100 kb in human pluripotent stem cells (hPSCs). Using REWRITE, we deleted 105–209 kb of the HLA locus and installed synthetic 24 kb or 100 kb HLA haplotypes, and a 62 kb antigen-processing locus. This uncovered a persistent, heritable "epigenetic ghost" — an active state lingering despite genetic removal — whose resolution to a silenced default state is driven by native intergenic DNA. These loci restored inducible expression in key lineages, sparing cells from NK-mediated killing and establishing HLA-matched T-cell tolerance, enabling off-the-shelf cell therapies. REWRITE facilitates extensible programming of multigenic functions in allogeneic human cells — from immune design to genome architecture discovery.

Main Text: Genome writing technologies have enabled the construction of fully synthetic chromosomes in single-celled organisms, integrating DNA synthesis, in vitro assembly, and in vivo editing (*1*). In mammalian systems, "bigDNA" integration strategies in mouse embryonic stem cells (mESCs) have installed synthetic loci >100 kb in size, preserving native regulatory elements for developmental or disease modeling (*2-9*). However, translating these approaches to human pluripotent stem cells (hPSCs) remains a major challenge. Compared to mESCs, hPSCs are more fragile: they have lower transfection efficiency, reduced homologous recombination, increased p53-mediated apoptosis, and anoikis when cultured as single cells (*10-12*). While site-specific integration of smaller constructs (~10 kb) has been reported in hPSCs (*13, 14*), and one report of a 173 kb BAC integration (*15*), no current platform enables iterative, scar-minimized rewriting of native loci >100 kb in hPSCs. This remains a key barrier for advancing synthetic genomics and regenerative engineering in human cells.

One of the most consequential regions for human genome rewriting is the 3.6 megabase HLA super-locus on chromosome 6. This region encodes "immune-identity" through the most polymorphic genes in the genome (*16*), including the class-I genes *HLA-A*, *-B*, and *-C*, expressed by nearly all nucleated cells (*17*). This diversity creates a compatibility barrier for allogeneic cells, as mismatched HLAs impair tolerance by CD8+ T cells and natural killer (NK) cells, undermining applications requiring precise immunological interaction.

hPSCs offer a renewable chassis for generating immune-interfacing cell types — such as macrophages for tissue surveillance, dendritic cells for antigen presentation, and thymic epithelial cells for T-cell education (*18-20*) — all of which require a compatible, programmable HLA interface. A strategy to directly rewrite class-I HLAs in hPSCs could enable the design of broadly accessible, lineage-compatible antigen-presenting cells.

Here we describe "REWRITE" (REcombinase Writing by Repeatable Integrations and Tag Excision), a system for multi-round, locus-scale engineering in hPSCs. REWRITE supports installation of large synthetic constructs (>100 kb) at native loci using recombinase-mediated cassette exchange (RMCE), removable markers, and iterative payload extension. We applied REWRITE to reconfigure the human HLA locus, deleting up to 209 kb and installing synthetic class-I haplotypes of 24 kb or 100 kb. This represents the first scar-minimized refactoring of a native human genomic region at this scale. The resulting constructs restore inducible HLA expression across lineages and enable HLA-matched immune recognition. By coordinating multigenic engineering, REWRITE lays the foundation for designing allogeneic, programmable antigen-presenting cells across diverse immune-engaged lineages.

REWRITE enables modular, iterative genome rewriting in hPSCs

To enable reproducible construction and integration of >100 kb synthetic payloads in hPSCs, we developed REWRITE: a four-stage process comprising (1) genomic deletion, (2) payload assembly, (3) site-specific integration, and (4) marker excision. REWRITE uses Cre-recombinase RMCE to install large payloads without double-strand breaks, which activates p53 and compromises genome stability in hPSCs (*11*).

The system begins with insertion of a recombinase-compatible landing pad (LP) via CRISPR-assisted homologous recombination, which functions as a genomic "docking-hub" (Fig. 1A). Our architecture incorporates a fixed EF1α promoter, start-codon, and loxP site to serve as a reusable "promoter-dock" — activating expression of all serially integrated payloads without reintroducing regulatory sequences. This is followed by exchangeable selection markers (e.g., blasticidin, mScarlet) and a heterotypic mutant lox site (loxM1). Payload vectors (PVs) act as "dock-

cassettes": they lack a promoter and start-codon but contain alternative selection markers (e.g., puromycin, mNeonGreen) and a third recombinase site (loxM2), becoming transcriptionally active only upon in-frame RMCE. Alternatively, terminal PVs also encode Flp-ERT2 and a second FRT site in place of the loxM2, enabling tamoxifen-inducible excision of selection markers and generation of scar-minimized final payloads (one FRT, one loxM), placed in genomic regions unlikely to affect gene expression (Fig. 1A).

We benchmarked LP designs at the ROSA26 locus in mESCs (fig. S1A–B). An LP-expressed double ERT2-Cre failed to integrate a 10 kb payload, likely due to weakened recombinase activity. Co-nucleofection with codon-optimized iCre increased colony yield 8-fold compared to wild-type Cre (fig. S1C–E). RMCE with loxP/loxFAS (*21*) outperformed loxP/lox2272 by 15-fold (fig. S1F). LP silencing without blasticidin thus motivated incorporation of a Universal Chromatin Opening Element (UCOE) (*22*), which maintained mScarlet expression for >6 days without selection (fig. S2).

Based on these results, we built an optimized LP called "LPneo" (Fig. 1A), incorporating iCre under a TRE3G promoter in reverse orientation to EF1α, which drives rtTA for inducible expression. LPneo includes Golden Gate sites for modular homology arm assembly and rapid deployment to other genomic sites. "LPcore", a variant lacking TRE3G-iCre, utilizes plasmid-expressed iCre-recombinase.

We integrated LPneo at four transcriptionally silent loci in PGP1 hPSCs: *CCR5*, ROGI1, *CD40LG*, and *TRBC1*(Fig. 1B, S3A). CRISPR-assisted HDR yielded 221–301 selected colonies per locus. Junction PCR of 30 colonies revealed HDR efficiencies of 3%–22%. RT-qPCR confirmed pluripotency (fig. S3B), and after 29 days without selection, LP silencing was minimized (fig. S3C).

To demonstrate iterative rewriting, we developed a series of payloads (Zero1-PV, Zero2-PV) with incompatible loxM variants (loxFAS, lox5171) (Fig. 1A). Zero1-PV integrated at 122 colonies per million cells nucleofected, with 10 of 12 clones PCR-validated (Fig. 1C). After full RMCE, Zero2-PV yielded ~200 colonies per million cells, with second-round RMCE confirmed in 3 of 4 clones (Fig. 1D).

We used ROGI1-LPneo hPSCs to test "FlpOut"-PV (terminal PV) (Fig. 1A). Co-nucleofecting anti-apoptosis plasmid *BCL-XL* (*23*) increased yields from 24 to 104 colonies per million cells. Junction PCR confirmed LP loss and correct FlpOut integration (Fig. 1E). To test marker removal, we induced Flp-ERT2 with tamoxifen. Fluorescent cells dropped from 62% to 10.5% in 5 days, indicating ~84% excision (Figs. 1F, S4A–B). Negative selection against residual markers further enriched for excised clones. PCR showed marker removal, with 8 of 11 clones losing puromycin resistance, and directly confirmed by Sanger sequencing (fig. S4C–D).

