Systematic characterization of existing and novel inducible transgenic systems in human pluripotent stem cells after prolonged differentiation

Michael D. Gallagher[1,*,#], Andrew S. Khalil[1,2,3,*], Qi Liu[1], Tenzin Lungjangwa[1], Antar Drews[1], Blake F. Hanan[2,3], Moritz List[1], Henry Hardart[1], David J. Mooney[2,3,#,†], Rudolf Jaenisch[1,4,#,†]

[1]Whitehead Institute for Biomedical Research, Cambridge, MA, USA
[2]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
[3]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA
[4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA
*contributed equally
[#]co-corresponding authors
[†]co-senior authors

ABSTRACT

The ability to control transgene expression both temporally and quantitatively in human-relevant cells and tissues is a cornerstone of biomedical research. Additionally, precise transgene control is crucial for optimizing human cell-based gene therapies. Human pluripotent stem cells (hPSCs) have facilitated major advances in disease modeling and the potential for regenerative medicine. Still, they are significantly limited by the lack of inducible transgenic systems that avoid silencing but maintain robust inducibility after differentiation to defined cell lineages. Here we systematically characterize the leakiness, inducibility, and tunability of multiple existing and novel transgenic systems in hPSCs and differentiated macrophages and microglia. Notably, we report the application of a small molecule-mediated splicing switch (X^{on}) that allows for tunable transgene expression both before and after differentiation, without the large protein tags required for current state-of-the-art degron-based methods. We use X^{on} to achieve tight control of reporter genes, overexpression of multiple neurodegeneration-associated genes, and Cas9-mediated genome editing. We also characterize the current limitations of this system and describe approaches that can alleviate some of these limitations. By assessing multiple transgenic systems whose inducibility spans the transcriptional, post-transcriptional, and post-translational levels, we highlight and improve upon a major technical challenge that hinders basic, translational, and clinical research in physiological human-based systems.

INTRODUCTION

Conditional gene expression systems are fundamental to precise molecular interrogation of cellular function. Classical platforms such as Tet-On/Off and Cre-Lox provide temporally resolved, gene-specific perturbations in defined experimental contexts[1, 2] . These systems enable both the timed induction or silencing of genes and their associated pathways, and orthogonal manipulation of distinct cellular subsets within heterogeneous populations, thereby facilitating investigations into gene function, lineage tracing, circuit analysis, and other areas. The same principles extend to the inducible expression of artificial or engineered proteins, and conditional expression has the potential to improve gene therapies[3]. However, translating these tools to human-based platforms, particularly those involving pluripotent stem cells (hPSCs) or fully differentiated cells and tissues, remains hampered by insufficient transgene control and challenges in efficient and homogeneous delivery.

Incomplete repression, poor induction fidelity, and lack of dosage control are persistent problems with conditional gene expression technologies. These limitations often result in ambiguous phenotypic readouts, especially when studying potent cell-state regulators such as transcription factors (TFs)[4-6]. Importantly, functional genomic approaches such as CRISPR/Cas9, transposase-mediated integration, and pooled library screens now underpin the majority of systematic gene function studies. While these systems have dramatically expanded the capacity to link genetic perturbations to phenotypes at scale, they rely on robust and tunable delivery of such perturbations, which can then be coupled to phenotypic outputs. However, most screening platforms remain limited by the lack of conditional systems that provide tight, temporally controlled perturbations in primary or hPSC-derived settings[5, 7, 8]. Leaky or unstable transgene induction not only reduces screen resolution but can also introduce selective pressures and bias the interpretation of results.

hPSCs, including both embryonic (ESCs) and induced pluripotent stem cells (iPSCs), are a uniquely enabling platform for basic and translational research in human systems. Their capacity for self-renewal and directed differentiation enables the generation of standardized, genetically defined, and reproducible sources of human cells. The ability to engineer hPSCs at the pluripotent stage, thoroughly characterize and standardize them, and subsequently differentiate them into various cell types and tissues offers a powerful means to generate isogenic, experimentally tractable cells[5]. In this way, hPSC-based approaches circumvent canonical challenges in homogeneous and efficient delivery to primary and terminally differentiated cells. The net outcome is reduced inter-experimental variability, achieved by obviating the need for repeated virus production and transduction, or transient delivery of genetic tools, thereby expanding the scope for both basic and translational research. Additionally, complex multicellular assemblies, including hPSC-derived organoids, now offer some of the most sophisticated in vitro models of human development, disease, and therapeutic response[9-11]. While genetic background contributes to heterogeneity in hPSC-based studies[12, 13], this heterogeneity reflects, at least in part, genetic variation between individuals that may affect therapeutic responses in clinical trials and the broader population. Thus, therapies that are first validated across diverse hPSC genetic backgrounds may reduce the risk of downstream clinical development.

Among the challenges in well-controlled conditional expression systems in hPSC-based systems, perhaps the most persistent problem is the inability to maintain tightly controlled inducible transgene expression after prolonged differentiations[4]. Control “tightness” in this context refers to basal activity or “leakiness” of the conditional expression system. Even low-level transgene leakiness undermines the generation of conditional lines, particularly when the transgene encodes a TF or potent effector, where minimal expression can disrupt lineage fidelity or cell identity or state. Additionally, constitutive presence of genome-wide perturbing moieties (e.g., continuous Cas9 or transposase activity) can result in toxicity, preventing the establishment of stable lines[5]. Misexpression of regulatory genes or perturbation strategies such as Cas9

during differentiation further complicates the interpretation of phenotypic screens intended to measure the function of mature cells[5]. Finally, the introduction of transgenes into differentiated cell types via viral transduction or non-viral methods such as electroporation is often toxic, inefficient, or attainable only with unwanted cellular responses that complicate the interpretation of results[5]. As a result, study designs spanning individual gene perturbations to high-throughput screening in hPSC systems remain constrained relative to immortalized lines, both in tractability and reliability[7].

Stability challenges of these systems generally apply to employing the strategy of engineering hPSCs, differentiating them into the target cell type, and then the capacity to induce transgene expression. During differentiation, hPSCs undergo substantial epigenetic remodeling. During this process, epigenetic silencing of the synthetic promoters that undergird the most common conditional transgene systems presents an additional and frequently underestimated barrier[4, 5]. Both synthetic and viral promoters are susceptible to DNA methylation and/or heterochromatin formation, and these effects are often irreversible. Moreover, the unpredictable onset of silencing during extended differentiation windows, precisely when temporal control is most needed, further limits the utility of these systems[4]. Overcoming these hurdles is central to establishing hPSCs as a truly programmable and robust experimental platform.

Finally, precise control of transgene dosage is important for studying physiological gene function, dose-dependent effects, and gene therapies[3, 14, 15]. While often overlooked in traditional overexpression approaches, concentration-dependent functions have been widely reported[6, 16-18]. Additionally, the recent discovery of biomolecular condensates that partition functionally-related molecules into membrane-less organelles demonstrates further the critical role of macromolecule concentration in many cellular pathways[19, 20]. From a therapeutic perspective, the success of gene or engineered cell therapies critically relies on appropriate transgene expression, with haploinsufficient disorders as a prime example, and in cases where excessive transgene expression can lead to toxicity or unwanted effects[3].

Here, we have systematically characterized the strengths and limitations of several prevalent hPSC conditional expression systems. The systems examined span transcriptionally regulated (Tet-On), post-translationally regulated (DHFR degron), and post-transcriptionally regulated (X^{on}) systems. We additionally evaluated new optimization strategies to improve their function, specifically in hPSC-derived myeloid cells. We chose myeloid cells (macrophages and microglia) due to their sensitive and context-dependent functions in tissue homeostasis, immune surveillance, and inflammation[21, 22], as well as a general field interest in applying functional genomic studies to better understand their roles in health and disease. Additionally, their sensitive phenotypic responses necessitate tight transgene control, and their derivation requires long culture protocols (~4-12 weeks)[21, 23], providing a stringent test case for this study. We highlight the X^{on} system as a novel hPSC-based system that overcomes important limitations of existing systems, and allows for tight control of reporter genes, overexpression of neurodegeneration-associated genes in microglia, and Cas9-based genome editing. Collectively, these efforts establish a foundation for building next-generation gene expression platforms optimized for application with hPSC-based technologies.

RESULTS

Transposon-integrated Tet-On transgenes are not functional in hPSCs or hPSC-derived myeloid cells.

We selected six candidate master microglial transcription factors (TFs) that are downregulated or silenced in cultured microglia, as reactivating these TFs may generate important insights into microglial function and improve current *in vitro* microglial cell models[24-27]. TFs were cloned into validated Sleeping Beauty (SB) transposons (pSBtet-GP and pSBtet-RP) that have been optimized for balancing robust constitutive reporter expression and tight control of the Tet-On-driven transgene[28] **(Fig. 1A)**. Three TFs were cloned into each construct,

separated by 2A cis-acting hydrolase element (CHYSEL) sequences that encode for ribosome skipping peptides[29, 30]. Co-transfection of each transposon with the SB100X transposase[31] into HEK293Ts **(Fig. 1B)** resulted in persistent GFP+ and tdTomato+ (tdT) cells, suggesting transposon integration. Fluorescence-activated cell sorting (FACS) isolated pure GFP+ and tdT+ populations, and doxycycline (dox) treatment robustly increased expression of the first two TFs in each transposon **(Fig. 1C–D)**, but not the third. Two bands near the expected size of TAL1 were detected in cells lacking the transgene, and did not increase in intensity in dox-treated transgenic cells **(Fig. S1A)**. KLF2 blotting revealed dox-specific high molecular weight species, potentially representing translational read-through across 2A peptides[30] **(Fig. S1B)**. HEK293Ts with stable integrations of both pSBtet-GP-TF and pSBtet-RP-TF displayed dox-inducible TF expression from both transgenes **(Fig. 1E)**. After demonstrating induction of the first two TFs in each SB transgene in HEK293Ts, we integrated the same transposons into hPSCs that we have used to generate hPSC-derived microglia-like cells (hMGLs)[25, 32] **(Fig. 1F)**. Persistent GFP+ and tdT+ cells were observed and sorted 7 days after transfection, though at lower frequencies and with weaker fluorescence than what was observed in HEK293Ts. GFP and tdT expression diminished with passaging, and dox treatment failed to induce expression of SALL1 in pSBtet-GP-TF hPSCs or myeloid progenitors (MPs) **(Fig. 1G)**. These MPs also died within 24h of puromycin (puro) treatment (data not shown), indicating silencing of both inducible and constitutive transgenes in SB transposons.

To test whether stronger constitutive promoters might prevent transgene silencing[4, 5], we transfected hPSCs with EGFP plasmids driven by CMV, CAG and PGK promoters, and compared them to the pSBtet-GP-TF plasmid, in which EGFP is driven by a chimeric ribosomal (RPBSA) promoter[28]. CMV and PGK plasmids produced more robust GFP expression than RPBSA and CAG **(Fig. S1C)**, although CMV has been shown to be prone to silencing in hPSCs[5]. We therefore replaced the RPBSA promoter in the original pSBtet constructs with PGK, and additionally included EF1A, which is one of the strongest known promoters across diverse cell types[33] **(Fig. S1D)**. EF1A-driven plasmids (pSBtet-GP-EF1A and pSBtet-RP-EF1A) produced the brightest GFP and tdT expression, respectively, so we used them to generate stable GFP- and tdT-expressing hPSCs **(Fig. S1E–F)**. However, puro treatment killed nearly all cells in both populations, and the small number of surviving cells died 6-10 days after initiating myeloid differentiation **(Fig. S1G–H)**.

Safe harbor-targeting does not rescue Tet-On function in hPSC-derived myeloid cells.

To test whether semi-random integration of SB transposons contributed to the silencing of Tet-On-driven transgenes, we generated hPSCs with a validated Tet-On transgene[34] targeted to the AAVS1 safe harbor locus, based on approaches developed in our lab[35, 36]. We replaced the existing transgene with EGFP (pUCM-EGFP) for simple readout of inducibility, and observed robust EGFP 48h after dox treatment **(Fig. 2A)**, irrespective of time in culture after AAVS1 targeting. However, upon differentiation to hPSC-derived macrophages (hMacs), we observed limited-to-no ability to induce expression of EGFP **(Fig. 2A)**. Similar results were observed with hMGLs and with other AAVS1-targeted Tet-On transgenes, including those driven by the supposed silencing-resistant CAG promoter (data not shown). Previous reports of Tet-On transgene silencing in pluripotent stem cells suggest that DNA methylation of the tetracycline-responsive element (TRE) promoter is responsible, and recruitment of TRE-binding domains fused to the Tet1 demethylase domain rescued transgene silencing in mouse embryonic stem cells[37, 38]. To test this approach in hPSCs, we synthesized new Tet-On donor constructs that employed dual reverse tetracycline transcriptional activator (rtTA) transgenes and the TRE-binding domain of the tetracycline transcriptional silencer (tTS) in a single polycistronic CAG-driven cassette, with or without fusion of the TET1 demethylase domain **(Fig. 2B)**. In these systems, the tTS or tTS-TET1 fusion proteins should bind the TRE promoter in the absence of dox, and potentially prevent DNA methylation via steric hindrance and/or demethylase activity.

Upon binding dox, the tTS-based proteins should be evicted from the TRE promoter and replaced by rtTA, resulting in transgene expression. As a control, we synthesized a standard 3rd generation Tet-On (TetO3G) transgene with rtTA alone in the same format **(Fig. 2B)**.

To test the new constructs' inducibility, we performed a time course of induction in hPSCs. We observed variable but robust EGFP induction across all dual rtTA-tTS constructs, both with and without TET1 domains **(Fig. 2C)**. However, induction kinetics were delayed relative to pUCM-EGFP **(Fig. 2A)** and our TetO3G control **(Fig. S2B)**. Among the three dual systems, the C-terminal fusion demonstrated the highest level of transgene induction, the N-terminal fusion demonstrated the most uniform induction, and the non-fusion system demonstrated both moderately variable and the lowest level of inducibility **(Fig. 2C)**. With respect to leakiness, the C-terminal fusion showed a marked increase in population level green median fluorescence intensity (MFI) and the percentage of cells expressing EGFP, compared to non-transgenic (WT) cells. Via these assays, no other transgenes exhibited leaky expression **(Fig. S2A)**. To test the impact of the dual rtTA-tTS transgenes on conferring resistance to the transgene silencing observed in **Fig. 2A**, we subjected all hPSCs engineered with modified Tet-On systems alongside WT cells to hMac differentiation, followed by 144h of dox induction. However, we observed no EGFP expression via flow cytometry in any systems. **(Fig. 2D)**. In summary, Tet-On systems fail to induce transgene expression after hPSC myeloid differentiation, regardless of transgene delivery method, constitutive promoter choice or transgene vector.

DHFR degron-based systems are active in hPSC-derived myeloid cells but not tightly regulated.

Having observed silencing with multiple Tet-On systems and the inability to prevent silencing with TET1 demethylase fusion proteins, we tested whether other existing inducible methods allow for robust, tight transgene control in hPSC-derived myeloid cells. We selected the dihydrofolate reductase (DHFR) degrons, which target C-terminally fused proteins for degradation by the ubiquitin-proteasome pathway (UPP), and have been used for a variety of hPSC, immortalized cell line, and in vivo applications[39]. The small molecule trimethoprim (TMP) inhibits DHFR-mediated degradation, and can be added when stable protein levels are desired **(Fig. 3B)**. We fused optimized N-terminal and C-terminal DHFR degrons[40] to a CAG-driven EGFP transgene, generating single and dual degron-tagged transgenes (**Fig. 3A)**. We additionally generated a separate dual-degron-tagged EGFP using the exact degron sequences previously used for diverse CRISPR screening applications in hPSC-derived neurons and microglia[41]. All constructs were targeted to the AAVS1 locus in hPSCs, as described above.

To test the inducibility of the different degron designs, we performed a dose-response study of TMP from 1.1 to 8.6µM. We performed flow cytometry of treated hPSCs 24h after induction, and observed robust EGFP induction in the dual, N-terminal, and previously published (PP) dual DHFR degron-tagged proteins, but not the C-terminally-tagged protein **(Fig. 3C)**. Within this dose range we did not observe tunable levels of transgene expression in any system **(Fig. S3A),** and we observed evidence of significant leakiness in the form of increased population level green MFI for the robustly inducible constructs (DHFR-EGFP, DHFR-EGFP-DHFR, and DHFR-EGFP-DHFR (PP)), as well as increases in EGFP+ cells for the N-terminal- and dual-tagged designs **(Fig. S3B)**. To test transgene inducibility after myeloid differentiation, we differentiated the degron hPSCs to hMacs, as described above, and treated them with 2.5uM TMP, a plateau concentration observed in the hPSC dose response study **(Fig. S3A)**. We observed reduced inducibility across all degron systems, but similar relative performance as observed in hPSCs, including lack of inducibility of the C-terminal degron design. Notably, substantial proportions of EGFP- cells were present in all systems **(Fig. 3D)**. Thus, DHFR degrons have limited utility in hPSC-derived myeloid cells, and do not allow for robust or tunable transgene induction.

X^{on} inducible splicing system allows for reproducible, robust and highly tunable transgene expression.

