

# Gene syntax defines supercoiling-mediated transcriptional feedback

Christopher P. Johnstone<sup>1</sup>, Kasey S. Love<sup>2</sup>, Sneha R. Kabaria<sup>1</sup>, Ross Jones<sup>3,4</sup>, Albert Blanch-Asensio<sup>5,6</sup>, Deon S. Ploessl<sup>1</sup>, Emma L. Peterman<sup>1</sup>, Rachel Lee<sup>1</sup>, Jiyoung Yun<sup>3,4</sup>, Conrad G. Oakes<sup>7</sup>, Christine L. Mummery<sup>5,6</sup>, Richard P. Davis<sup>5,6</sup>, Brandon J. DeKosky<sup>1,8</sup>, Peter W. Zandstra<sup>3,4</sup>, and Kate E. Galloway<sup>\*1</sup>

<sup>1</sup>*Department of Chemical Engineering, MIT, 25 Ames St., Cambridge, MA, 02139, USA*

<sup>2</sup>*Department of Biological Engineering, MIT, 25 Ames St., Cambridge, MA, 02139, USA*

<sup>3</sup>*School of Biomedical Engineering, UBC, 6088 University Boulevard, Vancouver, BC, V6T 1Z3, Canada*

<sup>4</sup>*Michael Smith Laboratories, UBC, 2185 East Mall, Vancouver, BC, V6T 1Z4, Canada*

<sup>5</sup>*Department of Anatomy and Embryology, Leiden University Medical Center, 2300RC Leiden, the Netherlands*

<sup>6</sup>*The Novo Nordisk Foundation Center for Stem Cell Medicine, reNEW, Leiden University Medical Center*

<sup>7</sup>*Department of Bioengineering, California Institute of Technology, Pasadena, CA, 91125, USA*

<sup>8</sup>*The Ragon Institute of Mass General, MIT, and Harvard, 600 Main St., Cambridge, MA, 02139, USA*

## Abstract

Gene syntax—the order and arrangement of genes and their regulatory elements—shapes the dynamic coordination of both natural and synthetic gene circuits. Transcription at one locus profoundly impacts the transcription of nearby adjacent genes, but the molecular basis of this effect remains poorly understood. Here, using integrated reporter circuits in human cells, we show that supercoiling-mediated feedback regulates expression of adjacent genes in a syntax-specific manner. Using Region Capture Micro-C, we measure induction-dependent formation of supercoiled plectonemes and syntax-specific chromatin structures in human induced pluripotent stem cells. Using syntax as a design parameter, we built compact gene circuits, tuning the mean, variance, and stoichiometries of expression across diverse delivery methods and cell types. Integrating supercoiling-mediated feedback into models of gene regulation will expand our understanding of native systems and enhance the design of synthetic gene circuits.

## 1 Main

Native gene circuits coordinate transcriptional programs to set diurnal rhythms, pattern cell fate, and orchestrate immune responses [1–3]. The non-random organization of genomes suggests that specific patterns of gene syntax—the relative order and orientation of genes and their regulatory elements—support coordinated regulation of co-localized genes [4–6]. Native gene circuits that require precise transcriptional coordination such as *Hox* genes [7, 8] and segmentation clocks [9] co-localize multiple transcriptional units within tens of kilobases. Potentially, gene syntax may constrain transcriptional noise and couple expression of adjacent genes in native and synthetic gene circuits (fig. 1a) [10]. The enrichment and depletion of syntaxes in the human genome and other genomes—which occur on the length scale of synthetic gene circuits—may suggest motifs for organizing pairs of adjacent genes for coordinated expression (figs. 1b and S1).

---

\*Corresponding author and lead contact, katiegal@mit.edu

Synthetic gene circuits offer programmable control of therapeutic cargoes, genome editors, and cell fate [11–19]. However, predictable forward design of gene circuits remains challenging, requiring iterative “design-build-test” loops to achieve desired functions. Harnessing syntax as an explicit design parameter may improve the predictability and performance of genome-integrated synthetic circuits [10]. While synthetic circuits are regularly integrated into the genome for cellular engineering, the reciprocal feedback between transcription and local chromatin structure remains unexplored in human cells. On the length scale of gene circuits (~ 10 kb), biophysical forces such as DNA supercoiling are predicted to rapidly couple the expression of colocalized genes [10, 20].

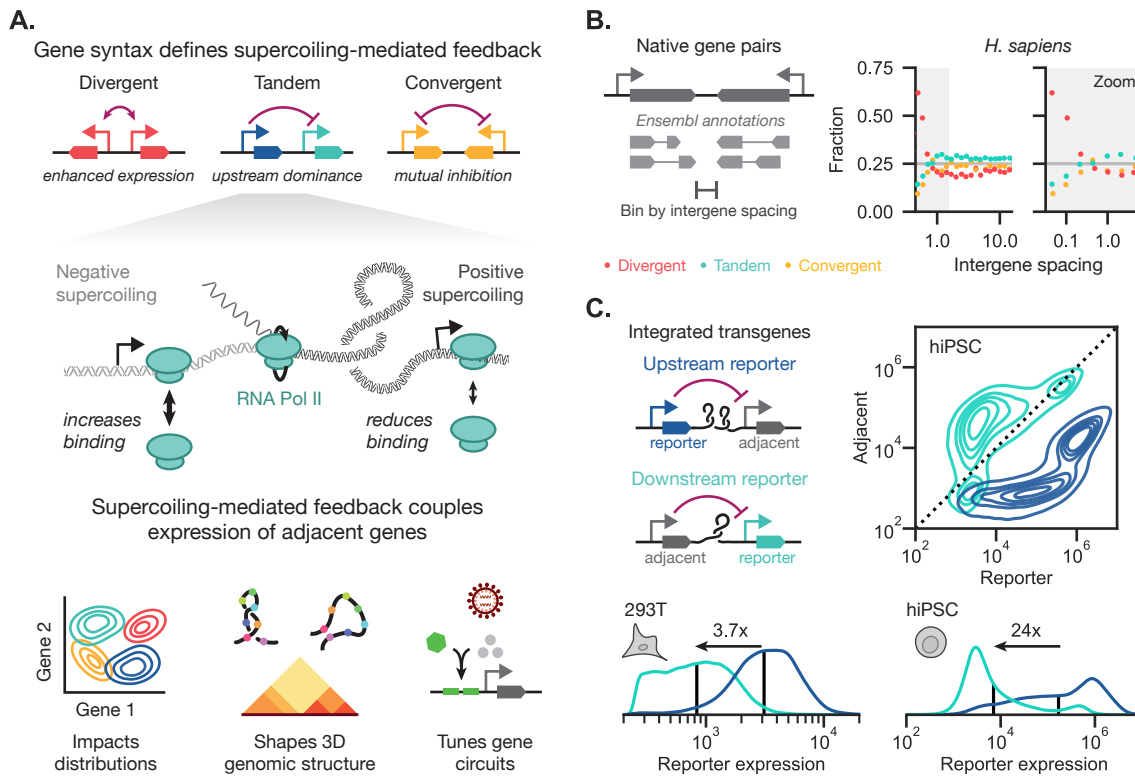
DNA supercoiling—the over- and under-twisting of DNA—influences a host of genomic processes [21], including transcriptional bursting [22–24], topoisomerase activity [25–30], chromatin folding [31, 32], and chromosome segregation [33]. By melting the double helical DNA polymer to read the underlying base pairs, transcribing RNA polymerases induce waves of DNA supercoiling. In yeast and human cells, supercoiling demarcates gene activity [34, 35]. Supercoiling alters RNA polymerase binding and changes the biochemical landscape of gene regulation [36, 37]. As supercoiling diffuses, transcription dynamically reshapes the structure and regulation of nearby genes, forming a feedback loop that we define as supercoiling-mediated feedback [10]. As described by the twin domain model [38, 39], transcribing polymerases generate positive supercoiling downstream and leave a wake of negative supercoiling upstream (fig. 1a). Negative supercoiling facilitates polymerase binding by reducing the binding energy, while positive supercoiling decreases binding rates by increasing this energy barrier. Thus, supercoiling-mediated feedback emerges through the directionality of transcription, which sets the energy landscape for subsequent polymerase binding events (fig. 1a)[10]. Models of supercoiling predict that transcription-induced changes in chromatin structure feed back into changes in transcriptional activity at adjacent genes [40–43], a phenomenon observed in bacteria [44], yeast [22, 45] and human cells [26, 35, 46]. However, while these forces shape gene regulation across species [22, 32, 34, 35, 45, 46], the impact of transcriptionally induced supercoiling on fine-scale chromatin structure and on the activity of gene circuits remains undefined in human cells.

DNA supercoiling is often studied through broad perturbations including loss and inhibition of topoisomerase and polymerases [22, 26, 35, 46, 47]. However, broad inhibition of transcription and topoisomerase activity can induce large changes in cellular physiology, limiting observations to acute treatments and short timescales. Alternatively, changing the transcriptional activity of a single transgene within a genetically uniform background offers the precise control required to investigate the predictions of supercoiling-mediated feedback.

Here, we use synthetic two-gene circuits as a model system to examine how transcription of a single gene couples the expression and folding of adjacent genes in human cells. By integrating inducible systems, we demonstrate that transcription-induced coupling generates syntax-specific profiles of expression across a range of human cell types and integration methods. Integrating circuits at a genomic safe harbor in human induced pluripotent stem cells (hiPSCs), we use Region Capture Micro-C [48] to characterize folding across the region surrounding the locus of circuit integration. Using the control and orthogonality of synthetic circuits, we identify transcription-induced changes in chromatin structure, demonstrating the predicted coupling of syntax-specific chromatin folding and patterns of expression. Induction of transcription perturbs chromatin structure hundreds of kilobases away, substantially altering the insulation and connectivity across the locus. Using principles of supercoiling-mediated feedback, we design compact synthetic gene circuits for efficient delivery and induction across a variety of cells. These techniques allow us to optimize production of a therapeutic antibody without substitution of genetic parts. Overall, we demonstrate how supercoiling-mediated feedback influences expression of adjacent genes, providing insights into native gene regulation and informing the design of synthetic systems.

## Upstream dominance defines expression profiles of constitutive tandem transgenes

In native genes, upstream transcription can reduce expression of downstream genes [49–51]. Models of supercoiling-mediated feedback predict that positive supercoiling generated by transcription at the upstream gene reduces the rate of transcription and thus expression at the downstream locus, resulting in upstream dominance (fig. 1a) [10]. To examine upstream dominance in a synthetic system, we constructed two-gene systems in tandem. Each gene is paired with a promoter and polyadenylation signal (PAS) to form a transcriptional unit. The modularity of these synthetic systems allows us to independently switch gene positions and regulatory elements.



**Figure 1:** Supercoiling-mediated feedback couples transcription and genome folding of adjacent genes.

a) Supercoiling modifies the energy required for polymerases to bind and locally melt DNA, leading to increased or decreased polymerase initiation. This biophysical feedback loop (purple arrows) generalizes across integration method and cell type, opening new engineering capabilities.

b) Using Ensembl annotations for the human genome, gene extents were identified using the maximum extent of all annotated exons. For each pair of adjacent genes, the relative orientation and intergene spacing was computed, split into equal-sized quantile bins, and summarized by orientation.

c) Due to accumulated positive supercoiling at the downstream promoter, expression from an upstream gene is predicted to decrease expression of a downstream gene. Two-gene constructs expressing fluorescent proteins from PGK promoters were integrated using PiggyBac in HEK293T cells. For a representative biological replicate, the distribution of the reporter is shown as a function of position in the circuit. In the hiPSCs, the adjacent gene is expressed from the weak PGK promoter and the reporter gene is expressed from the strong EF1a promoter.

To measure expression, we integrated these tandem two-gene systems into two common human cell lines, HEK293Ts and hiPSCs, via PiggyBac transposase. We switch the positions of the tandem genes to isolate the effect of position on expression level. We quantified expression of the fluorescent reporter genes by flow cytometry, providing single-cell resolution needed to measure expression distributions. Using identical promoters in both positions in HEK293Ts, we found that gene position strongly influences the

expression of the reporter. The gene in the upstream position expresses at levels nearly four-times higher than when placed downstream (fig. 1c). Even when pairing a strong promoter with a weak promoter in hiPSCs, we saw clear upstream dominance with a large, 24-fold shift in expression based on position (fig. 1c). Potentially, genetically encoded sequences—such as binding sites for the CCCTC-binding factor (CTCF) [52, 53] and the cHS4 insulator [54]—that restrict chromatin-mediated interactions may reduce upstream dominance. To examine this hypothesis, we inserted tandem-oriented CTCF binding sites that were previously reported to reduce enhancer-promoter interactions [53]. Flanking the upstream gene, the downstream gene, or the entire two-gene construct with these sites does not eliminate upstream dominance (fig. S2). Instead, addition of these sites generally reduces expression of one or both genes.

The identity of regulatory elements such promoters and PAS may influence coupling between genes [55, 56]. In testing a panel of common constitutive promoters, we consistently observed upstream dominance in PiggyBac-integrated HEK293Ts (figs. S3a and S3b), lentivirally integrated HEK293Ts (fig. S3c), and PiggyBac-integrated hiPSCs (fig. S4). In exchanging the PAS, we found that the choice of PAS has a minimal effect on the shape of expression distributions, mildly tuning the levels of gene expression (fig. S5). Thus, trends in syntax-specific expression are robust across a range of genetic parts for the tandem syntax.

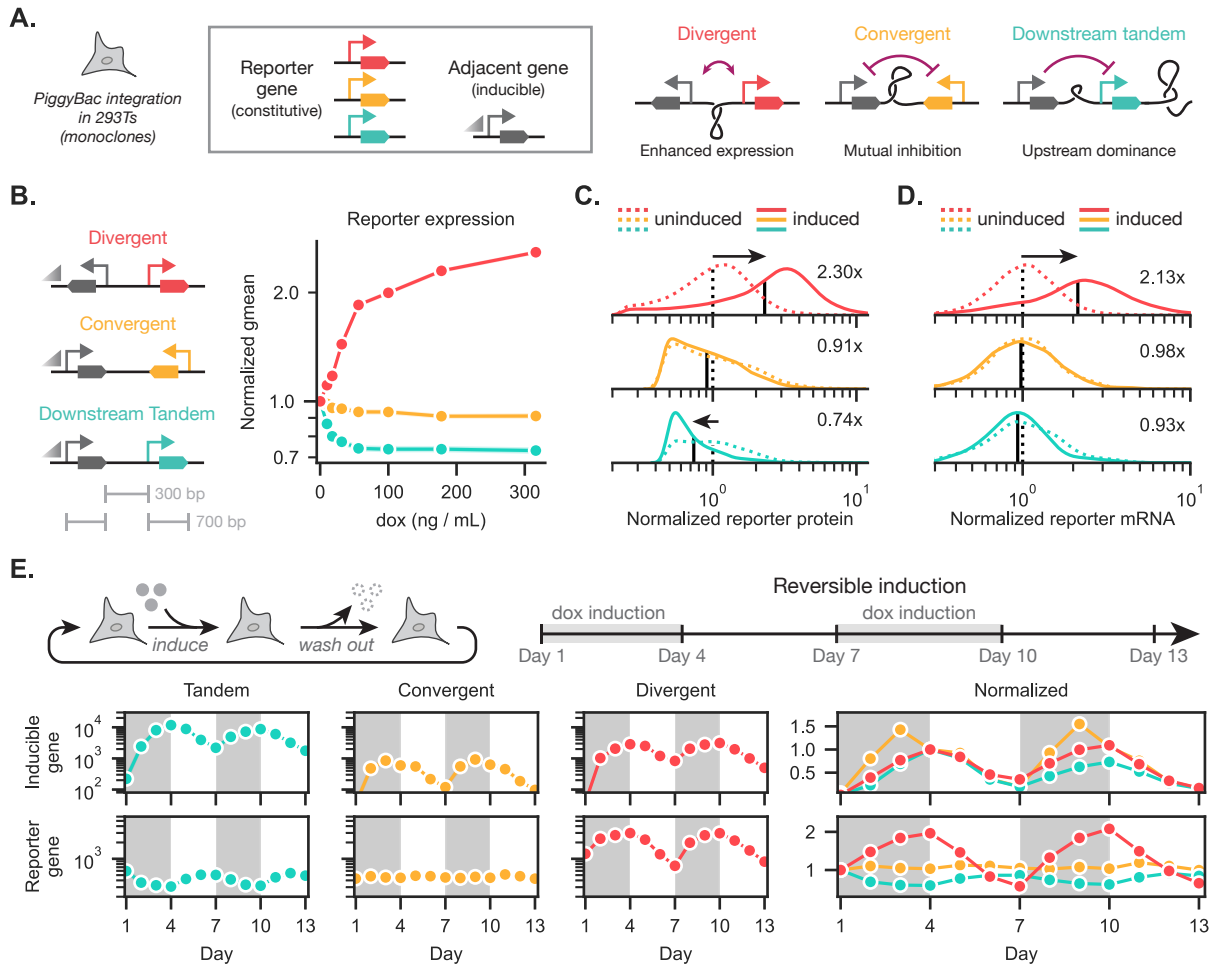
### **Transcription of an adjacent gene induces syntax-specific coupling**

Expression patterns of constitutively expressed tandem gene pairs suggest that syntax influences expression. However, these constitutive systems do not support dynamic control of transcription of a single adjacent gene, making it difficult to parse the transcription-driven mechanism of supercoiling-mediated feedback. To allow controlled induction of a single adjacent gene, we generated monoclonal HEK293T cell lines containing a doxycycline (dox)-inducible two-gene system in different syntaxes (fig. 2). To build these lines, we placed a constitutive reporter gene under the control of a strong constitutive promoter and an adjacent inducible gene under the control of the dox-inducible promoter TRE (fig. 2a). Using PiggyBac, we delivered the dox-inducible two-gene systems encoded in tandem, convergent, and divergent syntaxes, which are predicted to show different transcription-induced couplings [10]. The constitutively expressed dox-responsive activator, rtTA, was integrated from a separate PiggyBac donor. We sorted single cells to establish monoclonal lines of each syntax.

Upon dox addition, all syntaxes show strong induction of the TRE-driven inducible gene (figs. S6a and S6b). To quantify changes in expression of the constitutive reporter upon induction of the adjacent gene, we normalized reporter expression to the uninduced condition for each line. In the tandem syntax, induction of the upstream gene reduces expression of the downstream reporter gene (fig. 2b, figs. S6c and S6d). Conversely, induction of the divergent syntax strongly upregulates expression from the constitutive reporter, matching predictions of amplification in divergent syntax [10]. The convergent syntax shows a near-invariant profile of reporter expression. For the tandem and divergent syntaxes, induction of the adjacent gene results in a unimodal shift in the geometric mean (fig. 2c). Unimodal shifts indicate a general mechanism of regulation, such as changes in the transcription rate, that is not restricted to a subpopulation of cells.

As a transcription-based process, supercoiling-mediated feedback should manifest in the distributions of mRNAs. To measure the mRNA distributions, we used single-cell hybridization chain reaction RNA-FISH [56, 57] to quantify the transcriptional profiles of both the constitutive reporter gene and the dox-inducible gene (fig. 2d, figs. S6e, S6f and S7). As expected, mRNA profiles generally match the syntax-specific profiles of proteins (fig. 2c). Intriguingly, the convergent syntax shows substantial bimodality in mRNA expression, matching modeling predictions of bimodality that may be obscured by stable protein reporters [10]. Overall, the mRNA profiles align with the models of supercoiling-mediated feedback that predict that syntax influences transgene expression by altering rates of transcription.