These results show that REWRITE can support sequential integrations and facile marker-scar removal. Additional rounds would use other validated lox-mutants (*24*), partially used for limited small-payload serial integration in CHO cells (*25*), but here unified into a single system for rewriting human loci. This capacity for serial payload integration could establish hPSC chassis for iterative synthetic circuits, establishing REWRITE as a framework for staged genome reconfiguration (fig. S1G).

synHLA: a refactored class-I locus for designer HLA haplotypes

To enable single-step installation of full class-I HLA haplotypes, we engineered a synthetic, refactored locus ("synHLA") consolidating targeted sequences into a contiguous-insert,

overcoming their dispersal across a megabase (*26*). In the native genome, *HLA-B* and *-C* are separated by 80 kb, and *HLA-A* lies 1.3 megabases upstream (Fig. 2A). We captured the 93 kb *HLA-B/C* region from a linearized BAC into the FlpOut-PV backbone by yeast recombineering, then inserted *HLA-A* (PCR-amplified in two 5 kb fragments) 36 kb downstream of *HLA-B* (Fig. 2B–C). Deep sequencing verified the assembly (fig. S5A). The resulting synHLA, a fully assembled 100 kb construct (115 kb with vector), encoded a rare haplotype (Fig. 2D).

To test whether intergenic DNA was required for function, we designed a compact variant ("mini-synHLA") retaining only core regulatory loci. Enhancer boundaries were defined using ENCODE predictions. PCR-amplified genes (3–5 kb) were fused with native-sequence linkers and recombineered into FlpOut-PV, generating a 24 kb core insert (38 kb total) (Fig. 2E–F).

To customize haplotypes, we developed distinct "allele-swapping" strategies for synHLA and mini-synHLA (Fig. 2G). For synHLA, whose size and repeat density precluded standard cloning (*27*), we used in-yeast recombineering with *URA3* counter-selection and CRISPR/Cas9-editing, yielding two rare haplotypes (fig. S5C). For mini-synHLA, we used a modular system of three vectors encoding full-length HLA loci, each edited via yeast recombineering, and then amalgamated by DNA assembly using linker fragments. Final assemblies were validated by nanopore sequencing, generating three additional rare haplotypes (Figs. 2G, S5B).

These strategies enable programmable synthesis of class-I HLA haplotypes in full-length or compact formats, with modular control over regulatory architecture and allele composition. To assess design constraints across populations, we analyzed HLA-typed records from over 10,000 individuals (fig. S6). A few common haplotypes matched >50% of individuals in some groups (e.g., Northern European, Southern Chinese), but broader representation — particularly among African, admixed, or Middle Eastern ancestries — required rarer alleles. These findings highlight the limitations of fixed donor panels and underscore the value of on-demand synthesis for modeling population-specific immune interfaces.

Engineering biallelic class-I and class-II HLA deletions in hPSCs

To enable repeatable reconstitution of on-demand HLA haplotypes, we generated a panel of "blank" human pluripotent stem cell (hPSC) lines with biallelic deletions of class-I or class-II HLA gene clusters (Fig. 3A–B).

We first deleted the ~8 kb *HLA-A* locus using a Cpf1-homolog, Mad7 (fig. S7A–D). Screening four gRNA pairs with 100 bp ssODN HDR donors, transient puromycin enrichment, and *BCL-XL* co-expression (*23*), yielded two sets with clean monoallelic deletions. We recovered a scarless biallelic deletion clone (PGP1-ΔA) from 168 colonies and validated by whole genome sequencing (WGS).

After establishing the *HLA-A* deletion, we targeted the 93 kb *HLA-B/C* region, which was challenging due to repeat content and low gRNA specificity. To improve biallelic deletion efficiency, we co-delivered Cas9 and Mad7 gRNAs, a HygTK "deleter" cassette flanked by mutant FRT sites (fig. S7E), LPcore for allelic integrations, *BCL-XL*, and the NHEJ inhibitor Nedisertib. Nucleofection of a million PGP1 or PGP1-ΔA cells yielded 30–50 colonies, the majority with at least monoallelic deletions. Across multiple experiments, we obtained several clones achieving full biallelic deletion (PGP1-ΔBC), and multiple (six) independent fully "blank"-HLA lines (PGP1-ΔABC) (Figs. 3C–D, S7F–G, S8A). Downstream studies focused on two of those clones.

In parallel, we biallelically deleted a 209 kb class-II HLA segment (*HLA-DRB5*, *-DRB1*, *-DQA1*, *-DQB1*) in both NCRM2 and PGP1-ΔA hPSCs (Figs. 3F–G, S8C). Nucleofection of a million cells

per locus yielded 32–94 colonies. Junction PCR revealed monoallelic deletion efficiencies of 10–20%, and biallelic deletions in 2–10% of clones, named PGP1-ΔAΔ209 and NCRM2-Δ209 (figs. S8B, S8D).

We validated genomic integrity via 30x depth WGS in NCRM2-Δ209, PGP1-ΔAΔ209, and four distinct PGP1-ΔABC clones. All showed the expected deletions, no structural abnormalities or recurrent pathogenic *TP53* mutations (*28*) (Figs. 3I, S8E, S8G, S7H). Euploidy in PGP1-ΔABC, NCRM2-Δ209, and PGP1-ΔAΔ209 was independently confirmed by WGS (Figs. 3E, 3H, S8F). To evaluate the impact of large biallelic deletions on pluripotency, we performed RT-qPCR across multiple lines. All retained high expression of pluripotency markers (Fig. 3J).

Thus, we established a validated panel of class-I or partial class-II, HLA-blank hPSC chassis lines suitable for programmable reconstitution of immune-identity with synthetic HLAs.

REWRITE enables precise integration of >100 kb synthetic loci at native and ectopic sites

To test whether synthetic, refactored HLA loci could restore class-I expression in its native genomic context, we delivered both mini-synHLA (38 kb) and synHLA (115 kb) constructs into fully blank-HLA hPSC line PGP1-ΔABC. The parental PGP1 line is homozygous for HLA-A*02:01, -B*51:01, and -C*05:01 based on publicly available WGS. Both constructs were introduced into REWRITE landing pads by iCre-mediated recombination, and the marker-cassette was completely excised by FlpOut (Figs. 3L, S9A–B).

Across multiple experiments and independent users in different HLA-deleted hPSC lines, we consistently observed a strong size-dependent integration effect: ~100 selected colonies per million nucleofected cells for 38 kb mini-synHLA, but only ~1.5–3 colonies per million cells for 115 kb synHLA, though routinely 5 million are nucleofected (Fig. 3K). This drop-off likely reflects lower transfectability and nuclear import, increased toxicity, and fewer DNA molecules per microgram as size increases, reflecting biophysical constraints on bigDNA (*29*) not addressable at present. Tiled internal PCR and WGS verified full-length integration, and absence of genomic abnormalities or off-targets (Fig. 3M–N, S9C). High pluripotency was retained, also upon high passages in culture (Figs. 3O–Q, S9D-E) and embryoid bodies were generated as expected (fig. S9F).