Given the persistent and unresolvable epigenetic silencing of the Tet-On systems, and lack of tunability and limited inducibility of DHFR-based degrons in hPSC-derived myeloid cells, we explored the recently reported inducible splicing (X^{on}) system[42]. X^{on} consists of a ~1.2kb sequence from the 5' end of the human *SF3B3* gene, and includes a pseudoexon (PE) that is only spliced into the transgene mRNA in the presence of LMI070 (branaplam), a small molecule that increases U1 snRNP recruitment to 5' splice sites[43]. Previous studies have employed this system by moving the start codon of the gene of interest (GOI) into the PE, which results in LMI070-specific translation of the GOI with a 24 amino acid N-terminal tag[42] **(Fig. 4A)**.

To test this system, we targeted X^{on}-regulated EGFP (X^{on}-EGFP) constructs to the AAVS1 locus in hPSCs and examined inducibility of expression and leakiness in hPSCs, hMacs and hMGLs **(Fig. 4B)**. In all cell types, X^{on}-EGFP was robustly inducible and highly tunable, with expression levels strongly and linearly correlated with LMI070 dose (hPSC R^2=0.938, P=0.0067; hMGL R^2=0.932, R=0.0077; hMac R^2=0.958, P=0.021) **(Fig. 4C–K, S4)**. Uniform EGFP+ hPSC populations were obtained with as low as 5nM drug dose after 24h; however, doses of 30nM or higher resulted in morphological changes reminiscent of differentiation **(Fig. 4C, F)**. For hMGLs, 5nM and 20nM doses achieved 92% and 98% GFP+ cells, respectively **(Fig. 4D, G)**. Significant autofluorescence in hMacs, consistent with a highly phagocytic monocyte-derived macrophage identity[44], made it difficult to distinguish GFP+ cells from background signal. However, >90% of cells were unambiguously GFP+ at the 20nM dose **(Fig. 4E, H)**. These results were consistent across two different hPSC sources and an alternative, embryoid body-based hMGL differentiation protocol (hPSC #2 R^2=0.941, P<0.0001) **(Fig. S4)**. In addition, neither hMGLs or hMacs exhibited observable changes in cell morphology or apparent cell death with doses up to 100nM **(Fig. 4D-E, S4C)**. While minor EGFP leakiness was observed by flow cytometry **(Fig. 4F, S4)**, the levels observed were orders of magnitude below those seen in cells treated with even the lowest drug doses. In summary, X^{on} overcomes multiple limitations of existing inducible hPSC-based systems, and allows for highly tunable and silencing-resistant transgene expression.

Incorporating N-terminal 2A peptides can improve inducible X^{on} systems.

To test the utility of X^{on} for genes relevant to myeloid cell function, we replaced EGFP with SALL1, an important microglial regulator[45] **(Fig. 1A)**, downstream of X^{on}. Western blotting revealed that X^{on}-SALL1 hPSCs expressed SALL1 protein at higher levels than WT hPSCs, even in the absence of LMI070, and with no further induction after drug treatment. In contrast, EGFP exhibited dose-dependent and drug-specific expression **(Fig. 5A)**, consistent with the flow cytometry results in **Fig. 4**. We observed similar results with hPSCs from a different background (**Fig. S6A)** and in HEK293Ts transfected with the X^{on}-SALL1 donor construct **(Fig. 5B)**. To assess LMI070-mediated mRNA splicing, we designed qPCR primers to detect the expected mRNA species, and compared the results with those from X^{on}-EGFP hPSCs **(Fig. 5C)**. Two primer pairs verified the presence of the smaller, PE-excluded mRNA in untreated X^{on}-EGFP cells, and the larger, PE-included mRNA in treated cells **(Fig. 5D-E)**. However, untreated X^{on}-SALL1 cells contained an unexpected ~700bp product, which was sequence-confirmed to be the unspliced pre-mRNA. In treated X^{on}-SALL1 cells, only the expected splice product was detected **(Fig. 5E)**. When using primers outside the PE, we detected the expected mRNAs in both untreated and treated X^{on}-SALL1 cells, but failed to detect the larger unspliced RNA **(Fig. 5D)**. These results suggest that unspliced X^{on}-SALL1 pre-mRNA accumulates and is translated into full-length SALL1 protein, despite the absence of the PE and start codon.

We tested several strategies to ameliorate the apparent X^{on}-SALL1 RNA processing defect, including codon optimization and addition of P2A peptides, and tested these constructs in HEK293Ts. While codon optimization modestly reduced the leaky expression **(Fig. S6B)**, P2A peptides did not **(Fig. S6E)**. Surprisingly, an X^{on}-EGFP-P2A-SALL1 transgene maintained EGFP

inducibility despite leaky SALL1 expression, and the same effect was observed with an X^{on}-EGFP-P2A-TREM2 transgene **(Fig. S6C-D)**. In contrast, both WT and Alzheimer's mutant TREM2 were well controlled when EGFP-P2A was absent **(Fig. S6D)**. To test whether non-start ATG codons within EGFP-P2A allow for drug-independent translation, we mutated all five ATG codons (all located within EGFP) to isoleucine, but this did not prevent leakiness of either SALL1 or TREM2 **(Fig. S6B,E).** Finally, we tested SMAD3, another important microglial TF[46] **(Fig. 1)**, and GRN, a secreted growth factor linked to lysosomal function and multiple neurodegenerative diseases[47]. Both genes displayed LMI070-inducible expression in HEK293Ts, and addition of an upstream P2A peptide improved both systems **(Fig. 6A-B)**. AAVS1 targeting of X^{on}-P2A-GRN in hPSCs allowed for tightly controlled, tunable GRN overexpression **(Fig. 6C)**. We then differentiated these hPSCs to hMGLs and treated with LMI070, which resulted in a ~76% increase in secreted GRN **(Fig. 6D-E)**. In summary, X^{on} allows for tunable transgene expression in hPSCs, hMacs and hMGLs, with leakiness and inducibility depending on the transgene. Addition of a P2A peptide upstream of the transgene can improve transgene control, and this system can inducibly express disease-associated genes with established therapeutic relevance[47, 48] after hPSC differentiation to hMGLs.

X^{on}-Cas9 genome editing performance depends on sgRNA delivery strategy.

As we observed utility in N-terminal 2A-modified use cases of transgene overexpression in hMGLs **(Fig. 6)**, but detected leaky EGFP expression in the initial X^{on} validation and drug response studies **(Fig. S5**), we aimed to test the impact of potentially low levels of leaky expression on assays detecting irreversible events, such as genome editing. For the first assay, we employed a terminal genomic assay based on Cas9 double-strand breaks using hPSCs engineered with AAVS1-targeted X^{on}-Cas9 and transduced with a lentiviral sgRNA **(Fig. 7A)**. In a second assay, we used X^{on}-Cas9 hPSCs containing a constitutive fluorescent reporter targeted to the second AAVS1 allele, and transfected them with transient plasmid-based sgRNA targeting the reporter **(Fig. 7B)**.

First, we transduced WT and X^{on}-Cas9 hPSCs with a lentivirus containing a validated sgRNA targeting the transcriptional start site (TSS) of the endogenous B2M gene[49], differentiated these cells to hMacs and hMGLs, and treated with LMI070 **(Fig. 7A)**. Using a T7 endonuclease cleavage assay, we observed that while no editing occurred in WT cells, ~30-40% of alleles were edited in X^{on}-Cas9 cells, regardless of drug treatment **(Fig. 7B)**. These results suggest that X^{on}-regulated Cas9 is sufficiently leaky to achieve maximal genome editing in untreated cells with a constitutively expressed sgRNA. As an alternative strategy to address this limitation, we tested whether transient expression of the sgRNA would reduce Cas9 activity in the absence of LMI070. For this assay, we employed a dual AAVS1 targeting approach with X^{on}-Cas9 and an mScarlet (mSc) reporter protein inserted into each allele **(Fig. 7C)**. After targeting, we expanded the cells and then transiently transfected EGFP-containing plasmids containing a U6 promoter to drive sgRNA expression. We delivered the empty plasmid (pEGFP) and three sgRNA plasmids (pEGFP-sgRNA (3)), each containing a sgRNA targeting near the chromophore residue of the mScarlet protein, with and without LMI070 treatment. 48h post-transfection, we employed FACS to enrich EGFP-positive (sgRNA-containing) cells, and then cultured for an additional 5 days in 20nM LMI070. We then expanded the cells for an additional 5 days and performed flow cytometry for mSc expression. While we observed no difference in the mSc-negative population between cells transfected with pEGFP (with or without LMI070 treatment) and cells transfected with pEGFP-sgRNA (3) that were not treated with LMI070, sgRNA-transfected cells treated with LMI070 showed a significant increase in mSc-negative cells **(Fig. 7D-E)**. Thus, X^{on}-Cas9 generates inducible mScarlet knockout in cells containing mScarlet-targeting sgRNAs, with no evidence of leaky editing. Taken together, these results demonstrate that the X^{on} system can be used for diverse applications relevant to disease-associated gene

function studies, gene therapy, and genome editing, in hPSCs and fully differentiated microglia and macrophages.

DISCUSSION

Conditional control of transgenes provides substantial flexibility and powerful experimental benefits for relating genes to phenotypic function. Additionally, programmable genome perturbation technologies based on Cas9 allow for entire gene subsets or genome-wide functional screening to better understand human biology. hPSCs and their derivatives offer several unique advantages for developing transgenic studies to understand human biology[5]. First, these cells provide an unlimited and standardized supply of cells that can be differentiated into any human cell type – increasingly with native in vivo-like physiological relevance[25, 50]. Additionally, these cells are amenable to genetic engineering strategies, including Cas9-based targeted knock-in/out, as well as integrating viral transduction methods. When combined with conditional expression systems, these technologies are poised to offer a unique platform for experimentally connecting gene function to phenotype in any human cell type as well as new functional genomics and synthetic biology applications. However, persistent concerns are reported about the fidelity of these conditional expression systems as well as the unique challenges of epigenetic silencing of transgenes during hPSC differentiation. Here we systematically compared transcriptional (Tet-On), post-translational (DHFR degron), and post-transcriptional (X^{on} alternative splicing) systems across pluripotent and differentiated myeloid cell states. Our findings delineate both practical limitations and mechanistic barriers that collectively define the landscape of conditional expression in hPSC–based platforms.

As a proof-of-principle, we initially sought to conditionally express transcription factors (TFs) that have been implicated in microglial function based on epigenomic characterizations of microglia freshly isolated from human patients[51, 52]. Microglia and other tissue-resident macrophages (TRM) are highly sensitive to their environment, and independent lines of evidence suggest that current in vitro TRM cell models lack critical TFs that drive proper function *in vivo*[53]. Reactivation of these TFs in hPSC-derived microglia-like cells (hMGLs) and macrophages (hMacs) would thus allow for mechanistic insights into the roles of specific TFs in TRM function, how alterations in these TFs might be involved in human disease, and would generate an improvement over current hMGL and hMac cell models. As there are many candidate microglial TFs for reactivation[51], and unwanted expression of even a single powerful TF in hPSCs can induce unwanted differentiation, this experimental paradigm was a valuable test case for assessing the current limitations of inducible transgenic systems in hPSCs.

The Tet-On system is one of the most widely employed approaches to regulate expression of a transgene[54]. Via this approach, a transgene is placed under a synthetic promoter called a Tet Operator that contains serial repeats of a Tet-response Element (TRE). The modified cells also constitutively express the rT, reverse tetracycline-dependent DNA-binding protein, that binds to the TRE in the presence of doxycycline or tetracycline. In the Tet-On system, the rT is fused to a VP64 transactivating domain (rtTA). Upon treatment with tetracycline or doxycycline, the engineered cells begin transcription of the Tet Operator-driven cassette in a dose-responsive manner. To test this approach, we adopted a strategy of semi-random genomic integration of the Tet-On machinery via a Sleeping Beauty (SB) transposon and demonstrated stable, inducible expression in HEK293Ts **(Fig. 1A–E)**. However, the same approach failed to sustain consistent transgene functionality in hPSCs and displayed zero inducibility in hPSC-derived myeloid cells, as shown by weak-to-absent reporter fluorescence, loss of antibiotic resistance, and inability to detect dox-inducible TF expression **(Fig. 1F–G)**. Exploration of stronger promoters to drive the rtTA, antibiotic resistance and fluorescent reporter genes showed some improvement in early expression **(Fig. S1E–G)** but did not prevent silencing during differentiation **(Fig. S1H)**. These data are in line with prior studies and underscore oft-underappreciated challenges with applying Tet-On conditional expression systems in hPSC technology[5, 38, 55].

Epigenetic silencing of the TRE element has been previously reported to underlie the observed loss of inducibility. Specifically, previous studies have shown DNA methylation-mediated repression of the repeat elements in both the pluripotent state and during lineage commitment[37, 38]. As genome-wide regulation of promoter methylation status is expected to be region- and locus-dependent, we explored Cas9-based targeting of Tet-On constructs to safe-harbor loci known to retain expression during differentiation with specific promoters such as CAG. hPSCs targeted with the CAG-driven rtTA and a Tet-On-driven EGFP showed stable integration and robust expression of the transgene at the pluripotent state but still resulted in silencing following a ~4-week differentiation to myeloid lineage cells **(Fig. 2D)**. In addition to this limitation, we observed limited ability to maintain stable hPSC colonies (data not shown) when the EGFP transgene was replaced with TFs, suggesting low-level leaky expression of the Tet-On system, in line with previous studies[56-58].

Prior studies have shown that fusions of Tet-responsive DNA-binding domains to demethylating enzymes can be used to restore inducibility of cells differentiated from mouse pluripotent stem cells delivered in vivo[37]. Additionally, dual Tet-DNA binding systems that constitutively express both the rtTA protein and the canonical Tet-DNA binding protein that binds in the absence of tetracycline or doxycycline fused to a silencing KRAB domain (tTS) have been reported to reduce low-level leakiness of the Tet-On-driven transgene[59]. Here, we attempted to leverage both approaches in hPSCs and their derivatives and generated tTS fusion proteins containing a fusion of TET1 demethylase domains at either the N- or C-terminus of the tTS DNA-binding domain **(Fig. 2B)**. The selection of the tTS domain for the TET1 fusion was based on the assumption that during differentiation, in the absence of doxycycline, the tTS occupancy and thus the TET1 domain would be constant, potentially amplifying the demethylating activity at the Tet Operator. In hPSCs, all dual-regulated rtTA-tTS showed inducibility at the hPSC state, albeit with lower expression levels than the commonly used third-generation Tet-On (TetO3G) for the N- and no-TET1 fusion variations. All approaches also resulted in slower kinetics relative to the TetO3G system that lacked a tTS domain, indicating competition for the Tet Operator was present **(Fig. 2C)**. However, none of these variants resulted in inducibility in the myeloid state **(Fig. 2D)**, suggesting that the TET1 domain failed to prevent methylation of the TRE promoter during differentiation, or that silencing occurred via an alternative method.

As we and others have demonstrated robust and stable expression of safe-harbor transgenes via the constitutively active CAG promoter[5, 35], we explored additional systems where regulation does not occur via synthetic drug-inducible transcription factors that commonly feature repeat elements like the Tet Operator. The DHFR degron-based approach relies on post-translational stabilization of constitutively active proteins via introduction of C-, N-, or dual-terminal DHFR tags. In this iteration, in the absence of the stabilizing small molecule trimethoprim, the tagged protein is rapidly targeted for degradation via the ubiquitination–proteasomal pathway. Treatment with TMP prevents this process and rapidly results in active protein accumulation, as transcription and translation are constitutive, not rate-limiting[39]. Here, we explored similar EGFP constructs with N-, C-, and dual DHFR tags using amino acid sequences previously optimized for both termini[40], as well as other published sequences[41] **(Fig. 3A)**. All constructs were driven by the CAG promoter and inserted at the AAVS1 locus **(Fig. 3A–B)**. The N-, dual-, and previously published approaches resulted in TMP-dependent induction in hPSCs and hMacs **(Fig. 3C-D)** but did not provide tunable drug-responsive control **(Fig. 3C-D, S3A)**. After differentiation to hMacs, overall induction was reduced **(Fig. 3D)**, indicating either diminished TMP stabilization efficiency or altered CAG-driven expression dynamics in this cell type. Moreover, all functional constructs tested demonstrated measurable background fluorescence in untreated conditions above that of WT cells **(Fig. S3B)**. These findings align with previous observations that degron efficacy depends degron sequence and tagging strategy[39, 40], and that degron-based strategies may not be ideal for producing uniform populations of transgene-expressing cells.