DNA supercoiling-mediated coupling is predicted to be rapid and reversible. To test the reversibility of syntax-specific coupling, we sequentially induced and removed dox for three-day periods over 13 days. We observed repeatable induction- and syntax-specific coupling of the inducible and reporter genes. Over 13 days we observed minimal hysteresis, indicating that the trends in coupling are reversible and are not due to irreversible changes in the chromatin state (fig. 2e).



**Figure 2:** Transcription induces syntax-specific coupling of expression of adjacent genes.

- We integrated two-gene systems consisting of an dox-inducible gene (TRE) and a constitutively expressed gene (EF1a). The resulting cell lines were flow sorted to single cells and expanded as monoclonal populations.
- The geometric mean reporter expression, normalized to the uninduced condition, is shown as a function of dox concentration for the three different syntaxes. Geometric mean and associated 95% confidence interval shown over merged distributions from three wells.
- Full reporter protein and mRNA distributions are shown in the uninduced case and the second-highest dox induction state. Fold change of the geometric means are shown in black.
- Dox was sequentially introduced and removed in order to measure the turn-on and turn-off dynamics of the integrated systems. The systems respond reversibly to the presence of dox. Geometric mean and associated 95% confidence interval shown over merged distributions from three wells.

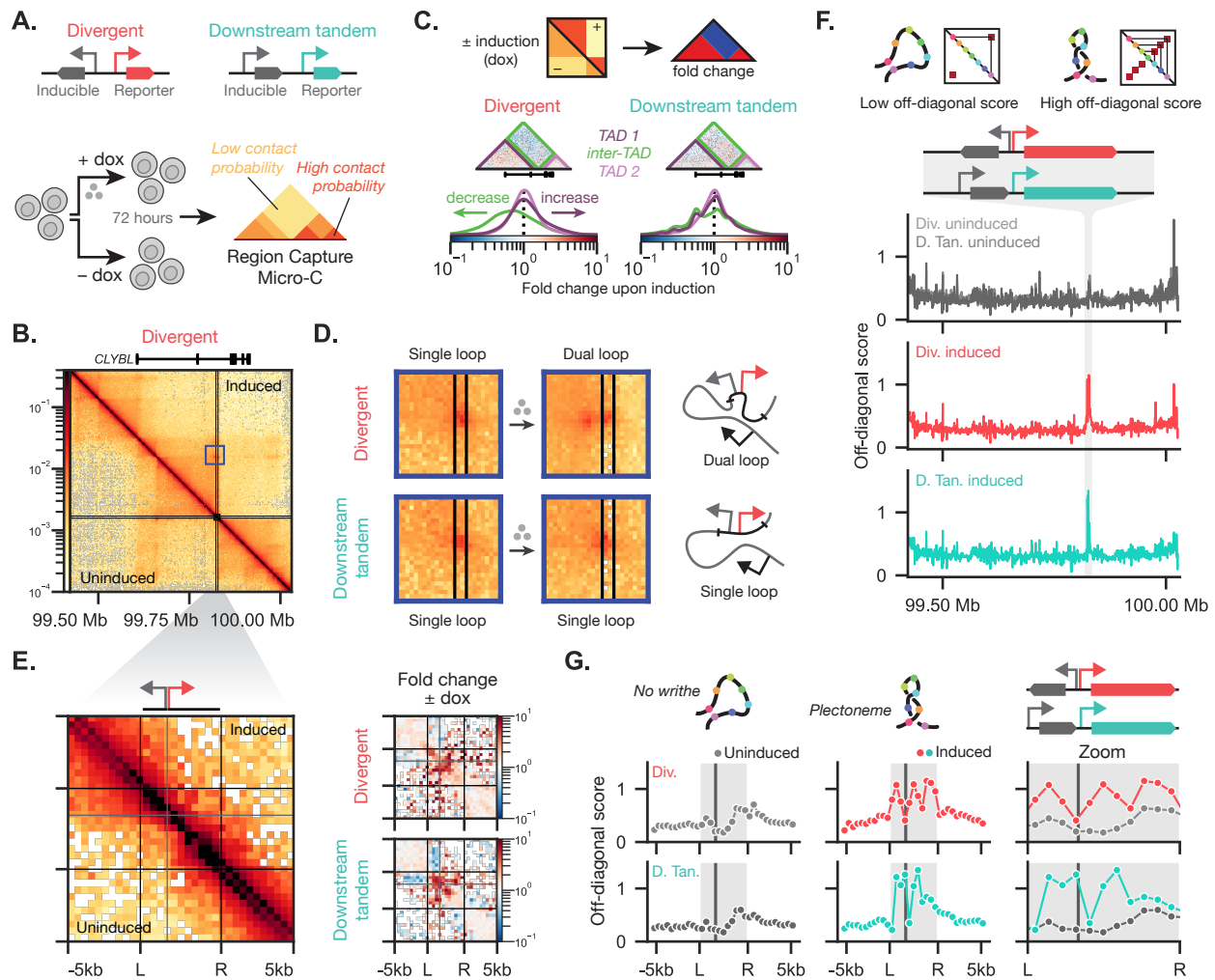
## Transcription induces syntax-specific structures across circuits and the surrounding locus

Supercoiling-mediated changes in gene expression occur alongside changes in chromatin structure. Transcription-induced supercoiling drives genome folding [58]. In bacteria and in yeast, chromatin structures correlate with supercoiling density [47, 59]. Inclusion of inducible promoters in synthetic circuits provides a facile mechanism to regulate transcriptional activity and measure transcriptionally induced chromatin structures. To understand how transcriptional activity and syntax affect the chromatin structure of our two-gene circuits, we used Region Capture Micro-C (RCMC) to measure the contact probability between genomic locations within a targeted region of interest (fig. 3) [48]. Using the STRAIGHT-IN Dual allele platform [60], we integrated dox-inducible circuits with tandem or divergent syntax into both alleles of a genomic safe harbor region located in intron 2 of the citrate lyase beta-like (*CLYBL*) gene [61] in hiPSCs, generating homozygous cell lines (fig. S8a). These circuits include the activator within the constitutively expressed gene. We confirmed that these tandem and divergent inducible circuits showed reversible, syntax-specific profiles of expression similar to those observed in the dox-inducible two-gene systems integrated into HEK293T cells (figs. 2 and S8b).

We cultured two biological replicates of each syntax with or without inducer and collected these cells for RCMC (figs. 3a and S9a). Reads from each replicate were merged and balanced (fig. S9b), resulting in matrices that quantify the probability that two chromatin regions are colocalized in 3D space. Two distal, unmodified capture regions did not show induction-dependent changes (fig. S9c). In contrast, across the locus of integration, we observed induction-specific structural changes for both the divergent syntax (fig. 3b) and downstream tandem syntax (fig. S10a).

The integration site is located at the boundary of two topologically associating domains (TADs) (fig. 3b). We quantified how induction of our small synthetic construct changes chromatin structure over the entire capture region by computing the fold-change in contact probability upon induction (fig. 3c). Induction of the divergent syntax reduces inter-TAD interactions and increases intra-TAD interactions. Using a common sliding-window insulation score [62, 63] to calculate TAD boundary strength, we observe a substantial weakening of this local boundary upon induction in the divergent syntax despite reduced inter-TAD interactions (fig. S11). Weakening of this boundary coincides with the emergence of a dual-loop domain. All conditions show a single “corner dot” that represents a loop domain between the integration site and the first intron of *CLYBL*. Induction in the divergent syntax generates a second corner dot anchored in a region 10 kb upstream of the integration locus (fig. 3d). This suggests that the weakening of the TAD boundary arises from interactions between these two loop domains.

Zooming into the 15-kB window around the circuit, we examined how induction affects local chromatin structure (figs. 3e and S10b). Plectonemes form as overwound (or underwound) DNA buckles, transferring twist into writhe. In addition to facilitating the loading of chromatin loop extruders [64], plectonemes should show higher intra-plectoneme contacts. Specifically, plectonemes should appear as small-scale off-diagonal regions of high contact probability (fig. 3f), similar in shape to the large-scale “jets” generated by cohesion loop extrusion [65]. Quantified as an off-diagonal score, we observe strong, induction-dependent plectonemic signals at the integrated locus (figs. 3f and 3g). The magnitude of these plectonemic signals is not replicated elsewhere in the capture region, strongly suggesting that induction of transcription reshapes chromatin folding around our synthetic circuit. Potentially, in addition to affecting polymerases, this local supercoiling density may affect the binding rates of other DNA-binding proteins.



**Figure 3:** Transcription induces syntax-specific chromatin structures across synthetic gene circuits and the surrounding locus. **a)** The resulting cell lines were split into two conditions, with half induced with doxycycline (dox) for 72 hours prior to harvesting. The cells were processed following the Region Capture Micro-C protocol in order to measure population chromatin structure around the region of integration. **b) - g)** Region Capture Micro-C data was binned at 500 bp and 2 kb resolution and iteratively balanced within the capture region. **b)** For the divergent cell line, only relatively small differences differentiate the induced and uninduced conditions across the  $\sim 700$  kb region around the integration site. **c)** For each condition, the fold change in contact probability was computed. The resulting distribution was binned by region. In the divergent case, the inter-TAD region shows reduced contact probability upon induction. **d)** Examining a corner dot representing a loop between the integration region and the first intron of *Clybl*, the induced divergent condition shows two corner-dots, suggesting an induction-dependent formation of a double-loop structure. This double loop does not appear in the downstream tandem syntax. **e)** For the divergent cell line, the surrounding 10 kb region around the integration site is shown at 500 bp resolution. **f)** The off-diagonal score is shown across the entire capture region. No other region shows strong increases in off-diagonal score upon induction. **g)** Local plectoneme formation can be quantified using the off-diagonal score. As opposed to a “corner dot” structure which indicates a loop domain, plectonemes should show contacts along the matrix off-diagonal. Examining the region immediately around the integration location, this off-diagonal score remains low in the uninduced case. However, upon induction, we see a strong increase in contact probability along the central off-diagonals and as measured by the off-diagonal score.

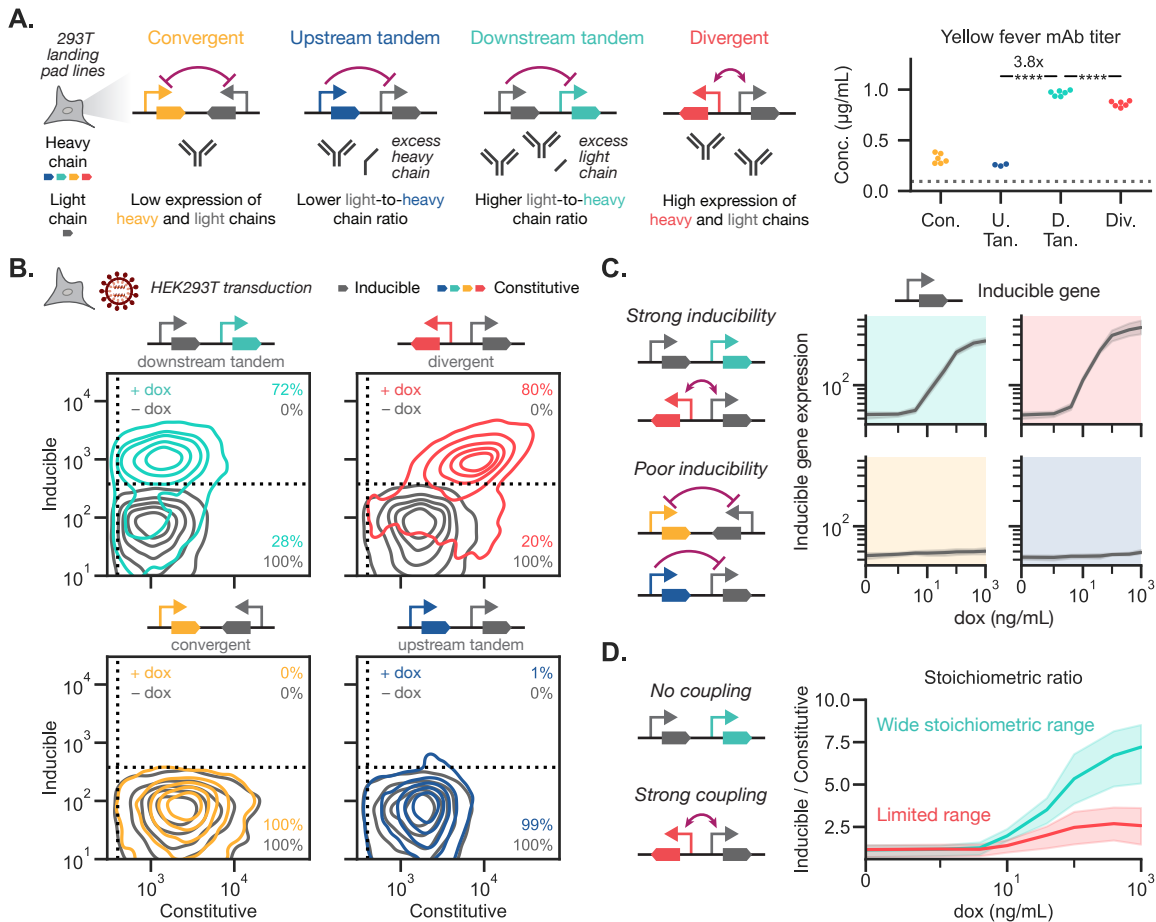
## Syntax-based tuning optimizes circuit expression and biologic production without part substitution

Synthetic circuits and other transgenic systems such as biologic producer lines are often optimized through selection of transcription factors, promoters, stoichiometric ratios, copy number, and integration locus [66–68]. However, as each element may affect the dynamics of gene expression, forward design through simple exchange of parts remains iterative. Given that syntax modulates expression to a similar degree as genetic element selection [56], we propose using syntax to tune the relative levels of expression without changing the sequences of these elements or their relative copy number. To test this syntax-based tuning scheme, we explored optimization of two common biotechnological tools: a monoclonal antibody producer line and an inducible lentivirus system.

Low-cost production of antibodies, especially for those against infectious and tropical diseases [69], can improve worldwide access to these antibody drugs. Increasing antibody titers offers a simple way to reduce the cost of production and enhance affordable access. To demonstrate the promise of syntax-based optimization, we integrated two-gene constructs encoding the heavy and light chains of an anti-yellow-fever monoclonal antibody into a landing pad HEK293T cell line. Previous reports [68, 70–72] suggest that excess light chain translation can increase titers. Thus, based on the principle of upstream dominance, we would expect that by setting a high ratio of light chain to heavy chain, the downstream tandem syntax would outperform the upstream tandem syntax. In measuring total human IgG titer via both sandwich ELISA (fig. 4a) and a bead agglutination assay (fig. S12), we found a nearly four-fold difference in antibody titer as a function of syntax, with the downstream tandem and divergent syntaxes providing the highest titers as expected.

Lentiviruses offer efficient delivery of transgenes to diverse primary cells for therapeutic applications such as *ex vivo* engineering of CAR-T therapies and *ex vivo* immune cell reprogramming [73]. Gene circuits and inducible systems offer safe, clinically guided control and the ability to target specific cell states [11]. However, ensuring robust co-expression of multiple genes in these systems remains challenging. To explore syntax-based tuning of expression from an inducible lentivirus, we tested all four possible two-gene syntaxes transduced into HEK293T cells. Only the divergent and downstream tandem syntaxes display an appreciable double-positive population at maximum induction (figs. 4b, S13 and S14). Expression of the constitutive gene in the upstream tandem and convergent syntaxes strongly inhibits induction across all inducer concentrations (fig. 4c). In the two syntaxes with strong induction, syntax sets the stoichiometric ratio between the two genes (fig. 4d). Weak coupling between the two genes in the tandem syntax allows a wide range of stoichiometries upon induction, varying stoichiometry seven-fold (fig. 4d). Conversely, the strong positive coupling between genes in the divergent syntax maintains a narrower ratio of expression. Tuning the ratio of expression between elements can significantly shift the behavior of gene circuits, potentially supporting or impeding desired functions [74, 75]. Together, our results demonstrate that syntax can tune expression levels in diverse synthetic circuits without requiring part substitution.





**Figure 4:** Syntax-based tuning optimizes circuit expression and biologic production without part substitution.

a) The light and heavy chains are expressed from two-gene constructs integrated at the *Rogi2* locus in HEK293T cells. Antibody titer, as measured via sandwich ELISA, differs across syntaxes. Points depict N=3-6 biological replicates. Statistics are two-sided student t-tests. \*\*\*\*:  $p < 0.0001$

b) Two-gene circuits, consisting of a constitutive gene and an inducible gene, were lentivirally transduced into HEK293T cells alongside a rtTA-expression lentivirus. Joint distributions in the absence (gray) or presence (colored) of 1 µg/mL dox are shown. Percentages refer to the proportion of cells expressing (top) or not expressing (bottom) the inducible gene.

c) The geometric mean of the inducible gene (gray) is shown as a function of dox concentration. Gray shading represents the 95% confidence interval across four biological replicates.

d) The stoichiometric ratio between the inducible and constitutive genes is shown as a function of dox concentration. Strong coupling reduces the change in this ratio. Colored shading represents the 95% confidence interval across four biological replicates.

## Syntax augments performance of compact gene circuits across cell types

Compact gene circuits support efficient delivery of therapeutic cargoes via size-restricted vectors such as lentiviruses and AAVs. However, the close proximity of multiple genes in these vectors introduces the potential for physical coupling between transcriptional units. To harness supercoiling-mediated feedback for improved circuit performance, we focused on optimizing a compact, lentivirally delivered “all-in-one” inducible circuit.

Unlike the inducible circuits in figs. 2 and 4, all-in-one designs include the dox-responsive activator on the same construct, resulting in both biophysical and biochemical coupling (fig. 5a). In the divergent syntax, positive supercoiling-mediated feedback should generate high, correlated expression. For both tandem syntaxes, negative feedback should reduce the degree of correlation between genes. Despite negative feedback, we expect that the downstream tandem syntax will support induction provided that activator levels remain sufficient [60].

Transducing hiPSCs, we observe robust induction from the divergent and downstream tandem syntaxes (figs. 5b and 5c). These trends are mirrored for transduction in HEK293T cells and mouse embryonic fibroblasts (fig. S15). Expression of the synthetic activator in the downstream tandem syntax is more than an order of magnitude lower than in the divergent syntax. However, for all syntaxes, expression of the activator increases relative to the uninduced case (fig. 5d). This increase may reflect a local increase in transcriptional resources that affects circuits with the activator in *cis* but not those in *trans* (i.e., with a separately integrated activator, as in figs. 2 and 4) (fig. S14) [60].