To assess whether fully assembled synHLA constructs were portable across genetic backgrounds and loci, we also delivered them into ectopic sites of the class-II region in deletion lines PGP1-ΔAΔ209 and NCRM2-Δ209 (fig. S10A, S10F). Integration frequencies were comparable (fig. S10B, S10G), pluripotency, genomic structure, and differentiation capability were preserved (fig. S10C–E), indicating that synHLA installation is robust to locus, background, and construct size, and establishing a foundation for future multi-site programming of hPSC genomes.

Architectural design of synthetic HLA loci governs their epigenetic memory

To dissect the regulatory consequences of our architectural designs, we first analyzed the basal HLA expression state at a per cell basis in hPSCs. As expected, wild-type hPSCs showed negligible HLA expression (*30*). In striking contrast, we discovered a persistent, heritable "epigenetic ghost" (Fig. 4A): following complete excision of the UCOE/EF1α promoter cassette (quantitatively confirmed by qPCR and WGS, Figs. 4B, S11A), the synHLA locus remained unexpectedly active. The expressing cell population remained stable at ~50% both before and after marker excision (Fig. 4A), contrary to the expectation that the locus would revert to its default silenced state. This active state was remarkably stable, persisting for over 20 population doublings

from clonal selection and with less than 5% change in expression over 14 additional passages (Fig. 4C), demonstrating a robust, heritable epigenetic memory.

To probe the architectural basis of this epigenetic memory, we directly compared our marker-excised mini- and full-synHLA constructs. This revealed a stark, architecture-dependent difference in ghost stability: two distinct mini-synHLA clones, lacking intergenic DNA, showed homogenous activity with ~78% of cells expressing surface HLA. In contrast, the full-synHLA line, containing the native intergenic regions, exhibited similarly fewer expressing cells as before at 51% (Fig. 4D). This population difference was not a result of clonal selection, as no changes in cell viability were observed (fig. S11B). Moreover, the mini-synHLA architecture resulted in higher mean fluorescence intensity of the expressing population (fig. S11C). These reproducible differences provide direct evidence that native intergenic DNA helps resolve this heritable active state. Importantly, both engineered loci remained fully responsive to IFN-γ induction, confirming the integrity of the core regulatory machinery (Figs. 4D, S11D).

We next performed RNA-seq after IFN-γ induction to understand how this epigenetic memory influences the marker-excised locus's transcriptional response. While gene expression profiles confirmed that all lines retained a pluripotent signature (Fig. 4E) (*31*), analysis of the HLA locus revealed three distinct, architecture-dependent outcomes (Fig. 4F). The wild-type locus remained largely refractory to induction, establishing the deeply silenced baseline. In stark contrast, compressed mini-synHLA had massive transcriptional output for all three HLAs. Crucially, the full-length construct demonstrated the repressive effects of the intergenic DNA; it buffered the epigenetic "ghost's" amplifying effect, reducing the induced expression of *HLA-A*, *-B*, and *-C* by approximately 4.8-, 3.9-, and 3.2-fold, respectively (Fig. 4F). These data provide robust functional evidence that native intergenic DNA contains essential architectural software that calibrates the epigenetic state of a genomic locus.

To determine if this active state is inherited to differentiated cells, we compared transcription before and after marker-excision in cells then differentiated into mesodermal endothelial-like cells induced with IFN-γ. Immunostaining showed expression of CD31 lineage-specific marker and uniform membrane localization of HLA proteins after induction, consistent with proper folding and surface presentation (Figs. 4G, S11E). In the pre-excised state, we observed increased HLA expression consistent with the UCOE/EF1α's influence. After marker removal, notably, expression returned to near-native levels, suggesting differentiation restores the native chromatin context (Fig. 4H).

Together, these data provide multi-layered validation that highly active regulatory regions, even after their complete genetic removal, can impose a heritable memory of an active state on the surrounding region, but this can be successfully erased by programs initiated during differentiation. More importantly, we show that the stability of this active state is silenced by architectural elements within non-coding DNA (Fig. 4I).

Immune engagement through synHLA and multi-locus augmentation

To test whether REWRITE could deliver additional antigen-processing genes beyond HLA, we assembled a 62 kb locus encoding *TAP1*, *TAP2*, *PSMB8*, *PSMB9*, and *BRD2* which form the peptide transporter complex and proteasomal machinery essential for antigen loading (Fig. 5A). This construct was derived from the NCRM2 background and then integrated into PGP1-ΔABC, which retains the endogenous copies of these genes. We generated ~7 clones per million nucleofected cells, twice the efficiency observed with synHLA, consistent with the smaller construct size (Figs. 5A, 3K). PCR and WGS confirmed full-length integration, scar-minimized,

resulting in triploid copies of all these genes (Figs. 5B–D, S12A–B). Furthermore, high pluripotency was retained (fig. S12C–D) and embryoid bodies were generated as expected (fig. S12E). These results demonstrate REWRITE's capacity for multi-locus installation of immune processing loci and lay the groundwork for reconstitution of the entire antigen presentation machinery.

Because thymic epithelial cells (TEC) orchestrate T-cell education and central tolerance via class-I HLA presentation, we next tested synHLA in this critical immune-educating lineage. We integrated our previously validated doxycycline-inducible *FOXN1* circuit (*32*), into wild-type, synHLA-, and mini-synHLA-engineered hPSCs and induced differentiation into TEC-like cells. These cells showed increased expression of the canonical TEC marker *DLL4* (Fig. 5E). Importantly, flow cytometry revealed that both wild-type and full-synHLA TECs exhibited minimal IFN-γ-inducible HLA expression (2–3%), consistent with the epigenetic ghost being successfully reset during differentiation (Fig. 5F). In contrast, mini-synHLA TECs showed significantly higher inducible expression (>20%), confirming its more robust anti-silencing architecture even in a differentiated context. These results demonstrate that synHLA remains transcriptionally competent after developmental transition, with architectural design dictating expressivity, supporting future applications in thymic antigen presentation, TCR education, and tolerance modeling.

To verify whether synHLA constructs restored functional immune recognition, we purified NK-cell and CD3+ T-cell populations from HLA-typed primary peripheral blood mononuclear cells (PBMCs) (fig. S13A–C). Each population was then co-cultured with wild-type PGP1, mini-synHLA (03:07:07), or blank engineered hPSCs matched or mismatched at class-I HLA alleles and assayed by flow cytometry and colorimetric analysis for cell death or damage. Parental PGP1-ΔABC cells lacking HLA class-I triggered NK-mediated cytotoxicity, consistent with missing-self recognition, whereas mini-synHLA engineered cells were spared (Figs. 5G, S13D). Moreover, in matched conditions for mini-synHLA, T-cell tolerance was preserved, whereas mismatches led to CD8+ mediated cell death (Figs. 5H, S13E).

Discussion

Our work introduces REWRITE, a platform for the iterative, scar-minimized rewriting of multi-genic regions in hPSCs (*26*). By demonstrating the first locus-scale refactoring of the genetically complex and immunologically consequential HLA region, REWRITE establishes a new and stringent benchmark for synthetic genomics in human cells (*17*). It provides a foundational technology for both the rational design of programmable immune interfaces, and the discovery of epigenetic principles governing human genome architecture. By integrating validated technologies like CRISPR and RMCE, REWRITE enables the precise engineering of complex loci without the toxic double-strand breaks (*33*) that hinder hPSC engineering (*10-12*). While scaling to even larger payloads will require further advances overcoming DNA biophysical constraints (*29*), REWRITE establishes a needed process for building the next generation of synthetic biological systems (*1*).