As transcriptional and post-translational systems yielded insufficient inducibility and tight regulation in hPSC-derived myeloid cells here, despite optimization strategies, we explored the middle domain of protein production and recently described post-transcriptional inducible splicing. The X^{on} inducible splicing system employs constitutively active expression of an mRNA transcript, enabling the use of the validated CAG promoter from the AAVS1 safe-harbor locus, but with the coding region of the transgenic protein encoding region lacking an ATG start site[42]. This approach eliminates the repeated response elements of synthetic promoters designed for drug-inducible DNA-binding proteins like rtTA as well as the unfavorable dynamics of constitutively produced protein degradation/stabilization. With X^{on}, a drug-responsive pseudoexon containing a start codon can be alternatively spliced to produce an mRNA that restores the ATG start site in frame with the deficient transgene protein coding region of the transcript. Here, we integrated an ATG-deficient EGFP transgene in the AAVS1 locus, driven by the CAG promoter, and downstream of the drug-switchable pseudoexon **(Fig. 4A)**. Unlike the Tet-On and DHFR degron systems, the X^{on} approach resulted in robust inducibility to generate uniformly EGFP+ cells at both the hPSC and myeloid states (both hMGL and hMac), across two hPSC backgrounds, and with low nanomolar drug doses **(Fig. 4)**. Additionally, the system was highly tunable in all cell types and across similar LMI070 concentrations **(Fig. 4C–K)**, highlighting its reproducibility relative to transcriptional or post-translational approaches. However, low-level basal fluorescence was detectable in untreated cells **(Fig. 4F, S5A)**, indicating residual leakiness that, while modest in magnitude, can have significant implications for irreversible or cumulative assays as discussed below.

Transgene-specific performance further revealed mechanistic constraints. Large or complex open reading frames, such as the ~4kb SALL1, disrupted efficient splicing, leading to pre-mRNA accumulation and unintended protein expression, even in the absence of drug **(Fig. 5A–E, S6A)**. By contrast, smaller transgenes such as TREM2 and GRN were tightly and robustly inducible **(Fig. 6A-E)**. Codon optimization, 2A sequence inclusion, and mutation of upstream non-start methionines variably improved tightness, with P2A-linked GRN demonstrating particularly strong induction with minimal background **(Fig. 6B–E)**. These results highlight that transgene features such as length, complexity, and perhaps structure, affect splicing switch performance.

Despite achieving robust, tunable, and stable transgene induction, the X^{on} system's residual leakiness was sufficient to interfere with genome-editing assays that rely on irreversible molecular outcomes, a feature central to Cas9-enabled functional genomic screens. In hPSCs containing X^{on}-controlled Cas9 and constitutive sgRNA expression, we observed maximal editing even in the absence of LMI070 **(Fig. 7A-B)**. This basal activity of conditional induction systems, often undetectable by standard biochemical assays **(Fig. 5A–B)**, produced measurable genetic modification, illustrating that apparent repression in reporter assays may mask functionally relevant leakiness in irreversible contexts. When sgRNAs were delivered transiently to X^{on}-Cas9/mScarlet dual-targeted hPSCs, we observed clean LMI070-dependent DNA cleavage and loss of reporter function **(Fig. 7C–E)**, suggesting that the X^{on} leakiness level may be tolerable for Cas9-based assays featuring transiently delivered sgRNA. Alternatively, recently developed methods can overcome challenges to viral transduction of differentiated myeloid cells, such as Vpx-augmented hMGL (Gallagher et al, medRxiv, 2025) and hMac lentiviral transduction, reported by our lab and others[60, 61]. These findings demonstrate that even minimal leakiness of inducible systems can compromise the interpretability of functional genomic experiments, and highlight that common protein-based readouts of transgenic induction have limitations that should not be overlooked.

The results of the systematic comparison of conditional expression systems reveal persistent limitations when combined with hPSC technology that should be considered when deploying them in experimental studies. Broadly, the implications of our results show that hPSCs themselves introduce distinct challenges that impact individual system performance. In

transcriptionally regulated Tet-On systems, persistent promoter silencing **(Fig. 1–2)** that occurs with extended differentiation results in a near-complete inability to apply this approach in hPSCs, even when using safe harbor loci and transgene demethylation strategies. For post-translational DHFR degron-based strategies, incomplete protein degradation and lack of tunability limit their application where fine control of a transgene of interest, such as protein-function dose–response studies **(Fig. 3)**, is necessary. Furthermore, inducibility was modest after differentiation to myeloid cells, which constrains experiments that require robust transgene expression or overexpression. The recently reported X^{on} alternative splicing approach yielded consistent, robust, and tunable transgene expression across pluripotent and differentiated cell states in multiple hPSC backgrounds and differentiation protocols **(Fig. 4)**, albeit with detectable basal transgene expression **(Fig. S4)**. The results from our high-sensitivity terminal assay of Cas9-mediated indel detection **(Fig. 7)** underscore that even low-level transgene leakiness can result in biologically meaningful effects. Therefore, use of conditional transgenic systems in hPSC-based functional genomic studies requires careful assessments of leakiness and inducibility, particularly during and after prolonged differentiation protocols, and experimental controls to account for pre-activation or selection bias in phenotypic enrichment or depletion.

Within the systematic comparisons of the prevalent conditional expression technologies in this study, none likely afford the fidelity required for genome-wide, inducible perturbation studies in cells derived from hPSCs in their present form. Recently reported modified X^{on} systems, which improve splice site specificity and drug-regulated activity, and could further minimize cryptic splice activation and unspliced transcript accumulation, represent promising efforts in this regard[62, 63]; however, their performance in high sensitivity assays in hPSC-based systems has not yet, to our knowledge, been described. Still, even with these efforts aimed at individual system refinements, achieving such functionality will likely necessitate multi-layered strategies that operate across transcriptional, post-transcriptional, and post-translational conditional expression systems. Hybridization across these existing technologies, given their orthogonal control levels in the regulation of transgenic protein production, may provide an avenue for reducing leakiness through combined regulatory control. For example, coupling X^{on}-based splicing regulation with ligand-dependent DHFR degron tagging or co-expression of inducible repressors may allow for superior performance than any individual system. Further combination of demethylating fusions with chromatin insulators or alternative strategies to recruit VP64-DNA-binding domains to synthetic promoters without the use of repeat response elements might afford a third access point to further tighten conditional expression.

In summary, this study provides a comprehensive evaluation of conditional gene expression systems in hPSCs and their differentiated derivatives. The findings here collectively highlight the functional capabilities and pitfalls of these systems in their application within hPSC technologies as well as establish a framework for designing the most robust inducible expression platforms currently available. While no system in this study fully met the demands of complete repression, strong induction, and tunable output regardless of cell state, the X^{on} splicing platform, particularly augmented with the 2A modification described here, currently represents the most tractable solution in hPSCs. Still, its unpredictable alternative splicing, translational readthrough, and residual leakiness all highlight a substantial need for further refinement or alternative solutions. The results here and proposed future advancements described provide implications relevant to research efforts employing human pluripotent stem cell-derived models, advancing their application to functional genomics, their use in regenerative medicine, and in synthetic-biology-dependent gene therapy development.

METHODS

Molecular cloning: Molecular cloning was carried out by inserting DNA fragments into restriction enzyme-digested backbone plasmid vectors using New England Biolabs (NEB) NEBuilder HiFi DNA Assembly Master Mix. DNA fragments were generated using a mixture of NEB Q5 High-Fidelity 2X Master Mix PCR with oligonucleotide primers synthesized from Millipore Sigma-Aldrich, Integrated DNA Technologies (IDT) synthesized double-stranded DNA fragments, and/or additional restriction enzyme-generated DNA fragments from other plasmid inserts. Detailed descriptions of insert generation are described in each construct design methods section. HiFi assembly reactions were transformed into NEB 5-alpha, 10-beta, and Stable cells, with LB agar plates incubated at 30C for 16-24h. Colonies were grown in LB with 50ug/mL ampicillin at 30C overnight, and plasmids were purified with the Omega-TEK E.Z.N.A. Plasmid DNA Mini Kit. Open reading frames (ORFs) were validated by comprehensive Sanger sequencing (Eton Bioscence) or whole plasmid Nanopore sequencing (Plasmidsaurus). Correct clones were expanded for plasmid isolation with E.Z.N.A. Endo-Free Plasmid DNA Maxi kits, which were used for transfections or cloning into lentiviral plasmids.

Generation of transposon constructs: SALL1, TAL1 and MEF2A ORFs were PCR amplified from GenScript cDNA ORF clones. SMAD3, MEF2C and KLF2 were amplified from CS2 Flag-Smad3 (Addgene #14052), MSCV-IRES-Tomato MEF2C (Addgene #89715) and FUW-TetO-loxP-hKLF2 (Addgene #60850) plasmids. All amplicons were generated with the Roche KAPA HiFi HotStart ReadyMix PCR kit, with 5% DMSO for GC-rich templates. PCR primers contained 20bp overhangs complementary to either SfiI-digested pSBtet-GP or pSBtet-RP, P2A or T2A sequences. PCR products were verified by agarose gel electrophoresis and column purified with the Zymo Research DNA Clean & Concentrator kit. PCR reactions with unwanted products were gel purified with the Zymo Research Zymoclean Gel DNA Recovery kit before column purification. NEBuilder HiFi assembly kits were used to combine SALL1, SMAD3 and TAL1 fragments with pSBtet-GP (Addgene #60495), and MEF2A, MEF2C and KLF2 with pSBtet-RP (Addgene #60497). To replace the RPBSA promoter in pSBtet plasmids, EF1A and PGK promoters were PCR amplified from pLOX-EWgfp and PGK-H2B-mCherry plasmids, respectively. Purified products were ligated into KpnI-HF/Bsu36I-digested pSBtet-GP and pSBtet-RP with NEB T4 DNA ligase, followed by transformation, plasmid purification, and Sanger sequencing, as described above.

Generation of Tet-On constructs: We generated four Tet-On-based systems. These constructs included the previously published 3rd generation TetO3G system[64] and a combination Tet-Tight-based approach that expresses the 3rd generation rtTA transactivating domain (rT fused to the VP64 transcriptional activator) and tTS (tT fused to the KRAB repressor domain) as a polycistronic cassette (pMF-TetTight, Addgene #193288)[65]. In the Tet-Tight system, the tTS was cloned downstream of the rtTA and separated by a P2A sequence. The remaining two systems were based on the Tet-Tight construct and contained N- or C-terminal fusions with the human TET1 demethylase domain. All Tet-On-based systems were synthesized as dsDNA fragments (IDT) and cloned into the AAVS1-CAG-EGFP donor construct, such that all rtTA and tTS-based transgenes were driven constitutively via the CAG promoter, downstream of the splice acceptor and puromycin selection cassette. In each of the four Tet-On systems, a seven repeat Tet-response element operator was cloned downstream of the rtTA/tTS transgene β-globin polyA sequence in reverse, with an open cloning site followed by an SV40 polyA sequence using PCR amplification of the pUCM-AAVS1-TO-hNGN2 (Addgene #105840) plasmid. HiFi assembly was performed to generate the final products, as described above. The AAVS1 CAG-driven EGFP reporter was again synthesized as a dsDNA fragment (IDT) with overhangs (left: gcgctagcacgcgtcagctgactagaggatccccgggtac, right: gtaccgagctcgaattctccaggcgatctgacggttcact) following KpnI digestion of the parent Tet-system plasmid.

Generation of DHFR degron constructs: Synthesized dsDNA fragments (IDT) were used to generate EGFP with N-terminal, C-terminal or dual-DHFR tags, and were assembled into the AAVS1-CAG-EGFP donor construct. We used previously established trimethoprim (TMP)-responsive degron sequences[40] and a separate dual degron strategy employed in hPSC-based CRISPR screens[41, 66-68].

Generation of X^{on} constructs: An X^{on}-GFP piggyBac plasmid was kindly gifted by Jonathan Weissman's lab at the Whitehead Institute. We amplified the X^{on}-GFP cassette from this plasmid with primers containing 20bp overhangs for HiFi assembly with AgeI-HF/EcoRV-HF-digested AAVS1-tdTomato donor construct (Addgene #159275). The resulting X^{on}-GFP donor construct was used to test this transgene in hPSCs. To facilitate rapid cloning of different transgenes, an additional X^{on} donor construct (X^{on} "open") was generated with an AflII cut site for cloning in a gene of interest, followed by a subsequent round of HiFi assembly with synthesized DNA fragments containing plasmid homology arms: AATTGTTGTTTCCCGTGGGTAAGAGGATCC (left) and AATTGTTGTTTCCCGTGGGTAAGAGGATCC (right). For specific transgenes, (EGFP, SALL1, SMAD3, TREM2, GRN, Cas9) the corresponding fragments were synthesized without the start codon, but with 5' and 3' overhangs to allow for assembly into the digested donor construct. The R47H TREM2 mutation was tested due to its association with Alzheimer's disease risk[69]. P2A sequences were added by incorporating the necessary sequences into the overhangs, and EGFP-P2A was added upstream of the genes of interest with a separate synthesized dsDNA fragment. All transgenes were cloned into either the X^{on} "open" donor or the original AAVS1-tdTomato-based donor.

Generation of sgRNA constructs: For transient mScarlet sgRNA experiments, the mScarlet sgRNA was cloned into a modified px330-Cas9 plasmid in which Cas9 was replaced with EGFP (px330-EGFP), such that only sgRNA and EGFP would be expressed in transfected cells. sgRNAs were inserted into px330-EGFP using the BbsI-HF restriction enzyme, as previously described[70].

HEK293T cell culture and transfections: HEK293Ts were cultured in DMEM supplemented with 10% inactivated cosmic calf serum, 2mM L-glutamine and 1% penicillin/streptomycin. Cells were passaged every 2-3 days with trypsin at a 1:8 split ratio. Transfections were performed with X-tremeGENE 9 reagent, with up to 4ug of total DNA per well of a 6 well plate, and 3uL of X-tremeGENE 9 per 1ug of DNA.

hPSC cell culture and transfections: Previously published female iPSCs[25] and H1 ESCs were cultured and passaged as previously described (Gallagher et al, medRxiv, 2025). Transfections were performed with ViaFect (Promega), with up to 4ug of total DNA per well of a 6 well plate, and 2uL of ViaFect per 1ug of DNA. ViaFect and DNA were diluted in Opti-MEM, combined, and added to cells with E8 media in a total volume of 2-3mL per well of a 6 well plate. For cells maintained in mTeSR Plus or StemFlex, E8 was replaced with maintenance media the next morning. For promoter testing, pmax-GFP, PGK-H2B-GFP, pLOX-Ewgfp (Addgene #12242), pcag-GFP (Addgene #11150) were transfected into H1 ESCs and assessed by FACS or fluorescence microscopy.

HEK293T transposon experiments: Cells were transfected with pSBtet-based transposons and pCMV(CAT)T7-SB100 (Addgene #34879), which encodes the SB100X transposase. Mass ratios of 1:3 to 1:20 were tested, with 1:20 producing the largest number of integrations. GFP+ or tdT+ cells were isolated by FACS 7-10 days after transfection, and these populations were frozen in liquid nitrogen and treated with 0.75-7.5ug/mL doxycycline (dox) to assess dox-inducible expression of transcription factors. Cell lysates were extracted 2-4 days after dox treatment for

Western blot analysis. For generation of cells with integrations of both pSBtet-GP-TF and pSBtet-RP-TF transposons, cells were co-transfected with both transposon constructs and SB100X.

hPSC transposon experiments: iPSCs were transfected with pSBtet-based constructs and pCMV(CAT)T7-SB100 using the ViaFect transfection protocol described above. As with HEK293Ts, excess SB100X was used to maximize integration efficiency, and FACS-based isolation of cells with integrations was performed 7-10 days after transfection. Dox was added at 0.2 or 2ug/mL to iPSCs or iPSC-derived myeloid progenitors for 2-4 days before lysate extraction for Western blotting. Puro resistance was assessed by adding 2ug/mL puro to pSBtet-GP-TF myeloid progenitors pSBtet-EF1A-GFP iPSCs before and throughout hMGL differentiation.

hMac and hMGL differentiation: hMacs and hMGLs were generated with two differentiation protocols, one using an embryoid body approach, and one using a monolayer approach, as previously described[25, 71], (Gallagher et al, medRxiv, 2025).

hPSC AAVS1 targeting: To generate the stable expression of constitutive or inducible transgene reporters, our previously published AAVS1-CAG-EGFP donor construct[35] was used as a starting template. Genome targeting was performed as previously described using PX458-AAVS1-sg (Addgene #194721) to express Cas9 and an AAVS1-targeting sgRNA (GGGGCCACTAGGGACAGGAT)[36], and the various AAVS1 donor constructs generated in this manuscript. Introduction into cells was either carried out via electroporation (Lonza 4D-Nucleofector System, according to the manufacturer's instruction) or transfection. Transfections were performed with ViaFect, using 2.67ug of donor construct and 1.33ug of PX458-AAVS1-sg per well of a 6 well plate. For transgenes with puromycin (puro) resistance, cells were treated with sequentially increasing doses of puro (typically 0.5, 1 and 2ug/mL) for 2 days each, and surviving colonies were frozen in liquid nitrogen and maintained without subcloning for subsequent experiments. 10mM Y-27632 was typically needed for the first two days of puro treatment to prevent transgenic cells from listing off in the presence of widespread death.