Supercoiling-mediated feedback is predicted to couple the probabilities of transcriptional bursting [10]. Even for stable protein reporters, changes in correlated bursting may be visible in the variance of co-expression. Variance between two genes across a population of cells can be decomposed into two components: intrinsic and extrinsic noise. Intrinsic noise quantifies the variability within individual single cells whereas extrinsic noise reflects differences such as cell size [76, 77]. Transcriptional co-bursting in the divergent syntax is predicted to reduce the intrinsic noise. Remarkably, across all transduced cell types, the divergent all-in-one syntax minimizes intrinsic noise (figs. 5e, 5f and S16a). Trends in intrinsic noise for these *cis* designs (i.e., all-in-one) match those for the *trans*-inducible circuits from fig. 4b (fig. 5f). However, unlike the *cis* designs, the *trans*-inducible circuit does not exhibit increased extrinsic noise in the divergent syntax relative to the tandem syntax (fig. S16b), aligning with model predictions (fig. S16c) [10]. Putatively, the lack of a biochemical positive feedback loop in the *trans* designs prevents an increase in extrinsic noise.

Insulator sequences can reduce coupling between integrated transgenes [78], potentially mitigating the effects of supercoiling-mediated feedback. To test this hypothesis, we added cHS4 insulator sequences to the intergenic region in the all-in-one circuit. For all syntaxes, addition of the cHS4 core or full insulator sequence did not substantially change expression levels or noise profiles (figs. 5g and S17). Together, these data indicate that syntax offers a powerful design parameter for coupling and tuning profiles of expression that can be harnessed to dampen or amplify noise [79, 80].



## Discussion

Transcription forms a dynamic feedback loop mediated by DNA supercoiling. Here we use two-gene synthetic gene circuits as model systems to probe the biophysical influence of adjacent transcription (fig. 1). Using inducible synthetic circuits, we isolate the effects of reversible biophysical coupling and examine the impact of transcriptionally induced coupling on gene expression (fig. 2) and chromatin structure. Applying fine-scale mapping of a genomic safe harbor using RCMC, we observe striking transcriptionally induced structures and perturbations of loop domains and TADs up to hundreds of kilobases away (fig. 3). Importantly, we observe formation of transcriptionally induced plectonemes for potentially the first time in living human cells. Integrating the predictions of supercoiling-mediated feedback into circuit design, we use syntax-based optimization to increase antibody production, to tune expression from lentiviruses, and to explore performance regimes of inducible circuits in diverse cell types (fig. 4). Overall, we find that syntax defines the profiles and performance of gene circuits across a range of integration methods, cargoes, and cell types, offering a design parameter for tuning the performance and predictability of circuits (fig. 5).

Syntax offers an orthogonal design variable that can be combined with library-based approaches. Growing compendiums of parts expand the potential for library-based approaches for circuit tuning [67, 81, 82], but syntax-based tuning can be used to optimize circuits even when parts are constrained [60]. When changing design features including promoters, polyadenylation sequences, cell type, and integration method, we consistently observed reproducible syntax-specific coupling of gene expression. Syntax-specific profiles persisted even in the presence of putative insulator sequences like *CHS4*, suggesting these features do not serve as barriers to transcriptionally induced coupling. However, exploration of additional intergene sequences, genetic elements, and locations within the genome may reveal elements that mitigate or amplify syntax-specific coupling.

In both constitutive and inducible systems, we find that upstream dominance strongly affects closely spaced tandem genes, resulting in reduced expression of the downstream gene (fig. 1). Previous work suggests that upstream dominance affects integrated transgenes [83], that choice of gene orientation and direction affects expression from adenovirus vectors [84], and that the divergent orientation offers an efficient “all-in-one” Cas9 editing circuit [85]. Increasing numbers of synthetic circuits employ divergent syntax for delivery to primary cells [11, 12, 86, 87]. As these applications require high rates of co-delivery, choice of divergent syntax may reflect the selection of functional circuits during the design process.

Supercoiling-mediated feedback provides an extremely rapid mechanism of transcriptional coupling [10]. In alignment with our predictions of supercoiling-mediated feedback, we find that transcription of an adjacent gene induces reversible, syntax-specific profiles of expression, tuning both the mean and variance (fig. 2). Parallel work in yeast showed increased burst coupling for *cis* divergent genes compared to *trans* at the *Gal1-Gal10* locus [22]. Potentially, harnessing the fast-timescale feedback of supercoiling may improve the performance of dynamic circuits such as pulse generators, toggle switches, and oscillators, which require coordinated expression of multiple genes. Supercoiling-mediated feedback may be especially useful to buffer noise in RNA-based control systems, supporting perfect adaptation and dosage control [12, 86, 88]. As noise and small changes in expression can direct cell fate [49, 79, 89, 90], syntax-based tuning offers a simple method to explore the stability of cell fate by perturbing native networks with levels of transgenes that vary in their mean and variability.

While our findings are consistent with both biophysical predictions of DNA-supercoiling-mediated feedback [10] and *in vitro* studies [91–94], we do not make direct supercoiling measurements [31, 47, 95, 96]. While the off-diagonal score in our maps of chromatin contact probability suggest the formation of plectonemes, direct measurements of supercoiling are challenging in living cells [22]. Genome-wide studies of supercoiling suggest that the twin-domain model of supercoiling describes

eukaryotic genomes [26, 47], with supercoiling affecting the performance of gene editing [97], bacterial chromosomal segregation [33], and nucleosome and structural maintenance protein placement and function [21]. Assays to measure supercoiling in living cells rely either on small molecule intercalators like psoralen that target negatively supercoiled DNA [31, 98, 99] or proteins that specifically bind to positively supercoiled DNA [26, 47]. Putatively, RCMC allows us to capture transcriptionally induced supercoils of either sign. Future work will be needed to combine region capture with assays of supercoiling to fully examine the hypothesis of DNA supercoiling-mediated feedback.

Engineered synthetic circuits are much more compact than native mammalian genomes and therefore are putatively more strongly affected by supercoiling. However, regulation of some native genes depend on adjacent gene expression. Tandem arrays of *Hox* genes show temporal activation during development that proceeds from downstream to upstream genes [100]. Potentially, supercoiling-mediated upstream dominance may reinforce other mechanisms driving the phenomenon of posterior dominance, which restricts the reactivation of downstream *Hox* genes. Transcription of non-coding RNAs tunes expression of native adjacent genes, both amplifying and attenuating expression of co-localized genes [49, 101]. Loss of transcription from an adjacent gene can induce significant developmental defects, suggesting adjacent transcription plays an essential role in tuning expression to control and coordinate cell fate [9, 49].

While supercoiling-mediated feedback offers a compelling explanation for the coupling of gene expression and syntax-specific profiles of intrinsic noise, alternative modes of regulation may influence two-gene circuits. For instance, collisions of RNA polymerases undergoing transcriptional readthrough may reduce expression in the convergent and tandem syntaxes [102]. However, rates of readthrough of single transgenes integrated into the genome are estimated to be low, representing less than one percent of polyadenylated transcripts [56], a possible underestimate if readthrough transcripts are less stable. Alternatively, methods of defining polymerase positions on single DNA fibers may resolve questions on readthrough and facilitated recruitment of RNAPII via supercoiling-mediated feedback [103]. Additionally, both transcription and supercoiling offer predictions for nucleosome positioning and histone modifications which may refine our understanding of the interlocking modes of gene regulation that control expression of synthetic and native genes.

Despite the clean, abstract way that synthetic circuits are often drawn, integration into the genome wraps synthetic circuitry in layers of native regulation. Biochemical interactions *and* biophysical forces combine to shape our genomes. By harnessing both layers of control, syntax and supercoiling-mediated feedback can enhance the predictability, performance, and functional range of engineered gene circuits.

## Acknowledgments

We would like to thank Mike Laub and Anders Hansen for their feedback. We thank Mary Ehmann and Maria Castellanos for reviewing the manuscript. Research reported in this manuscript was supported by the NIH NIGMS under award number R35-GM143033, by the National Science Foundation under the NSF-CAREER under award number 2339986, and with funding from Institute for Collaborative Biotechnologies (W911NF-19-2-0026), and by the Air Force Research Laboratory MURI (FA9550-22-1-03t16). K.S.L. is supported by the National Science Foundation Graduate Research Fellowship Program under grant No. 1745302. R.D.J. is supported by a Michael Smith Health Research BC trainee award (RT-2021-1946) and an MSL Pathway to Independence Award. This work was supported in part by the Koch Institute Support (core) Grant P30-CA14051 from the National Cancer Institute. Work in P.W.Z.'s lab was supported by the Stem Cell Network, the Canadian Institute for Health Research, and the Wellcome Leap Human Organs, Physiology, and Engineering (HOPE) program. P.W.Z. is the Canada Research Chair in Stem Cell Bioengineering. C.L.M., R.D. and A.B.A. were supported by Novo Nordisk Foundation grant (NNF21CC0073729; reNEW). B.J.D. is supported by the NIH (R01AI181684 and R21AI166396).

### *Author contributions*

Following the contributor role taxonomy, C.P.J. contributed to project conceptualization, methodology, validation, investigation/data acquisition, statistical analysis, writing (original draft, review, editing), and visualization. K.S.L. contributed to conceptualization and methodology (all-in-one circuits), validation, investigation/data acquisition, statistical analysis, writing (review, editing), and visualization. S.R.K. contributed to investigation/data acquisition (two-gene lentiviruses) and writing (review, editing). R.J. contributed to conceptualization and methodology (constitutive hiPSC circuits), data acquisition, and visualization. A.B.A. contributed to conceptualization and methodology (STRAIGHT-IN circuits), investigation/data acquisition, and writing (review). D.S.P. contributed to conceptualization and methodology (STRAIGHT-IN circuits), investigation/data acquisition. E.L.P. contributed to investigation/data acquisition (RNA-HCR-FISH) and writing (review, editing). R.L. contributed to investigation/data acquisition (constitutive PiggyBac cell lines) and writing (review). J.Y. contributed to investigation/data acquisition (constitutive hiPSC circuits). C.G.O. contributed to methodology (two-gene PiggyBac constructs) and writing (review). C.L.M. contributed to resources, writing (review), supervision, funding acquisition. R.P.D. contributed to resources, writing (review), supervision, funding acquisition. B.J.D. conceptualization and methodology (antibody experiments), resources, and writing (review). P.W.Z. contributed to writing (review), resources, supervision, project administration, and funding acquisition. K.E.G. contributed to conceptualization, methodology, validation, resources, writing (review, editing), supervision, project administration, and funding acquisition.

### *Declaration of interests*

None

### *Supplementary Materials*

- Materials and Methods
- Tables S1 to S5.
- Figs. S1 to S17.

## **References**

1. Lamouille, S., Xu, J. & Derynck, R. Molecular Mechanisms of Epithelial–Mesenchymal Transition. *Nature Reviews Molecular Cell Biology* **15**, 178–196. ISSN: 1471-0080 (Mar. 2014) (cit. on p. 1).
2. Ivashkiv, L. B. & Donlin, L. T. Regulation of Type I Interferon Responses. *Nature Reviews Immunology* **14**, 36–49. ISSN: 1474-1741 (Jan. 2014) (cit. on p. 1).
3. Mosig, R. A. *et al.* Natural Antisense Transcript of Period2, Per2AS, Regulates the Amplitude of the Mouse Circadian Clock. *Genes & Development* **35**, 899–913. ISSN: 0890-9369. pmid: [34016691](#) (June 2021) (cit. on p. 1).
4. Gherman, A., Wang, R. & Avramopoulos, D. Orientation, Distance, Regulation and Function of Neighbouring Genes. *Human Genomics* **3**, 143–156. ISSN: 1473-9542. pmid: [19164091](#) (Jan. 1, 2009) (cit. on p. 1).
5. Trinklein, N. D. *et al.* An Abundance of Bidirectional Promoters in the Human Genome. *Genome Research* **14**, 62–66. ISSN: 1088-9051, 1549-5469. pmid: [14707170](#) (Jan. 1, 2004) (cit. on p. 1).

6. Korbel, J. O., Jensen, L. J., von Mering, C. & Bork, P. Analysis of Genomic Context: Prediction of Functional Associations from Conserved Bidirectionally Transcribed Gene Pairs. *Nature Biotechnology* **22**, 911–917. ISSN: 1546-1696 (July 2004) (cit. on p. 1).
7. Darbellay, F. *et al.* The Constrained Architecture of Mammalian Hox Gene Clusters. *Proceedings of the National Academy of Sciences* **116**, 13424–13433. ISSN: 0027-8424, 1091-6490. pmid: [31209053](https://pubmed.ncbi.nlm.nih.gov/31209053/) (July 2, 2019) (cit. on p. 1).
8. Murphy, S. E. & Boettiger, A. N. Polycomb Repression of Hox Genes Involves Spatial Feedback but Not Domain Compaction or Phase Transition. *Nature Genetics* **56**, 493–504. ISSN: 1546-1718 (Mar. 2024) (cit. on p. 1).
9. Zinani, O. Q. H., Keseroğlu, K., Ay, A. & Özbudak, E. M. Pairing of Segmentation Clock Genes Drives Robust Pattern Formation. *Nature*, 1–6. ISSN: 1476-4687 (Dec. 23, 2020) (cit. on pp. 1, 13).
10. Johnstone, C. P. & Galloway, K. E. Supercoiling-Mediated Feedback Rapidly Couples and Tunes Transcription. *Cell Reports* **41**. ISSN: 2211-1247 (Oct. 18, 2022) (cit. on pp. 1–4, 10, 12, 42).
11. Li, H.-S. *et al.* Multidimensional Control of Therapeutic Human Cell Function with Synthetic Gene Circuits. *Science* **378**, 1227–1234 (Dec. 16, 2022) (cit. on pp. 2, 8, 12).
12. Love, K. S., Johnstone, C. P., Peterman, E. L., Gaglione, S. & Galloway, K. E. *Model-Guided Design of microRNA-based Gene Circuits Supports Precise Dosage of Transgenic Cargoes into Diverse Primary Cells* <https://www.biorxiv.org/content/10.1101/2024.06.25.600629v2> (2024). Pre-published (cit. on pp. 2, 12).
13. Cho, J. H. *et al.* Engineering Advanced Logic and Distributed Computing in Human CAR Immune Cells. *Nature Communications* **12**, 792. ISSN: 2041-1723 (Feb. 4, 2021) (cit. on p. 2).
14. Anzalone, A. V. *et al.* Search-and-Replace Genome Editing without Double-Strand Breaks or Donor DNA. *Nature* **576**, 149–157. ISSN: 1476-4687 (Dec. 2019) (cit. on p. 2).
15. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (Feb. 15, 2013) (cit. on p. 2).
16. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (Feb. 15, 2013) (cit. on p. 2).
17. Tousley, A. M. *et al.* Co-Opting Signalling Molecules Enables Logic-Gated Control of CAR T Cells. *Nature* **615**, 507–516. ISSN: 1476-4687 (Mar. 2023) (cit. on p. 2).
18. Allen, G. M. *et al.* Synthetic Cytokine Circuits That Drive T Cells into Immune-Excluded Tumors. *Science* **378**, eaba1624 (Dec. 16, 2022) (cit. on p. 2).
19. Gao, W. *et al.* Engineered T Cell Therapy for Central Nervous System Injury. *Nature* **634**, 693–701. ISSN: 1476-4687 (Oct. 2024) (cit. on p. 2).
20. Brahmachari, S., Tripathi, S., Onuchic, J. N. & Levine, H. Nucleosomes Play a Dual Role in Regulating Transcription Dynamics. *Proceedings of the National Academy of Sciences* **121**, e2319772121 (July 9, 2024) (cit. on p. 2).
21. Li, S., Vemuri, C. & Chen, C. DNA Topology: A Central Dynamic Coordinator in Chromatin Regulation. *Current Opinion in Structural Biology* **87**, 102868. ISSN: 1879-033X. pmid: [38878530](https://pubmed.ncbi.nlm.nih.gov/38878530/) (June 14, 2024) (cit. on pp. 2, 13).
22. Patel, H. P. *et al.* DNA Supercoiling Restricts the Transcriptional Bursting of Neighboring Eukaryotic Genes. *Molecular Cell* **83**, 1573–1587.e8. ISSN: 1097-2765. pmid: [37207624](https://pubmed.ncbi.nlm.nih.gov/37207624/) (May 18, 2023) (cit. on pp. 2, 12).

23. Ancona, M., Bentivoglio, A., Brackley, C. A., Gonnella, G. & Marenduzzo, D. Transcriptional Bursts in a Nonequilibrium Model for Gene Regulation by Supercoiling. *Biophysical Journal* **117**, 369–376. ISSN: 00063495 (July 2019) (cit. on p. 2).
24. Geng, Y. *et al.* A Spatially Resolved Stochastic Model Reveals the Role of Supercoiling in Transcription Regulation. *PLOS Computational Biology* **18**, e1009788. ISSN: 1553-7358 (Sept. 19, 2022) (cit. on p. 2).
25. Bhola, M. *et al.* RNA Interacts with Topoisomerase I to Adjust DNA Topology. *Molecular Cell* **84**, 3192–3208.e11. ISSN: 1097-2765. pmid: [39173639](https://pubmed.ncbi.nlm.nih.gov/39173639/) (Sept. 5, 2024) (cit. on p. 2).
26. Longo, G. M. C. *et al.* Type II Topoisomerases Shape Multi-Scale 3D Chromatin Folding in Regions of Positive Supercoils. *Molecular Cell*. ISSN: 1097-2765 (Oct. 31, 2024) (cit. on pp. 2, 13).
27. Lee, J. *et al.* Chromatinization Modulates Topoisomerase II Processivity <https://www.biorxiv.org/content/10.1101/2023.10.03.560726v1> (2023). Pre-published (cit. on p. 2).
28. Saha, S. & Pommier, Y. R-Loops, Type I Topoisomerases and Cancer. *NAR Cancer* **5**, zcad013. ISSN: 2632-8674 (Mar. 1, 2023) (cit. on p. 2).
29. Sutormin, D. *et al.* Interaction between Transcribing RNA Polymerase and Topoisomerase I Prevents R-loop Formation in *E. Coli*. *Nature Communications* **13**, 4524. ISSN: 2041-1723. pmid: [35927234](https://pubmed.ncbi.nlm.nih.gov/35927234/) (Aug. 4, 2022) (cit. on p. 2).
30. Le, T. T. *et al.* Synergistic Coordination of Chromatin Torsional Mechanics and Topoisomerase Activity. *Cell* **179**, 619–631.e15. ISSN: 0092-8674 (Oct. 17, 2019) (cit. on p. 2).
31. Yao, Q., Zhu, L., Shi, Z., Banerjee, S. & Chen, C. Topoisomerase-Modulated Genome-Wide DNA Supercoiling Domains Colocalize with Nuclear Compartments and Regulate Human Gene Expression. *Nature Structural & Molecular Biology*, 1–14. ISSN: 1545-9985 (Aug. 16, 2024) (cit. on pp. 2, 12, 13).
32. Corless, S. & Gilbert, N. Effects of DNA Supercoiling on Chromatin Architecture. *Biophysical Reviews* **8**, 245–258. ISSN: 1867-2469 (Sept. 1, 2016) (cit. on p. 2).
33. Martin-Gonzalez, A. *et al.* DNA Supercoiling Enhances DNA Condensation by ParB Proteins. *Nucleic Acids Research* **52**, 13255–13268. ISSN: 0305-1048 (Nov. 27, 2024) (cit. on pp. 2, 13).
34. Achar, Y. J., Adhil, M., Choudhary, R., Gilbert, N. & Foiani, M. Negative Supercoil at Gene Boundaries Modulates Gene Topology. *Nature* **577**, 701–705. ISSN: 1476-4687 (7792 Jan. 2020) (cit. on p. 2).
35. Naughton, C. *et al.* Transcription Forms and Remodels Supercoiling Domains Unfolding Large-Scale Chromatin Structures. *Nature Structural & Molecular Biology* **20**, 387–395. ISSN: 1545-9985. pmid: [23416946](https://pubmed.ncbi.nlm.nih.gov/23416946/) (3 Mar. 2013) (cit. on p. 2).
36. Ma, J. *et al.* Transcription Factor Regulation of RNA Polymerase’s Torque Generation Capacity. *Proceedings of the National Academy of Sciences* **116**, 2583–2588 (Feb. 12, 2019) (cit. on p. 2).
37. Qian, J. *et al.* Chromatin Buffers Torsional Stress During Transcription. *bioRxiv: The Preprint Server for Biology*, 2024.10.15.618270. ISSN: 2692-8205. pmid: [39464147](https://pubmed.ncbi.nlm.nih.gov/39464147/) (Oct. 18, 2024) (cit. on p. 2).
38. Tsao, Y.-P., Wu, H.-Y. & Liu, L. F. Transcription-Driven Supercoiling of DNA: Direct Biochemical Evidence from in Vitro Studies. *Cell* **56**, 111–118. ISSN: 0092-8674 (Jan. 13, 1989) (cit. on p. 2).
39. Janissen, R., Barth, R., Polinder, M., van der Torre, J. & Dekker, C. Single-Molecule Visualization of Twin-Supercoiled Domains Generated during Transcription. *Nucleic Acids Research* **52**, 1677–1687. ISSN: 0305-1048 (Feb. 28, 2024) (cit. on p. 2).



