
Highly multiplexed mammalian metabolic engineering with a shotgun approach

In the format provided by the
authors and unedited

Supplementary Note 1: Characterization of SGE-derived CHO-Val clones using ^{13}C -labeling

Clones CHO-Val-D1 and CHO-Val-D3 were cultured in a valine-free 'heavy' RPMI medium in which ^{12}C glucose was substituted for its fully labeled ^{13}C counterpart and to which 2 mM ^{13}C sodium pyruvate was added (**Extended Data Fig. 2B**). Both clones were cultured in this valine-free heavy medium over 3 passages while control cells were cultured in the equivalent medium but with valine added back to facilitate their growth on ^{13}C -labeled valine pathway substrates over the same timeframe. Using LCMS, we confirmed detection of $^{13}\text{C}_5$ -labeled valine in our engineered clones via extracted ion chromatogram, MS1, and MS2 analysis (shown for CHO-Val-D1 in **Extended Data Fig. 2C-E**). In quantifying ^{13}C -labeling levels of valine across cell lines, we considered only isotopologues of valine with at least 2 carbon-13 atom incorporations as being endogenously biosynthesized given that carbon-13 constitutes approximately 1% of all carbon. Therefore, the natural abundance of $^{13}\text{C}_1$ valine is 5% (5 carbons x 1% representation) while $^{13}\text{C}_2$ valine is only expected to represent 0.25% of the total pool (5% x 5%).

Supplementary Note 2: Characterization of SGE-derived CHO-Ile clones using ^{13}C -labeling

Clones CHO-Ile-10-H1 and CHO-Ile-10-H2 were cultured in an isoleucine-free 'heavy' RPMI medium in which ^{12}C glucose was substituted for its fully labeled ^{13}C counterpart and to which 2 mM ^{13}C sodium pyruvate was added (**Extended Data Fig. 4B**). Both clones were cultured in this isoleucine-free heavy medium over 3 passages while control cells were cultured in the equivalent medium but with isoleucine added back to facilitate their growth on ^{13}C -labeled isoleucine pathway substrates over the same timeframe. Using high resolution LCMS, we found accurate mass peaks corresponding to the retention time of both leucine and isoleucine, which share identical molecular weights, at 132.1010 m/z (M+0) in control cells, while only a peak corresponding to leucine was found in engineered clones. At 134.1086 m/z (M+2), the expected

labeling signature for endogenous biosynthesis of isoleucine from $^{13}\text{C}_3$ pyruvate + $^{13}\text{C}_0$ 2-oxobutanoate, mere background signals of leucine and isoleucine could be seen in control cells while a prominent peak corresponding to isoleucine could be seen in the engineered clones (**Fig. 4C**). With MS1 analysis, we observed only M+0 isoleucine and the spiked-in internal standard $^{13}\text{C}_6/^{15}\text{N}_1$ isoleucine in control cells while signatures of both M+0, M+2, and M+6 isoleucine and internal standard $^{13}\text{C}_6/^{15}\text{N}_1$ isoleucine could be found in the prototrophic clones (**Fig. 4D**). Using narrow isolation MS2 (± 0.2 da), we were further able to confirm that the fragmentation signatures for each M+0, M+2, and M+6 as well as internal standard $^{13}\text{C}_6/^{15}\text{N}_1$ isoleucine detected within CHO-Ile-10-H1 were consistent with expectations (**Extended Data Fig. 4C**). In quantifying ^{13}C -labeling levels of isoleucine across cell lines, we considered only isotopologues of isoleucine with at least 2 carbon-13 atom incorporations as being endogenously biosynthesized given that carbon-13 constitutes approximately 1% of all carbon. Therefore, the natural abundance of $^{13}\text{C}_1$ isoleucine is 6% (6 carbons x 1% representation) while $^{13}\text{C}_2$ isoleucine is only expected to represent 0.36% of the total pool (6% x 6%).

Supplementary Note 3: Characterization of barcode swapping, integration sites, and transcriptomic profiles of SGE-derived clones

Our primary aim was to identify functional pathway solutions that can be reconstituted on a single construct (**Fig. 3J**), rather than to deeply characterize every SGE-derived clone. However, discrepancies between genotype, expression, and phenotype in select cases prompted further investigation. We therefore evaluated whether SGE barcodes faithfully reflect their assigned TU identities, including the potential for barcode swapping during lentiviral packaging, the impact of TU integration sites on clonal phenotypes, and global transcriptomic characterization.

Across the 4 CHO clones (CHO-Val-D1, CHO-Val-D3, CHO-Ile-10-H1 and CHO-Ile-10-H2) for which we performed long-read genome sequencing, 95.1% and 93.7% of reads carried their expected promoter and CDS barcodes, respectively (**Supplementary Fig. 7A**), corresponding to one barcode swap per CHO clone (**Supplementary Table 2**). These results indicate that barcode swapping is infrequent and unlikely to substantially impact genotype assignment across populations. Further, this limitation can be completely bypassed in future screens by using a non-lentiviral delivery mechanism.

Remaining discordant calls are consistent with low-level clonal heterogeneity. TUs classified as background by amplicon-seq were detectable at low abundance by long-read sequencing (**Extended Data Fig. 5B-C; Supplementary Fig. 7B**), and long-read data revealed more integration sites than predicted by amplicon-seq in some cases (17 and 15 sites in CHO-Ile-10-H1 and CHO-Ile-10-H2, respectively; **Supplementary Table 3**).

We mapped TU integration sites across CHO-Val-D1, CHO-Val-D3, CHO-Ile-10-H1, and CHO-Ile-10-H2 using long-read DNA sequencing. Reads supporting these sites frequently contained target site duplications (TSDs), consistent with lentiviral integration (**Supplementary Fig. 7C**). Across both valine and isoleucine prototrophic clones, integrations occurred predominantly in intergenic and intronic regions (**Supplementary Fig. 7D-E; Supplementary Table 3**).