DHFR degron experiments: To compare performance of the different degron designs, we targeted H1 hPSCs at the AAVS1 locus with the above degron constructs as described above. We then performed a trimethoprim (TMP) (Thermo Fisher Scientific, cat. #J67175.XF) dose-response curve via serial dilution of 2.5µg/ml (8.6 µM) to 0.312 µg/mL (1.1µM) diluted in mTeSR Plus medium. 24h post-treatment, EGFP intensity was measured via flow cytometry and compared that of WT (non-engineered) hPSCs, as well as each engineered degron without TMP treatment. For silencing outcome comparisons, the engineered H1 hPSCs were differentiated into hMacs and treated with 2.5µM TMP for 24h, with similar flow cytometry-based assessment of induction capacity.

Tet-On experiments: To compare the different Tet-On design strategies, we targeted H1 hPSCs at the AAVS1 locus with the various Tet-On systems described above. After puro-based selection, we performed a time course of induction in hPSCs using 1µg/mL of doxycycline in mTeSR Plus medium, and assessed EGFP fluorescence by flow cytometry. Measurements of induction efficiency were taken at 48, 96, and 144h post-treatment, with fresh dox-supplemented media replaced at each timepoint. For silencing experiments, hPSCs were differentiated into hMacs without prior dox exposure. hMacs were then subjected to 1µg/mL doxycycline for 144h and EGFP fluorescence was assessed by flow cytometry.

X^{on} transgene induction experiments: To evaluate X^{on}-EGFP in hPSCs, we performed a LMI070 (MedChemExpress, cat #HY-19620) dose-response curve via dilution of 100nM to 2nM in mTeSR Plus medium. After 24-48h, we assessed fluorescence intensity via flow cytometry and/or

Western blot, and compared to WT (non-engineered) hPSCs untreated (DMSO only) X^{on}-EGFP cells. For differentiated cells, dose response curves were performed after differentiation of X^{on}-EGFP hPSCs to hMGLs and hMacs, using both embryoid body[71] and monolayer[25] protocols. GFP was assessed by flow cytometry after 24h of treatment. X^{on}-EGFP, X^{on}-SALL1, X^{on}-TREM2, X^{on}-SMAD3 and X^{on}-GRN, including variations of these transgenes (see Fig. S3, 5 and 6), were tested by transfecting HEK293Ts with AAVS1 donor constructs containing each indicated transgene, followed by treatment with 20nM LMI070 or DMSO for 24h-48h. Western blots were used to assess expression of the transgenes. X^{on}-SALL1 and X^{on}-P2A-GRN were additionally targeted to AASVS1 in hPSCs, and tested for inducibility in hPSC-derived myeloid progenitors and hMGLs made with the monolayer hMGL protocol. X^{on}-SALL1 hPSCs and myeloid progenitors were treated with 5-20nM LMI070 or DMSO for 24h-48h, and SALL1 expression was determined by Western blot. X^{on}-P2A-GRN hMGLs were treated with 25nM LMI070 or DMSO for 72h, and levels of secreted GRN were determined by ELISA. In all experiments, WT cells lacking X^{on} transgenes were treated with the same LMI070 or DMSO doses as controls. For the GRN ELISA experiments, X^{on}-P2A-GFP hPSCs were differentiated into hMGLs with the monolayer protocol. Myeloid progenitors were plated at 1 million cells in 4mL microglia maturation media per well of a 6 well plate, with two biological replicates per condition. Five days later cells were treated with DMSO or 25nM LMI070, and the media was collected three days afterwards. Media from each well was split into two 1.5mL tubes, filtered through a 0.45um mesh filter, and frozen in 1mL aliquots at -80C.

<u>Integrated sgRNA X^{on} Cas9 induction experiments:</u> To generate the AAVS1 X^{on}-Cas9 targeting construct, a spCas9 dsDNA fragment, minus a start codon, was synthesized and cloned into the AAVS1 X^{on}-Open construct, and targeted to AAVS1 in hPSCs, as described above. The AAVS1 X^{on}-Cas9 donor template contains a promoter-less neomycin resistance cassette upstream of the CAG promoter, with a splice acceptor for endogenous expression of the selection marker. sgRNA lentivirus was made as previously described (Gallagher et al, medRxiv, 2025). Briefly, HEK293Ts were transfected with pBA904 BFP-expressing Perturb-seq plasmid (Addgene #122238) containing a validated B2M-targeting sgRNA, pCMV-VSV-G and psPAX2-chp6, followed by virus collection 48h-72h later. WT and X^{on}-Cas9 hPSCs cultured in 300uL mTeSR Plus in a 24 well matrigel-coated plate were transduced with 120uL of B2M lentivirus, and 300uL IMDM was added to bring the total volume in each well to 600uL. Cells received a full mTeSR media change the next day, and were passaged and expanded one day later. Transduced cells were selected with puro, frozen in liquid nitrogen, and differentiated to hMacs and hMGLs with the embryoid body protocol. After 7 days in hMac or hMGL maturation media, cells were treated with 0, 10 or 30nM LMI070, then treated a second time with fresh media 3 days later. Two days after the second treatment, BFP expression was analyzed by flow cytometry and gDNA was purified with the QIAgen DNeasy Blood & Tissue kit. gDNA concentration was determined by QuBit, and Cas9 editing efficiency was quantified by T7 endonuclease I digestion of heteroduplexes with the NEB EnGen Mutation Detection Kit. Digested PCR products were run on a bioanalyzer to determine the ratio of edited to total product. Data were analyzed and plotted with GraphPad Prism.

<u>Transient sgRNA X^{on} Cas9 induction experiments:</u> X^{on}-Cas9 hPSCs were targeted with a constitutively active CAG-mScarlet-3XNLS transgene into the AAVS1 locus, containing with a splice-acceptor driven puromycin-resistant cassette for selection. Dual puro/neo antibiotic selection yielded a population of inducible Cas9/mScarlet fluorescent hPSCs. For transient sgRNA expression, these hPSCs were transfected with either the modified px330-EGFP plasmid (lacking Cas9 and sgRNA) or three similar plasmids containing different sgRNAs targeting near the mScarlet chromophore residue (sgRNA #1: ATATCCCAGCTAAACGGCAG, sgRNA #2: CGGCTGCCATACATAAACTG, sgRNA #4: CTGAGCCCGCAGTTTATGTA). Cells were transfected concurrently with or without 20nM LMI070, and we performed FACS to isolate EGFP+

cells after 48h to isolate sgRNA-expressing cells. 5 days post-FACS and 7 days post-transfection, these cells were subjected to flow cytometry to assess mScarlet expression.

Western blots: Cell lysates were extracted with RIPA, and protein concentrations were quantified with the Pierce BCA Protein Assay kit. Lysates were denatured at 95C for 10 min, loaded onto 4-12% and 10% NuPAGE Bis-Tris Mini protein gels with 1X NuPAGE LDS Sample Buffer, and run at 100V in XCell SureLock Mini-Cell electrophoresis units. Protein was transferred from the gels to PVDF membranes in Bio-Rad Mini Trans-Blot Cells at 350mA for 2.5h. Blots were rinsed in TBST and TBS, blocked with 5% non-fat dry milk in TBST, and incubated with primary antibodies diluted in 5% milk with 0.02% sodium azide. Primary antibody incubations were performed overnight at 4C on a shaker. Secondary antibodies (Sigma-Aldrich goat anti-rabbit, goat anti-mouse and rabbit anti-goat IgG H&L Chain Specific Peroxidase Conjugate) were applied to blots at 1:10,000 dilutions in 5% milk for 1.5-2h at room temperature. Blots were developed with Western ECL or Western PicoPlus reagents either on X-ray films or with a Bio-Rad ChemiDoc Imaging System. SeeBlue Plus2 Pre-Stained Protein Standard and Precision Plus Protein Dual Color Standard were used as markers. Primary antibodies and dilutions included: abcam ab125247 anti-GAPDH (1:10,000), abcam ab6556 anti-GFP (1:5000), R&D AF2420 anti-GRN (1.25ug/mL), Novus Biologicals NBP2-61812 anti-KLF2 (1:365), Santa Cruz Biotechnology sc-17785 anti-MEF2A (1:1000), Sigma HPA005533 anti-MEF2C (1:250-1:1000), R&D PP-K9814-00 anti-SALL1 (1:1000), abcam ab40854 anti-SMAD3 (1:5000), Thermo Fisher Scientific PA5-46840 anti-TAL1 (1:1000), Cell Signaling Technology #91068S anti-TREM2 (1:1000).

Flow cytometry: HEK293Ts were dissociated with Trypsin, hPSCs with ReLeSR or Versene, hMGLs with Trypsin (if attached) or by collection and pelleting (if floating), and hMacs collected with Versene and cell scraping. Cells were washed in FACS buffer (PBS with 1% HyClone FBS, 2mM EDTA and 25mM HEPES), stained with 0.2ug/mL DAPI and analyzed and sorted with a BD FACSAria II or BD LSRFortessa. Tet-On induction time course experiments were calibrated from day to day using Rainbow Calibration Beads (Thermo Fisher). Sorted hPSCs were plated in maintenance media containing Stem Cell Technologies CloneR or CloneR2 to facilitate survival of single cells. All flow cytometry data were analyzed with FlowJo and GraphPad Prism.

Microscopy: All images were taken with a Nikon Eclipse Ti-U or Nikon Eclipse TE2000-U microscope and analyzed with ImageJ.

PCR analysis of X^{on} splicing events

PCR primers were designed to detect amplicons that increase in size (pair 1), or are unique (pair 2), when the pseudoexon is spliced into the mRNA. X^{on}-GFP primer pair 1 (F: GCTTTGCCATTCTTGGAAAC, R: cagacacgctgaacttgt), X^{on}-GFP primer pair 2 (F: GAGTCAATCCAAGTgccAc, R: cagacacgctgaacttgt), X^{on}-SALL1 primer pair 1 (F: CCACTGGCATCAGCTTTG, R: cggattggaaatgttgaggc), X^{on}-SALL1 primer pair 2 (F: GAGTCAATCCAAGTgccAc, R: ttagtagggcgactcggt). RNA was extracted from X^{on}-GFP and X^{on}-SALL1 hPSCs 48h after treatment with 20nM LMI070 or DMSO, using the QIAGEN miRNeasy Mini Kit. 1ug of each RNA sample was DNase-treated (Thermo Fisher Scientific, cat. #EN0521), and cDNAs were generated with Quantabio qScript cDNA SuperMix (cat. #95048-025), according to manufacturer instructions. Primers were tested for efficacy and specificity by qPCR on an Applied Biosystems QuantStudio 6 Flex, using Fast SYBR Green 2X master mix (Thermo Fisher Scientific, cat. #4385616). Amplicons were then generated with NEB Q5 PCR master mix and analyzed by agarose gel electrophoresis, with NEB 100bp DNA ladder. Identify of amplicons was confirmed by Sanger sequencing (Eton Bioscience) after column purification.

GRN ELISA

GRN ELISA was performed with the AdipoGen Life Sciences Progranulin (human) ELISA kit (cat. # AG-45A-0018YEK-KI01), according to the manufacturer's instructions. Frozen media from LMI070-treated X^{on}-P2A-GRN hMGLs was thawed and diluted 1:50 in microglia maturation media before performing the ELISA. Each frozen media aliquot from the same well were analyzed, but were averaged and not considered as biological replicates. Biological replicates were different wells of the same cell population treated with the same LMI070 dose.

Statistical analyses

Simple linear regressions were performed to test for correlations between LMI070 dose and EGFP expression in X^{on}-EGFP hPSCs, hMacs and hMGLs. R^2 and p-values are reported in the results section. Effects of Tet-On and DHFR degron transgenes on leaky EGFP expression were analyzed with 1-way ANOVAs. Effects of the X^{on}-EGFP transgene on leaky EGFP expression were analyzed with 2-tailed t-tests. GRN ELISA results were analyzed with two-tailed t-tests to determine effects of LMI070 dose on GRN levels in hMGLs. X^{on}-Cas9 editing efficiency in treated and untreated B2M sgRNA-expressing hMGLs and hMacs was analyzed with a 2-way ANOVA. mScarlet knockout in X^{on}-Cas9 hPSCs transfected with mScarlet-targeting sgRNAs was analyzed with a 1-way ANOVA. All statistical analyses were performed in GraphPad Prism.

REFERENCES

1. Ryding, A.D., M.G. Sharp, and J.J. Mullins, *Conditional transgenic technologies.* J Endocrinol, 2001. **171**(1): p. 1-14.
2. Maltzman, J.S. and L.A. Turka, *Conditional gene expression: a new tool for the transplantologist.* Am J Transplant, 2007. **7**(4): p. 733-40.
3. Butterfield, G.L., et al., *Gene regulation technologies for gene and cell therapy.* Mol Ther, 2025. **33**(5): p. 2104-2122.
4. Cabrera, A., et al., *The sound of silence: Transgene silencing in mammalian cell engineering.* Cell Syst, 2022. **13**(12): p. 950-973.
5. Balmas, E., et al., *Manipulating and studying gene function in human pluripotent stem cell models.* FEBS Lett, 2023. **597**(18): p. 2250-2287.
6. Noviello, G., *Transcriptional regulation as a dose-dependent process: insights from transcription factor tuning.* Open Biol, 2025. **15**(8): p. 240328.
7. Zhou, H., et al., *Genome-scale CRISPR-Cas9 screening in stem cells: theories, applications and challenges.* Stem Cell Res Ther, 2024. **15**(1): p. 218.
8. Li, K., et al., *CRISPR-based functional genomics screening in human-pluripotent-stem-cell-derived cell types.* Cell Genom, 2023. **3**(5): p. 100300.
9. Escopete, S., et al., *Human cardiac organoids for disease modeling and drug discovery.* Trends Mol Med, 2025.
10. Faravelli, I., et al., *Brain Organoids: Tools for Understanding the Uniqueness and Individual Variability of the Human Brain.* Annu Rev Genomics Hum Genet, 2025. **26**(1): p. 299-320.
11. Sljukic, A., et al., *Advances in liver and pancreas organoids: how far we have come and where we go next.* Nat Rev Gastroenterol Hepatol, 2025.
12. Yamanaka, S., *Pluripotent Stem Cell-Based Cell Therapy-Promise and Challenges.* Cell Stem Cell, 2020. **27**(4): p. 523-531.
13. Volpato, V. and C. Webber, *Addressing variability in iPSC-derived models of human disease: guidelines to promote reproducibility.* Dis Model Mech, 2020. **13**(1).
14. Basilicata, M.F. and C.I. Keller Valsecchi, *The good, the bad, and the ugly: Evolutionary and pathological aspects of gene dosage alterations.* PLoS Genet, 2021. **17**(12): p. e1009906.
15. Lupski, J.R., *Biology in balance: human diploid genome integrity, gene dosage, and genomic medicine.* Trends Genet, 2022. **38**(6): p. 554-571.
16. Naqvi, S., et al., *Precise modulation of transcription factor levels identifies features underlying dosage sensitivity.* Nat Genet, 2023. **55**(5): p. 841-851.
17. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.* Nat Genet, 2000. **24**(4): p. 372-6.
18. Liu, W., et al., *Dissecting the impact of transcription factor dose on cell reprogramming heterogeneity using scTF-seq.* Nat Genet, 2025. **57**(10): p. 2522-2535.
19. Banani, S.F., et al., *Biomolecular condensates: organizers of cellular biochemistry.* Nat Rev Mol Cell Biol, 2017. **18**(5): p. 285-298.
20. Alberti, S., *The wisdom of crowds: regulating cell function through condensed states of living matter.* J Cell Sci, 2017. **130**(17): p. 2789-2796.
21. Tiwari, S.K., et al., *Induced pluripotent stem cell-derived macrophages as a platform for modelling human disease.* Nat Rev Immunol, 2025. **25**(2): p. 108-124.
22. Hoeksema, M.A. and C.K. Glass, *Nature and nurture of tissue-specific macrophage phenotypes.* Atherosclerosis, 2019. **281**: p. 159-167.
23. Hasselmann, J. and M. Blurton-Jones, *Human iPSC-derived microglia: A growing toolset to study the brain's innate immune cells.* Glia, 2020. **68**(4): p. 721-739.