A key discovery of our work is the "epigenetic ghost," a potent, heritable memory imposed by an excised active regulatory region. This phenomenon represents a form of epigenetic inertia: a persistent de-repression of genetic sequence that resists returning to its silenced default state even after the inducing element is removed. Crucially, our direct comparison of full-length and compact constructs provided a definitive functional test, demonstrating that native intergenic DNA contains the essential architectural software to resolve this ghost state. This perhaps occurs by spreading of

heterochromatin from repressed repetitive elements (*34*) (Fig. 4I) or through dedicated insulator sequences (*35*). One might test for this with targeted approaches like ChIP-qPCR for histone modifications to circumvent the challenges of sequencing and mapping at the repetitive HLA locus (*36*). This discovery establishes non-coding DNA not as passive filler, but as an active manager of long-term epigenetic fidelity — a new cautionary design principle for synthetic genomics and minimal genomes (*37*). While differentiation's robust chromatin remodeling erases this ghost (Figs. 4H, 5F), its persistence provides a model for the incomplete epigenetic resets seen in iPSC somatic memory, offering an experimental system to dissect mechanisms of epigenetic fidelity.

This platform provides a powerful new paradigm for engineering the human immune interface. In contrast to "immune cloaking" strategies that sever immunological communication (*38-42*), REWRITE enables the rational programming of "immune-visibility" to preserve it. This has profound therapeutic implications, from creating custom haplotypes for diverse populations to unlocking a suite of future applications like high-throughput TCR discovery and the generation of off-the-shelf antigen-presenting cells (*43*). Furthermore, this hPSC-based platform represents a major shift toward more accurate, human-specific immune modeling, moving beyond the limitations of murine systems in alignment with recent funding priorities (*44*).

Looking forward, the bigDNA capacity of REWRITE enables allogeneic hPSC chassis to carry complex, programmable synthetic circuits (*45*). Its support for sequential integration facilitates the additive design-build-test cycles required to build increasingly sophisticated living therapies with multi-modal features like kill-switches, logic gates, and adaptive effector programs. By providing a system to both construct and understand complex, immune-interfacing systems, REWRITE helps unlock the genome and epigenome as programmable design spaces. This work thus lays the foundation for the next era of synthetic genomics: the architectural rewriting of entire megabase multi-genic loci to program human cell function.

References and Notes

1. J. S. James, J. Dai, W. L. Chew, Y. Cai, The design and engineering of synthetic genomes. *Nat Rev Genet*, (2024).
2. H. A. Wallace *et al.*, Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell* **128**, 197-209 (2007).
3. E. C. Lee *et al.*, Complete humanization of the mouse immunoglobulin loci enables efficient therapeutic antibody discovery. *Nat Biotechnol* **32**, 356-363 (2014).
4. L. A. Mitchell *et al.*, De novo assembly and delivery to mouse cells of a 101 kb functional human gene. *Genetics* **218**, (2021).
5. R. Brosh *et al.*, A versatile platform for locus-scale genome rewriting and verification. *Proc Natl Acad Sci U S A* **118**, (2021).
6. S. Pinglay *et al.*, Synthetic regulatory reconstitution reveals principles of mammalian Hox cluster regulation. *Science* **377**, eabk2820 (2022).
7. R. Brosh *et al.*, Synthetic regulatory genomics uncovers enhancer context dependence at the Sox2 locus. *Mol Cell* **83**, 1140-1152 e1147 (2023).
8. W. Zhang *et al.*, Mouse genome rewriting and tailoring of three important disease loci. *Nature* **623**, 423-431 (2023).
9. R. Ordonez *et al.*, Genomic context sensitizes regulatory elements to genetic disruption. *Mol Cell* **84**, 1842-1854 e1847 (2024).
10. T. P. Zwaka, J. A. Thomson, Homologous recombination in human embryonic stem cells. *Nat Biotechnol* **21**, 319-321 (2003).
11. R. J. Ihry *et al.*, p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat Med* **24**, 939-946 (2018).
12. M. Ohgushi *et al.*, Molecular pathway and cell state responsible for dissociation-induced apoptosis in human pluripotent stem cells. *Cell Stem Cell* **7**, 225-239 (2010).
13. N. Chaudhari, A. M. Rickard, S. Roy, P. Droge, H. Makhija, A non-viral genome editing platform for site-specific insertion of large transgenes. *Stem Cell Res Ther* **11**, 380 (2020).
14. F. Zhu *et al.*, DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. *Nucleic Acids Res* **42**, e34 (2014).
15. A. Blanch-Asensio *et al.*, STRAIGHT-IN enables high-throughput targeting of large DNA payloads in human pluripotent stem cells. *Cell Rep Methods* **2**, 100300 (2022).
16. J. Robinson *et al.*, The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res* **43**, D423-431 (2015).
17. J. K. Kulski, S. Suzuki, T. Shiina, Human leukocyte antigen super-locus: nexus of genomic supergenes, SNPs, indels, transcripts, and haplotypes. *Hum Genome Var* **9**, 49 (2022).
18. C. R. Perez, M. De Palma, Engineering dendritic cell vaccines to improve cancer immunotherapy. *Nat Commun* **10**, 5408 (2019).
19. Y. R. Na, S. W. Kim, S. H. Seok, A new era of macrophage-based cell therapy. *Exp Mol Med* **55**, 1945-1954 (2023).
20. A. Zeleniak *et al.*, De novo construction of T cell compartment in humanized mice engrafted with iPSC-derived thymus organoids. *Nat Methods* **19**, 1306-1319 (2022).
21. A. M. Lanza, T. J. Dyess, H. S. Alper, Using the Cre/lox system for targeted integration into the human genome: loxFAS-loxP pairing and delayed introduction of Cre DNA improve gene swapping efficiency. *Biotechnol J* **7**, 898-908 (2012).