Importantly, integration of TUs within endogenous genes was not associated with substantial changes to target gene expression. The only exceptions were *Brinp3*, a mitochondrially localized protein involved in neuronal differentiation¹, and *Ccdc112*, a protein implicated in centriolar organization and sperm mitochondrial sheath formation^{2,3}. *Brinp3* was upregulated in CHO-Val-D1 (in which a TU insertion occurred), but not in CHO-Val-D3, which had no insertion. In CHO-Val-D1, reads mapping to *Brinp3* were observed exclusively downstream of the integrated TU, in the same orientation as the TU transcript, and were restricted to exon 7 rather than spanning the full gene. This pattern suggests that the apparent increase in *Brinp3*

expression may reflect transcriptional readthrough from the integrated TU captured in RNA-seq, rather than a bona fide upregulation of the endogenous gene. (**Supplementary Fig. 7F-G**)

Ccdc112 was marginally downregulated in CHO-Ile-10-H1 (in which a TU insertion occurred) but marginally upregulated in CHO-Ile-10-H2, which had no insertion, despite similar prototrophic phenotypes, further suggesting that these effects reflect local genomic disruptions rather than primary drivers of the respective phenotypes (**Supplementary Fig. 7H**).

To further assess global effects of TU integration, we profiled the transcriptomes of prototrophic clones across both isoleucine and valine screens via RNA-seq. Specifically, we analyzed CHO-Val-D1 cells cultured in valine-free or valine-replete medium and CHO-Ile-10-H1 cells cultured in isoleucine-free or isoleucine-replete medium, together with parental CHO cells grown in nutrient-replete conditions. Across all comparisons, transcriptomic differences relative to parental cells were modest under both nutrient-replete and dropout conditions (**Supplementary Fig. 8A-E; Supplementary Fig. 9A-E; Supplementary Tables 4-5**), indicating that multiplexed TU integration imposed a limited metabolic burden and did not elicit major transcriptional responses.

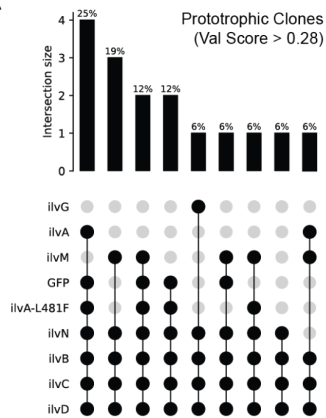
For Jurkat-Val-31, TU integrations mapped to 7 genomic loci, all within introns. At two loci, we observed tandem TU integrations, a previously documented phenomenon for high MOI lentiviral screens^{4,5}; however, none of the intronic disruptions correlated to notable changes in expression of affected genes (**Supplementary Table 6-7**). We performed RNA-seq on Jurkat-Val-31 cultured in valine-free and valine-replete medium, respectively, as well as on parental control cells cultured in valine-replete medium only. Global transcriptomic differences across all pairwise comparisons were modest, indicating minimal perturbation from TU integration (**Supplementary Fig. 10A-E; Supplementary Table 7**).

Supplementary Note 4: Characterization of barcode signatures underlying single vs. double prototrophic clones

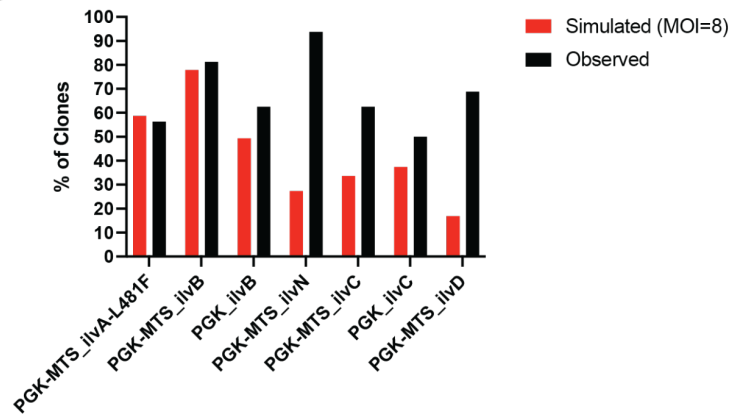
When looking at the genotypes underlying each double prototrophic clone, we found *ilvB*, *ilvC*, and *ilvD* to be universally present with *ilvM* additionally present in 76.5% of clones (**Extended Data Fig. 6B**). However, we noticed that CHO-Val clones and CHO-Ile clones exhibited differential underlying CDS usage suggesting that analyzing both populations on aggregate might conflate signal given that each population was not specifically selected to be doubly prototrophic (**Extended Data Fig. 6C**). We therefore focused on parsing out differences within clones amongst the different phenotypes present within each population.

Among CHO-Val clones, 6 clones were prototrophic for both valine and isoleucine (Val+Ile+), 10 were prototrophic for valine only (Val+Ile-) and none were prototrophic for isoleucine only (Val-Ile+). Among CHO-Ile clones, 12 clones were Ile+Val+, 7 clones were Ile+Val- and 21 clones were Ile-Val+ (**Extended Data Fig. 6D**). When looking at the gene content underlying each phenotype, CHO-Val clones looked to have similar CDS representation irrespective of phenotype. On the other hand, CHO-Ile cells showed notable differences in *ilvG*, *ilvA*, and *ilvA-L481F* content across phenotypes (**Extended Data Fig. 6E**). In particular, *ilvG* was seen in 100% of Ile+Val- clones, 75% of Ile+Val+ clones and in <10% of Ile-Val+ clones indicating it is a strong determinant of isoleucine prototrophic function and possibly detrimental to valine prototrophic function. This is consistent with the known catalytic function of *ilvG*, which exhibits 57-180X greater preference for isoleucine biosynthetic intermediates over valine biosynthetic intermediates (**Table 1**). *ilvA* was most likely to be found in the Ile+Val- population suggesting it supports isoleucine prototrophic function but not valine prototrophic function, while *ilvA-L481F* was most used in the Ile+Val+ population suggesting that a feedback mechanism may be constraining the manifestation of the doubly prototrophic phenotype.