24. Gosselin, D., et al., *Screen-Printed Polyaniline-Based Electrodes for the Real-Time Monitoring of Loop-Mediated Isothermal Amplification Reactions.* Anal Chem, 2017. **89**(19): p. 10124-10128.
25. Svoboda, D.S., et al., *Human iPSC-derived microglia assume a primary microglia-like state after transplantation into the neonatal mouse brain.* Proc Natl Acad Sci U S A, 2019. **116**(50): p. 25293-25303.
26. Nott, A., et al., *Brain cell type-specific enhancer-promoter interactome maps and disease-risk association.* Science, 2019. **366**(6469): p. 1134-1139.
27. Hasselmann, J., et al., *Development of a Chimeric Model to Study and Manipulate Human Microglia In Vivo.* Neuron, 2019. **103**(6): p. 1016-1033 e10.
28. Kowarz, E., D. Loscher, and R. Marschalek, *Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines.* Biotechnol J, 2015. **10**(4): p. 647-53.
29. de Felipe, P., *Skipping the co-expression problem: the new 2A "CHYSEL" technology.* Genet Vaccines Ther, 2004. **2**(1): p. 13.
30. Liu, Z., et al., *Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector.* Sci Rep, 2017. **7**(1): p. 2193.
31. Mates, L., et al., *Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates.* Nat Genet, 2009. **41**(6): p. 753-61.
32. Muffat, J., et al., *Efficient derivation of microglia-like cells from human pluripotent stem cells.* Nat Med, 2016. **22**(11): p. 1358-1367.
33. Qin, J.Y., et al., *Systematic comparison of constitutive promoters and the doxycycline-inducible promoter.* PLoS One, 2010. **5**(5): p. e10611.
34. Fernandopulle, M.S., et al., *Transcription Factor-Mediated Differentiation of Human iPSCs into Neurons.* Curr Protoc Cell Biol, 2018. **79**(1): p. e51.
35. Hockemeyer, D., et al., *Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases.* Nat Biotechnol, 2009. **27**(9): p. 851-7.
36. Ma, H., J.F. Jeppesen, and R. Jaenisch, *Human T Cells Expressing a CD19 CAR-T Receptor Provide Insights into Mechanisms of Human CD19-Positive beta Cell Destruction.* Cell Rep Med, 2020. **1**(6): p. 100097.
37. Godecke, N., et al., *Controlled re-activation of epigenetically silenced Tet promoter-driven transgene expression by targeted demethylation.* Nucleic Acids Res, 2017. **45**(16): p. e147.
38. Ordovas, L., et al., *Efficient Recombinase-Mediated Cassette Exchange in hPSCs to Study the Hepatocyte Lineage Reveals AAVS1 Locus-Mediated Transgene Inhibition.* Stem Cell Reports, 2015. **5**(5): p. 918-931.
39. Natsume, T. and M.T. Kanemaki, *Conditional Degrons for Controlling Protein Expression at the Protein Level.* Annu Rev Genet, 2017. **51**: p. 83-102.
40. Iwamoto, M., et al., *A general chemical method to regulate protein stability in the mammalian central nervous system.* Chem Biol, 2010. **17**(9): p. 981-8.
41. Tian, R., et al., *CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons.* Neuron, 2019. **104**(2): p. 239-255 e12.
42. Monteys, A.M., et al., *Regulated control of gene therapies by drug-induced splicing.* Nature, 2021. **596**(7871): p. 291-295.
43. Palacino, J., et al., *SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice.* Nat Chem Biol, 2015. **11**(7): p. 511-7.
44. Knab, A., et al., *Illuminating Immunity: A Systematic Review of Immune Cell Autofluorescence.* J Biophotonics, 2025. **18**(6): p. e202400576.
45. Buttgereit, A., et al., *Sall1 is a transcriptional regulator defining microglia identity and function.* Nat Immunol, 2016. **17**(12): p. 1397-1406.
46. Fixsen, B.R., et al., *SALL1 enforces microglia-specific DNA binding and function of SMADs to establish microglia identity.* Nat Immunol, 2023. **24**(7): p. 1188-1199.

47. Life, B.E. and B.R. Leavitt, *Progranulin function and regulation in the CNS.* Trends Neurosci, 2025. **48**(7): p. 523-537.
48. Weiner, H.L., *Immune mechanisms and shared immune targets in neurodegenerative diseases.* Nat Rev Neurol, 2025. **21**(2): p. 67-85.
49. Horlbeck, M.A., et al., *Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation.* Elife, 2016. **5**.
50. Ma, H., et al., *The nuclear receptor THRB facilitates differentiation of human PSCs into more mature hepatocytes.* Cell Stem Cell, 2022. **29**(5): p. 795-809 e11.
51. Gosselin, D., et al., *An environment-dependent transcriptional network specifies human microglia identity.* Science, 2017. **356**(6344).
52. Troutman, T.D., E. Kofman, and C.K. Glass, *Exploiting dynamic enhancer landscapes to decode macrophage and microglia phenotypes in health and disease.* Mol Cell, 2021. **81**(19): p. 3888-3903.
53. Depp, C., et al., *Microglia transcriptional states and their functional significance: Context drives diversity.* Immunity, 2025. **58**(5): p. 1052-1067.
54. Das, A.T., L. Tenenbaum, and B. Berkhout, *Tet-On Systems For Doxycycline-inducible Gene Expression.* Curr Gene Ther, 2016. **16**(3): p. 156-67.
55. Klatt, D., et al., *Differential Transgene Silencing of Myeloid-Specific Promoters in the AAVS1 Safe Harbor Locus of Induced Pluripotent Stem Cell-Derived Myeloid Cells.* Hum Gene Ther, 2020. **31**(3-4): p. 199-210.
56. Pinto, R., et al., *Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression.* Nucleic Acids Res, 2017. **45**(13): p. e123.
57. Costello, A., et al., *Leaky Expression of the TET-On System Hinders Control of Endogenous miRNA Abundance.* Biotechnol J, 2019. **14**(3): p. e1800219.
58. Meyer-Ficca, M.L., et al., *Comparative analysis of inducible expression systems in transient transfection studies.* Anal Biochem, 2004. **334**(1): p. 9-19.
59. Zhu, Z., et al., *Use of the tetracycline-controlled transcriptional silencer (tTS) to eliminate transgene leak in inducible overexpression transgenic mice.* J Biol Chem, 2001. **276**(27): p. 25222-9.
60. Dolan, M.J., et al., *Exposure of iPSC-derived human microglia to brain substrates enables the generation and manipulation of diverse transcriptional states in vitro.* Nat Immunol, 2023. **24**(8): p. 1382-1390.
61. Navarro-Guerrero, E., et al., *Optimized protocol for CRISPR knockout of human iPSC-derived macrophages.* STAR Protoc, 2024. **5**(1): p. 102903.
62. Chi, Y., et al., *A compact, versatile drug-induced splicing switch system with minimal background expression.* Cell Rep Methods, 2024. **4**(9): p. 100842.
63. Weber, J.A., et al., *Temporal restriction of Cas9 expression improves CRISPR-mediated deletion efficacy and fidelity.* Mol Ther Nucleic Acids, 2024. **35**(2): p. 102172.
64. Zhou, X., et al., *Optimization of the Tet-On system for regulated gene expression through viral evolution.* Gene Ther, 2006. **13**(19): p. 1382-90.
65. Friesen, M., et al., *Development of a physiological insulin resistance model in human stem cell-derived adipocytes.* Sci Adv, 2022. **8**(24): p. eabn7298.
66. Drager, N.M., et al., *A CRISPRi/a platform in human iPSC-derived microglia uncovers regulators of disease states.* Nat Neurosci, 2022. **25**(9): p. 1149-1162.
67. Cooper, Y.A., et al., *Functional regulatory variants implicate distinct transcriptional networks in dementia.* Science, 2022. **377**(6608): p. eabi8654.
68. Tian, R., et al., *Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to ferroptosis.* Nat Neurosci, 2021. **24**(7): p. 1020-1034.
69. Zhang, L., et al., *TREM2 and sTREM2 in Alzheimer's disease: from mechanisms to therapies.* Mol Neurodegener, 2025. **20**(1): p. 43.

70. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.
71. Richards, A., et al., *SARS-CoV-2 infection of human pluripotent stem cell-derived liver organoids reveals potential mechanisms of liver pathology.* iScience, 2022. **25**(10): p. 105146.

FIGURE LEGENDS

Figure 1: Transposon-integrated Tet-On transgenes are not functional in hPSCs or hPSC-derived myeloid cells. (A) Diagrams of pSBtet-GP-TF and pSBtet-RP-TF transposons for dox-inducible TF expression. **(B)** Experimental schematic of transposon testing in HEK23T cells. **(C,D)** Western blot analyses of HEK293Ts with stable integrations of pSBtet-GP-TF (panel C) and pSBtet-RP (panel D). "-" and "+" signs indicate absence and presence of transposon and dox treatment (0.75 or 7.5ug/mL). Blotted proteins are indicated to the left of the blots. **(E)** Western blot analysis of cells with dual integrations of pSBtet-GP-TF and pSBtet-RP-TF (indicated as pSBtet-FP-TF). Expected molecular weight bands for blotted proteins are indicated to the right. **(F)** Experimental schematic of transposon testing in hPSC-derived myeloid progenitors. **(G)** Western blot analysis of SALL1 in hPSCs and hPSC-derived myeloid cells with stable pSBtet-GP-TF integrations. Blotted proteins are indicated to the right of each blot.

Figure 2: Safe harbor integration does not rescue Tet-On function in hPSC-derived myeloid cells. (A) Comparison of the commonly used pUCM TetO3G induction platform to drive EGFP expression in hPSCs and hMacs illustrating lack of transgene inducibility at the myeloid stage. **(B)** Schematic of the control TetO3G (top) and dual rtTA/tTS-based designs (bottom). **(C)** Representative flow cytometry plots of time-course dox-induction in rtTA-tTS designs, showing retention of inducibility in hPSCs. **(D)** Representative flow cytometry plots of time-course dox-induction in all rtTA-tTS designs, showing lack of inducibility in hMacs.

Figure 3: DHFR degron-based systems are functional in hPSC-derived myeloid cells but not tightly regulated. (A) Schematic for DHFR-tagging of N-, C-, and dual termini of EGFP in a CAG-driven format in the AAVS1 locus. **(B)** TMP-mediated stabilization of the DHFR-tagged fusion proteins to prevent degradation via the ubiquitin-proteasome pathway. **(C)** Representative flow cytometry plots of TMP dose-response induction in all DHFR degron designs, showing inducibility in hPSCs in all but the C-terminally tagged EGFP. WT represents unengineered hPSCs. **(D)** Representative flow cytometry plots of 2.5uM TMP induction in all DHFR degron designs, showing retention of inducibility in hMacs with all but the C-terminally tagged EGFP. WT represented unengineered hMacs.

Figure 4: X^{on} inducible splicing system allows for reproducible, robust and highly tunable transgene expression. (A) Schematic of X^{on} inducible splicing mechanism. Start codon is removed from the gene of interest (GOI) so that translation requires LMI070-mediated inclusion of ATG in the pseudoexon, resulting in a 24 amino acid N-terminal tag. e1=exon 1, e2=exon 2, PE=pseudoexon. **(B)** AAVS1 targeting strategy for testing X^{on}-EGFP in hMacs and hMGLs. SA=splice acceptor. PuroR=puro resistance. **(C-E)** Fluoresence microscopy images of X^{on}-EGFP hPSCs, hMGLs and hMacs, treated with various doses of LMI070. **(F-H)** Quantification of EGFP by flow cytometry in dose-response experiments. "Parental" samples are WT cells that lack the X^{on}-EGFP transgene. LMI070 dose is indicated to the right of each image. **(I-K)** Scatter plot representation of data from panels F-H.

Figure 5: Incorporating N-terminal 2A peptides can improve inducible X^{on} systems. (A,B) Western blot analyses of X^{on}-EGFP and X^{on}-SALL1 transgenes in hPSCs (panel A) and HEK293T cells (panel B). Expected molecular weights are ~30kDa for EGFP (yellow arrow in panel B indicates expected product), ~140kDa for SALL1, and ~37kDa for GAPDH. **(C)** PCR strategy for detecting X^{on} splice products. Primer pairs are represented by red and blue arrows. **(D,E)** PCR results for both primer pairs, with detected mRNAs indicated by schematics to the right of each gel. Expected product sizes for panel D are 164bp and 223bp for X^{on}-EGFP with and without LMI070, and 125bp and 184bp for X^{on}-SALL1. Expected product sizes for panel E are 183bp for

X^{on}-EGFP and 202bp for X^{on}-SALL1 with LMI070. Unexpected X^{on}-SALL1 product is 684bp and represents unspliced pre-mRNA.

Figure 6: Tunable GRN overexpression with X^{on} in hPSCs and hMGLs. (A,B) Western blot analysis of HEK293Ts transfected with X^{on}-SMAD3 and X^{on}-GRN AAVS1 donor constructs. **(C)** Western blot analysis of hPSCs with AAVS1-targeted X^{on}-P2A-GRN transgene. **(D)** Schematic of X^{on}-mediated GRN overexpression in hMGLs. **(E)** ELISA quantification of secreted GRN from WT and X^{on}-P2A-GRN hMGLs, treated with 0nM or 25nM LMI070 for 72h. n.s.=not significant; *P<0.05.

Figure 7: X^{on}-Cas9 genome editing performance depends on sgRNA delivery strategy. (A) Schematic representation of high-sensitivity terminal Cas9-double strand break assay in X^{on}-Cas9 hPSCs and differentiated cells with constitutive sgRNA expression. **(B)** Editing efficiency at the B2M TSS in WT and X^{on}-Cas9 hMacs and hMGLs with and without LMI070 treatment. Individual circles represent biological replicates, and horizontal lines represent replicate mean values. **(C)** Schematic representation of high-sensitivity terminal Cas9-double strand break assay in X^{on}-Cas9/mScarlet hPSCs with transient sgRNA expression. **(D)** Representative flow cytometry plots of X^{on}-Cas9/mScarlet hPSCs with and without concurrent LMI070 induction at the time of transfection with non-sgRNA control (pEGFP) or mScarlet sgRNA (pEGFP-sgRNA (3)) plasmids. **(E)** Summary quantification of mScarlet expression in three replicate KO experiments. n.s.=not significant

SUPPLEMENTARY FIGURE LEGENDS

Figure S1: Effects of promoter replacement on pSBtet transposons in hPSCs and hPSC differentiation. (A,B) Western blot analyses of HEK293Ts transfected with pSBtet-GP-TF (panel A) and pSBtet-RP-TF (panel B). Panel A shows blot results for TAL1, with arrows indicating bands corresponding to the expected molecular weight of TAL1 (35-40kDa). Panel B shows blot results for KLF2, with arrow indicating dox-specific bands. As with Figure 1, dox doses are 0.75 and 7.5ug/mL. **(C)** Flow cytometry analysis of hPSCs transfected with plasmids expressing EGFP under control of RPBSA, CAG, PGK and CMV promoters. GFP+ cells are shown in blue, with the proportions indicated above the GFP+ gates. **(D)** Schematic of promoter pSBtet promoter replacement strategy. **(E,F)** Brightfield (top) and GFP (bottom) channels of hPSCs with stable integrations of pSBtet-GP-EF1A (panel E) and pSBtet-RP-EF1A (panel F). Ratios above images indicate transfection mass ratios of SB100X transposase and pSBtet transposons, with higher ratios improving integration efficiency. Scale bars are 500µm. **(G,H)** Brightfield and GFP channels from images of pSBtet-GP-EF1A hPSCs at days 5 (panel G) and 10 (panel H) of hMGL differentiation. Puro-induced cell death is visible at day 10 but not day 5. Scale bars are 500µm.

Figure S2: Quantification of leaky expression in Tet-On and tTS-based systems. (A) Geometric mean fluorescence intensity (MFI) at 488nm of all singlet gated hPSCs (left) and the percentage of EGFP+ cells when gated at ~0.9% against unengineered WT cells, all in the absence of doxycycline (DOX). **(B)** Time course of EGFP expression response represented in hours post-DOX exposure in hPSCs engineered with the standard third-generation Tet-On system, TetO3G.

Figure S3: Quantification of leaky expression in DHFR degron systems. (A) Dose-response of EGFP expression represented 24h post-trimethoprim (TMP) exposure in hPSCs engineered with all DHFR degron systems. **(B)** Geometric mean fluorescence intensity (MFI) at 488nm of all singlet gated hPSCs (left) and the percentage of EGFP+ cells when gated at ~0.9% against unengineered WT cells, all in the absence of TMP.

Figure S4: X^{on}-EGFP performance in alternative hPSCs and hMGLs generated with a different differentiation protocol. (A-B) hPSCs from a different genetic background as those in **Fig. 4** display tunable LMI070-mediated EGFP expression. **(C)** hMGLs produced from hPSCs shown in panel A, but with a different differentiation protocol[71] than those shown in Fig. 4, display LMI070 dose-dependent EGFP expression, with no obvious toxicity.

Figure S5: Quantification of leaky expression in X^{on} system. (A) Representative flow cytometry plot demonstrating distribution of 488nm fluorescence in unengineered WT hPSCs (left), overlayed with X^{on}-EGFP-engineered hPSCs (right), both with ~0.9% stringency gate illustrated. **(B)** Geometric mean fluorescence intensity (MFI) at 488nm of all singlet gated X^{on}-EGFP hPSCs (left) and the percentage of EGFP+ cells when gated at ~0.9% against unengineered WT cells, all in the absence of LMI070.

Figure S6: Approaches to ameliorate technical limitations of the X^{on} system. (A) Western blot analysis of AAVS1-targeted X^{on}-SALL1 in hPSCs from a different genetic background than those shown in Fig. 5A, indicating leakiness and poor inducibility in the presence of LMI070. **(B)** Western blot analysis of HEK293Ts transfected with codon-optimized X^{on}-SALL1 and Met→Ile mutated X^{on}-EGFP-P2A-TREM donor constructs. **(C)** Western blot analysis of HEK293Ts transfected with X^{on}-EGFP-P2A-SALL1 donor construct. **(D)** Western blot analysis of HEK293Ts transfected with X^{on}-TREM2 WT, X^{on}-TREM2 R47H, and X^{on}-EGFP-P2A-TREM2 donor

constructs. **(E)** Western blot analysis of HEK293Ts transfected with X^{on}-P2A-SALL1 and Met→Ile mutated X^{on}-EGFP-P2A-SALL1 donor constructs.