40. Boulas, I. *et al.* Assessing in Vivo the Impact of Gene Context on Transcription through DNA Supercoiling. *Nucleic Acids Research* **51**, 9509–9521. ISSN: 0305-1048 (Oct. 13, 2023) (cit. on p. 2).
41. Marko, J. F. Torque and Dynamics of Linking Number Relaxation in Stretched Supercoiled DNA. *Physical Review E* **76**, 021926. ISSN: 1539-3755, 1550-2376 (Aug. 29, 2007) (cit. on p. 2).
42. Sevier, S. A. & Levine, H. Mechanical Properties of Transcription. *Physical Review Letters* **118**, 268101 (June 27, 2017) (cit. on p. 2).
43. Hacker, W. C. & Elcock, A. H. Spotter: A Single-Nucleotide Resolution Stochastic Simulation Model of Supercoiling-Mediated Transcription and Translation in Prokaryotes. *Nucleic Acids Research*, gkad682. ISSN: 0305-1048 (Aug. 21, 2023) (cit. on p. 2).
44. Yeung, E. *et al.* Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks. *Cell Systems* **5**, 11–24.e12. ISSN: 2405-4712 (July 26, 2017) (cit. on p. 2).
45. Chong, S., Chen, C., Ge, H. & Xie, X. S. Mechanism of Transcriptional Bursting in Bacteria. *Cell* **158**, 314–326. ISSN: 0092-8674 (July 17, 2014) (cit. on p. 2).
46. Teves, S. S. & Henikoff, S. Transcription-Generated Torsional Stress Destabilizes Nucleosomes. *Nature structural & molecular biology* **21**, 88–94. ISSN: 1545-9993. pmid: 24317489 (1 Jan. 2014) (cit. on p. 2).
47. Guo, M. S., Kawamura, R., Littlehale, M. L., Marko, J. F. & Laub, M. T. High-Resolution, Genome-Wide Mapping of Positive Supercoiling in Chromosomes. *eLife* **10** (eds Berger, J. M. & Barkai, N.) e67236. ISSN: 2050-084X (July 19, 2021) (cit. on pp. 2, 6, 12, 13).
48. Goel, V. Y., Huseyin, M. K. & Hansen, A. S. Region Capture Micro-C Reveals Coalescence of Enhancers and Promoters into Nested Microcompartments. *Nature Genetics* **55**, 1048–1056. ISSN: 1546-1718 (6 June 2023) (cit. on pp. 2, 6, 24).
49. Ganesh, V. S. *et al.* Novel Syndromic Neurodevelopmental Disorder Caused by de Novo Deletion of CHASERR, a Long Noncoding RNA <https://www.medrxiv.org/content/10.1101/2024.01.31.24301497v1> (2024). Pre-published (cit. on pp. 3, 12, 13).
50. Rom, A. *et al.* Regulation of CHD2 Expression by the Chaserr Long Noncoding RNA Gene Is Essential for Viability. *Nature Communications* **10**, 5092. ISSN: 2041-1723 (Nov. 8, 2019) (cit. on p. 3).
51. Martens, J. A., Laprade, L. & Winston, F. Intergenic Transcription Is Required to Repress the *Saccharomyces Cerevisiae* SER3 Gene. *Nature* **429**, 571–574. ISSN: 1476-4687 (June 2004) (cit. on p. 3).
52. Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611–1627. ISSN: 0092-8674, 1097-4172. pmid: 26686651 (Dec. 17, 2015) (cit. on p. 4).
53. Liu, M. *et al.* Genomic Discovery of Potent Chromatin Insulators for Human Gene Therapy. *Nature Biotechnology* **33**, 198–203. ISSN: 1546-1696 (Feb. 2015) (cit. on p. 4).
54. Lu, X.-B., Guo, Y.-H. & Huang, W. Characterization of the cHS4 Insulator in Mouse Embryonic Stem Cells. *FEBS open bio* **10**, 644–656. ISSN: 2211-5463. pmid: 32087050 (Apr. 2020) (cit. on p. 4).
55. Calvo-Roitberg, E. *et al.* mRNA Initiation and Termination Are Spatially Coordinated <https://www.biorxiv.org/content/10.1101/2024.01.05.574404v1> (2024). Pre-published (cit. on p. 4).

56. Peterman, E. L. *et al.* *High-Resolution Profiling Reveals Coupled Transcriptional and Translational Regulation of Transgenes* <https://www.biorxiv.org/content/10.1101/2024.11.26.625483v2> (2024). Pre-published (cit. on pp. 4, 8, 13, 23, 27).
57. Choi, H. M. T., Beck, V. A. & Pierce, N. A. Next-Generation in Situ Hybridization Chain Reaction: Higher Gain, Lower Cost, Greater Durability. *ACS Nano* **8**, 4284–4294. ISSN: 1936-0851 (May 27, 2014) (cit. on p. 4).
58. Björkegren, C. & Baranello, L. DNA Supercoiling, Topoisomerases, and Cohesin: Partners in Regulating Chromatin Architecture? *International Journal of Molecular Sciences* **19**, 884. ISSN: 1422-0067 (3 Mar. 2018) (cit. on p. 6).
59. Fu, Z., Guo, M. S., Zhou, W. & Xiao, J. Differential Roles of Positive and Negative Supercoiling in Organizing the E. Coli Genome. *Nucleic Acids Research* **52**, 724–737. ISSN: 0305-1048 (Jan. 25, 2024) (cit. on p. 6).
60. Blanch-Asensio, A. *et al.* *STRAIGHT-IN Dual: A Platform for Dual, Single-Copy Integrations of DNA Payloads and Gene Circuits into Human Induced Pluripotent Stem Cells* <https://www.biorxiv.org/content/10.1101/2024.10.17.616637v1> (2024). Pre-published (cit. on pp. 6, 10, 12, 27, 28).
61. Cerbini, T. *et al.* Transcription Activator-like Effector Nuclease (TALEN)-Mediated CRYBL Targeting Enables Enhanced Transgene Expression and One-Step Generation of Dual Reporter Human Induced Pluripotent Stem Cell (iPSC) and Neural Stem Cell (NSC) Lines. *PLoS One* **10**, e0116032. ISSN: 1932-6203. pmid: 25587899 (2015) (cit. on p. 6).
62. Open2C *et al.* Cooltools: Enabling High-Resolution Hi-C Analysis in Python. *PLOS Computational Biology* **20**, e1012067. ISSN: 1553-7358 (May 6, 2024) (cit. on p. 6).
63. Zufferey, M., Tavernari, D., Oricchio, E. & Ciriello, G. Comparison of Computational Methods for the Identification of Topologically Associating Domains. *Genome Biology* **19**, 217. ISSN: 1474-760X (Dec. 10, 2018) (cit. on p. 6).
64. Narducci, D. N. & Hansen, A. S. Reeling It in: How DNA Topology Drives Loop Extrusion by Condensin. *Nature Structural & Molecular Biology* **29**, 623–625. ISSN: 1545-9985 (July 2022) (cit. on p. 6).
65. Guo, Y. *et al.* Chromatin Jets Define the Properties of Cohesin-Driven in Vivo Loop Extrusion. *Molecular Cell* **82**, 3769–3780.e5. ISSN: 1097-2765. pmid: 36182691 (Oct. 20, 2022) (cit. on p. 6).
66. Gam, J. J., DiAndreth, B., Jones, R. D., Huh, J. & Weiss, R. A ‘Poly-Transfection’ Method for Rapid, One-Pot Characterization and Optimization of Genetic Systems. *Nucleic Acids Research* **47**, e106. ISSN: 0305-1048. pmid: 31372658 (Oct. 10, 2019) (cit. on p. 8).
67. O’Connell, R. W. *et al.* *Ultra-High Throughput Mapping of Genetic Design Space* <https://www.biorxiv.org/content/10.1101/2023.03.16.532704v1> (2023). Pre-published (cit. on pp. 8, 12).
68. Schlatter, S. *et al.* On the Optimal Ratio of Heavy to Light Chain Genes for Efficient Recombinant Antibody Production by CHO Cells. *Biotechnology Progress* **21**, 122–133. ISSN: 1520-6033 (2005) (cit. on p. 8).
69. Hooft van Huijsduijnen, R. *et al.* Reassessing Therapeutic Antibodies for Neglected and Tropical Diseases. *PLoS Neglected Tropical Diseases* **14**, e0007860. ISSN: 1935-2727. pmid: 31999695 (Jan. 30, 2020) (cit. on p. 8).

70. Carver, J. *et al.* Maximizing Antibody Production in a Targeted Integration Host by Optimization of Subunit Gene Dosage and Position. *Biotechnology Progress* **36**, e2967. ISSN: 1520-6033 (2020) (cit. on p. 8).
71. Lee, Z., Wan, J., Shen, A. & Barnard, G. Gene Copy Number, Gene Configuration and LC/HC mRNA Ratio Impact on Antibody Productivity and Product Quality in Targeted Integration CHO Cell Lines. *Biotechnology Progress* **40**, e3433. ISSN: 1520-6033 (2024) (cit. on p. 8).
72. Ho, S. C. L. *et al.* Control of IgG LC:HC Ratio in Stably Transfected CHO Cells and Study of the Impact on Expression, Aggregation, Glycosylation and Conformational Stability. *Journal of Biotechnology* **165**, 157–166. ISSN: 0168-1656 (June 10, 2013) (cit. on p. 8).
73. Ascic, E. *et al.* In Vivo Dendritic Cell Reprogramming for Cancer Immunotherapy. *Science* **386**, eadn9083 (Sept. 5, 2024) (cit. on p. 8).
74. Chen, Z. *et al.* A Synthetic Protein-Level Neural Network in Mammalian Cells. *Science* **386**, 1243–1250 (Dec. 13, 2024) (cit. on p. 8).
75. Galloway, K. & Johnstone, C. Bringing Neural Networks to Life. *Science* **386**, 1225–1226 (Dec. 13, 2024) (cit. on p. 8).
76. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183–1186. ISSN: 0036-8075, 1095-9203. pmid: [12183631](https://pubmed.ncbi.nlm.nih.gov/12183631/) (Aug. 16, 2002) (cit. on p. 10).
77. Swaffer, M. P. *et al.* RNA Polymerase II Dynamics and mRNA Stability Feedback Scale mRNA Amounts with Cell Size. *Cell*. ISSN: 0092-8674 (Nov. 8, 2023) (cit. on p. 10).
78. Cabrera, A. *et al.* The Sound of Silence: Transgene Silencing in Mammalian Cell Engineering. *Cell Systems* **13**, 950–973. ISSN: 2405-4712 (Dec. 21, 2022) (cit. on p. 10).
79. Desai, R. V. *et al.* A DNA-repair Pathway Can Affect Transcriptional Noise to Promote Cell Fate Transitions. *Science*. ISSN: 0036-8075, 1095-9203. pmid: [34301855](https://pubmed.ncbi.nlm.nih.gov/34301855/) (July 22, 2021) (cit. on pp. 10, 12).
80. Johnstone, C. P. & Galloway, K. E. Engineering Cellular Symphonies out of Transcriptional Noise. *Nature Reviews Molecular Cell Biology*. ISSN: 1471-0080 (Mar. 15, 2021) (cit. on p. 10).
81. Tycko, J. *et al.* High-Throughput Discovery and Characterization of Human Transcriptional Effectors. *Cell* **183**, 2020–2035.e16. ISSN: 0092-8674 (Dec. 23, 2020) (cit. on p. 12).
82. Oesinghaus, L., Castillo-Hair, S., Ludwig, N., Keller, A. & Seelig, G. *Quantitative Design of Cell Type-Specific mRNA Stability from microRNA Expression Data* <https://www.biorxiv.org/content/10.1101/2024.10.28.620728v1> (2025). Pre-published (cit. on p. 12).
83. Beal, K. M. *et al.* The Impact of Expression Vector Position on Transgene Transcription Allows for Rational Expression Vector Design in a Targeted Integration System. *Biotechnology Journal* **18**, 2300038. ISSN: 1860-7314 (2023) (cit. on p. 12).
84. Fry, L. E. *et al.* Promoter Orientation within an AAV-CRISPR Vector Affects Cas9 Expression and Gene Editing Efficiency. *The CRISPR Journal* **3**, 276–283. ISSN: 2573-1599. pmid: [32833533](https://pubmed.ncbi.nlm.nih.gov/32833533/) (Aug. 1, 2020) (cit. on p. 12).
85. Castel, B., Tomlinson, L., Locci, F., Yang, Y. & Jones, J. D. G. Optimization of T-DNA Architecture for Cas9-mediated Mutagenesis in Arabidopsis. *PLOS ONE* **14**, e0204778. ISSN: 1932-6203 (Jan. 9, 2019) (cit. on p. 12).

86. Du, R., Flynn, M. J., Honsa, M., Jungmann, R. & Elowitz, M. B. *miRNA Circuit Modules for Precise, Tunable Control of Gene Expression* <https://www.biorxiv.org/content/10.1101/2024.03.12.583048v1> (2024). Pre-published (cit. on p. 12).
87. Kabaria, S. R. *et al.* *Programmable Promoter Editing for Precise Control of Transgene Expression* <https://www.biorxiv.org/content/10.1101/2024.06.19.599813v2> (2025). Pre-published (cit. on p. 12).
88. Aoki, S. K. *et al.* A Universal Biomolecular Integral Feedback Controller for Robust Perfect Adaptation. *Nature* **570**, 533–537. ISSN: 1476-4687 (June 2019) (cit. on p. 12).
89. Wang, N. B. *et al.* *Proliferation History and Transcription Factor Levels Drive Direct Conversion* <https://www.biorxiv.org/content/10.1101/2023.11.26.568736v1> (2023). Pre-published (cit. on pp. 12, 26).
90. Hansen, M. M. K. *et al.* A Post-Transcriptional Feedback Mechanism for Noise Suppression and Fate Stabilization. *Cell* **173**, 1609–1621.e15. ISSN: 0092-8674 (June 14, 2018) (cit. on p. 12).
91. Revyakin, A., Ebright, R. H. & Strick, T. R. Promoter Unwinding and Promoter Clearance by RNA Polymerase: Detection by Single-Molecule DNA Nanomanipulation. *Proceedings of the National Academy of Sciences* **101**, 4776–4780. ISSN: 0027-8424, 1091-6490. pmid: [15037753](https://pubmed.ncbi.nlm.nih.gov/15037753/) (Apr. 6, 2004) (cit. on p. 12).
92. Vanderlinden, W., Skoruppa, E., Kolbeck, P. J., Carlon, E. & Lipfert, J. DNA Fluctuations Reveal the Size and Dynamics of Topological Domains. *PNAS Nexus* **1**, pgac268. ISSN: 2752-6542 (Nov. 1, 2022) (cit. on p. 12).
93. Kim, E., Gonzalez, A. M., Pradhan, B., van der Torre, J. & Dekker, C. Condensin-Driven Loop Extrusion on Supercoiled DNA. *Nature Structural & Molecular Biology*. ISSN: 1545-9985. pmid: [35835864](https://pubmed.ncbi.nlm.nih.gov/35835864/) (July 14, 2022) (cit. on p. 12).
94. Gao, X., Hong, Y., Ye, F., Inman, J. T. & Wang, M. D. Torsional Stiffness of Extended and Plectonemic DNA. *Physical Review Letters* **127**, 028101 (July 7, 2021) (cit. on p. 12).
95. Corless, S., Naughton, C. & Gilbert, N. Profiling DNA Supercoiling Domains in Vivo. *Genomics Data* **2**, 264–267. ISSN: 2213-5960 (Dec. 1, 2014) (cit. on p. 12).
96. Corless, S. & Gilbert, N. Investigating DNA Supercoiling in Eukaryotic Genomes. *Briefings in Functional Genomics* **16**, 379–389. ISSN: 2041-2649 (Nov. 1, 2017) (cit. on p. 12).
97. Newton, M. D. *et al.* Negative DNA Supercoiling Induces Genome-Wide Cas9 off-Target Activity. *Molecular Cell* **83**, 3533–3545.e5. ISSN: 1097-2765 (Oct. 5, 2023) (cit. on p. 13).
98. Visser, B. J. & Bates, D. in *Bacterial Chromatin: Methods and Protocols* (ed Dame, R. T.) 147–156 (Springer US, New York, NY, 2024). ISBN: 978-1-0716-3930-6 (cit. on p. 13).
99. Kouzine, F. *et al.* Transcription-Dependent Dynamic Supercoiling Is a Short-Range Genomic Force. *Nature Structural & Molecular Biology* **20**, 396–403. ISSN: 1545-9985 (3 Mar. 2013) (cit. on p. 13).
100. Mark, M., Rijli, F. M. & Chambon, P. Homeobox Genes in Embryogenesis and Pathogenesis. *Pediatric Research* **42**, 421–429. ISSN: 1530-0447 (Oct. 1997) (cit. on p. 13).
101. Engreitz, J. M. *et al.* Local Regulation of Gene Expression by lncRNA Promoters, Transcription and Splicing. *Nature* **539**, 452–455. ISSN: 1476-4687 (7629 Nov. 2016) (cit. on p. 13).
102. Shearwin, K. E., Callen, B. P. & Egan, J. B. Transcriptional Interference – a Crash Course. *Trends in Genetics* **21**, 339–345. ISSN: 0168-9525 (June 1, 2005) (cit. on p. 13).