A



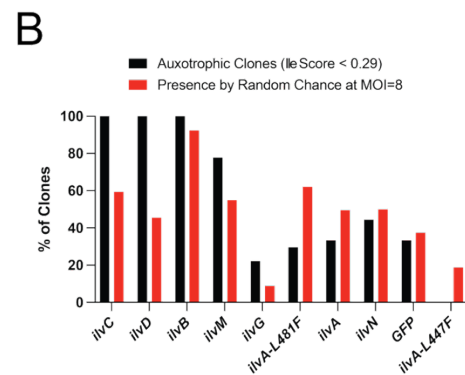
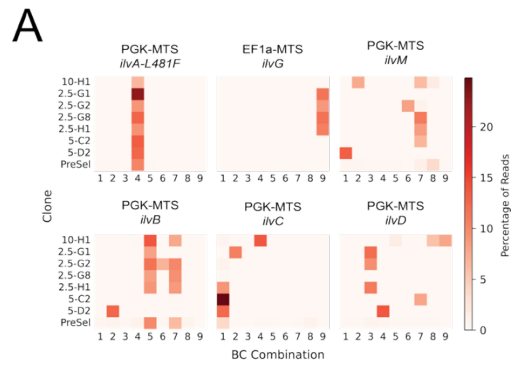
B



Supplementary Fig. 1. CDS and 'solution' TU usage across all valine prototrophic clones

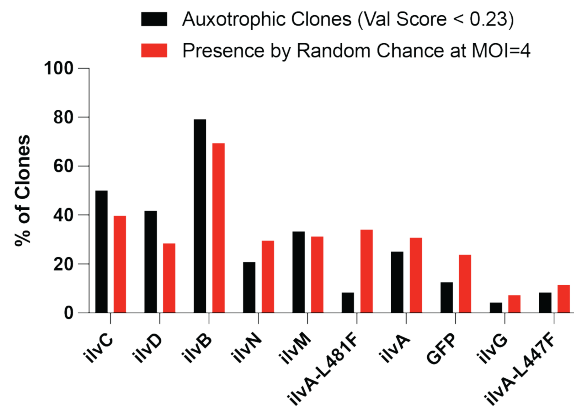
(A) UpsetPlot depicting which combinations of CDSes were determined to be present within each of the 16 valine prototrophic CHO clones

(B) Usage of 'pathway solution' TUs among all 16 valine prototrophic CHO clones

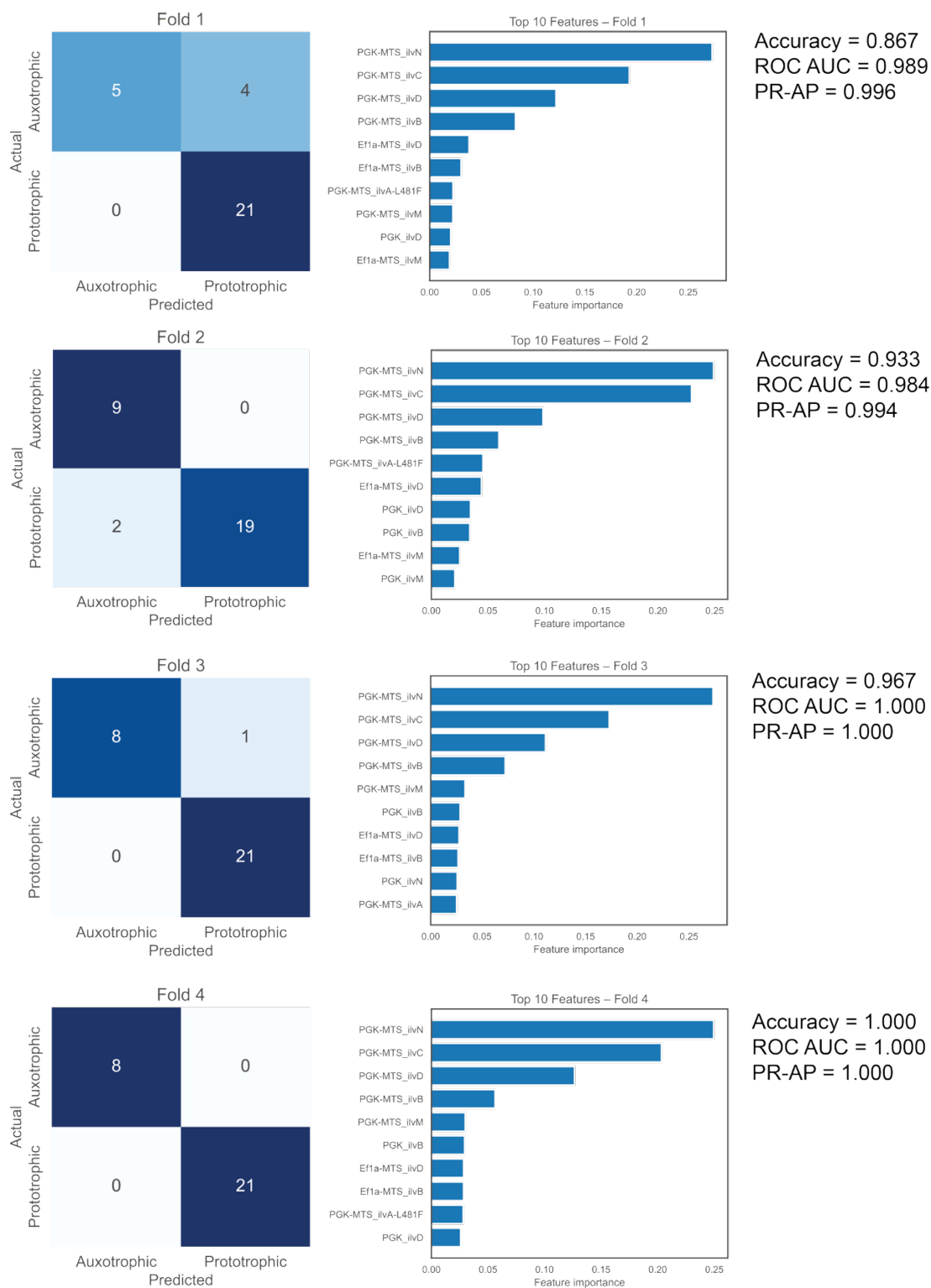


Supplementary Fig. 2. Barcode usage among isoleucine pathway ‘solution’ TUs and CDS usage in isoleucine auxotrophic clones