Figure 1

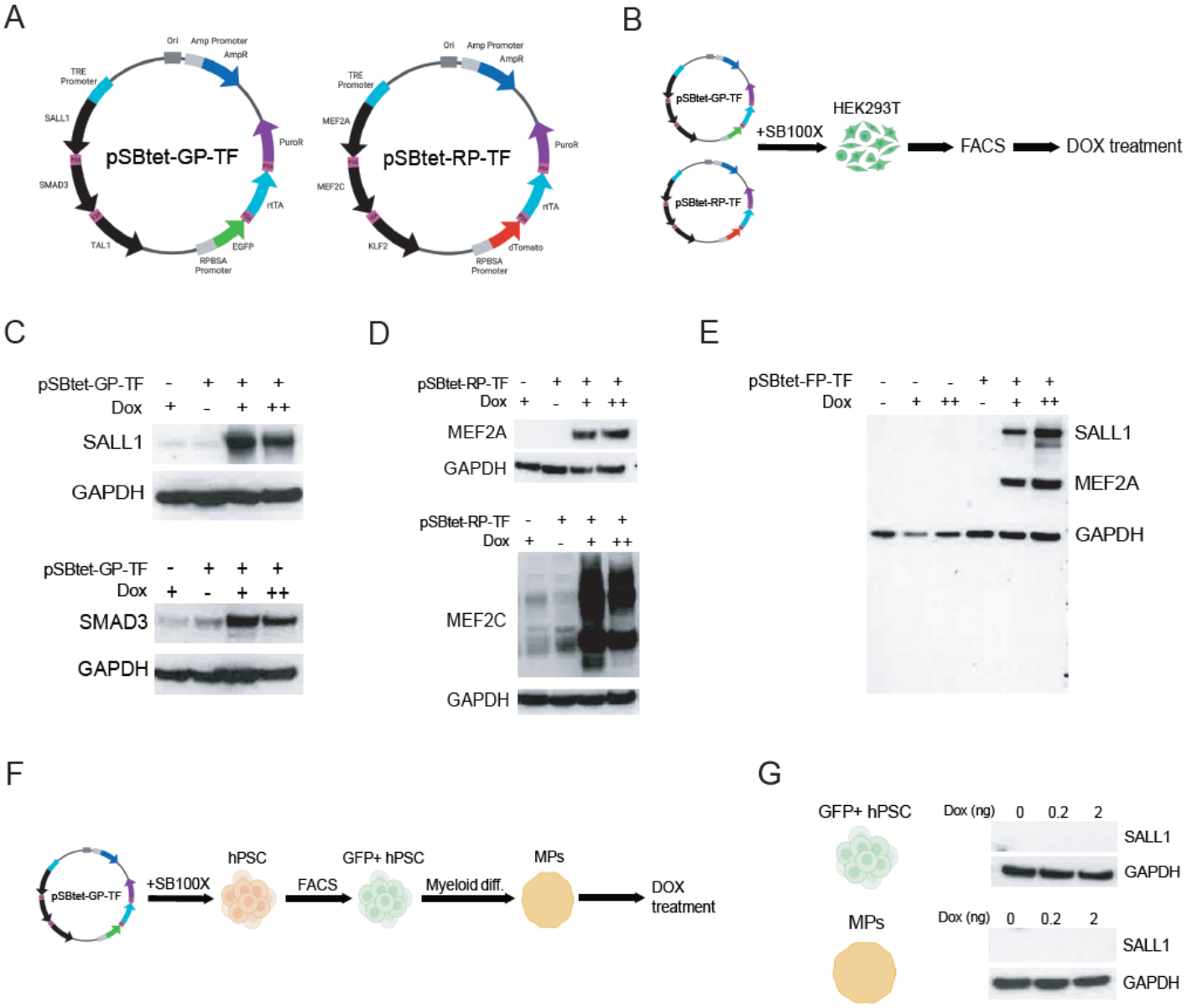

A
Ori
Amp Promoter
AmpR
TRE Promoter
SALL1
SMAD3
TAL1
RPBSA Promoter
EGFP
rtTA
PuroR
pSBtet-GP-TF
MEF2A
MEF2C
KLF2
dTomato
pSBtet-RP-TF
B
pSBtet-GP-TF
pSBtet-RP-TF
+SB100X
HEK293T
FACS
DOX treatment
C
pSBtet-GP-TF
Dox
SALL1
GAPDH
SMAD3
D
pSBtet-RP-TF
MEF2A
MEF2C
E
pSBtet-FP-TF
F
hPSC
GFP+ hPSC
MPs
Myeloid diff.
DOX treatment
G
GFP+ hPSC
Dox (ng)
0
0.2
2
SALL1
GAPDH
MPs

Figure 2

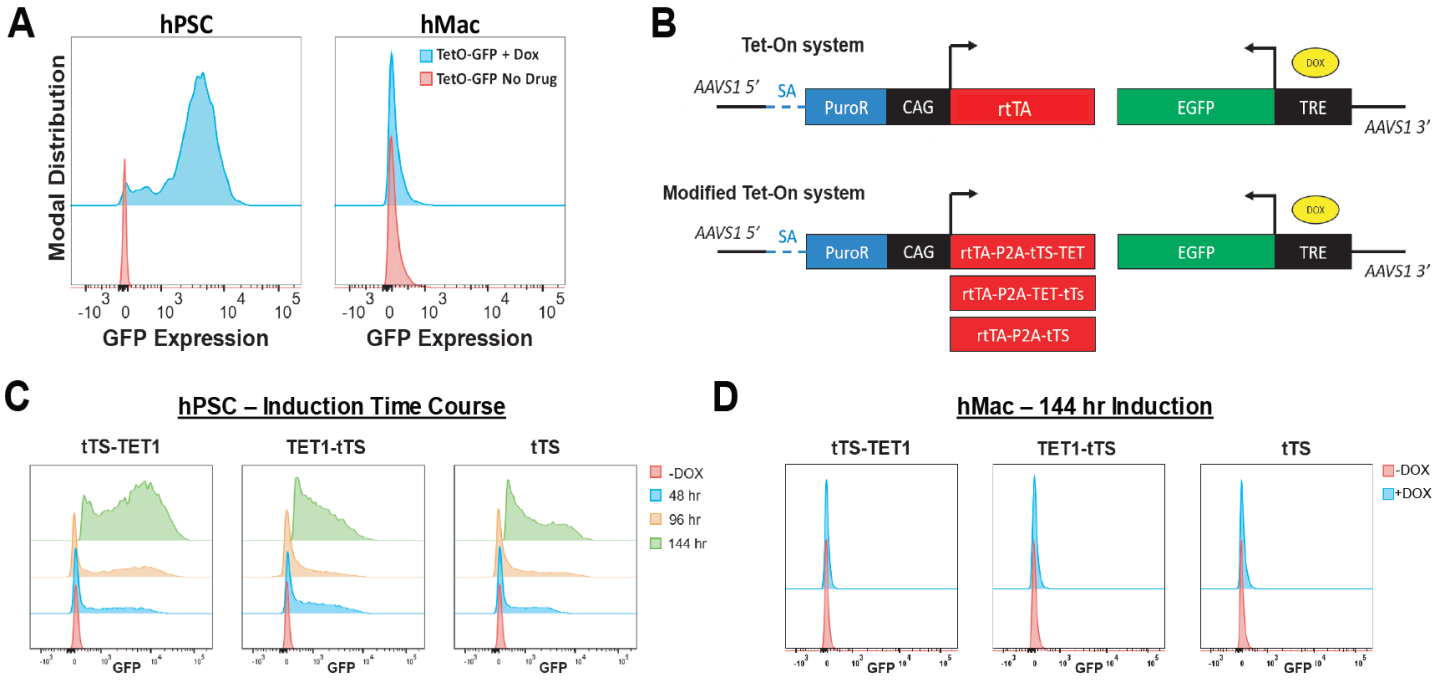

A
hPSC
hMac
TetO-GFP + Dox
TetO-GFP No Drug
Modal Distribution
GFP Expression
GFP Expression
B
Tet-On system
AAVS1 5'
SA
PuroR
CAG
rtTA
EGFP
TRE
DOX
AAVS1 3'
Modified Tet-On system
AAVS1 5'
SA
PuroR
CAG
rtTA-P2A-tTS-TET
rtTA-P2A-TET-tTs
rtTA-P2A-tTS
EGFP
TRE
DOX
AAVS1 3'
C
hPSC – Induction Time Course
tTS-TET1
TET1-tTS
tTS
-DOX
48 hr
96 hr
144 hr
GFP
D
hMac – 144 hr Induction
tTS-TET1
TET1-tTS
tTS
-DOX
+DOX
GFP

Figure 3

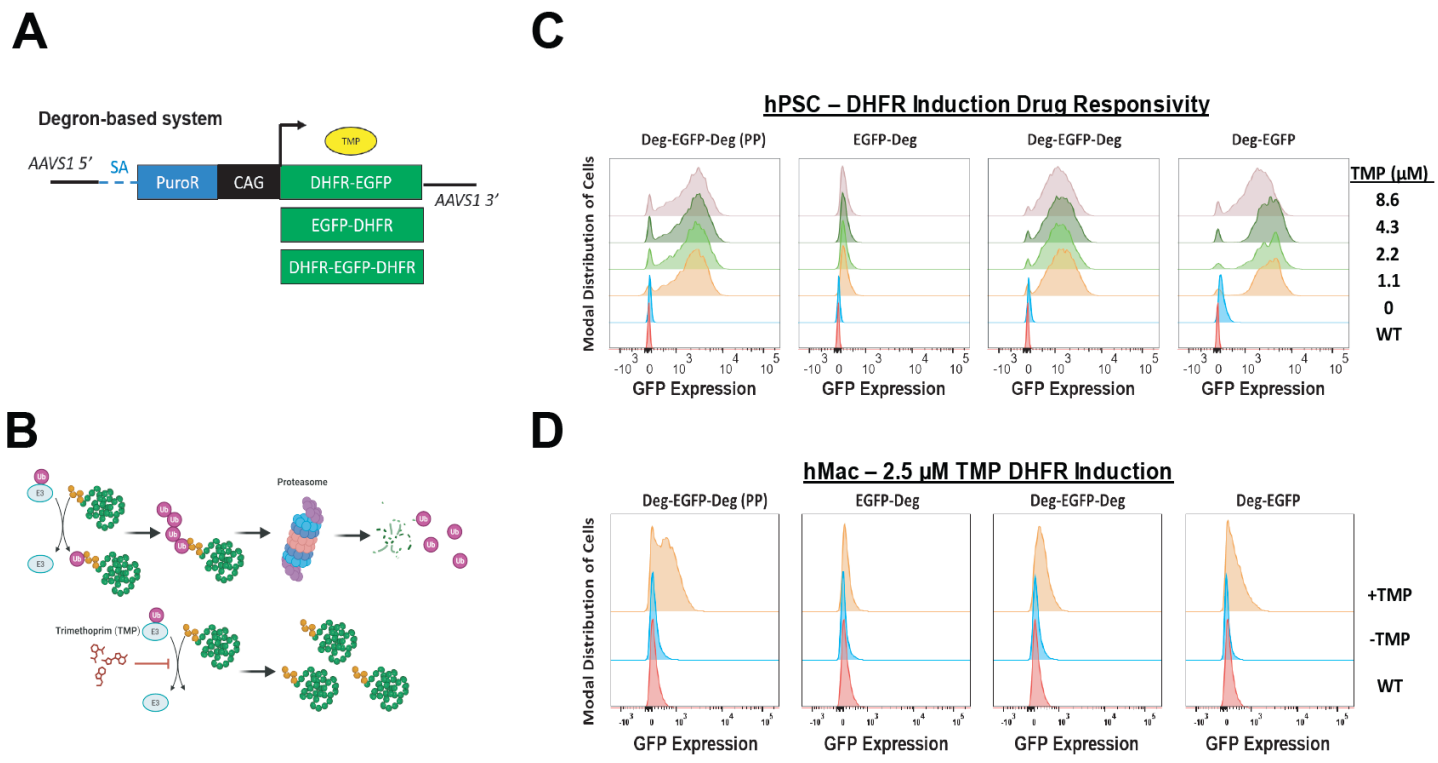

A
Degron-based system
AAVS1 5'
SA
PuroR
CAG
TMP
DHFR-EGFP
EGFP-DHFR
DHFR-EGFP-DHFR
AAVS1 3'
B
Proteasome
Trimethoprim (TMP)
C
hPSC – DHFR Induction Drug Responsivity
Deg-EGFP-Deg (PP)
EGFP-Deg
Deg-EGFP-Deg
Deg-EGFP
Modal Distribution of Cells
GFP Expression
TMP (μM)
8.6
4.3
2.2
1.1
0
WT
D
hMac – 2.5 μM TMP DHFR Induction
Deg-EGFP-Deg (PP)
EGFP-Deg
Deg-EGFP-Deg
Deg-EGFP
Modal Distribution of Cells
GFP Expression
+TMP
-TMP
WT

Figure 4

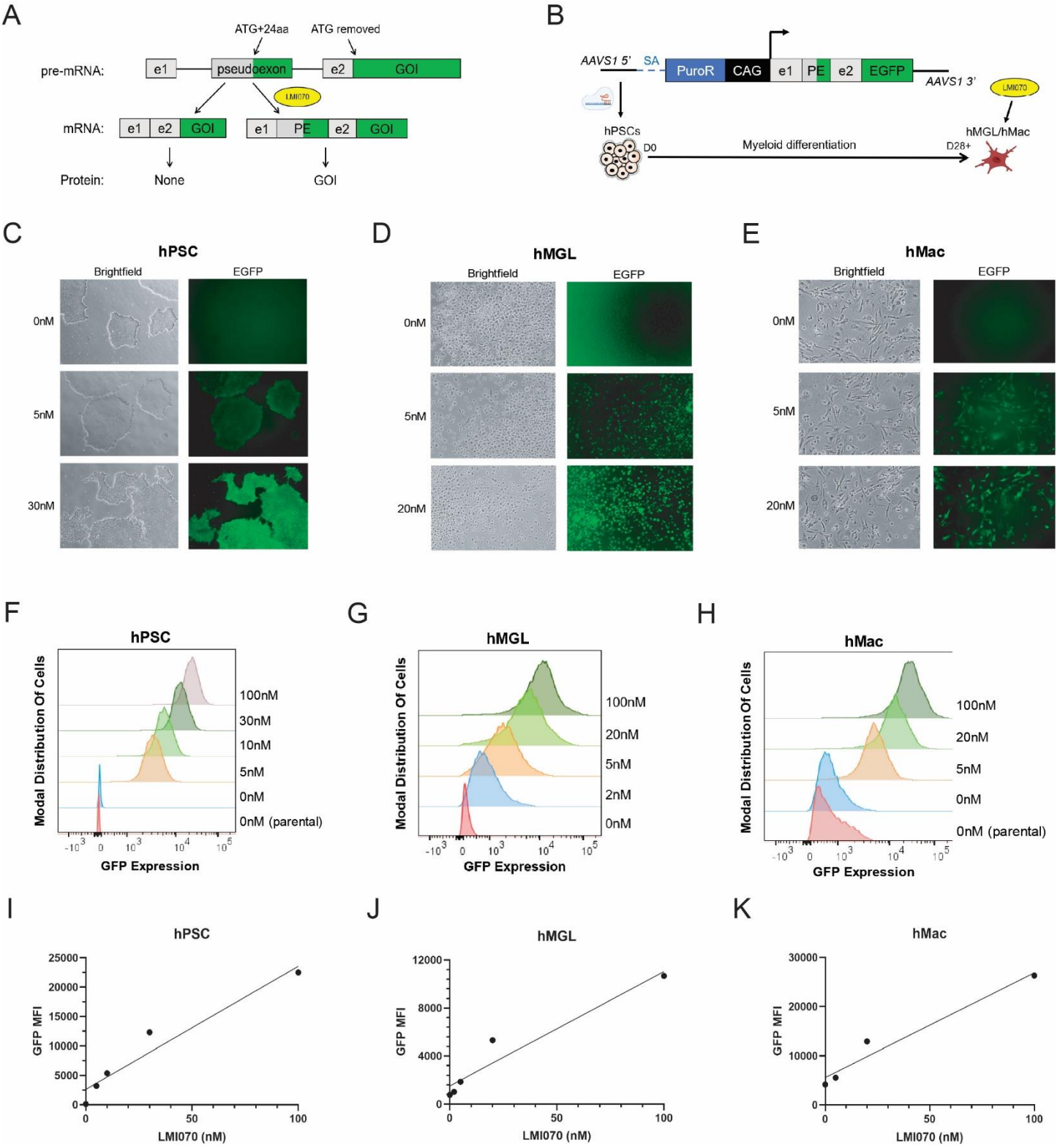

A
ATG+24aa
ATG removed
pre-mRNA:
e1
pseudoexon
e2
GOI
LMI070
mRNA:
e1
e2
GOI
e1
PE
e2
GOI
Protein:
None
GOI
B
AAVS1 5'
SA
PuroR
CAG
e1
PE
e2
EGFP
AAVS1 3'
LMI070
hPSCs
D0
Myeloid differentiation
hMGL/hMac
D28+
C
hPSC
Brightfield
EGFP
0nM
5nM
30nM
D
hMGL
Brightfield
EGFP
0nM
5nM
20nM
E
hMac
Brightfield
EGFP
0nM
5nM
20nM
F
hPSC
Modal Distribution Of Cells
100nM
30nM
10nM
5nM
0nM
0nM (parental)
GFP Expression
G
hMGL
Modal Distribution Of Cells
100nM
20nM
5nM
2nM
0nM
GFP Expression
H
hMac
Modal Distribution Of Cells
100nM
20nM
5nM
0nM
0nM (parental)
GFP Expression
I
hPSC
GFP MFI
LMI070 (nM)
J
hMGL
GFP MFI
LMI070 (nM)
K
hMac
GFP MFI
LMI070 (nM)