22. S. Williams *et al.*, CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. *BMC Biotechnol* **5**, 17 (2005).
23. X. L. Li *et al.*, Highly efficient genome editing via CRISPR-Cas9 in human pluripotent stem cells is achieved by transient BCL-XL overexpression. *Nucleic Acids Res* **46**, 10195-10215 (2018).
24. J. Sheren, S. J. Langer, L. A. Leinwand, A randomized library approach to identifying functional lox site domains for the Cre recombinase. *Nucleic Acids Res* **35**, 5464-5473 (2007).
25. Y. Kameyama, Y. Kawabe, A. Ito, M. Kamihira, An accumulative site-specific gene integration system using Cre recombinase-mediated cassette exchange. *Biotechnol Bioeng* **105**, 1106-1114 (2010).
26. J. D. Boeke *et al.*, GENOME ENGINEERING. The Genome Project-Write. *Science* **353**, 126-127 (2016).
27. D. L. Neil *et al.*, Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. *Nucleic Acids Res* **18**, 1421-1428 (1990).
28. F. T. Merkle *et al.*, Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature* **545**, 229-233 (2017).
29. L. L. Lesueur, L. M. Mir, F. M. Andre, Overcoming the Specific Toxicity of Large Plasmids Electrotransfer in Primary Cells In Vitro. *Mol Ther Nucleic Acids* **5**, e291 (2016).
30. M. Drukker *et al.*, Characterization of the expression of MHC proteins in human embryonic stem cells. *Proc Natl Acad Sci U S A* **99**, 9864-9869 (2002).
31. A. M. Tsankov *et al.*, A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells. *Nat Biotechnol* **33**, 1182-1192 (2015).
32. T. Yanagi *et al.*, Termination sequence between an inducible promoter and ubiquitous chromatin opening element (UCOE) reduces gene expression leakage and silencing. *J Biol Eng* **19**, 29 (2025).
33. T. Ohno *et al.*, Biallelic and gene-wide genomic substitution for endogenous intron and retroelement mutagenesis in human cells. *Nat Commun* **13**, 4219 (2022).
34. N. Stamidis, J. J. Zylicz, RNA-mediated heterochromatin formation at repetitive elements in mammals. *EMBO J* **42**, e111717 (2023).
35. C. K. Y. Hong *et al.*, Massively parallel characterization of insulator activity across the genome. *Nat Commun* **15**, 8350 (2024).
36. C. Xie *et al.*, Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. *Proc Natl Acad Sci U S A* **114**, 8059-8064 (2017).
37. A. Cabrera *et al.*, The sound of silence: Transgene silencing in mammalian cell engineering. *Cell Syst* **13**, 950-973 (2022).
38. G. G. Gornalusse *et al.*, HLA-E-expressing pluripotent stem cells escape allogeneic responses and lysis by NK cells. *Nat Biotechnol* **35**, 765-772 (2017).
39. X. Han *et al.*, Generation of hypoimmunogenic human pluripotent stem cells. *Proc Natl Acad Sci U S A* **116**, 10441-10446 (2019).
40. T. Deuse *et al.*, Hypoimmunogenic derivatives of induced pluripotent stem cells evade immune rejection in fully immunocompetent allogeneic recipients. *Nat Biotechnol* **37**, 252-258 (2019).
41. X. Hu *et al.*, Hypoimmune induced pluripotent stem cells survive long term in fully immunocompetent, allogeneic rhesus macaques. *Nat Biotechnol* **42**, 413-423 (2024).

42. R. Lanza, D. W. Russell, A. Nagy, Engineering universal cells that evade immune detection. *Nat Rev Immunol* **19**, 723-733 (2019).
43. C. A. Klebanoff, S. S. Chandran, B. M. Baker, S. A. Quezada, A. Ribas, T cell receptor therapeutics: immunological targeting of the intracellular cancer proteome. *Nat Rev Drug Discov* **22**, 996-1017 (2023).
44. NIH, NIH to prioritize human-based research technologies. *New initiative aims to reduce use of animals in NIH-funded research.* 2025.
45. C. J. Bashor, I. B. Hilton, H. Bandukwala, D. M. Smith, O. Veiseh, Engineering the next generation of cell-based therapeutics. *Nat Rev Drug Discov* **21**, 655-675 (2022).
46. H. H. Kuo *et al.*, Negligible-Cost and Weekend-Free Chemically Defined Human iPSC Culture. *Stem Cell Reports* **14**, 256-270 (2020).
47. M. Gu, Efficient Differentiation of Human Pluripotent Stem Cells to Endothelial Cells. *Curr Protoc Hum Genet* **98**, e64 (2018).

Acknowledgments: We would like to acknowledge members of the Neochromosome team for their contributions on assembling the synHLA variants. Following an initial draft written by the authors, AI-assisted technologies (ChatGPT-4o/5 and Gemini Pro 2.5) were used by the corresponding author (DMT) to aid in condensing the manuscript and to generate adversarial critiques to improve clarity and rigor. The authors reviewed and edited all AI-generated suggestions and are solely responsible for the final content of the manuscript.

Funding:

National Institutes of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) grant R43AI148008 (DMT, LAM)

National Institutes of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) grant DP2AI154417 (DMT)

Author contributions:

Conceptualization: DMT

Methodology: SFG, SL, SJ, MK, SD, SFP, BY, TY, TLD, NA, MH, DMT

Investigation: SFG, SL, SJ, MK, SD, SFP, BY, TY, TLD, NA, MH, DMT

Visualization: SFG, SL, SJ, MK, SD, SFP, BY, TY, DMT

Funding acquisition: DMT, LAM

Project administration: DMT

Supervision: DMT

Writing – original draft: SFG, SL, SJ, SD, SFP, TY, DMT

Writing – review & editing: SFG, DMT

Competing interests: This work has been filed for patent rights under application US20230295668A1. DMT and NA are former employees of Opentrons/Neochromosome Inc, and own private shares. LAM, TD, and MH are employees of Opentrons/Neochromosome Inc.

Data and materials availability: All raw fastq files from whole-genome sequencing and RNA-seq are available at BioProject PRJNA1327383. Plasmids and cell lines are available through a materials transfer agreements (MTAs). All other data are available in the main text or the supplementary materials.