- (A) Relative abundance of TUs, which comprise the isoleucine prototrophic pathway ‘solution’ and their 9 underlying barcode combinations (3 per promoter/OLS, 3 per CDS) across all strongly isoleucine prototrophic clones (Ile Score > 0.75).
- (B) CDS usage amongst 27 isoleucine auxotrophic clones (Ile Score < 0.29) is compared to the simulated expectation.

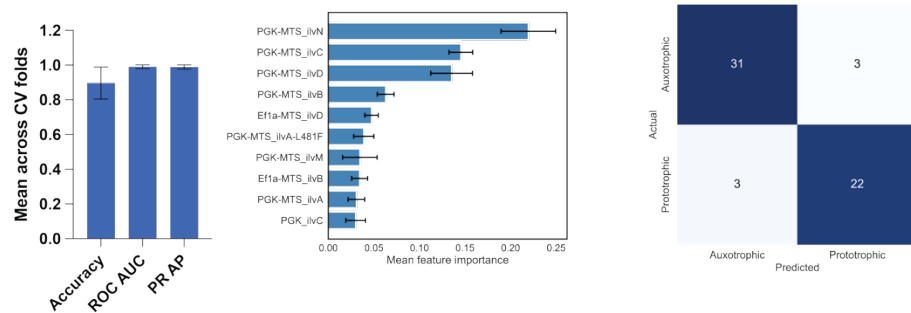


Supplementary Fig. 3. CDS usage amongst all valine auxotrophic clones carrying at least 1 TU integration is compared to the simulated expectation.

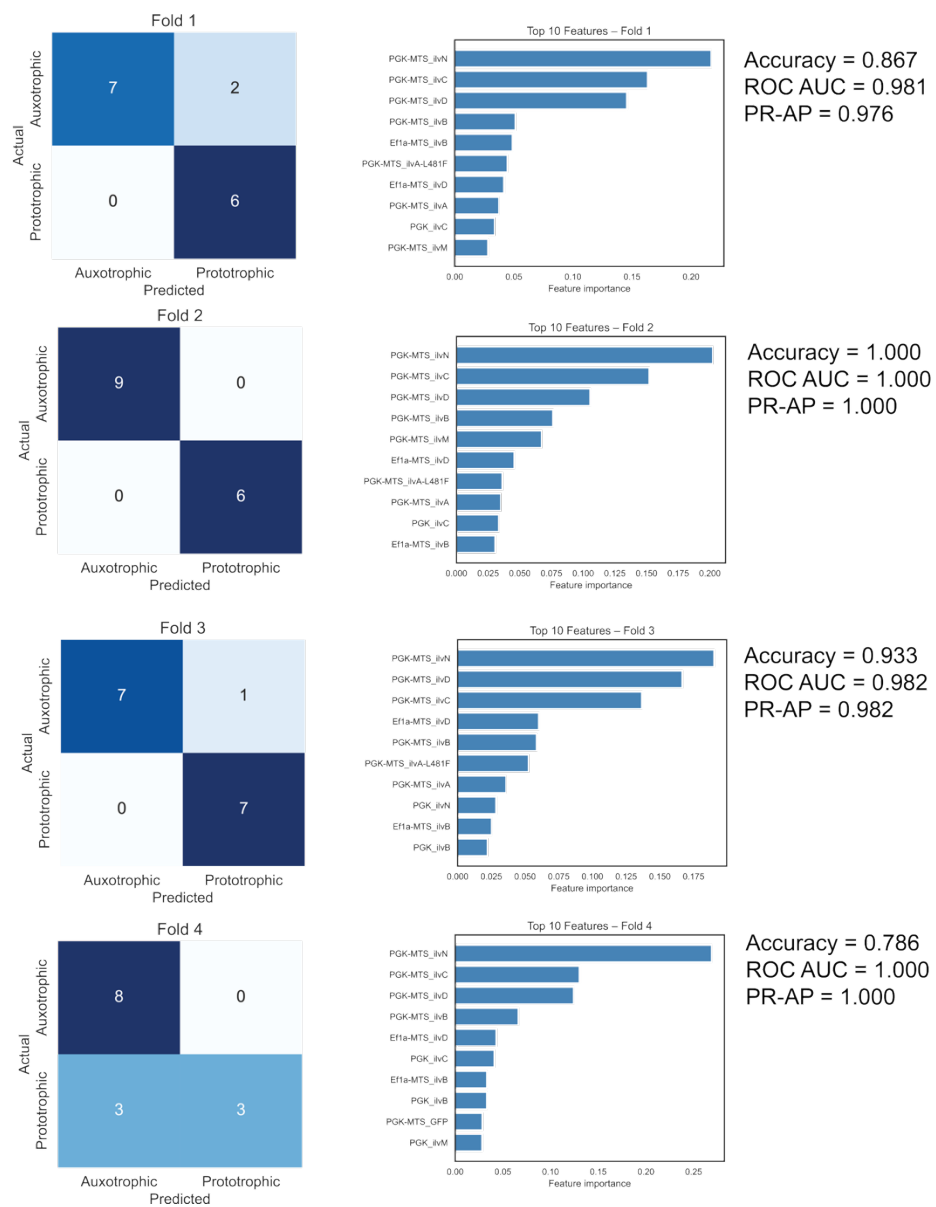


Supplementary Fig. 4. 4-fold cross-validation to validate Random Forest model stability across different train:test compositions. Classifications (left), top 10 features (middle) and metrics (right).