Figure 5

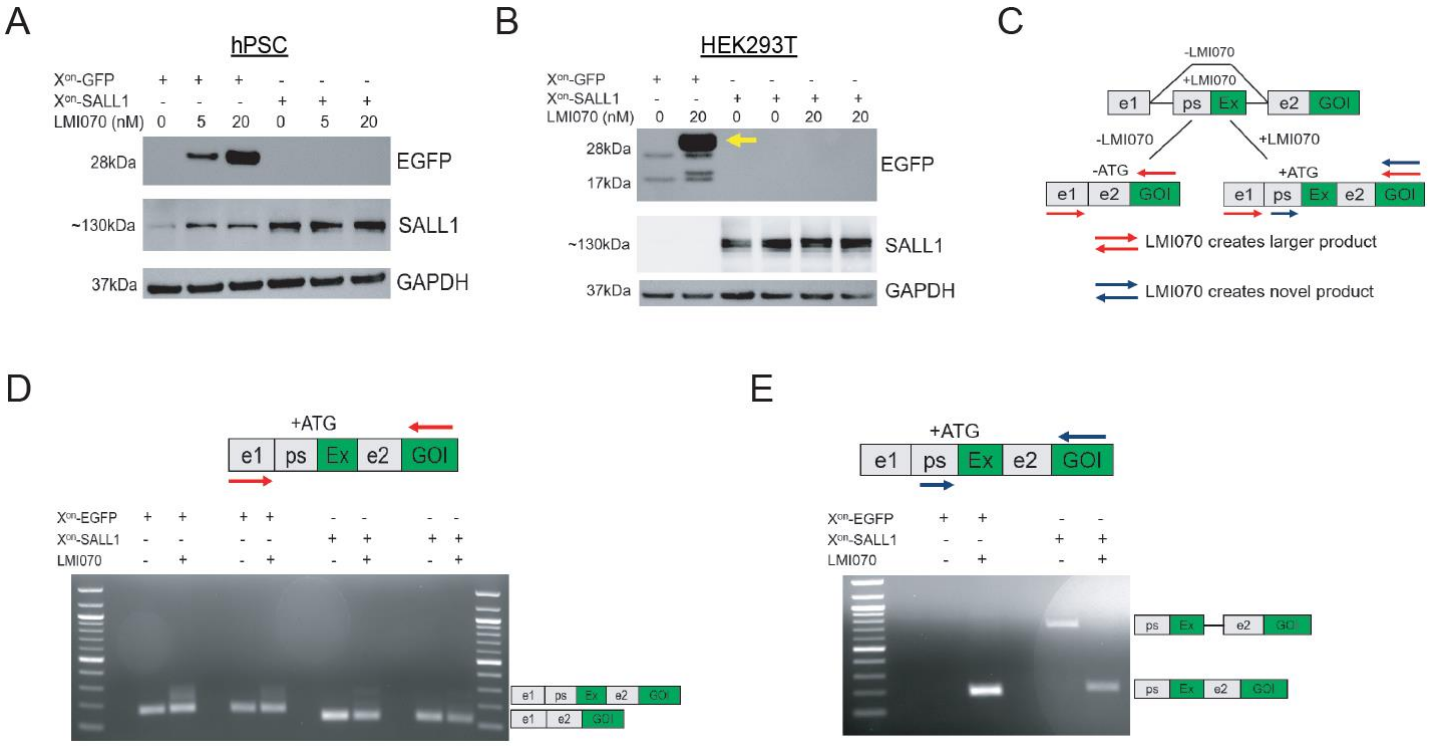

A
hPSC
Xon-GFP + + + - - -
Xon-SALL1 - - - + + +
LMI070 (nM) 0 5 20 0 5 20
28kDa
EGFP
~130kDa
SALL1
37kDa
GAPDH
B
HEK293T
Xon-GFP + + - - - -
Xon-SALL1 - - + + + +
LMI070 (nM) 0 20 0 0 20 20
28kDa
17kDa
EGFP
~130kDa
SALL1
37kDa
GAPDH
C
-LMI070
+LMI070
e1 ps Ex e2 GOI
-LMI070
+LMI070
-ATG
+ATG
e1 e2 GOI
e1 ps Ex e2 GOI
LMI070 creates larger product
LMI070 creates novel product
D
+ATG
e1 ps Ex e2 GOI
Xon-EGFP + + + + - - - -
Xon-SALL1 - - - - + + + +
LMI070 - + - + - + - +
e1 ps Ex e2 GOI
e1 e2 GOI
E
+ATG
e1 ps Ex e2 GOI
Xon-EGFP + + - -
Xon-SALL1 - - + +
LMI070 - + - +
ps Ex e2 GOI
ps Ex e2 GOI

Figure 6

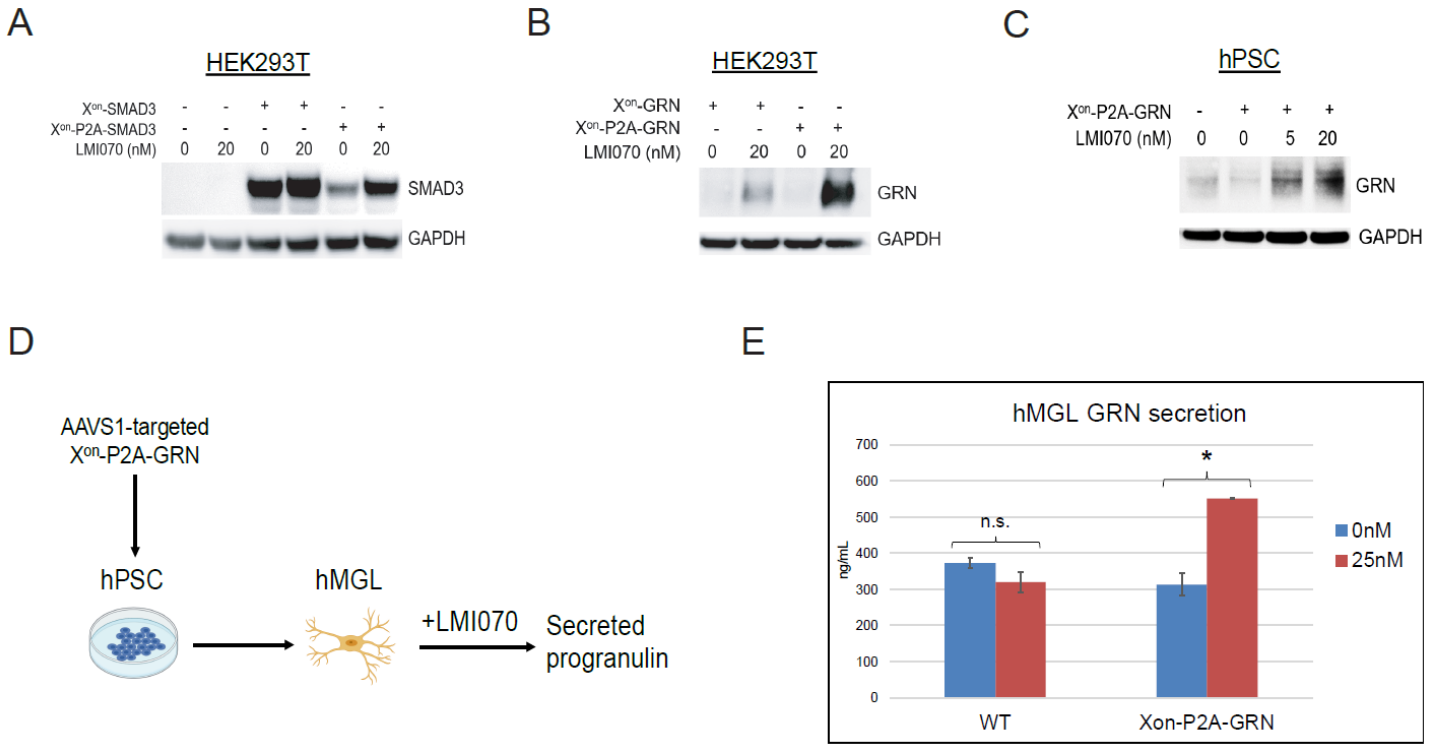

A
HEK293T
Xon-SMAD3 - - + + - -
Xon-P2A-SMAD3 - - - - + +
LMI070 (nM) 0 20 0 20 0 20
SMAD3
GAPDH
B
HEK293T
Xon-GRN + + - -
Xon-P2A-GRN - - + +
LMI070 (nM) 0 20 0 20
GRN
GAPDH
C
hPSC
Xon-P2A-GRN - + + +
LMI070 (nM) 0 0 5 20
GRN
GAPDH
D
AAVS1-targeted
Xon-P2A-GRN
hPSC
hMGL
+LMI070
Secreted
progranulin
E
hMGL GRN secretion
ng/mL
700
600
500
400
300
200
100
0
n.s.
*
0nM
25nM
WT
Xon-P2A-GRN

Figure 7

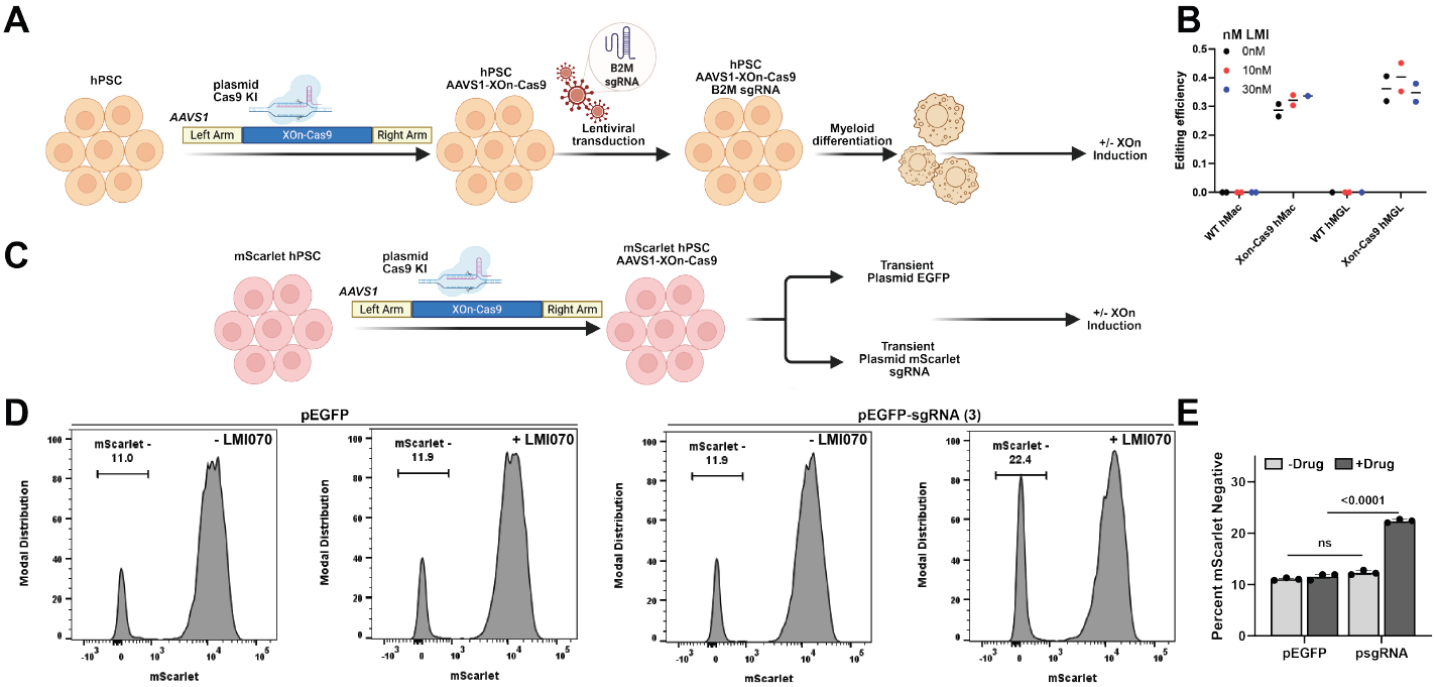

A
hPSC
plasmid
Cas9 KI
AAVS1
Left Arm
XOn-Cas9
Right Arm
hPSC
AAVS1-XOn-Cas9
B2M
sgRNA
Lentiviral
transduction
hPSC
AAVS1-XOn-Cas9
B2M sgRNA
Myeloid
differentiation
+/- XOn
Induction
B
nM LMI
0nM
10nM
30nM
Editing efficiency
WT hMac
Xon-Cas9 hMac
WT hMGL
Xon-Cas9 hMGL
C
mScarlet hPSC
plasmid
Cas9 KI
AAVS1
Left Arm
XOn-Cas9
Right Arm
mScarlet hPSC
AAVS1-XOn-Cas9
Transient
Plasmid EGFP
Transient
Plasmid mScarlet
sgRNA
+/- XOn
Induction
D
pEGFP
pEGFP-sgRNA (3)
mScarlet -
11.0
- LMI070
mScarlet -
11.9
+ LMI070
mScarlet -
11.9
- LMI070
mScarlet -
22.4
+ LMI070
Modal Distribution
mScarlet
E
-Drug
+Drug
<0.0001
ns
Percent mScarlet Negative
pEGFP
psgRNA

Figure S1

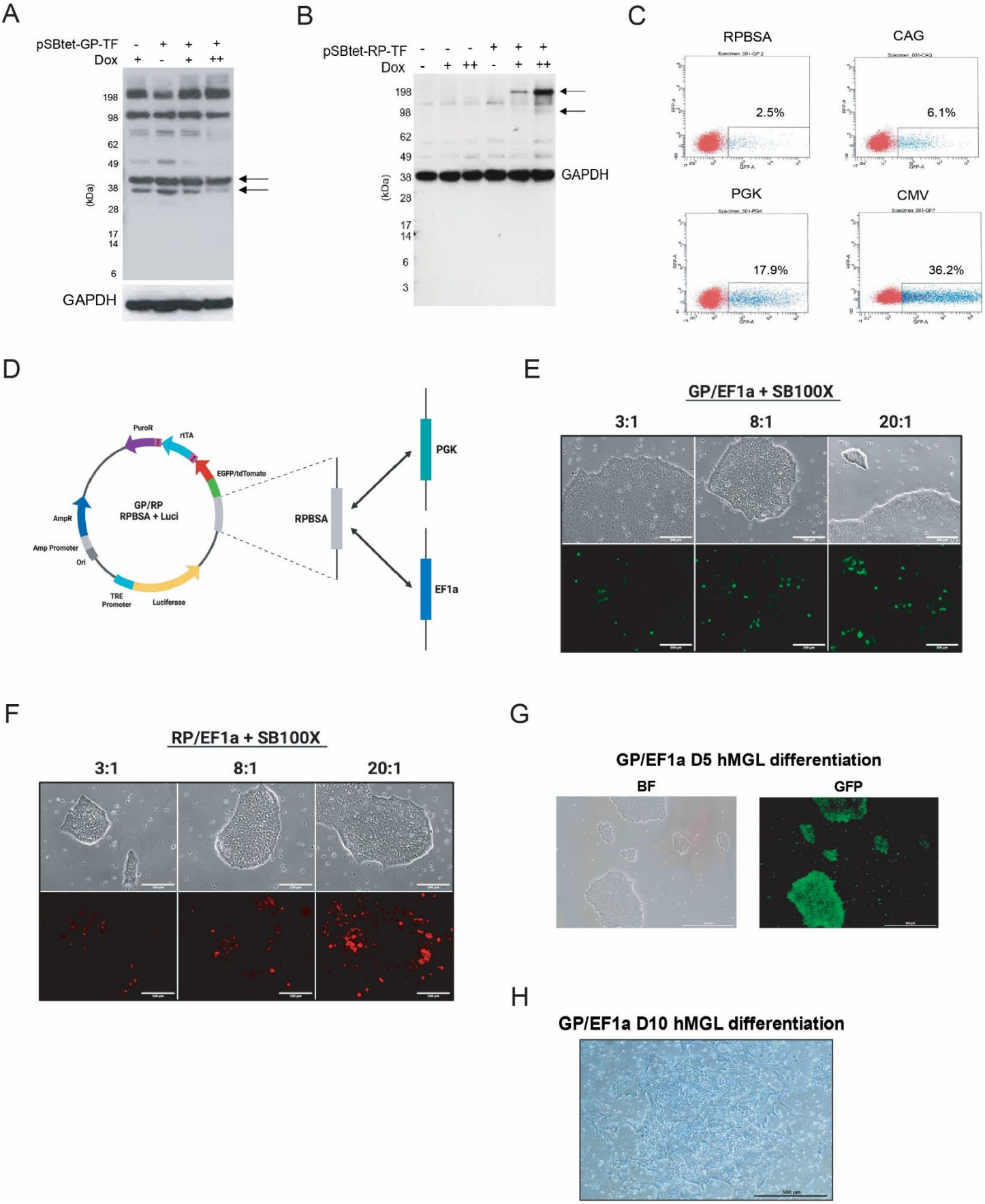

A
pSBtet-GP-TF - + + +
Dox + - + ++
198
98
62
49
38
28
17
14
6
(kDa)
GAPDH
B
pSBtet-RP-TF - - - + + +
Dox - + ++ - + ++
198
98
62
49
38
28
17
14
6
3
(kDa)
GAPDH
C
RPBSA
2.5%
CAG
6.1%
PGK
17.9%
CMV
36.2%
D
PuroR
rtTA
EGFP/tdTomato
GP/RP
RPBSA + Luci
AmpR
Amp Promoter
Ori
TRE
Promoter
Luciferase
RPBSA
PGK
EF1a
E
GP/EF1a + SB100X
3:1
8:1
20:1
F
RP/EF1a + SB100X
3:1
8:1
20:1
G
GP/EF1a D5 hMGL differentiation
BF
GFP
H
GP/EF1a D10 hMGL differentiation