103. Swanson, E. G. *et al.* *Deaminase-Assisted Single-Molecule and Single-Cell Chromatin Fiber Sequencing* <https://www.biorxiv.org/content/10.1101/2024.11.06.622310v1> (2025). Pre-published (cit. on p. 13).
104. Yang, T. *et al.* HiCRep: Assessing the Reproducibility of Hi-C Data Using a Stratum-Adjusted Correlation Coefficient. *Genome Research* **27**, 1939–1949. ISSN: 1549-5469. pmid: [28855260](https://pubmed.ncbi.nlm.nih.gov/28855260/) (Nov. 2017) (cit. on p. 25).
105. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A Hyperactive piggyBac Transposase for Mammalian Applications. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1531–1536. ISSN: 0027-8424. pmid: [21205896](https://pubmed.ncbi.nlm.nih.gov/21205896/) (Jan. 25, 2011) (cit. on p. 26).
106. Blanch-Asensio, A., Grandela, C., Mummery, C. L. & Davis, R. P. STRAIGHT-IN: A Platform for Rapidly Generating Panels of Genetically Modified Human Pluripotent Stem Cell Lines. *Nature Protocols*, 1–44. ISSN: 1750-2799 (Aug. 23, 2024) (cit. on p. 28).

# Supplemental information

## Materials and Methods

### Bioinformatic analysis of intergene spacing across organisms

Genome annotations for *Saccharomyces cerevisiae*, *Mus musculus*, *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Danio rerio* were downloaded from ENSEMBL, Release 110 (July 2023). Annotations with the type gene, ncRNA\_gene, and pseudogene were selected and gene pair information (intergene spacing and orientation) was calculated for all gene pairs and gene trios. For fig. S1d, a null hypothesis distribution for gene trios was generated by randomly selecting gene pairs with intergene spacings and orientations ( $o_l, x_l$ ) and ( $o_r, x_r$ ) and treating this as a gene trio with the given left and right spacings and orientations.

### Cloning of genetic constructs

All plasmids were constructed using a combination of scarless NEB HiFi assembly and an in-house Golden Gate cloning scheme that allows for facile multi-level assembly of gene cassettes and multi-cassette plasmids. The fluorescent proteins used for each construct are listed in table S1. All construct designs and source plasmid identifiers for every figure panel are listed in table S2. Plasmid descriptions are found in tables S3 to S5. Plasmids maps for all plasmids are included as supplemental files.

### Derivation of cell lines with matrix of constitutive promoters

Freshly passaged HEK293T cells maintained in DMEM + 10% FBS were seeded at 160k cells per well onto 0.1% gelatin-coated 24-well plates. The following day (0 dpt), cells were transfected with a 4:1 mass ratio of donor plasmid to PiggyBac supertransposase, with a total transfected plasmid mass of 450 ng. Donor plasmids express two genes: one mRuby2 transcript, and one PuroR-T2A-mGreenLantern transcript. At 1 day post transfection (dpt), cells were media changed into fresh media containing 1  $\mu$ g / mL puromycin. Selection was maintained for two days (until 3 dpt), and surviving cells were passaged onto 6-well plates. After around a week of outgrowth (until each cell line was ~30% confluent), 1  $\mu$ g / mL puromycin was maintained for around two days, until no visible surviving cells in an untransfected control well were observed.

One day prior to sorting, single color control plasmids were transfected into fresh HEK293T cells for use in compensation. A Sony MA-900 was used to polyclonally sort cells. The majority of cell lines were double-positive sorted, using gates that exclude an untransfected control. Cell lines with extremely low expression (PGK-PGK) were sorted in a single-positive manner, where a cell was included if it was positive for mRuby2 or mGreenLantern. Cells were sorted into Pen-Strep supplemented conditioned media, containing 50% 0.2 micron-filtered media from a confluent plate of HEK293T cells and 50% fresh DMEM + 10% FBS. The resulting cells were outgrown and tested negative for mycoplasma.

### Generation of monoclonal 293T cell lines

Freshly passaged HEK293T cells were seeded at 20k cells per well onto 0.1% gelatin-coated 96-well plates. The following day, cells were transfected with a total of 340 ng of DNA, with a 1:1:1 mass ratio of PiggyBac supertransposase : two-gene tandem rtTA and PuroR donor : two-gene donor. On 1 dpt, 2 dpt, and 3 dpt, cells were media changed into fresh media containing 1  $\mu$ g / mL puromycin. On 4 dpt, cells were passaged to 24-well scale and left to outgrow in media without puromycin. On 14 dpt, cells were passaged onto a 6-well plate. On 17 dpt, cells were passaged at high dilution to a new 6-well plate and maintained for the next ten days in dox-containing media.

One day prior to sorting, single color control plasmids were transfected for compensation. A Sony MA-900 was used to monoclonally sort double-positive cells, using gates that exclude an untransfected control. Specifically, individual cells were sorted into wells of a fresh 96-well plate with Pen-Strep supplemented conditioned media as described in the previous sections. Cells were allowed to outgrow over a period of

two weeks. The resulting monoclonal cell lines tested negative for mycoplasma.

#### **Characterization of fluorescent protein expression in monoclonal lines**

30k cells per well of each monoclonal cell line were seeded onto 0.1% gelatin coated 96-well plates. The next day, media was changed into DMEM + 10% FBS containing dox concentrations from zero to 316 ng / mL dox. Two days later, cells were prepared for flow cytometry.

#### **Characterization of mRNA profile in monoclonal lines**

150k cells per well of each monoclonal line were seeded, onto 0.1% gelatin coated 12-well plates. The next day, media was changed into media, with half the wells supplemented with 316 ng / mL dox. Two days later, cells were resuspended in PBS for HCR RNA-FISH.

Here, we use Molecular Instruments probe sets for TagBFP (compatible with B1 amplifiers conjugated to Alexa Fluor™ 647) and mRuby2 (compatible with B2 amplifiers conjugated to Alexa Fluor™ 514). We used the optimized FISH protocol as described in Peterman *et al.* [56]. Briefly, after suspension in PBS, cells were transferred to 96-well v-bottom plate for HCR Flow-FISH. After each resuspension, spins were performed at 500 rcf for 5 minutes with default settings, unless otherwise noted. Cells were first fixed through incubation in 4% PFA for 15 minutes at room temperature. After spinning, cells were then permeabilized using 0.5% Tween-20 for 15 minutes at room temperature. Next, cells were spun and resuspended in hybridization buffer for 30 minutes at 37°C. During this incubation, probe set stock solution was diluted in hybridization buffer to a concentration of 14 nM for transfected cells or 28 nM for integrated cell lines. Cells were spun and resuspended in this probe solution for incubation overnight at 37°C. Due to the viscosity of the hybridization buffer, these spins were performed with reduced deceleration speed to minimize cell loss.

Following hybridization, cells were spun and resuspended in wash buffer for 15 minutes at 37°C. Then, cells were spun and resuspended in 5X SSCT for five minutes at room temperature. After these washes, cells were spun and resuspended in amplification buffer for 30 minutes at room temperature. Amplifier solution was prepared by combining separately snap-cooled hairpins h1 and h2 at a concentration of 130 nM in amplification buffer. Cells were spun and then incubated with this amplifier solution overnight at room temperature.

Following amplification, cells were spun and resuspended in 5X SSCT for one 30 minute incubation and one five minute incubation at room temperature. Finally, cells were spun and resuspended in PBS for flow cytometry.

#### **Dox titration series and time course induction**

The monoclonal cell lines as described above were seeded at low confluency, 20k cells per well, onto 0.1% gelatin coated 24-well plates, in DMEM + 10% FBS containing 316 ng/mL dox. Independent wells were seeded for every experimental day. On each subsequent day from day 1 to day 13, cells were disassociated with trypsin and flowed as described in the Flow Cytometry section. Cells were media changed into media without dox at day 4. At day 7, the remaining relatively confluent cells were passaged onto fresh 24-well plates, at 12.5k cells per well, in media containing dox. At day 10, cells were media changed into media without dox.

#### **Derivation of integrated rtTA cell lines**

Freshly passaged HEK293T cells were seeded at 180k cells per well onto 0.1% gelatin coated 24-well plates. The following day (0 dpt), cells were transfected with a total of 90 ng of PiggyBac supertransposase and 225 ng of plasmid 37, a plasmid encoding for rtTA, SNAP-tag, and a zeocin resistance gene. The following day (1 dpt), cells were split onto a 6-well plate and media changed into media containing 1000  $\mu$ g / mL zeocin. The following day (2 dpt), the selection media was replaced. On 3 dpt, cells were media changed into fresh media not containing zeocin. On 4 dpt, cells were passaged to a fresh 6-well plate and allowed to outgrow, then were frozen down. Later, the selected lines were unfrozen, allowed a passage to recover and were SNAP-stained immediately prior to sort (at 500:1 dilution). A Sony MA-900 was used

to monoclonally sort SNAP-positive cells into individual wells of a 96-well plate containing Pen-Strep supplemented conditioned media. Six monoclonal lines were outgrown and tested negative for mycoplasma.

For each of these lines, a test plasmid containing a TRE-inducible gene was transfected into each line alongside doxycycline. Across the six monoclonal lines, we selected the monoclonal with a medium level of tight expression, as measured from the inducible gene. This monoclonal became our integrated rtTA line.

### **Lentivirus production**

To produce all lentiviral vectors, HEK293T Lenti-X cells (Takara 632180) were seeded at 7.0 to 7.5 million cells per 10cm dish coated with 0.1% gelatin. The following day, each plate was transfected with 6  $\mu$ g transfer plasmid, 6  $\mu$ g packaging plasmid (psPAX2, Addgene #12260), and 12  $\mu$ g envelope plasmid (VSVG, Addgene #12259). Six hours later, each 10cm dish was media changed into 6.5 mL of HEPES-buffered (25 mM HEPES at pH 7.0) DMEM + 10% FBS. The following day, this media was collected, and fresh buffered media was added. The following day, these two aliquots were combined, filtered through a 0.45  $\mu$ m PES filter, and combined with Lenti-X Concentrator (Takara 631232) overnight and concentrated following manufacturer instructions.

### **Transduction of tandem syntax lentivirus**

Freshly passaged HEK293T cells were seeded at 15k cells per well onto 0.1% gelatin coated 96-well plates. The following day, cells were media changed into media containing 5  $\mu$ g/mL polybrene (hexadimethrine bromide, Sigma-Aldrich, H9268-5G) and a serial dilution over 3.5 orders of magnitude (highest concentration: 5.0  $\mu$ L concentrated virus per well). After two days, cells were disassociated and flowed. From the titration, representative concentrations were chosen with an MOI of approximately 0.5 for plotting in fig. S3c.

### **Region Capture Micro-C**

Three capture regions of interest, each roughly 0.5 Mb around three common landing pad integration sites were selected: *CLYBL* (chr13, 99,422,000-100,031,000), *AAVS* (chr19, 54,810,000-55,715,000), and *Rogi2* (chr3, 22,542,000-22,964,000). A custom probe panel of 80-mer probes targeting these regions and all synthetic parts inserted into these loci was designed and ordered from Twist Bioscience. Probes were selected based on a medium-stringency filter to reduce off-target pulldown, with some key areas near native TAD boundaries included using a low-stringency filter.

The homozygous hiPSC lines were expanded and outgrown at 6-well scale, with each of the eight conditions (two orientations, two induction conditions, and two bioreplicates) outgrown on an entire six-well plate. Cells were disassociated with Gentle Cell Dissociation Reagent (STEMCELL Technologies, 100-1077) according to the manufacturer's instructions and counted. From each of the eight conditions, 15M cells were collected and treated as described in Goel *et al.* [48]. An additional 25M cells were collected and for the MNase titration. Briefly, cells were washed in PBS and crosslinked using DSG (ThermoFisher 20593) and formaldehyde. After quenching, cells were washed, resuspended in PBS and counted, then resuspended in Micro-C buffer #1. Cell counts per condition varied between 6M and 13M cells.

A key variable in Micro-C is the ratio of MNase to cells. To identify ideal digestion concentrations, an MNase (Worthington Biochem LS004798) titration was performed on the separated aliquot. Increasing amounts of MNase were added and samples were purified and gel-separated. The optimal amount of MNase digests to primarily mononucleosomal fragments, with few but visible dinucleosomal and trinucleosomal bands. Using the optimal ratio, samples were digested with MNase, end-repaired, blunted, and labeled with biotinylated nucleotides. The biotin-labeled chromatin was then proximity ligated overnight. After enzymatic cleanup steps, the chromatin was then reverse crosslinked overnight. Dinucleosomes were selected using a gel extraction and T1 Streptavidin beads (Invitrogen 65601) were used to purify labeled fragments.

In order to determine the minimum number of cycles required to reach the required library concen-



tration while minimizing PCR duplicates, a test amplification was performed with a small aliquot of each sample for a range of cycle numbers. The resulting PCRs were run on an agarose gel and quantified. Using this quantification, each sample was amplified to reach a target mass of 200 ng. NEB Multiplex Oligos for Illumina Primer Set 1 (NEB E7335) and NEBNext® High-Fidelity 2X PCR Master Mix (NEB M0541) were used for all PCRs, and sample barcodes were selected following the NEB recommendations. The resulting libraries were purified using AmPure XP beads (Beckman Coulter A63880) quantified via Bioanalyzer and qPCR using the NEBNext Library Quant kit for Illumina (NEB E7630).

The resulting libraries were mixed in equimass proportions and region-captured following Twist Bioscience's Standard Hybridization Target Enrichment Protocol, with the modification that a test PCR was used using the same reagents as above to identify an appropriate number of amplification cycles. The resulting region-capture libraries were purified and quantified via Bioanalyzer and qPCR. The pooled library was sequenced via paired-end 2x75 cycle sequencing using Element's AVITI Cloudbreak Freestyle flowcell (e.g. direct Illumina anchor sequencing without index conversion).

### **RCMC bioinformatics**

For each integrated orientation, custom human genomes were generated by editing the GRCh38 genome build to contain the inserted synthetic construct at the *CLYBL* landing pad. Paired-end reads were aligned to these genomes using bowtie2 using the `-very-sensitive-local` preset. Then, pairtools was used to identify Hi-C pairs (with `-walks-policy mask` and `-min-mapq 2`) and deduplicate them (with `-max-mismatch 1`). The resulting reads were converted to .mcool format using cooltools. The resulting matrices were ICE balanced (within each capture region for the majority of this work, and globally for fig. S9b) and binned into bins at both the 500 bp and 2000 bp resolution.

To evaluate the reproducibility of this method, the stratum-corrected correlation coefficient as implemented in Yang *et al.* [104] was used to evaluate the similarity of all samples, both within the capture region of interest fig. S9a and in distal regions fig. S9c. Based on the high reproducibility observed, reads from the two bioreplicates were merged for the rest of this work.

Using the balanced contact matrices, fold-changes in contact probability upon induction are calculated by dividing these matrices by each other. The sliding-window insulation score in figs. S11a and S11b is calculated using the `insulation` function in cooltools and uses a sliding window of 50 kb. The off-diagonal score presented in figs. 3f and 3g sums the contact probability along the off-diagonal up to a given distance, centered at a given genomic coordinate. For the off-diagonal score, we use the 500 bp binned matrices and sum bins within 5 kb of the target location.

### **Lentiviral transduction of two-gene cassettes**

To calculate lentiviral titer, a two-fold serial dilution of the concentrated lentivirus was combined with 5  $\mu\text{g}/\text{mL}$  polybrene (hexadimethrine bromide, Sigma-Aldrich H9268) and 20,000 HEK293T cells per well in a 0.1% gelatin-coated 96-well plate. The following day, cells were changed into fresh DMEM + 10% FBS. Two days later, cells were flowed. The fraction of expressing cells in each dilution condition was used to compute viral titer, assuming a Poisson process for infection. Viral titers were then used to calculate the volume of concentrated virus needed to infect cells at the desired MOI.

On the day of transduction, 5k HEK293T cells in suspension were co-infected with virus expressing the two-gene cassette of different syntax at MOI of 0.08, and 3  $\mu\text{L}$  of lentivirus expressing rtTA-P2A-mGreenLantern at high MOI. Cells were plated into a 96-well with DMEM + 10% FBS containing 5  $\mu\text{g}/\text{mL}$  polybrene (hexadimethrine bromide, Sigma-Aldrich H9268). The following day, media was replaced with fresh DMEM + 10% FBS. At three days post infection, cells were passaged into 6-well plates. Once confluent, cells were flow-sorted on the Sony-MA900 for double-positive constitutive reporter gene and rtTA-mGreenLantern. The polyclonal sorted lines were then re-plated post-sorting and passaged consistently to maintain cells at 80% confluence or lower.

For evaluating effect of doxycycline addition, the post-sorted, integrated HEK293T cell lines were

re-plated into a 96-well plate at 39k cells/well with DMEM + 10% FBS. The day after plating, media was replaced with DMEM + 10% FBS contained various concentration of doxycycline from 1  $\mu\text{g}/\text{mL}$  to 0  $\mu\text{g}/\text{mL}$ . Cells were flowed at three days post doxycycline treatment. Bioreplicates represent different passages of the post-sorted, integrated cell lines.

#### **Human iPSC culture and transfection**

iPS11 cells (Alstem, episomal HFF-derived) were maintained in mTeSR1, mTeSR Plus, or eTeSR media (StemCell Technologies) supplemented with 0.5% penicillin/streptomycin (GIBCO). Cells were grown in normoxia conditions (20% O<sub>2</sub>, 37°C, 5% CO<sub>2</sub>) on tissue culture-treated plastic plates coated for 30-60 min with Geltrex (Gibco) or Cultrex (Bio-Techne). For routine passaging, iPSCs were dissociated to small clusters by (1) incubating in TrypLE Express (Gibco) for 2-4 minutes at 37°C, followed by quenching with mTeSR or eTeSR and light pipetting, or (2) incubating in ReLeSR (STEMCELL Tech) according to manufacturer's instructions. On thaw or passage with TrypLE, cells were treated with 5  $\mu\text{M}$  ROCK inhibitor (ROCKi) Y-27632 (STEMCELL Tech).

For transfections, iPSCs were collected and seeded at 75-100 x 10<sup>3</sup> cells/mL in mTeSR1, mTeSR Plus, or eTeSR supplemented with 5  $\mu\text{M}$  ROCKi Y-27632. The following day, cells were fed with fresh media and transfected with 80 ng (96w), 480 ng (24w), or 960 ng (12w) DNA complexed at a 3:1 ratio of  $\mu\text{g}$  DNA: $\mu\text{L}$  Lipofectamine Stem Reagent (ThermoFisher STEM00015). The DNA mixtures typically comprised 6:1:1:1 ratio (by mass) of plasmids respectively encoding the PiggyBac transposon, hyperactive PiggyBac transposase (hyPBBase [105]), Puromycin resistance gene (pac), and constitutive transfection marker (mK02). On days 1 and 2 after transfection, cells were fed with fresh media supplemented with 0.5  $\mu\text{g}/\text{mL}$  Puromycin (Sigma P8833) to select for transfected cells transiently expressing pac from a non-integrating plasmid, thereby enriching for cells with PiggyBac integrations up to 90% purity. On subsequent days, cells were fed and passaged normally without Puro, and subsampled during passages for flow cytometry.