A



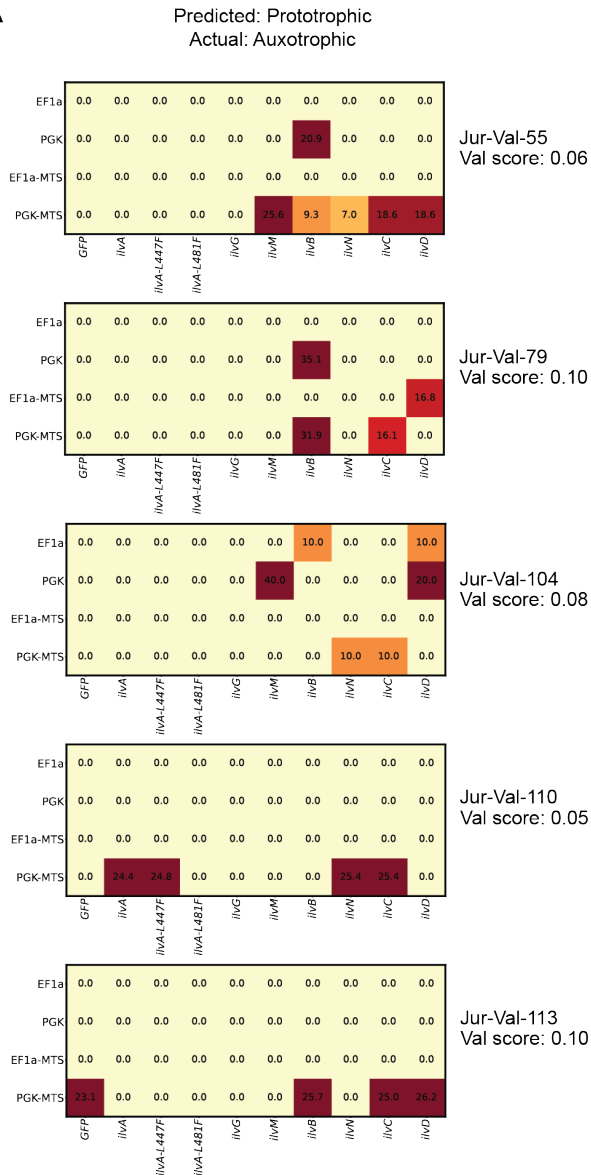
B



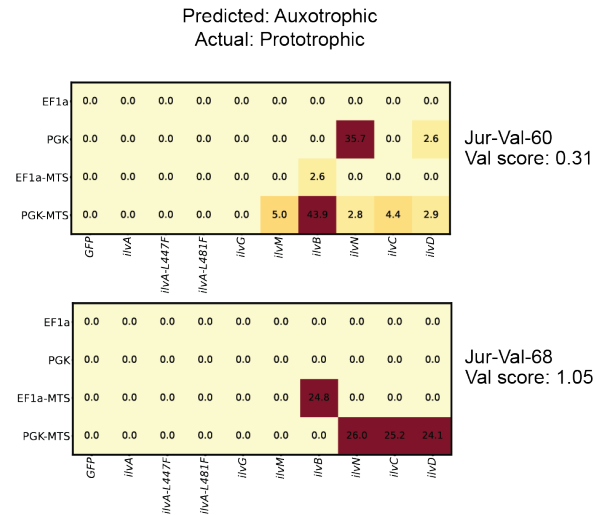
Supplementary Fig. 5. Training of an RF classifier after removal of duplicate ‘sibling’ clones

- (A) 4-fold cross-validation was used to ensure stability of the RF model without duplicate ‘sibling’ clones. Metrics (left), top 10 features (middle), and classifications (right) across the 4 folds are summarized. Error bars represent the standard deviation across folds.
- (B) Classifications (left), top 10 features (middle) and metrics (right) for each of the 4 cross-validation folds.