Figure S2

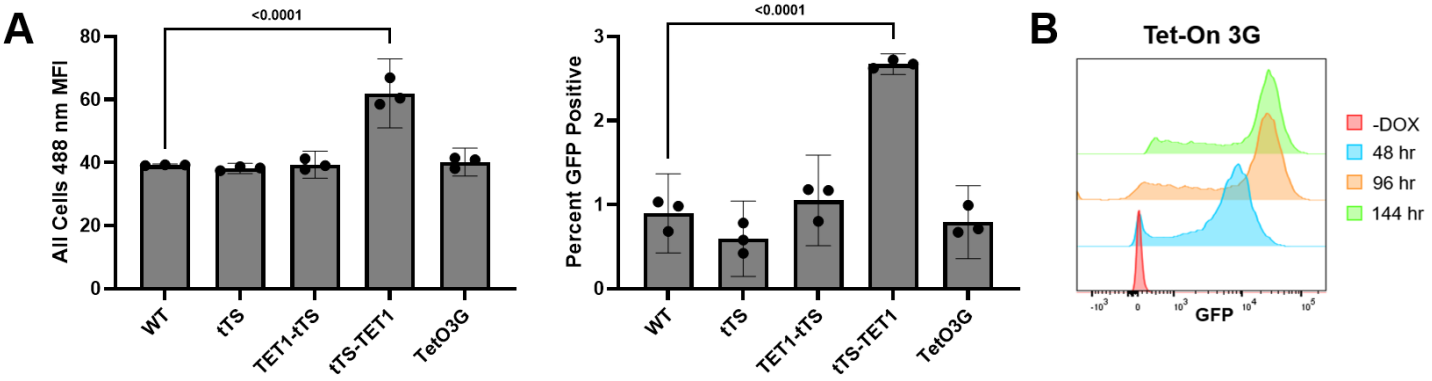

A
<0.0001
All Cells 488 nm MFI
80
60
40
20
0
WT
tTS
TET1-tTS
tTS-TET1
TetO3G
<0.0001
Percent GFP Positive
3
2
1
0
WT
tTS
TET1-tTS
tTS-TET1
TetO3G
B
Tet-On 3G
-DOX
48 hr
96 hr
144 hr
GFP

Figure S3

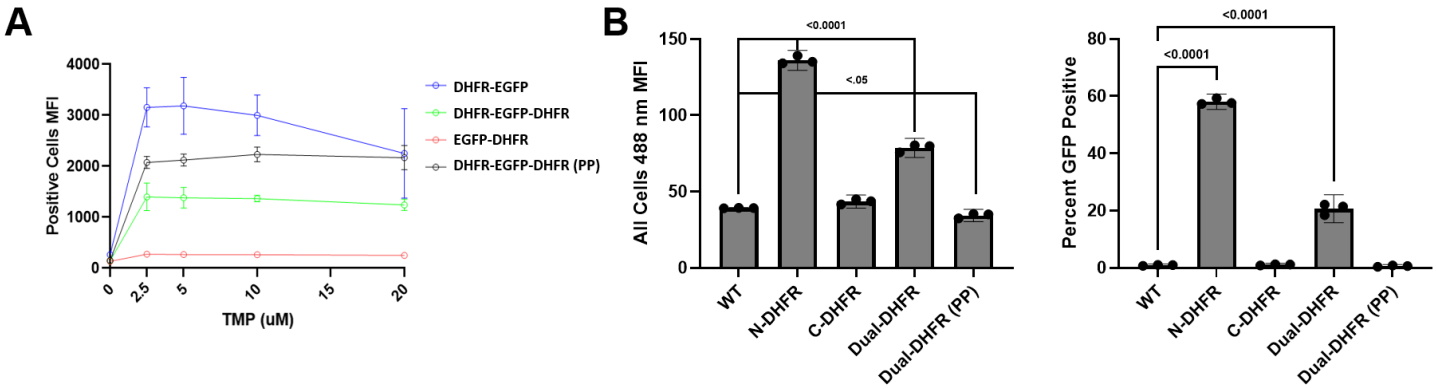

A
Positive Cells MFI
4000
3000
2000
1000
0
0
2.5
5
10
15
20
TMP (uM)
DHFR-EGFP
DHFR-EGFP-DHFR
EGFP-DHFR
DHFR-EGFP-DHFR (PP)
B
All Cells 488 nm MFI
150
100
50
0
<0.0001
<.05
WT
N-DHFR
C-DHFR
Dual-DHFR
Dual-DHFR (PP)
Percent GFP Positive
80
60
40
20
0
<0.0001
<0.0001
WT
N-DHFR
C-DHFR
Dual-DHFR
Dual-DHFR (PP)

Figure S4

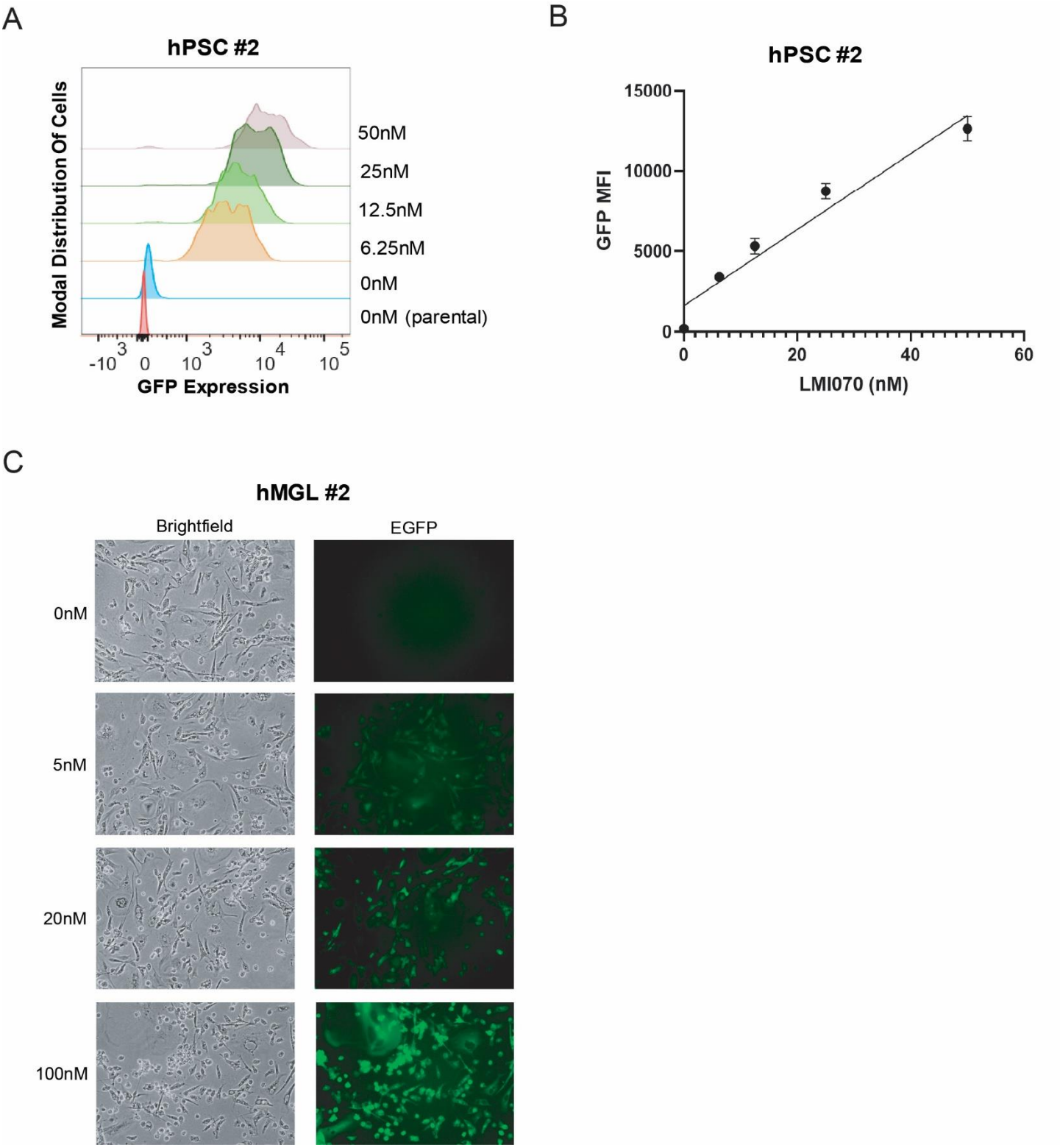

A
hPSC #2
Modal Distribution Of Cells
50nM
25nM
12.5nM
6.25nM
0nM
0nM (parental)
-10^3
0
10^3
10^4
10^5
GFP Expression
B
hPSC #2
GFP MFI
15000
10000
5000
0
0
20
40
60
LMI070 (nM)
C
hMGL #2
Brightfield
EGFP
0nM
5nM
20nM
100nM

Figure S5

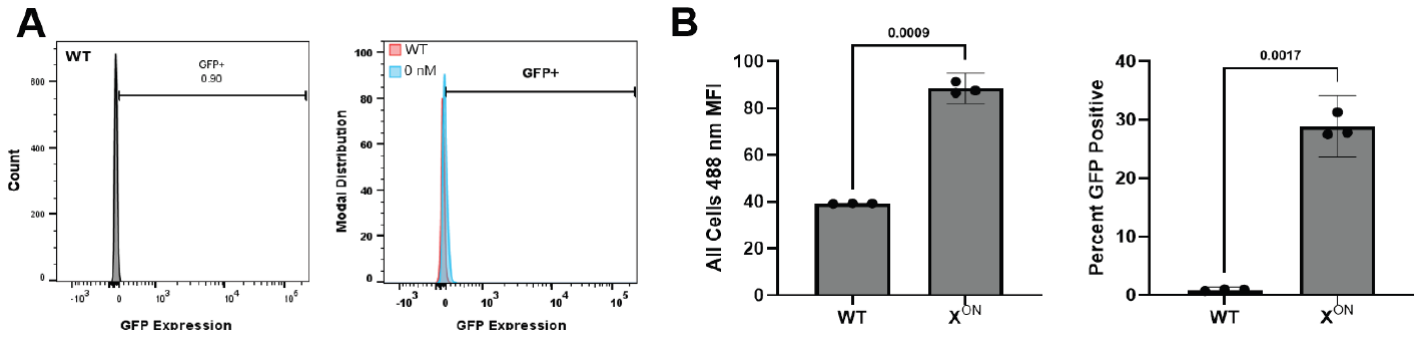
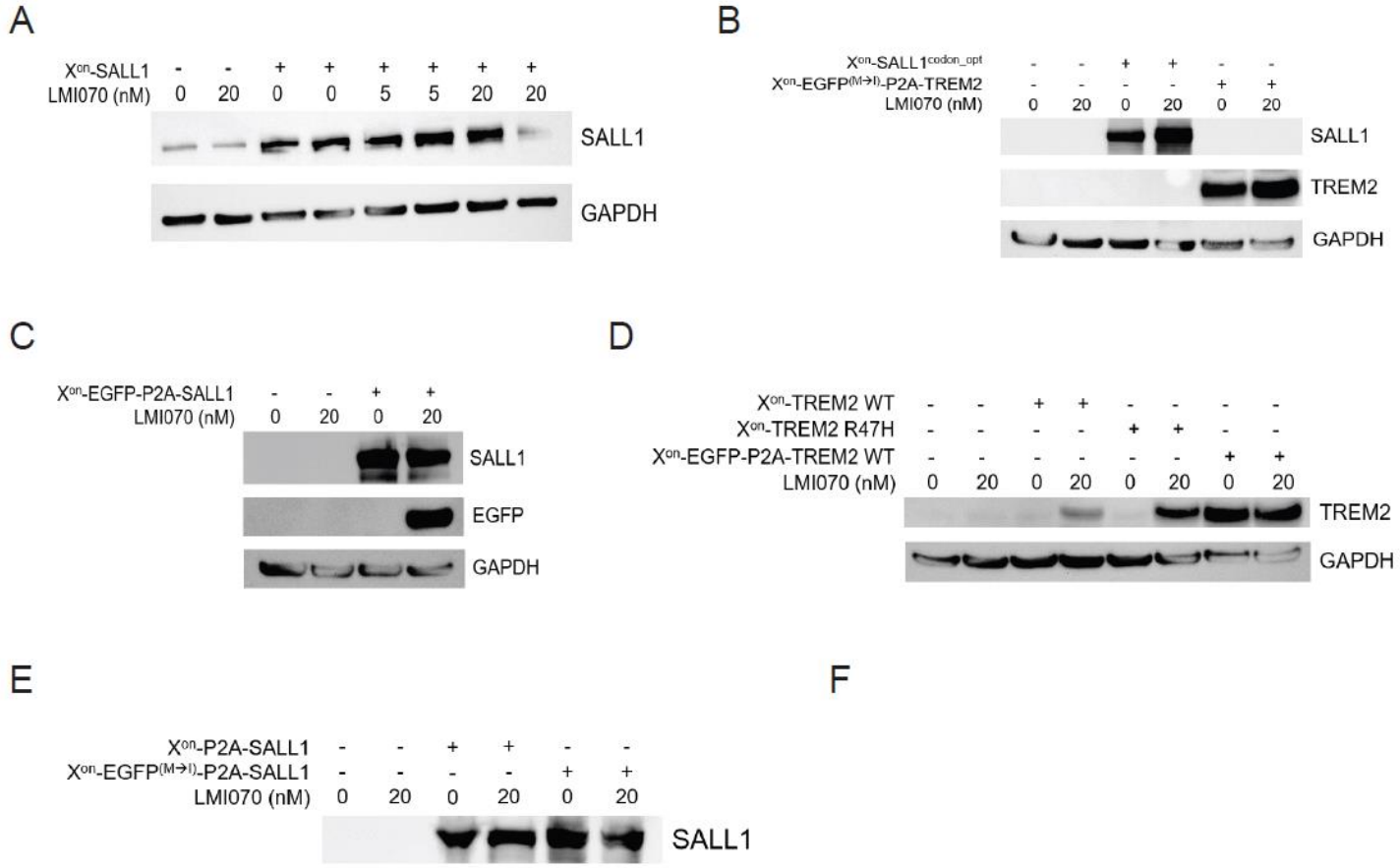

A
WT
GFP+
0.90
Count
GFP Expression
WT
0 nM
GFP+
Modal Distribution
GFP Expression
B
0.0009
All Cells 488 nm MFI
WT
X^ON
0.0017
Percent GFP Positive
WT
X^ON

Figure S6


A
Xon-SALL1 - - + + + + + +
LMI070 (nM) 0 20 0 0 5 5 20 20
SALL1
GAPDH
B
Xon-SALL1codon_opt - - + + - -
Xon-EGFP(M→I)-P2A-TREM2 - - - - + +
LMI070 (nM) 0 20 0 20 0 20
SALL1
TREM2
GAPDH
C
Xon-EGFP-P2A-SALL1 - - + +
LMI070 (nM) 0 20 0 20
SALL1
EGFP
GAPDH
D
Xon-TREM2 WT - - + + - - - -
Xon-TREM2 R47H - - - - + + - -
Xon-EGFP-P2A-TREM2 WT - - - - - - + +
LMI070 (nM) 0 20 0 20 0 20 0 20
TREM2
GAPDH
E
Xon-P2A-SALL1 - - + + - -
Xon-EGFP(M→I)-P2A-SALL1 - - - - + +
LMI070 (nM) 0 20 0 20 0 20
SALL1
F