Supplementary Materials

Materials and Methods

Figs. S1 to S13

Tables S1 to S5

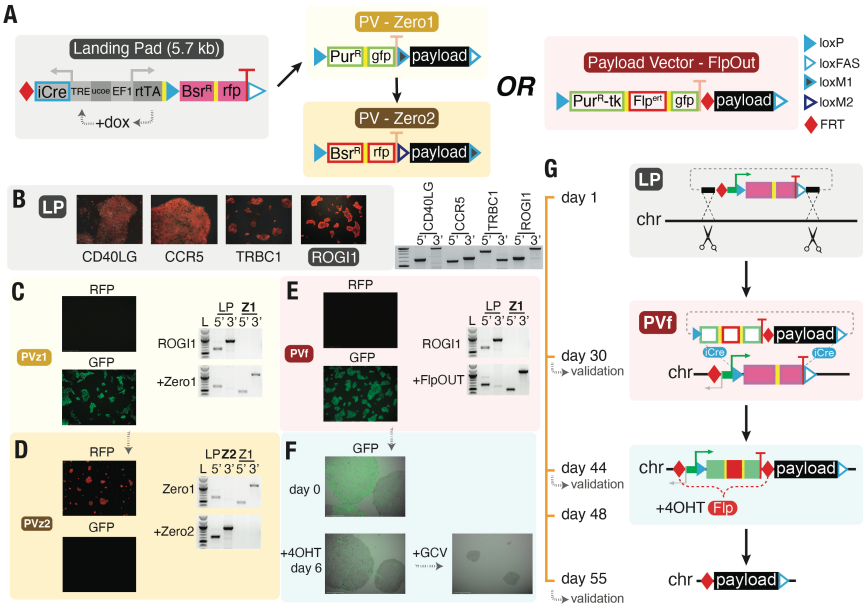

Fig. 1. REWRITE platform architecture and sequential integration in hPSCs.
(**A**) Schematic of the Landing Pad (LPneo) and terminal Payload Vector (FlpOut/PVf). LPneo contains a doxycycline-inducible iCre (via TRE-rtTA), Blasticidin resistance (Bsr^R), mScarlet (RFP), and a wild-type loxP site, flanked by an FRT site for marker removal. An alternate version (LPmin) lacks TRE-iCre. Golden Gate cloning is used to append locus-specific homology arms (fig. S1). FlpOut vectors encode a puromycin–thymidine kinase fusion (Pur^R-TK), tamoxifen-inducible FlpERT2, mNeonGreen (GFP), and a second FRT site. Alternatively, orthogonal payloads (Zero1/PVz1, Zero2/PVz2) enable iterative integrations via loxM1 and loxM2. All systems use wild-type loxP for initial cassette exchange.
(**B**) Targeted integration of LPneo into four transcriptionally silent loci in hPSCs, with junction PCR confirming successful LPneo insertion.
(**C**) Integration of PVz1 into LPneo at ROGI1, with GFP expression and junction PCR confirmation.
(**D**) Iterative RMCE of PVz2 into the PVz1 cassette using loxM1 (while introducing loxM2), with RFP marker and validated junctions.
(**E**) RMCE-mediated integration of FlpOut into the LPneo-ROGI1 locus, showing RFP-to-GFP marker conversion and junction PCR validation.
(**F**) Tamoxifen induction of FlpERT2 results in >80% marker excision by day 6, with additional negative-selection using ganciclovir (GCV) to remove residual Pur^R-TK–positive cells.
(**G**) Timeline schematic illustrating sequential LP insertion, FlpOut integration, and inducible marker removal.

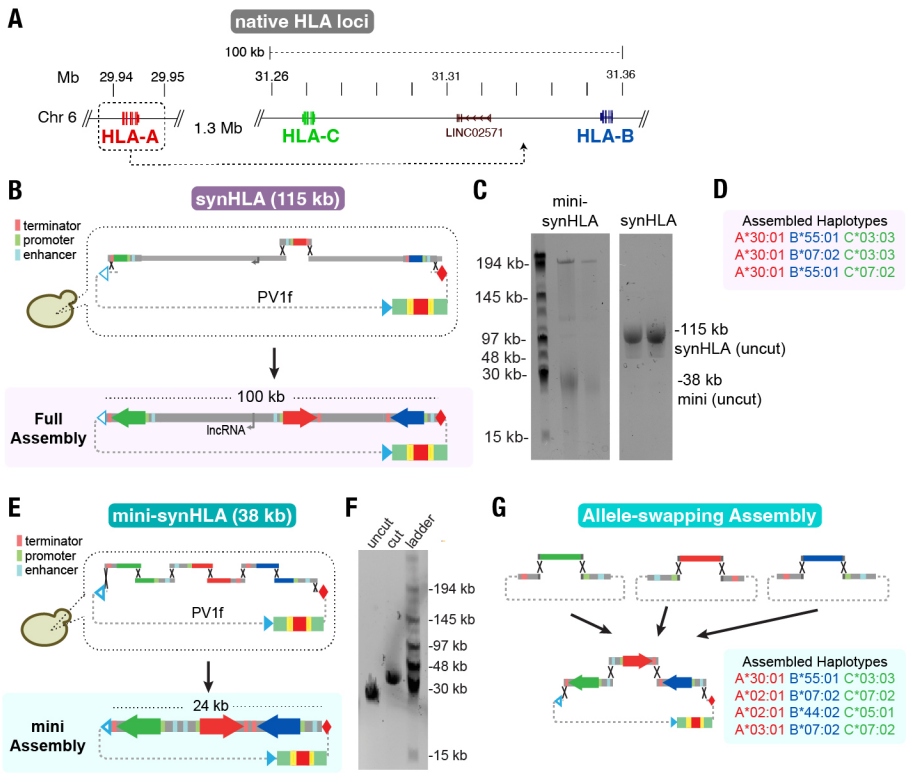

Fig. 2. Synthetic HLA haplotypes assembled by modular bigDNA engineering.
(**A**) Genomic organization of the native class-I HLA loci on chromosome 6, showing the 1.3 Mb separation between *HLA-A* and *HLA-B*. REWRITE enables repositioning of *HLA-A* into a contiguous synthetic haplotype with *HLA-B* and *HLA-C*.
(**B**) Modular assembly of full-length synHLA (115 kb) by in-yeast recombineering using BAC-derived segments and synthetic junctions. FlpOut PV vector backbone shown.
(**C**) Field-inversion gel electrophoresis (FIGE) of uncut synHLA and mini-synHLA constructs, confirming correct assembly and plasmid size integrity.
(**D**) Example haplotypes generated by allele-swapping for class-I *HLA-A*, *-B*, and *-C* genes. Colors correspond to native loci shown in (A).
(**E**) Modular assembly of compact mini-synHLA (38 kb) using defined enhancer, promoter, and coding segments for each gene.
(**F**) FIGE validation of mini-synHLA construct size (uncut and restriction-digested forms).
(**G**) Schematic of allele-swapping strategy using modular fragments to generate multiple synthetic haplotypes across different genetic backgrounds.