A



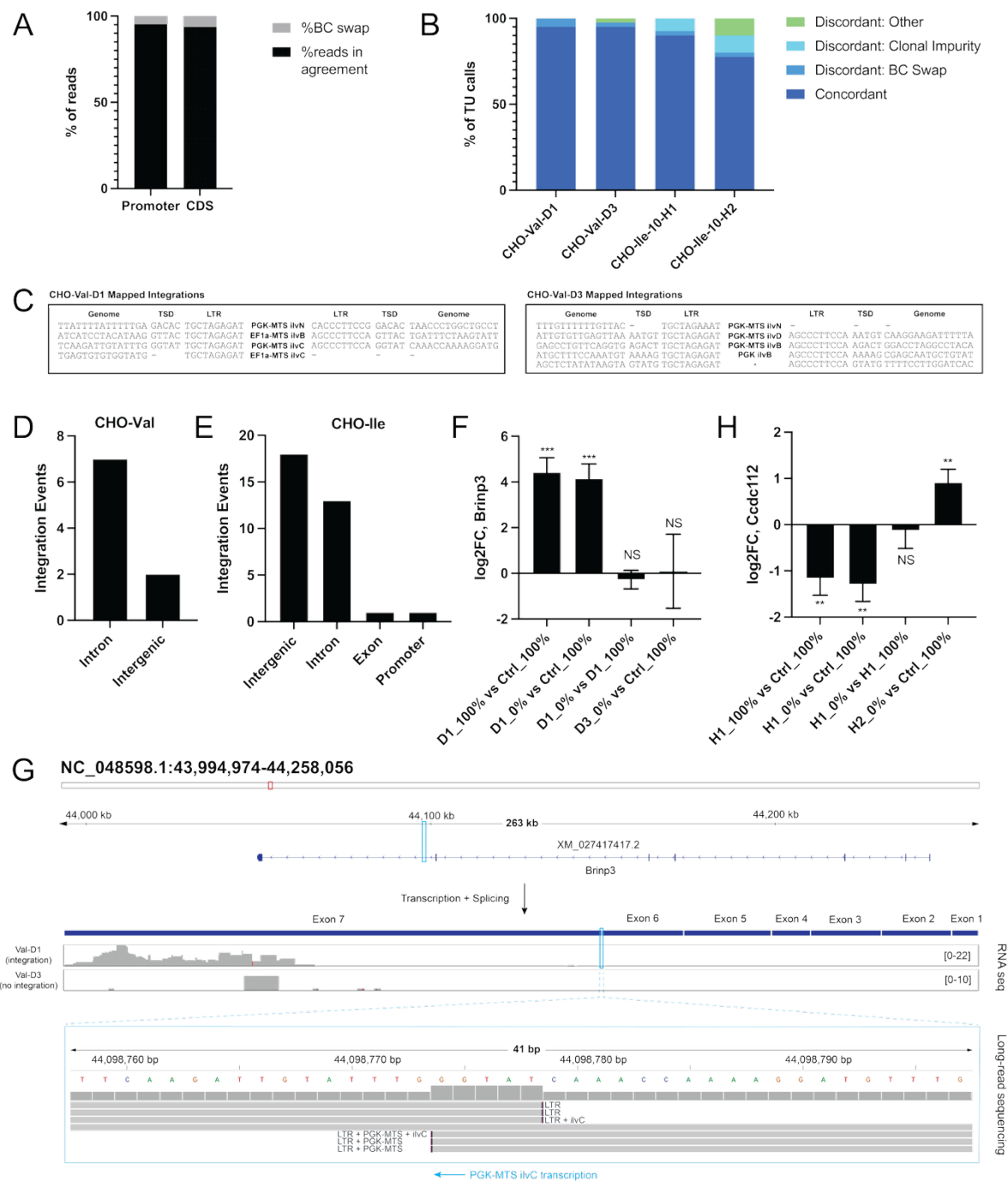
B



Supplementary Fig. 6. SGE barcode heatmaps for Jurkat clones wrongly classified by the trained RF model.

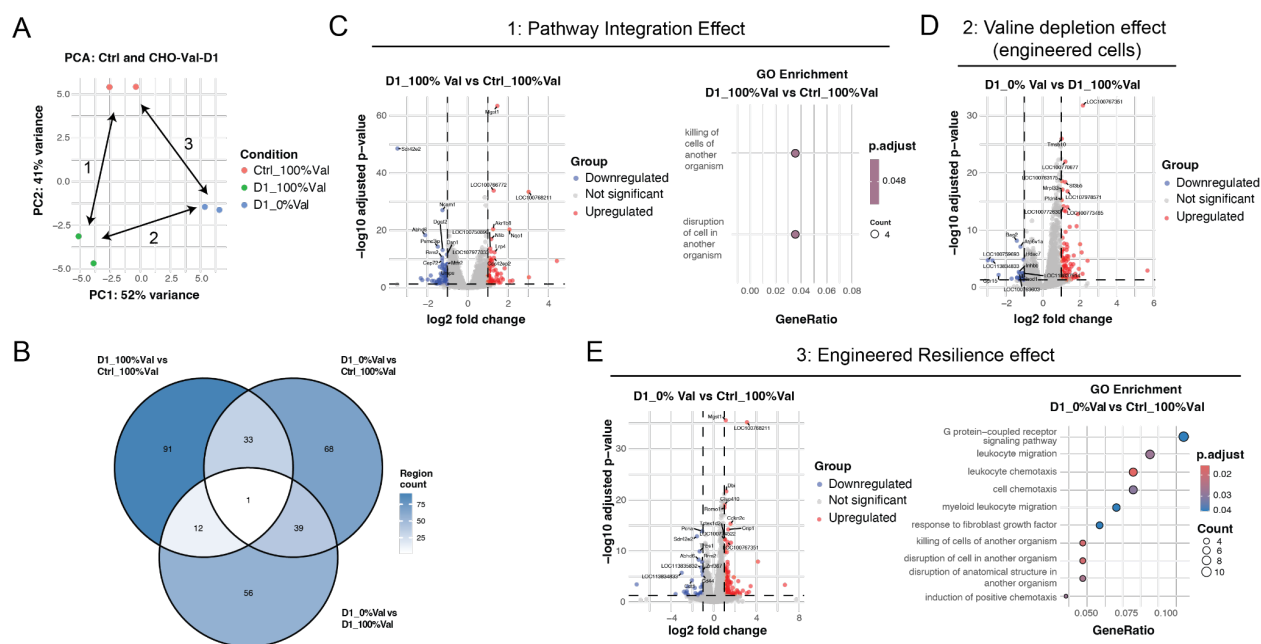
(A) Percentage of amplicon-seq reads corresponding to each TU within clones wrongly predicted to be prototrophic by the RF model

(B) Percentage of amplicon-seq reads corresponding to each TU within clones wrongly predicted to be auxotrophic by the RF model