#### **Lentiviral transduction of the engineered all-in-one inducible circuits**

To calculate lentiviral titer, a two-fold serial dilution of the concentrated lentivirus was combined with 5  $\mu\text{g}/\text{mL}$  polybrene (hexadimethrine bromide, Sigma-Aldrich H9268) and 20k HEK293T cells per well in a 0.1% gelatin-coated 96-well plate. The following day, cells were changed into fresh DMEM + 10% FBS containing 1  $\mu\text{g}/\text{mL}$  doxycycline. Two days later, cells were flowed. The fraction of expressing cells in each dilution condition was used to compute viral titer, assuming a Poisson process for infection. Viral titers for iPS11 cells and mouse embryonic fibroblasts (MEFs) were estimated as 2-fold greater or four-fold lower than for HEK293T cells. Viral titers were then used to calculate the volume of concentrated virus needed to infect cells at an MOI of 0.3.

To transduce HEK293T cells, 20k cells were combined with 5  $\mu\text{g}/\text{mL}$  polybrene and the calculated amount of concentrated virus. The next day, cells were changed into fresh DMEM + 10% FBS with or without 300 ng/mL dox. Two days later, cells were flowed.

To transduce iPS11 cells, cells were dissociated with Gentle Cell Dissociation Reagent according to manufacturer's instructions, then seeded at 15k cells/well in a Geltrex-coated 96-well plate with mTeSR-Plus and 5  $\mu\text{M}$  ROCKi. The next day, the media was replaced with fresh mTeSR-Plus containing the calculated amount of concentrated virus and 5  $\mu\text{g}/\text{mL}$  polybrene. Plates were then spun at 1500 x g for 90 minutes. The following day, media was changed to fresh mTeSR-Plus with or without 300 ng/mL dox. Two days later, cells were imaged and flowed.

To transduce mouse embryonic fibroblasts (MEFs), passage 1 primary MEFs—isolated as described in Wang *et al.* [89]—were thawed and allowed to recover for 1-2 days in DMEM + 10% FBS. Cells were then dissociated with Trypsin+EDTA and seeded at 10k cells per well in a 0.1% gelatin-coated 96-well plate. The following day, the media was replaced with fresh DMEM + 10% FBS containing the calculated amount of concentrated virus and 5  $\mu\text{g}/\text{mL}$  polybrene. Plates were then spun at 1500 x g for 90 minutes.

The following day, media was changed to fresh DMEM + 10% FBS with or without 300 ng/mL dox. Two days later, cells were flowed. Biological replicates include cells from multiple independent isolations.

#### **Generation of antibody producer cell lines**

To integrate two-gene antibody production cassettes, we used a Bxb1-mediated landing pad (LP) line with attP receptor site at R<sub>ogi2</sub> in HEK293Ts for parallel integration of each cassette Peterman *et al.* [56]. The landing pad functions analogously to the STRAIGHT-IN landing pad Blanch-Asensio *et al.* [60]. The landing pad contains a truncated puromycin resistance gene missing the promoter and start codon. Upon Bxb1-mediated recombination between the attB site on the donor plasmid and attP site in the LP, an EF1a promoter and start codon is placed in-frame of the resistance gene, conferring recombinant cells resistance to puromycin.

To integrate the donor plasmids, the LP HEK293T line was seeded at 100k cells per well onto 0.1% gelatin coated 24-well plates. The following day, the cells were transfected using 450 ng of two-gene donor plasmid and 300 ng of CAG-Bxb1 (gift from the Wong Lab at Boston University). At 1dpt, the cells were media changed. At 2 dpt, the cells were passaged to 0.1% gelatin coated 6-well plates. At 3 dpt, the cells were media changed into media containing 1  $\mu$ g / mL puromycin. Selection was maintained until an untransfected well was fully selected against (around 1.5 weeks). Once confluent (five to six days of puromycin administration), cells were passaged at a split ratio of 1:10 to dilute out residual donor plasmid, at which point cells were ready for use in downstream analyses.

#### **Titering of antibody production yields**

After outgrowth and passaging of the antibody producer lines, triplicates were seeded on 6-well plates at 500k cells per 6-well. The day after, cells were media changed into 2 mL of fresh DMEM + 10% FBS. After six days of outgrowth, 1.1 mL of supernatant was collected from every condition and centrifuged at 10k g for 10 minutes to remove cell debris. The top 1.0 mL of clarified supernatant was transferred to Pierce 35 kDa PES protein concentrator columns (Thermo Scientific 88502). Each sample was centrifuged to a final volume less than 75  $\mu$ L. Each concentrated sample was diluted to 75  $\mu$ L with pH 7.4 PBS.

The resulting concentrated samples were processed following manufacturer instructions using a Human IgG (Total) ELISA kit (Thermo Scientific BMS2091) and a Easy-Titer Human IgG (H+L) kit (Thermo Scientific 23310). Absorbance was measured on a Tecan Infinite M1000 Pro at 450nm (ELISA) and 340nm (Easy-Titer).

#### **Flow cytometry**

HBSS supplemented with 2% FBS (iPS11) or PBS (others) was used as Flow buffer. Cells were washed once with PBS, then dissociated with Trypsin+EDTA (HEK293T, MEFs), TrypLE Express (iPS11) or Gentle Cell Dissociation Reagent (iPS11). After dissociation, cells were quenched with Flow buffer or media, then spun down for 5 min @ 300-500 x g. Cells were resuspended in Flow buffer and acquired on CytoFLEX LX N3-V5-B3-Y5-R3-I0 (Beckman Coulter) or Attune NxT Flow Cytometer (Thermo Fisher). Data were gated on FlowJo software (v10.X; BD Biosciences) or CytExpert (v2.6; Beckman Coulter), then exported for plotting and additional analysis.

#### **Generation of homozygous hiPSC lines integrated into *CLYBL* loci using STRAIGHT-IN Dual platform**

We used the STRAIGHT-IN Dual platform to integrate the all-in-one inducible circuits at both alleles of *CLYBL* [60]. A single copy of the divergent and downstream tandem syntaxes for the all-in-one inducible circuits was integrated into the GT allele of *CLYBL*. Briefly, for the GT allele, 600 ng of the donor plasmids, Bxb1-GT\_AIO-TetOn\_mScarlet\_Divergent and Bxb1-GT\_AIO-TetOn\_mScarlet\_Down-Tandem (Addgene, #229794) were transfected along with 400 ng of Bxb1-expressing plasmid (Addgene, #51271) into 1x10<sup>5</sup> STRAIGHT-IN Dual hiPSCs using Lipofectamine Stem Reagent (ThermoFisher). A second copy of the all-in one inducible circuits was integrated into the GA allele of *CLYBL*. For this, the Bxb1-GA\_AIO-

TetOn\_mScarlet\_Divergent (Addgene # 229791) and Bxb1-GA\_AIO-TetOn\_mScarlet\_Down-Tandem (Addgene # 229792) donor plasmids were transfected following the STRAIGHT-IN Dual procedure [60]. The selection and excision of the lines was performed using the of the STRAIGHT-IN protocol described in detail here [106].

## A Supplemental figures and tables

Figure	Reporter / Constitutive	Adjacent / Inducible
fig. 1c (left), figs. S3a and S3b	mGreenLantern (mGL)	
fig. 1c (right), fig. S2	TagBFP	mNeonGreen
figs. 2, 4b to 4d and S16c	mRuby2	TagBFP
figs. 5, S15, S16a, S16b and S17	mGreenLantern (mGL)	mRuby2

**Table S1:** Mapping between described genes and fluorophores

Figure	Cell type	Plasmids	Integration	Selection information
fig. 1c, left	293T	1 (UT), 2 (DT)	PiggyBac	Puro, polyclonal sort
fig. 1c, right	iPS11	3, 4	PiggyBac	Transient puro
figs. 2 and S6	293T	5 + {6 (DT), 7 (C), 8 (D)}	PiggyBac	Puro, mono-clonal sort
figs. 3b to 3g and S8 to S11	STRAIGHT-IN hiPSC	9, 10 (DT); 11, 12 (D)	Bxb1 LP	STRAIGHT-IN
figs. 4a and S12	Rogi2 LP 293T	13 (UT), 14 (DT), 15 (C), 16 (D)	Bxb1 LP	Puro
figs. 4b to 4d, S13 and S16c	293T	17/18 + 19 (UT), 20 (DT), 21 (C), 22 (D)	Lentivirus	Polyclonal sort
fig. 5	iPS11	23 (UT), 24 (DT), 25 (Div)	Lentivirus	None
fig. S2, downstream	iPS11	44 (UT), 45 (DT)	PiggyBac	Transient puro
fig. S2, upstream	iPS11	46 (UT), 47 (DT)	PiggyBac	Transient puro
fig. S2, full	iPS11	48 (UT), 49 (DT)	PiggyBac	Transient puro
figs. S3a and S3b, EFS-EFS	293T	29 (UT), 30 (DT)	PiggyBac	Puro, polyclonal sort
figs. S3a and S3b, EF1a-EF1a	293T	31 (UT), 32 (DT)	PiggyBac	Puro, polyclonal sort
figs. S3a and S3b, EF1a-PGK	293T	33 (UT), 34 (DT)	PiggyBac	Puro, polyclonal sort
figs. S3a and S3b, PGK-EF1a	293T	35 (UT), 36 (DT)	PiggyBac	Puro, polyclonal sort
fig. S3c	293T	26, 27, 28	Lentivirus	None
fig. S4	iPS11	50 - 73	PiggyBac	Transient puro
fig. S5, Div	iPS11	50, 74, 75	PiggyBac	Transient puro
fig. S5, UT	iPS11	55, 76, 77	PiggyBac	Transient puro
fig. S5, DT	iPS11	3, 78, 79	PiggyBac	Transient puro
figs. S15, S16a and S16b	293T, MEF	25 (Div), 24 (DT), 23 (UT)	Lentivirus	None
fig. S17, no insulator	iPS11, 293T, MEF	25 (D), 24 (DT), 23 (UT)	Lentivirus	None
fig. S17, cHS4 core	iPS11, 293T, MEF	38 (D), 39 (DT), 40 (UT)	Lentivirus	None
fig. S17, cHS4 full	iPS11, 293T, MEF	41 (D), 42 (DT), 43 (UT)	Lentivirus	None

**Table S2:** Constructs, cell type, and integration method is listed for every data panel.

Ref	Identifier	Orientation	Construct
1	pTA292	U Tan	PGK-Puro-T2A-mGL-SV40 PGK-mRuby2-bGH
2	pTA289	D Tan	PGK-mRuby2-bGH PGK-Puro-T2A-mGL-SV40
3	aRJ134	D Tan	PGK-mNeonGreen-syn_pa EF1a-TagBFP-syn_pa
4	aRJ135	U Tan	EF1a-TagBFP-syn_pa PGK-mNeonGreen-syn_pa
5	pTA187	Tan	CMV-rtTA-bGH CMV-PuroR-bGH
6	pTA129	D Tan	TRE-TagBFP-bGH EF1a-mRuby2-SV40
7	pTA130	Con	TRE-TagBFP-bGH (EF1a-mRuby2-SV40)
8	pTA131	Div	(TRE-TagBFP-bGH) EF1a-mRuby2-SV40
9	pABA03	D Tan	GA allele, TRE3G-mScarlet-WPRE CAG-rtTA-T2A-mTagBFP2-bGH
10	pABA04	D Tan	GT allele, TRE3G-mScarlet-WPRE CAG-rtTA-T2A-mTagBFP2-bGH
11	pABA01	Div	GA allele, (TRE3G-mScarlet-WPRE) CAG-rtTA-T2A-mTagBFP2-bGH
12	pABA02	Div	GT allele, (TRE3G-mScarlet-WPRE) CAG-rtTA-T2A-mTagBFP2-bGH
13	pTA515	U Tan	CMV-mAb17HC-bGH CMV-mAb17LC-P2A-mRuby2-bGH
14	pTA516	D Tan	CMV-mAb17LC-P2A-mRuby2-bGH CMV-mAb17HC
15	pTA517	Con	CMV-mAb17HC-bGH (CMV-mAb17LC-P2A-mRuby2-bGH)
16	pTA518	Div	(CMV-mAb17HC-bGH) CMV-mAb17LC-P2A-mRuby2-bGH
17	pKG2752	n/a	LX1, EFS-rtTA-P2A-mGL-WPRE
18	pTA169	n/a	LX1, CMV-rtTA-bGH
19	pTA444	U Tan	LX1, EF1a-mRuby2-bGH TRE-TagBFP-bGH
20	pTA443	D Tan	LX1, TRE-TagBFP-bGH EF1a-mRuby2-bGH
21	pTA445	Con	LX1, EF1a-mRuby2-bGH (TRE-TagBFP-bGH)
22	pTA446	Div	LX1, (TRE-TagBFP-bGH) EF1a-mRuby2-bGH
23	pKG3549	U Tan	EFS-rtTA-P2A-mGL-WPRE TRE-mRuby2-bGH
24	pKG1484	D Tan	TRE-mRuby2-bGH EFS-rtTA-P2A-mGL-WPRE
25	pKG1487	Div	(TRE-mRuby2-bGH) EFS-rtTA-P2A-mGL-WPRE

**Table S3:** Plasmids used in main-text figures. Text in parentheses indicates a construct placed in the antisense direction.

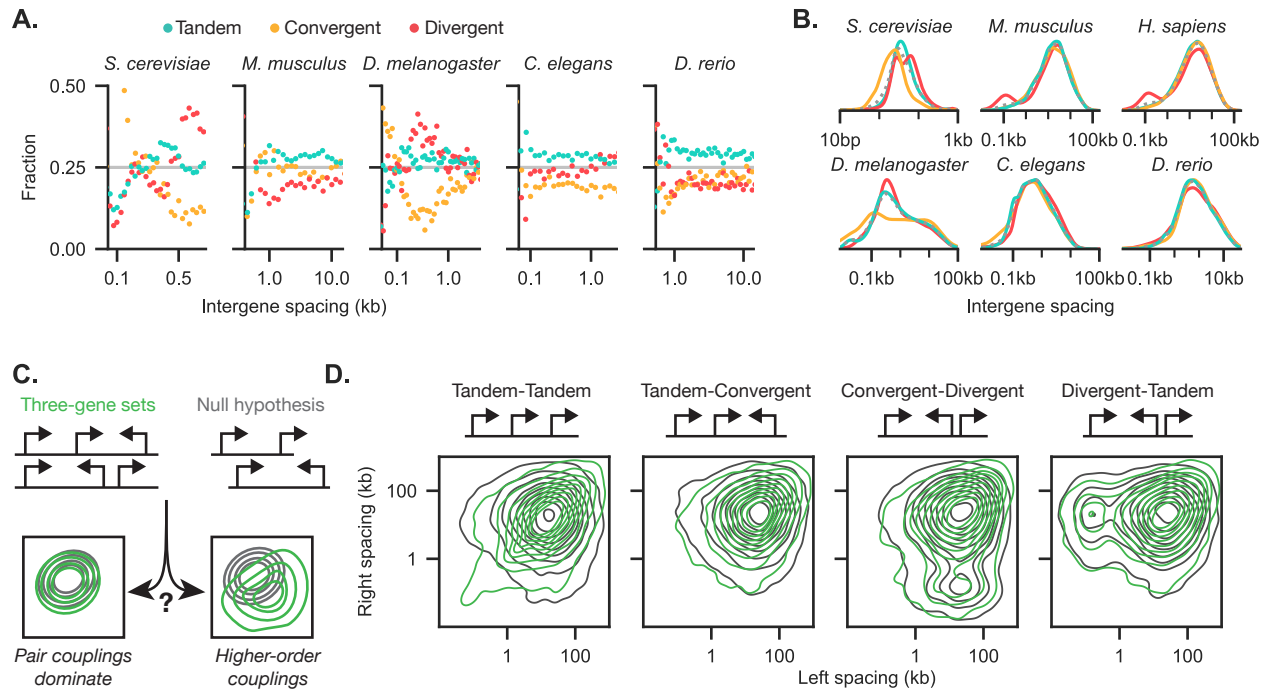
Ref	Identifier	Orientation	Construct
26	pGEEC270	D Tan	LX1, PGK-SNAPtag-bGH EF1a-mRuby2-WPRE
27	pGEEC272	D Tan	LX1, UbC-SNAPtag-bGH EF1a-mRuby2-WPRE
28	pGEEC276	n/a	LX1, EF1a-mRuby2-WPRE
29	pTA291	U Tan	EFS-Puro-T2A-mGL-SV40 EFS-mRuby2-bGH
30	pTA288	D Tan	EFS-mRuby2-bGH EFS-Puro-T2A-mGL-SV40
31	pTA293	U Tan	EF1a-Puro-T2A-mGL-SV40 EF1a-mRuby2-bGH
32	pTA290	D Tan	EF1a-mRuby2-bGH EF1a-Puro-T2A-mGL-SV40
33	pTA345	U Tan	PGK-Puro-T2A-mGL-SV40 EF1a-mRuby2-bGH
34	pTA342	D Tan	EF1a-mRuby2-bGH PGK-Puro-T2A-mGL-SV40
35	pTA344	U Tan	EF1a-Puro-T2A-mGL-SV40 PGK-mRuby2-bGH
36	pTA343	D Tan	PGK-mRuby2-bGH EF1a-Puro-T2A-mGL-SV40
37	pTA309	n/a	UbC-rtTA-P2A-SNAPtag-T2A-ZeoR-bGH
38	pKG3572	Div	(TRE-mRuby2-bGH) EFS-rtTA-P2A-mGL-WPRE
39	pKG3571	D Tan	TRE-mRuby2-bGH cHS4.core EFS-rtTA-P2A-mGL-WPRE
40	pKG3574	U Tan	EFS-rtTA-P2A-mGL-WPRE cHS4.core TRE-mRuby2-bGH
41	pKG3576	Div	(TRE-mRuby2-bGH) cHS4.full EFS-rtTA-P2A-mGL-WPRE
42	pKG3575	D Tan	TRE-mRuby2-bGH cHS4.full EFS-rtTA-P2A-mGL-WPRE
43	pKG3578	U Tan	EFS-rtTA-P2A-mGL-WPRE cHS4.full TRE-mRuby2-bGH

**Table S4:** Plasmids used in supplemental figures. Text in parentheses indicates a segment placed in the antisense direction.