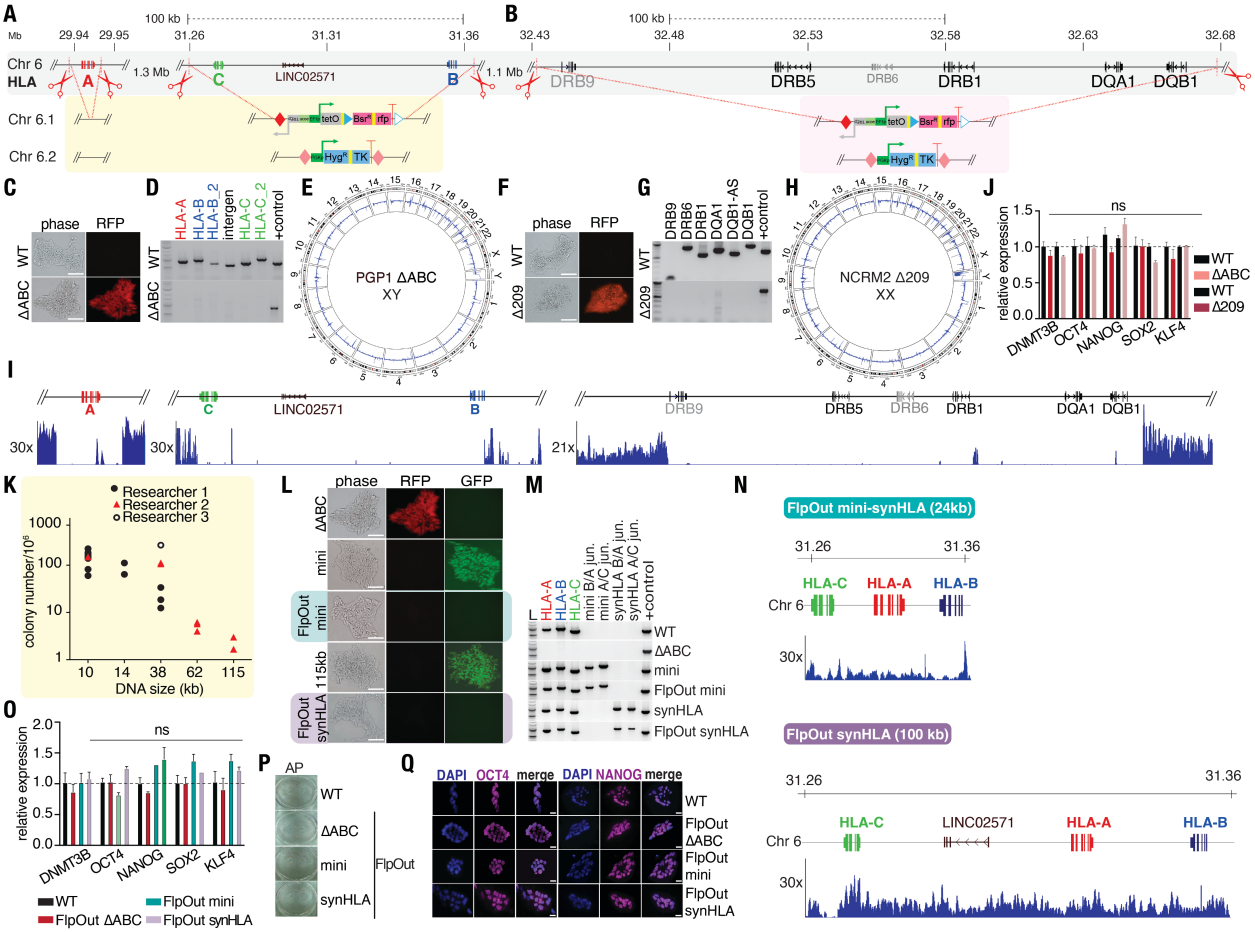

Fig. 3. Genome rewriting of class-I and class-II HLA regions in human pluripotent stem cells.

(**A**) CRISPR-based biallelic deletion of *HLA-A* (8 kb) and *HLA-B*/*C* locus (100 kb) in PGP1 hPSCs followed by landing pad and HygTK "deleter" cassette integration.

(B) CRISPR-based deletion of 209 kb class-II HLA cluster in female NCRM2 hPSCs.

(C) Representative images of WT PGP1 and ΔABC hPSCs in phase contrast and with the landing pad integration (RFP) (Scale bar = 200 µm)

(D) PCR gel verifying *HLA-A*, *-B*, *-C* class-I biallelic deletions in ΔABC. WT PGP1 as a control.

(E) Circos plot showing no chromosomal aberrations of PGP1-ΔABC.

(F) Representative images of WT NCRM2 and Δ209 kb class-II HLA hPSCs shown in phase contrast and with the landing pad integration (RFP). (Scale bar = 200 µm)

(G) PCR gel verifying *HLA-DRB1*, *-DQA1*, *-DQB1* class-II gene biallelic deletions in Δ209 kb hPSC. WT NCRM2 as a control.

(H) Circos plot showing no chromosomal aberrations of NCRM2-Δ209 hPSC.

(I) Biallelic deletion of class-I HLA and 209 kb class-II HLA genes is shown via WGS against hg38.

(J) Expression of pluripotency genes was measured by RT-qPCR in WT PGP1, ΔABC, WT NCRM2, Δ209kb class-II HLA. Gene expression was normalized to WT NCRM2 (dotted line).

Data represent mean ± SEM from n=3 independent experiments and p-values were calculated by two-way ANOVA t-test (* $p<0.05$, ns=not significant).

(K) Total integrated colonies obtained normalized per million cells for different DNA size (kb) payloads nucleofected using the REWRITE method. The graph shows experiments performed by three different researchers.

(L) Representative images of ΔABC, mini- and synHLA integration before and after flippase removal of the cassette are shown in phase contrast, with the landing pad integration (RFP), and the payload deliveries (GFP). (Scale bar = 200 μm).

(M) PCR gel verifying the integration of the class-I HLA genes upon delivery of mini- and synHLA payloads. WT PGP1 and ΔABC as controls. Specific junctions between HLA-B/A and HLA-A/C are amplified to show the integration of the synthetic cassette.

(N) The precise delivery of class-I HLA genes for both mini- and synHLA payload into the native site is shown via WGS against hg38.

(O) Expression of pluripotency genes was measured by RT-qPCR in WT PGP1, FlpOut ΔABC, FlpOut mini-synHLA and synHLA. Gene expression was normalized to WT PGP1 (dotted line). Data represent mean ± SEM from n=3 independent experiments. P-values were calculated by two-way ANOVA t-test (* $p<0.05$, ns = not significant).

(P) Images representing alkaline phosphatase staining for each marker-excised cell line. WT PGP1 as a control.

(Q) Immunostaining images of Oct4 and Nanog in WT PGP1, FlpOut ΔABC, mini- and synHLA. Cell nuclei were stained with DAPI. (Scale bar = 50 μm).

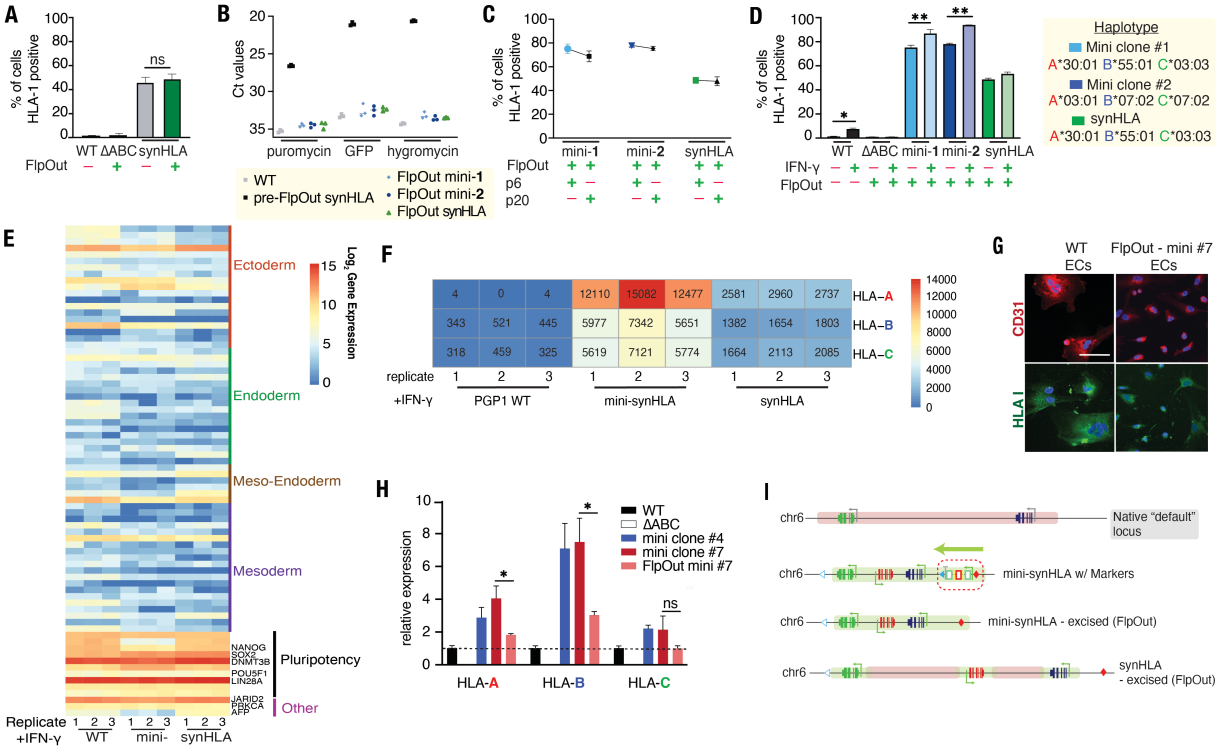

Fig. 4. Functional restoration of HLA expression reveals epigenetic memory.