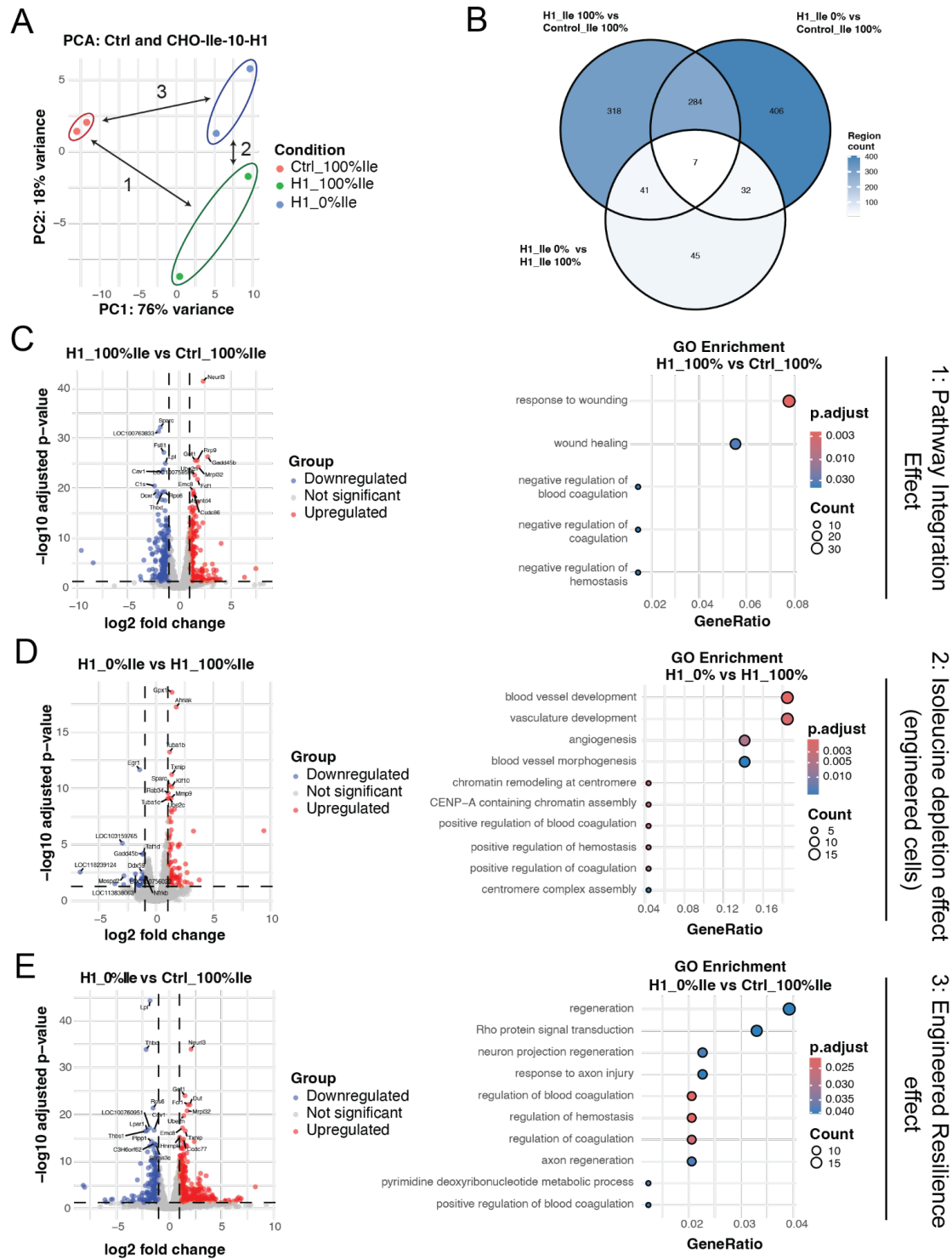
Supplementary Fig. 7. Characterization of barcode swapping and the genomic contexts of TU integration sites

- (A) Frequency of barcode swapping across CHO-Val-D1, CHO-Val-D3, CHO-Ile-10-H1 and CHO-Ile-10-H2 as detected by long-read sequencing
- (B) Concordance and discordance of TUs called by amplicon-seq and by long-read sequencing in clones CHO-Val-D1, CHO-Val-D3, CHO-Ile-10-H1 and CHO-Ile-10-H2. Discordant calls are split into categories of barcode swapping, clonal impurity, and other.
- (C) Long-read sequencing enabled mapping of lentiviral integration sites in clones revealing expected 5-bp target site duplications. Mapped sites in CHO-Val-D1 and CHO-Val-D3 are depicted.
- (D) Genomic context for integrations detected in clones CHO-Val-D1 and CHO-Val-D3
- (E) Genomic context for integrations detected in clones CHO-Ile-10-H1 and CHO-Ile-10-H2
- (F) *Brinp3* was significantly upregulated in CHO-Val-D1 cells relative to control cells but not in CHO-Val-D3 ($\log_2\text{FC} = \log_2 \text{Fold Change}$).
- (G) Schematic depicting the *Brinp3* disruption in intron 6 in CHO-Val-D1. A PGK-MTS *ilvC* transgene was integrated in the same direction as *Brinp3*. Sequencing reads from RNA-seq and long-read DNA sequencing that map to the locus are also depicted.
- (H) *Ccdc112* was significantly downregulated in CHO-Ile-10-H1 but significantly upregulated in CHO-Ile-10-H2 ($\log_2\text{FC} = \log_2 \text{Fold Change}$).