Ref	Identifier	Orientation	Construct
44	aRJ137	U Tan	EF1a-TagBFP-Syn_pA D1-insulator PGK-mNG-Syn_pA E1-insulator
45	aRJ136	D Tan	PGK-mNG-Syn_pA D1-insulator EF1a-TagBFP-Syn_pA E1-insulator
46	aRJ133	U Tan	D1-insulator EF1a-TagBFP-Syn_pA E1-Insulator PGK-mNG-Syn_pA
47	aRJ132	D Tan	D1-insulator PGK-mNG-Syn_pA E1-insulator EF1a-TagBFP-Syn_pA
48	aRJ139	U Tan	D1-insulator EF1a-TagBFP-Syn_pA PGK-mNG-Syn_pA E1-insulator
49	aRJ138	D Tan	D1-insulator PGK-mNG-Syn_pA EF1a-TagBFP-Syn_pA E1-insulator
50	aRJ210	Div	(PGK-mNeonGreen-Syn_pA) EF1a-TagBFP-bGH
51	aRJ211	Tan	PGK-mNeonGreen-Syn_pA PGK-TagBFP-Syn_pA
52	aRJ212	Tan	PGK-mNeonGreen-Syn_pA UbC-TagBFP-Syn_pA
53	aRJ213	Tan	PGK-mNeonGreen-Syn_pA CAG-TagBFP-Syn_pA
54	aRJ214	Tan	EF1a-mNeonGreen-Syn_pA EF1a-TagBFP-Syn_pA
55	aRJ215	Tan	EF1a-mNeonGreen-Syn_pA PGK-TagBFP-Syn_pA
56	aRJ216	Tan	EF1a-mNeonGreen-Syn_pA UbC-TagBFP-Syn_pA
57	aRJ217	Tan	EF1a-mNeonGreen-Syn_pA CAG-TagBFP-Syn_pA
58	aRJ218	Tan	UbC-mNeonGreen-Syn_pA EF1a-TagBFP-Syn_pA
59	aRJ219	Tan	UbC-mNeonGreen-Syn_pA PGK-TagBFP-Syn_pA
60	aRJ220	Tan	UbC-mNeonGreen-Syn_pA UbC-TagBFP-Syn_pA
61	aRJ221	Tan	UbC-mNeonGreen-Syn_pA CAG-TagBFP-Syn_pA
62	aRJ222	Tan	CAG-mNeonGreen-Syn_pA EF1a-TagBFP-Syn_pA
63	aRJ223	Tan	CAG-mNeonGreen-Syn_pA PGK-TagBFP-Syn_pA
64	aRJ224	Tan	CAG-mNeonGreen-Syn_pA UbC-TagBFP-Syn_pA
65	aRJ225	Tan	CAG-mNeonGreen-Syn_pA CAG-TagBFP-Syn_pA
66	aRJ242	Tan	PGK-mNeonGreen-Syn_pA Inert-TagBFP-Syn_pA
67	aRJ243	Tan	EF1a-mNeonGreen-Syn_pA Inert-TagBFP-Syn_pA
68	aRJ244	Tan	UbC-mNeonGreen-Syn_pA Inert-TagBFP-Syn_pA
69	aRJ245	Tan	CAG-mNeonGreen-Syn_pA Inert-TagBFP-Syn_pA
70	aRJ246	Tan	Inert-mNeonGreen-Syn_pA PGK-TagBFP-Syn_pA
71	aRJ247	Tan	Inert-mNeonGreen-Syn_pA EF1a-TagBFP-Syn_pA
72	aRJ248	Tan	Inert-mNeonGreen-Syn_pA UbC-TagBFP-Syn_pA
73	aRJ249	Tan	Inert-mNeonGreen-Syn_pA CAG-TagBFP-Syn_pA
74	aRJ228	Div	(pGK-mNeonGreen-Syn_pA) EF1a-TagBFP-bGH
75	aRJ229	Div	(pGK-mNeonGreen-bGH) EF1a-TagBFP-Syn_pA
76	aRJ230	U Div	EF1a-mNeonGreen-Syn_pA PGK-TagBFP-bGH
77	aRJ231	U Div	EF1a-mNeonGreen-bGH PGK-TagBFP-Syn_pA
78	aRJ232	D Div	PGK-mNeonGreen-Syn_pA EF1a-TagBFP-bGH
79	aRJ233	D Div	PGK-mNeonGreen-bGH EF1a-TagBFP-Syn_pA

**Table S5:** Plasmids used in iPSC supplemental figures. Text in parentheses indicates a segment placed in the antisense direction.





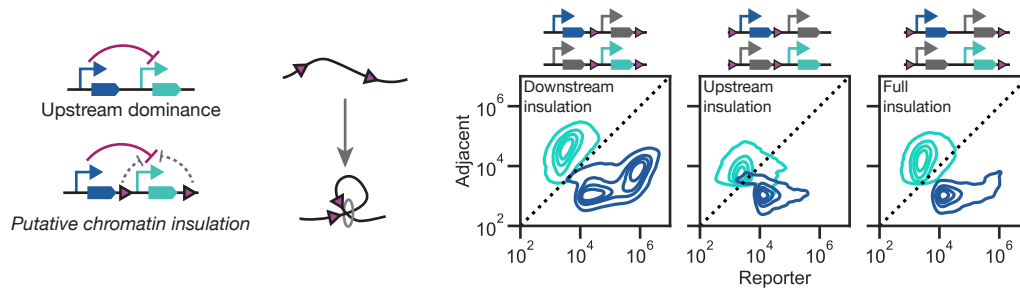
**Figure S1:** Bioinformatic analysis of gene pairs and trios in eukaryotic genomes.

a) Orientation distributions for gene pairs binned by intergene spacing are shown for *Saccharomyces cerevisiae*, *Mus musculus*, *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Danio rerio*.

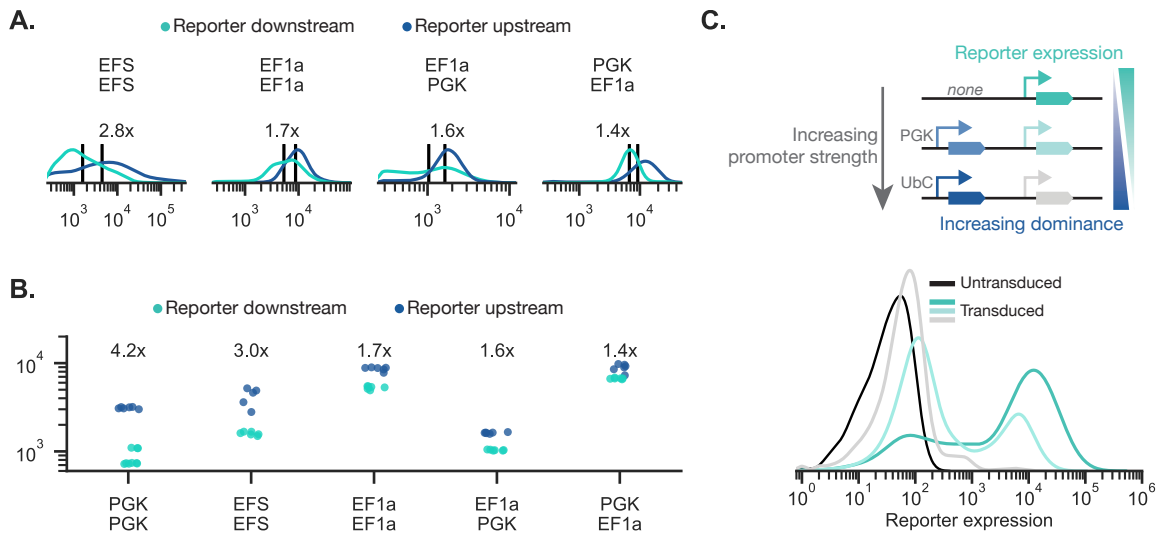
b) Instead of binning all gene pairs by intergene spacing as in fig. 1b, gene pairs were separated by orientation and the intergene spacing distributions were compared to the overall intergene distribution (gray, dashed).

c) Higher order couplings between trios of genes can be evaluated by comparing the joint distribution of the left and right intergene-spacing for all pairs of genes, compared to a null hypothesis where gene trios are synthetically created by combining gene pairs.

d) For the human genome, three-gene pairs can be split into four orientations. The distribution of intergene spacings are compared to the null hypothesis where gene pairs are randomly combined to form three-gene trios. The overlap of the distributions indicates that trios were largely explained by the pairs.



**Figure S2:** Flanking CTCF sites may insulate a chromatin region from its surroundings. Tandem-oriented CTCF binding sites were introduced flanking the upstream gene, the downstream gene, or the entire construct. Joint distributions of the two genes are shown.

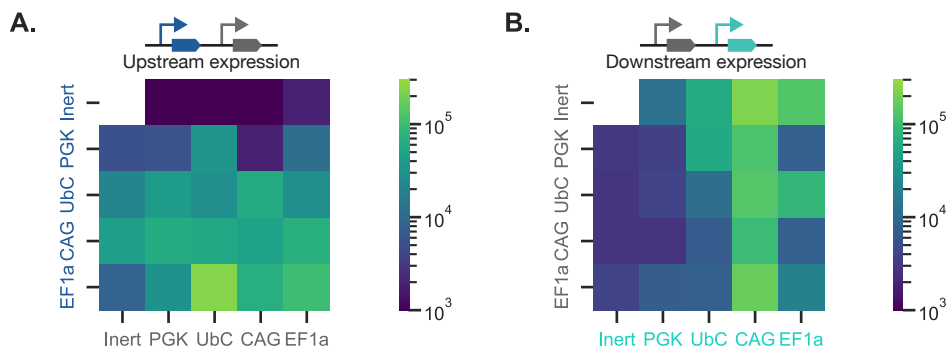


**Figure S3:** Two-gene constitutive gene pairs show upstream dominance.

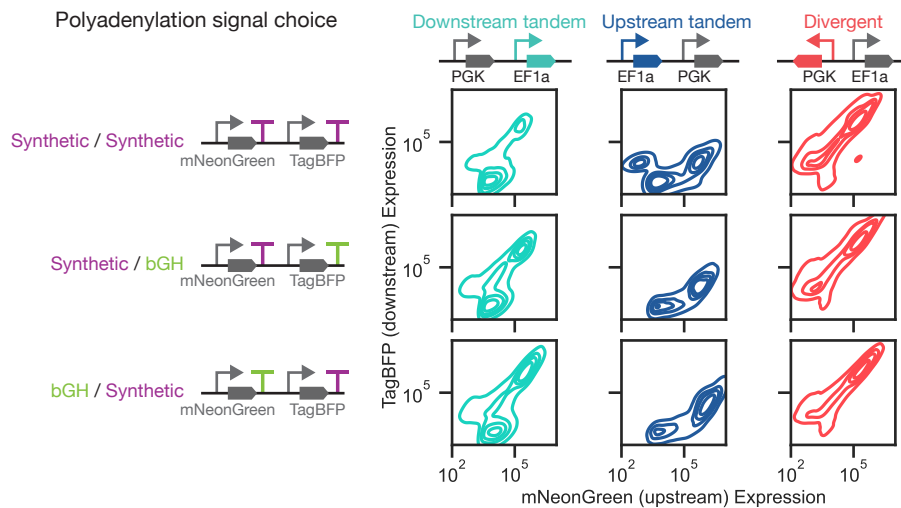
a) Representative distributions of a reporter gene in the upstream and downstream positions for different combinations of constitutive promoters PiggyBac-integrated into HEK293T cells. All promoter pairs demonstrate upstream dominance.

b) The geometric mean of the reporter expression for each biological replicate (N=5), with the median upstream dominance fold change shown above.

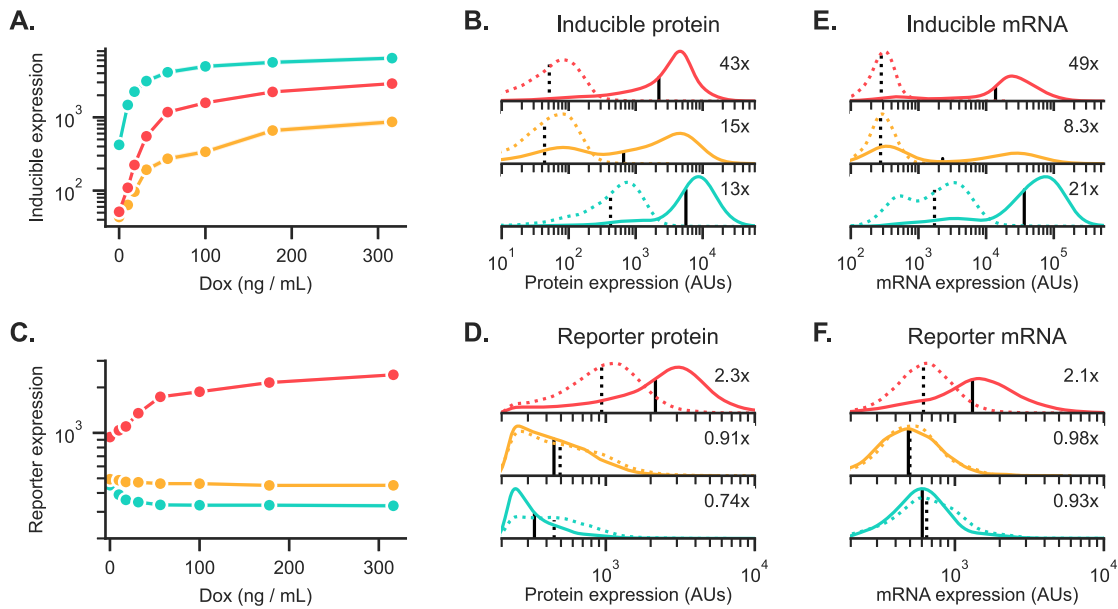
c) Three lentiviruses with the same constant downstream reporter but different upstream genes were transduced into HEK293T cells. Stronger upstream expression reduces downstream reporter expression.



**Figure S4:** Expression patterns of tandem two-gene constructs PiggyBac-integrated into hiPSC lines. A panel of constitutive promoters were placed in both the upstream and downstream positions and integrated into hiPSCs. Average of measurements from three biological replicates, (N=3).  
 a) Upstream expression for every tested promoter combination.  
 b) Downstream expression for every tested promoter combination.

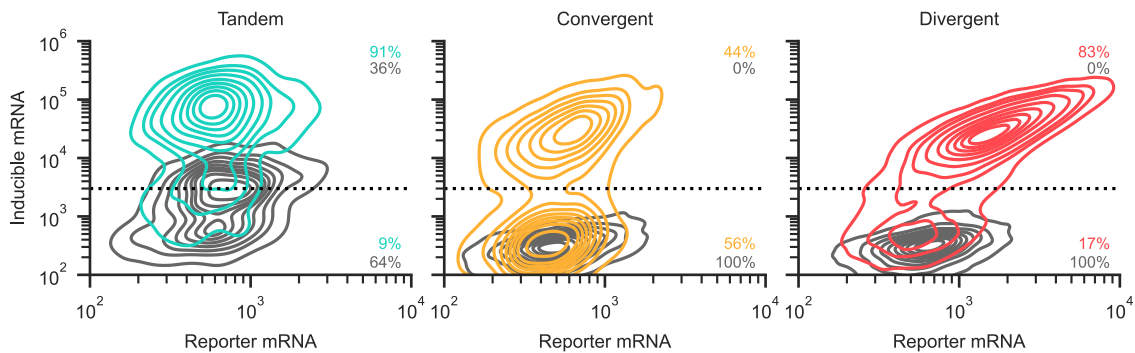


**Figure S5:** For three different combinations of polyadenylation signals, representative joint distributions for PiggyBac-integrated iPSCs are shown. The PAS signal choice only minimally affects the resulting distributions for each orientation. Each column shows similar joint distributions.

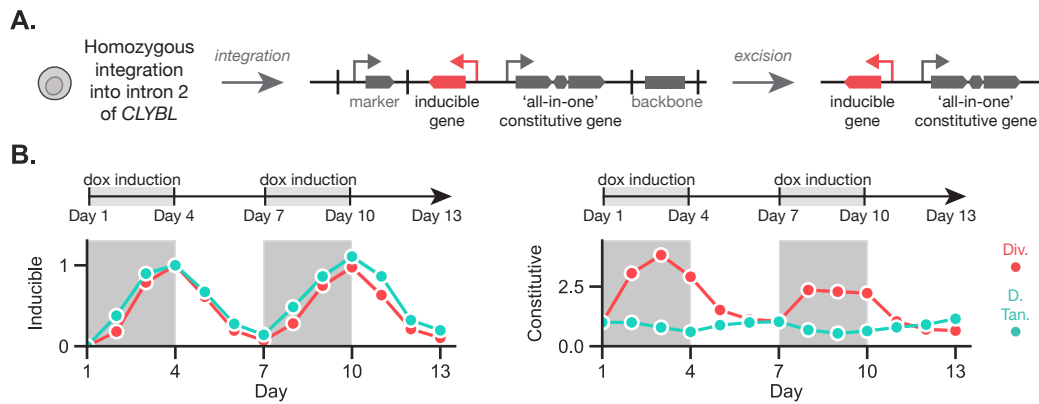


**Figure S6:** Circuit orientation affects both mRNA and protein levels.

- a) The geometric mean of the un-normalized inducible gene expression is shown as a function of dox induction. The geometric mean and 95% confidence interval are calculated over three merged wells.
- b) Representative protein distributions are shown for the inducible gene in both the uninduced (dashed) and induced (solid) cases.
- c) The geometric mean of the un-normalized reporter expression is shown as a function of dox induction. The geometric mean and 95% confidence interval are calculated over three merged wells.
- d) Representative protein distributions are shown for the reporter gene in both the uninduced (dashed) and induced (solid) cases.
- e),f) Representative mRNA distributions are shown for the inducible and reporter genes in both the uninduced (dashed) and induced (solid) cases.



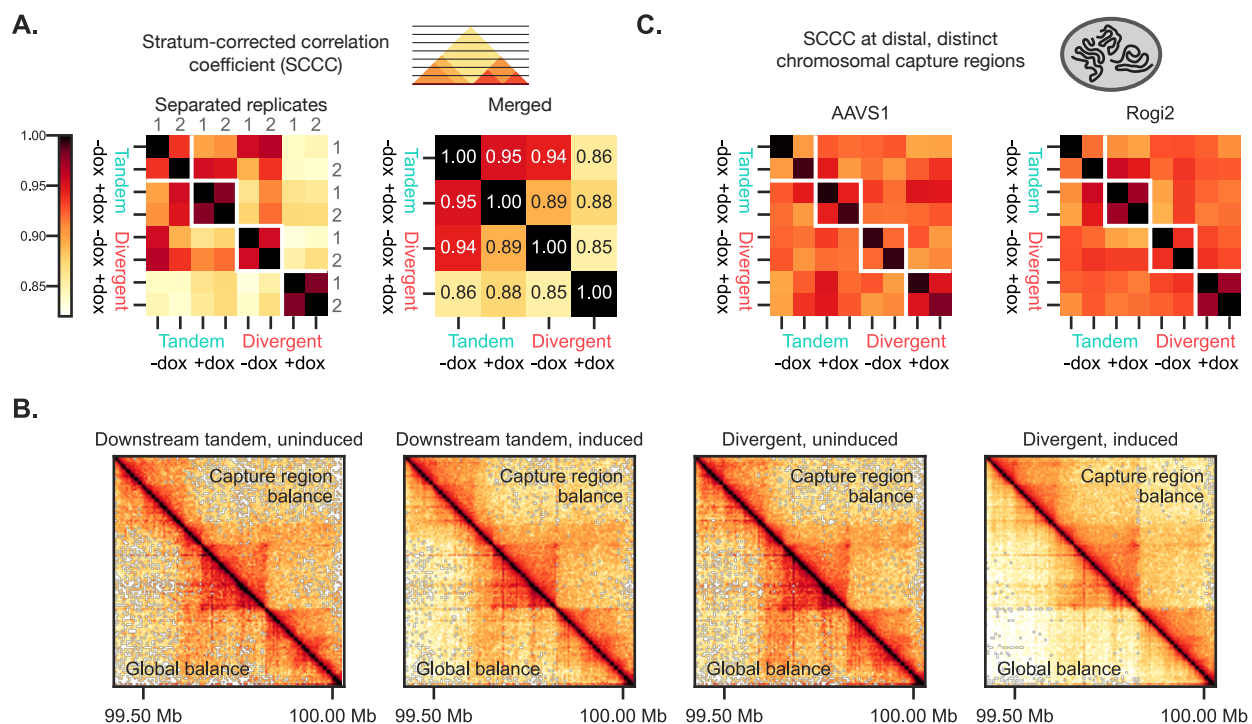
**Figure S7:** Joint mRNA distributions are shown for the monoclonal lines presented in fig. 2 for the uninduced (gray) and induced (colored) conditions.