(A) The expression of class-I HLA was determined by flow cytometry in WT PGP1, FlpOut ΔABC, and pre- and post-marker excised synHLA hPSCs. Class-I HLA is stained with anti-pan HLA class-I antibody.

(B) Verification of complete marker excision in all FlpOUT hPSCs by qPCR. Pre-excision synHLA hiPSCs with a Ct value in the 20s. All other samples showed non-specific detection. Data represent mean ± SEM from n = 3 independent experiments.

(C) Cell division does not resolve HLA expression. Post-excision synHLA hPSCs (P6, passage 6) grown for an additional 14 passages (P20, passage 20) show no statistically significant change in cells expressing HLA as analyzed by flow cytometry. Data represent mean ± SEM from n = 3 independent experiments.

(D) The expression of class-I HLA was determined by flow cytometry in WT PGP1, FlpOut ΔABC, FlpOut (marker-excised) mini- and synHLA hPSCs with & without IFN-γ induction. Class-I HLA is stained with anti-pan HLA class-I antibody. Data represent mean ± SEM from n = 3 independent experiments. P-values were calculated by unpaired two-tailed t-test ($p < *$ 0.05, ** 0.01, ns= not significant). Two different haplotypes for the mini-synHLAs were used.

(E) Log_2 gene expression heatmap of 80 pluripotency and lineage-specific markers from RNA-seq.

(F) Gene expression values of *HLA-A*, *HLA-B*, and *HLA-C* on WT PGP1, FlpOut mini-, and FlpOut synHLA (n = 3 biological replicates) from RNA-seq.

(G) Representative images of WT PGP1 and FlpOut mini synHLA clone #7 upon mesodermal endothelial-like cell differentiation. Cells were stained with CD31 marker (red) and anti-pan HLA class I antibody (green) upon IFN-γ induction. (scale bar=150 µm)

(H) Expression of class-I HLA genes was measured by RT-qPCR in WT PGP1, ΔABC, mini-synHLA clone#4 and #7, FlpOut mini-synHLA clone #7 upon mesodermal endothelial-like cell differentiation. Gene expression was normalized to WT PGP1 (dotted line). Data represent mean ± SEM from n=3 independent experiments. P-values were calculated by unpaired two-tailed t-

test (* $p < 0.05$, ns= not significant)
(I) Diagram of transcriptional memory observed at the locus. Native locus is silenced. The marker exerts a transcriptional opening effect even after being excised. The full synHLA with intergenic DNA silences expression more readily than without.

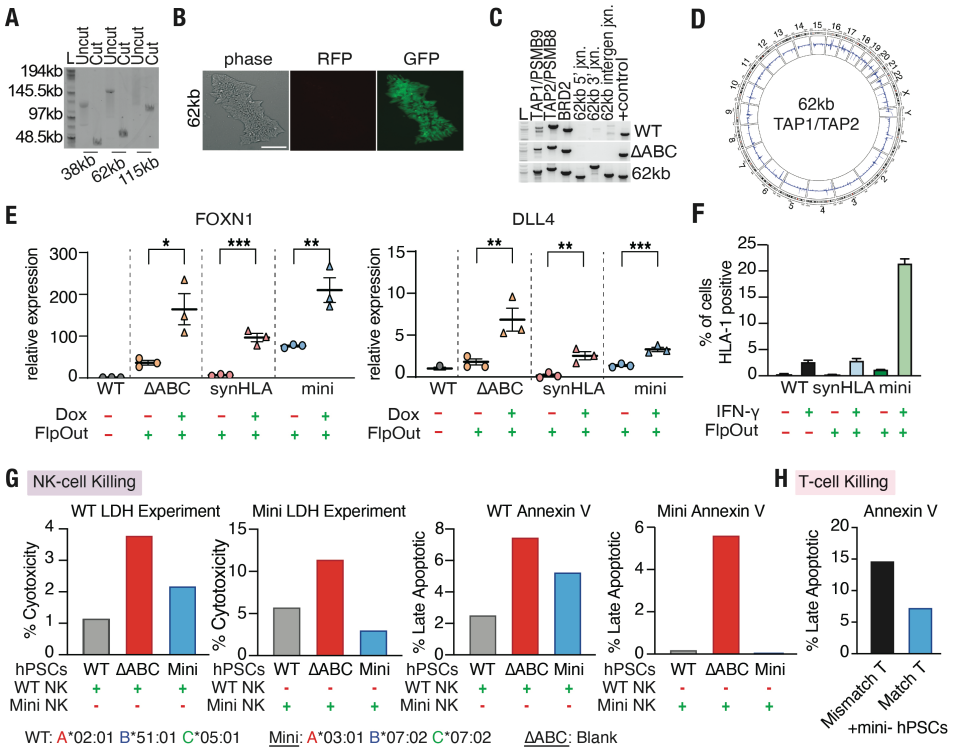

Fig. 5. Multi-locus programming in hPSCs and engineered immune recognition.
(**A** Field-inversion gel electrophoresis (FIGE) of uncut mini-synHLA, 62 kb TAP construct, and synHLA confirming correct assembly and plasmid size integrity.
(B) Representative images of 62 kb integration are shown in phase contrast, and with the payload deliveries (GFP). (Scale bar = 200 µm).
(C) PCR gel verifying the integration of *TAP1/PSMB9*, *TAP2/PSMB8*, and *BRD2* genes upon delivery of the 62 kb payloads. WT PGP1 as a control. Specific junctions between the synthetic genes are amplified.
(D) Circos plot showing no chromosomal aberrations of PGP1 62 kb. WT PGP1 as a control.
(E) Expression of *FOXN1* and *DLL4* genes was measured by RT-qPCR in WT PGP1, FlpOut ΔABC, FlpOut mini- and synHLA cell lines carrying the inducible *FOXN1* circuit. Cells were cultured with or without 1 µg/ml doxycycline (Dox) treatment. Data represents mean ± S.D. from n = 3 independent experiments. P-values were calculated by two-tailed Student's t-test (p < *0.05, **0.01, ***0.005).
(F) Flow cytometry for anti-pan class-I HLA presentation in thymic epithelial like- cells (TECs) with and without IFN-γ generated from WT PGP1, FlpOut synHLA, and FlpOut mini-synHLA. hPSCs were induced into TECs with Dox for 7 days. Data represent mean ± SEM from n = 3 independent experiments.
(G) Graphs showing cytotoxicity observed via Lactate Dehydrogenase (LDH) release assay and Annexin V staining in IFN-γ induced hPSCs following co-culture with WT-PGP1 or mini-synHLA HLA-typed NK cells sorted from PBMCs (n = 2 experiments for LDH release)
(H) Representative graph showing cytotoxicity observed via Annexin V and 7-AAD staining in IFN-γ -induced engineered mini-synHLA hPSCs following co-culture with A*03:01, B*07:02, C*07:02 and WT-PGP1 HLA-typed T cells sorted from PBMCs (n = 2 biological replicates).