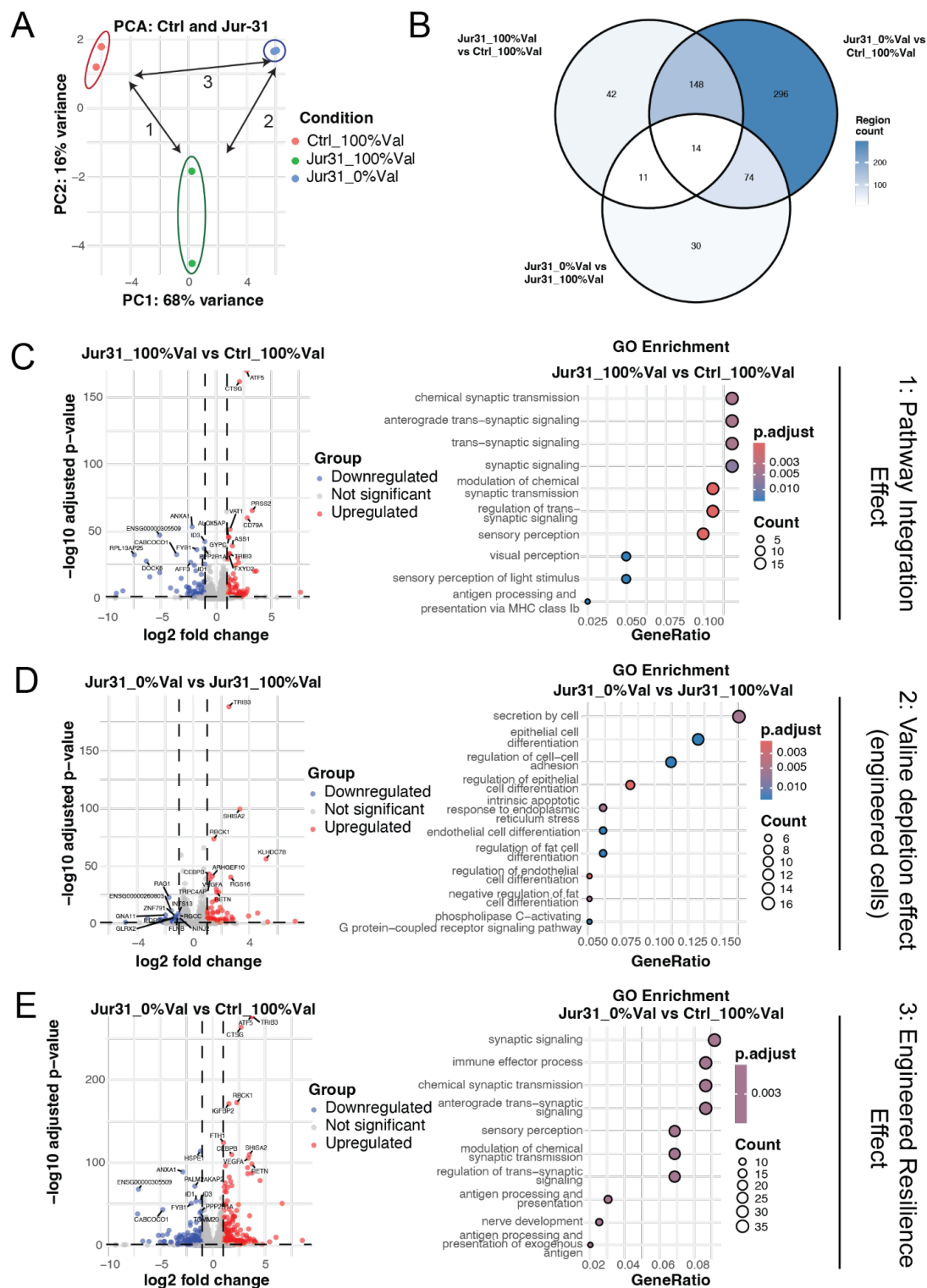
Supplementary Fig. 8. Global transcriptome profile comparing CHO-Val-D1 and control cells

- (A) Principal component analysis (PCA) depiction of RNA-seq data from CHO-Val-D1 cultured in valine-free and valine-replete medium, with parental control cells cultured in valine-replete medium.
- (B) Venn diagram depicting the number of differentially expressed genes (DEGs) across all pairwise sample comparisons and their overlap.
- (C) DEGs and gene ontology (GO) category enrichment between CHO-Val-D1 and parental cells, both cultured on valine-replete medium.
- (D) DEGs between CHO-Val-D1 cultured on valine-free medium and CHO-Val-D1 cultured on valine-replete medium. No GO categories were significantly enriched.
- (E) DEGs and GO category enrichment between CHO-Val-D1 cultured on valine-free medium and parental cells cultured on valine-replete medium.



Supplementary Fig. 9. Global transcriptome profile comparing CHO-Ile-10-H1 and control cells

- (A) PCA depiction of RNA-seq data from CHO-Ile-10-H1 cultured in isoleucine-free and isoleucine-replete medium, with parental control cells cultured in isoleucine-replete medium.
- (B) Venn diagram depicting the number of differentially expressed genes (DEGs) across all pairwise sample comparisons and their overlap.
- (C) DEGs and GO category enrichment between CHO-Ile-10-H1 and parental cells, both cultured on isoleucine-replete medium.
- (D) DEGs and GO category enrichment between CHO-Ile-10-H1 cultured on isoleucine-free medium and CHO-Ile-10-H1 cultured on isoleucine-replete medium.
- (E) DEGs and GO category enrichment between CHO-Ile-10-H1 cultured on isoleucine-free medium and parental cells cultured on isoleucine-replete medium.



Supplementary Fig. 10. Global transcriptome profile comparing Jurkat-Val-31 and control cells

- (A) PCA of RNA-seq data from Jurkat-Val-31 cells cultured in valine-free and valine-replete medium, with parental control cells cultured in valine-replete medium.
- (B) Venn diagram depicting DEGs across all pairwise sample comparisons and their overlap.
- (C) DEGs and GO category enrichment between Jurkat-Val-31 and parental cells, both cultured on valine-replete medium.
- (D) DEGs and GO category enrichment between Jurkat-Val-31 cultured on valine-free medium and Jurkat-Val-31 cultured on valine-replete medium.
- (E) DEGs and GO category enrichment between Jurkat-Val-31 cultured on valine-free medium and parental cells cultured on valine-replete medium.

REFERENCES

1. Berkowicz, S. R., Featherby, T. J., Whisstock, J. C. & Bird, P. I. Mice lacking Brinp2 or Brinp3, or both, exhibit behaviors consistent with neurodevelopmental disorders. *Front. Behav. Neurosci.* **10**, 196 (2016).
2. Gupta, G. D. *et al.* A dynamic protein interaction landscape of the human centrosome-cilium interface. *Cell* **163**, 1484–1499 (2015).
3. Graffeo, M. L. *et al.* A novel mechanism of sperm midpiece epididymal maturation and the role of CCDC112 in sperm midpiece formation and establishing an optimal flagella waveform. *Cell Commun. Signal.* **23**, 319 (2025).
4. Wanisch, K. & Yáñez-Muñoz, R. J. Integration-deficient lentiviral vectors: a slow coming of age. *Mol. Ther.* **17**, 1316–1332 (2009).
5. Winkler, T. *et al.* No evidence for clonal selection due to lentiviral integration sites in human induced pluripotent stem cells. *Stem Cells* **28**, 687–694 (2010).