**Figure S8:** hiPSC cell line generation and behavior

a) The homozygous lines were created by integrating a plasmid donor into both alleles of intron 2 of *CLYBL* in hiPSCs. The plasmid backbone components and auxiliary elements of the STRAIGHT-IN platform are excised to leave the all-in-one circuit.

b) Over a period of two weeks, cells were repeatedly induced and show similar time-course data to fig. 2e, showing both upstream dominance and divergent enhancement. The geometric mean and 95% confidence interval are calculated over three merged wells.

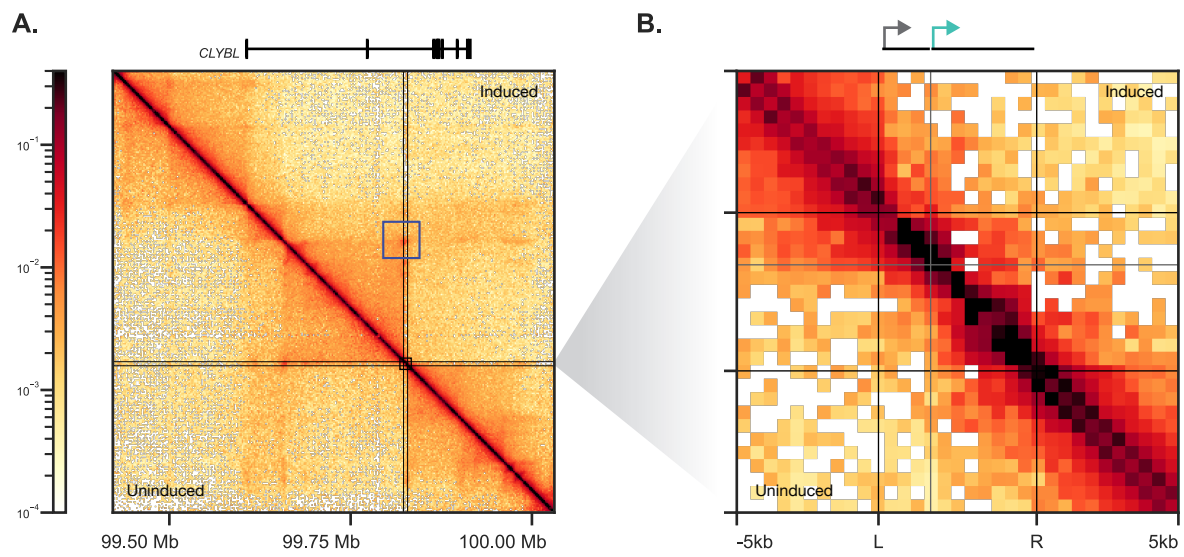


**Figure S9: Region Capture Micro-C validation.**

a) The stratum-corrected correlation coefficient (SCCC) measures the similarity between two interaction matrices while more heavily weighting short-distance genomic contacts. The SCCC is shown for each biological replicate and for the post-merged matrices.

b) The four resulting merged RCMC genomic contacts were iteratively balanced (e.g. normalized such that every row and column sums to 1 and can be interpreted as a probability distribution), both across the entire genome and only within the capture region. Capture-region balancing provided cleaner contact probability distributions.

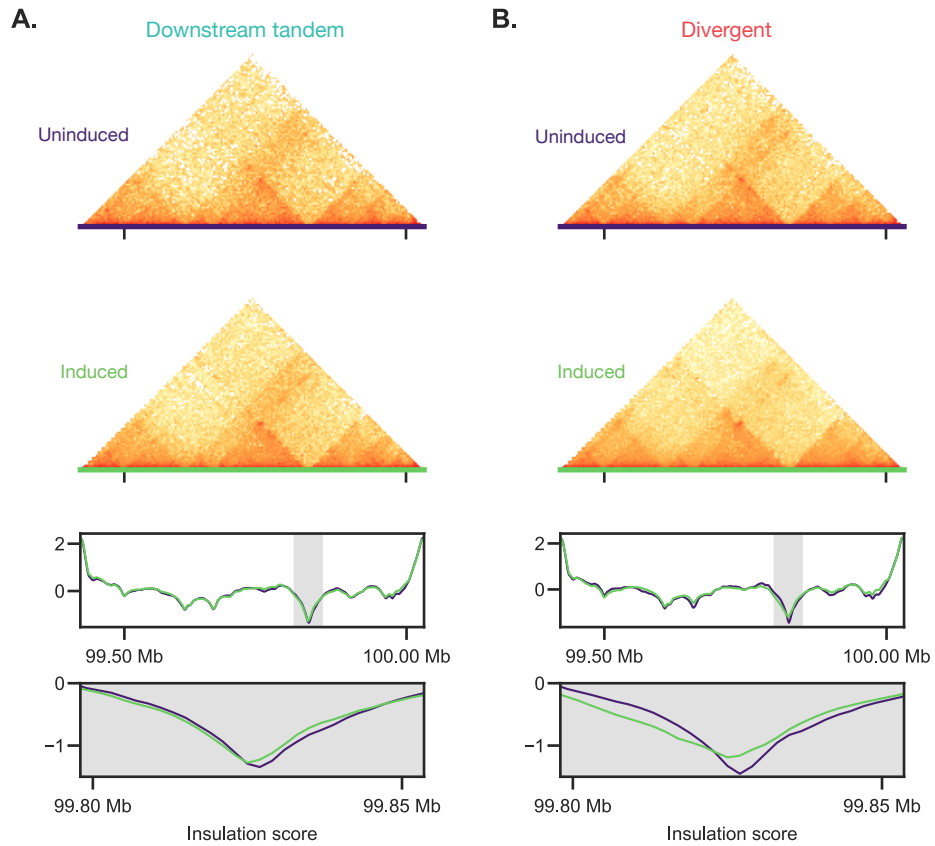
c) Two distal regions, AAVS1, and Rogi2, were largely unaffected by both integration of the synthetic construct at *CLYBL* and by dox induction, as measured with the SCCC.



**Figure S10:** Global and local chromatin structure surrounding the downstream tandem integration.

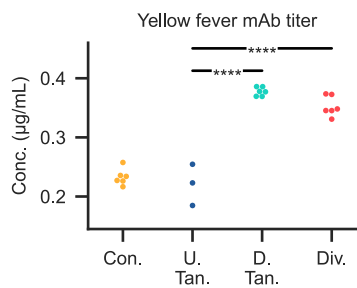
a) At 2000 bp resolution, the uninduced and induced chromatin structure across the entire capture region is shown.

b) At 500 bp resolution, the uninduced and induced chromatin structure in the local region surrounding the integration site are shown.



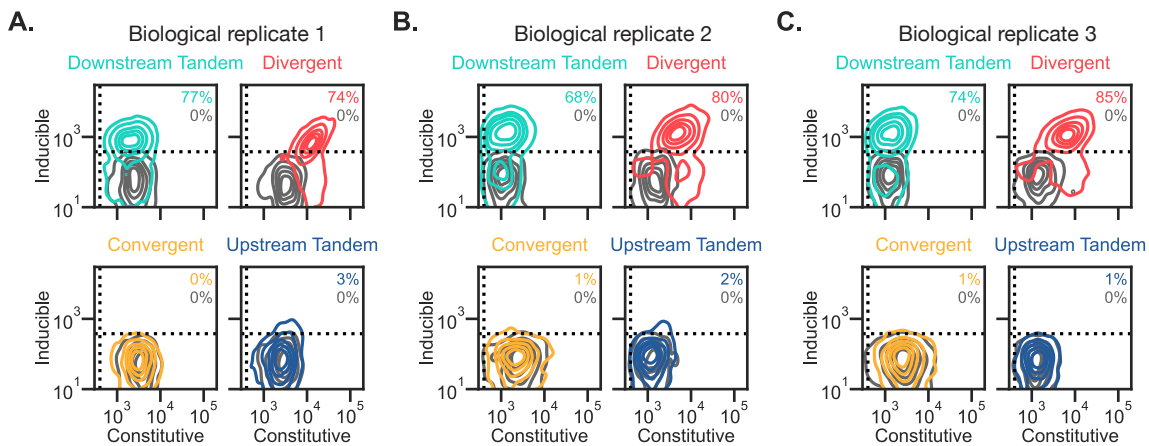
**Figure S11:** Region capture Micro-C insulation scores.

- a) The sliding-window insulation score is shown for both the uninduced and induced cases, for the downstream tandem orientation. No strong deviation is observed upon induction.
- b) The sliding-window insulation score for the divergent orientation is shown. Lower local insulation is seen at the TAD boundary upon induction.

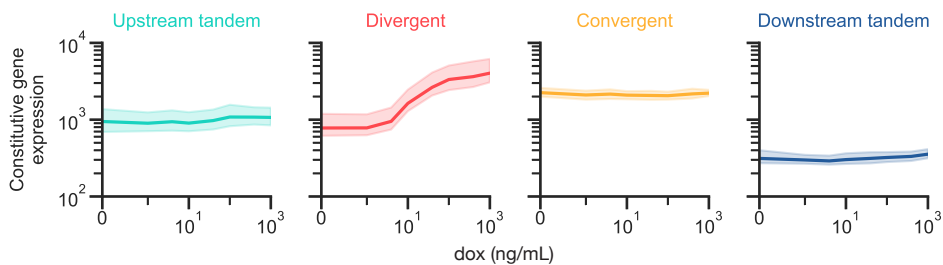


**Figure S12:** Antibody titer, as measured by a bead agglutination assay, for cell lines in fig. 4a expressing heavy and light chains with different syntaxes, N=3-6 biological replicates. Statistics are two-sided student t-tests. \*\*\*\* :  $p < 0.0001$

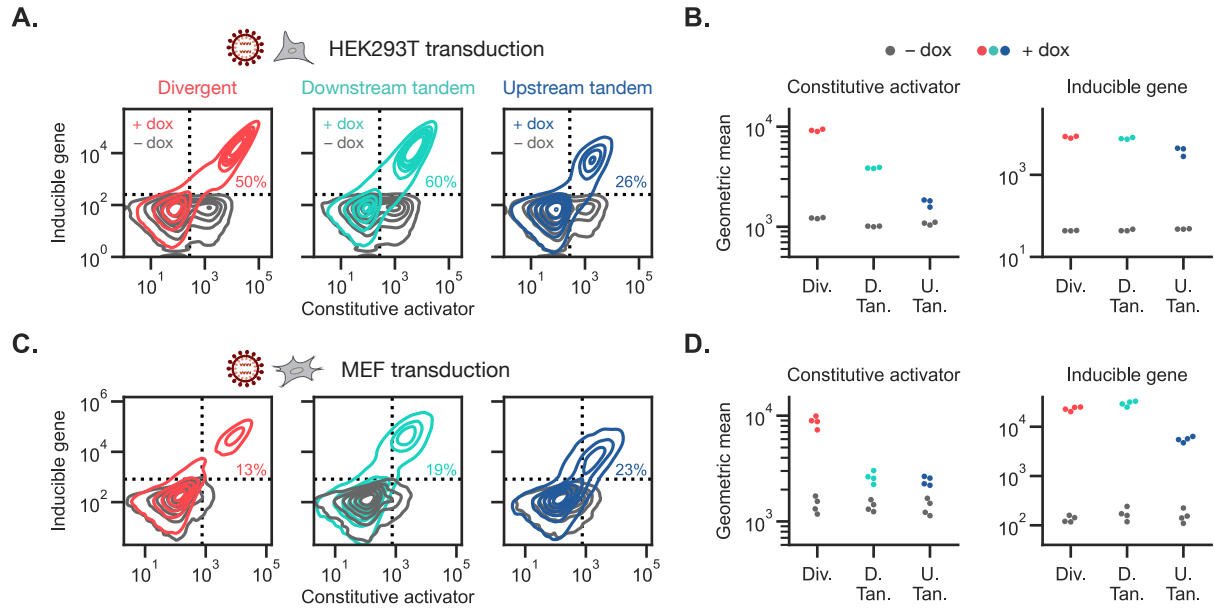




**Figure S13:** Joint distributions for three additional biological replicates of the lentiviral transductions in HEK293T cells presented in fig. 4b.

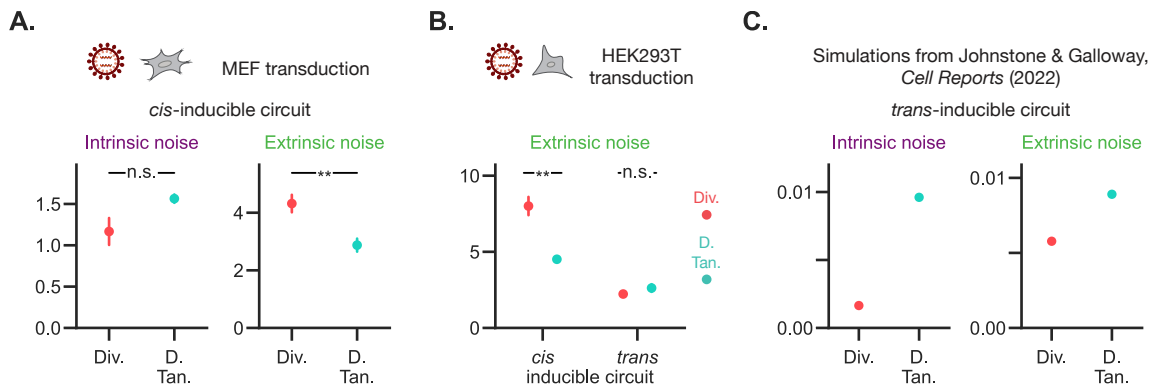


**Figure S14:** Constitutive gene expression as a function of inducer (dox) concentration for the lentiviral transductions of HEK293T cells presented in fig. 4b. Constitutive gene expression remains constant for the tandem and convergent syntaxes but increases with dox for the divergent syntax. Shading represents the 95% confidence interval across four biological replicates.



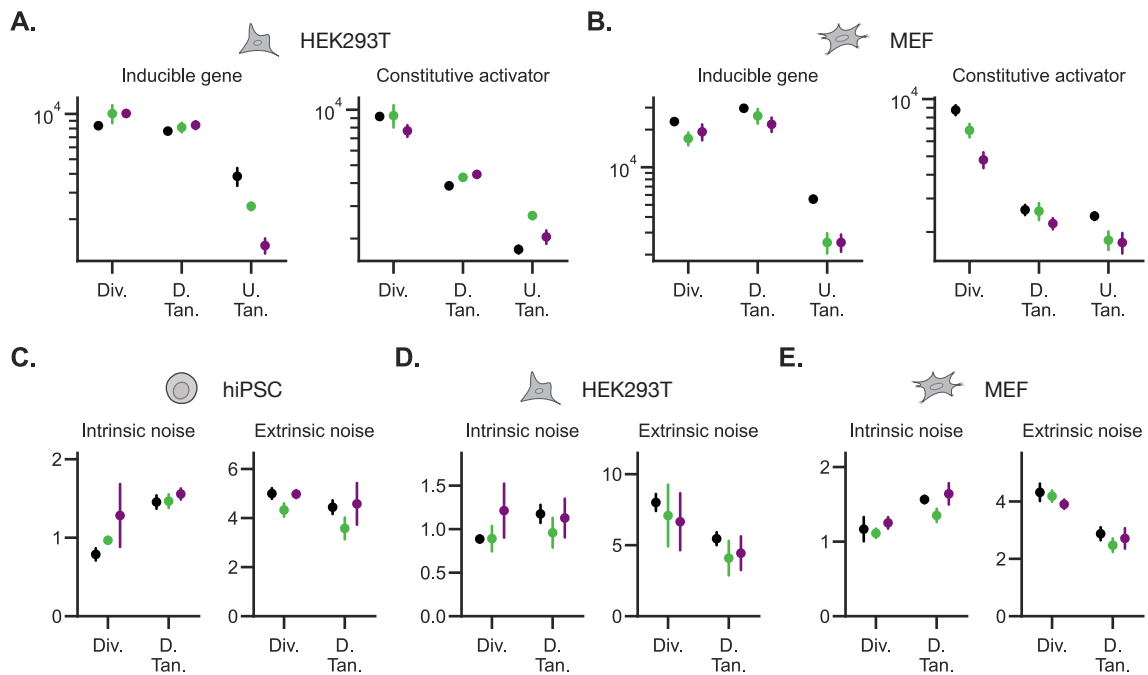
**Figure S15:** Performance of all-in-one inducible circuits across cell types.

- a) Joint distributions of the constitutive activator and inducible gene are shown for the inducible all-in-one circuits in fig. 5a. Circuits were lentivirally transduced into HEK293T cells. Distributions show one representative biological replicate.
- b) Geometric mean of constitutive activator and inducible gene expression for populations from a) gated on activator-positive cells. The dashed vertical line in a) shows this gate. Points depict three biological replicates.
- c) Representative joint distributions of the constitutive activator and inducible gene are shown for the same circuits lentivirally transduced into primary mouse embryonic fibroblasts (MEFs).
- d) Data from c) displayed as in b), N=4 biological replicates.



**Figure S16:** Characterization of noise in inducible circuits.

- a) Noise analysis for lentiviral transduction of the *cis*-inducible circuit in mouse embryonic fibroblasts (MEFs) induced with 300 ng/mL dox. Noise was calculated for the double-positive population. Points represent mean  $\pm$  standard error for N=4 biological replicates. Statistics are two-sided student t-tests. n.s.:  $p > 0.05$ , \*\*:  $p < 0.01$
- b) The extrinsic noise component is shown for the populations in fig. 5f. HEK293T cells were lentivirally transduced with the *cis*-inducible (all-in-one) circuit from fig. 5a or the *trans*-inducible circuit from fig. 4b and induced with dox, n=3-4 biological replicates.
- c) The same noise analysis for simulations of the *trans*-inducible circuit modeled in Johnstone & Galloway [10]. Points represent noise calculated for populations of 2,000 simulations per condition.



**Figure S17:** The cHS4 insulator does not mitigate the effects of syntax in lentiviral delivery of an inducible, all-in-one circuit. **a)** The circuits in fig. 5g were lentivirally transduced into HEK293T cells. Geometric mean expression of the inducible gene and constitutive activator are shown for populations induced with 300 ng/mL dox. Points represent mean  $\pm$  standard error for N=4 biological replicates.

**b)** The circuits in fig. 5g were also lentivirally transduced into mouse embryonic fibroblasts (MEFs), n=5 biological replicates.

**c)** Noise analysis for the circuits transduced into hiPSCs in fig. 5g. All pairwise comparisons between insulators for each syntax are not significant, two-sided student t-tests,  $p > 0.05$ .

**d)** Noise analysis for the circuits transduced into HEK293T cells in **a)**. All pairwise comparisons between insulators for each syntax are not significant.

**e)** Noise analysis for the circuits transduced into MEFs in **b)**. All pairwise comparisons between insulators for each syntax are not significant.