

Signalling pathways drive heterogeneity of ground state pluripotency

Kirsten R. McEwen*, Sarah Linnett, Harry G. Leitch, Prashant Srivastava, Lara Al-Zouabi, Tien-Chi Huang, Maxime Rotival, Alex Sardini, Thalia E. Chan, Sarah Filippi, Michael P.H. Stumpf, Enrico Petretto, Petra Hajkova*

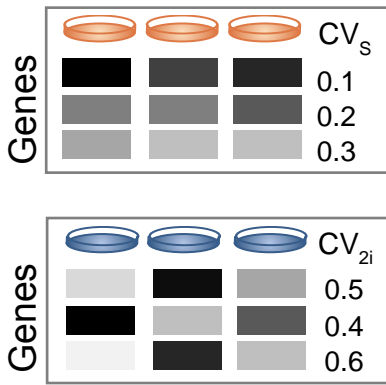
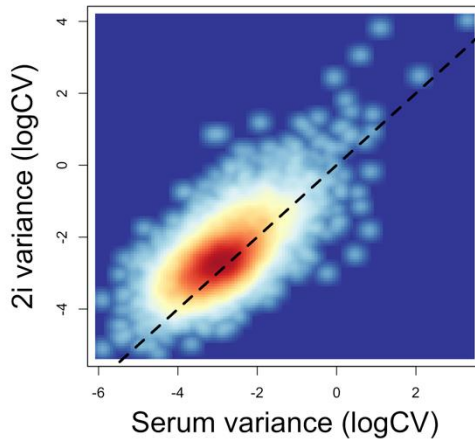
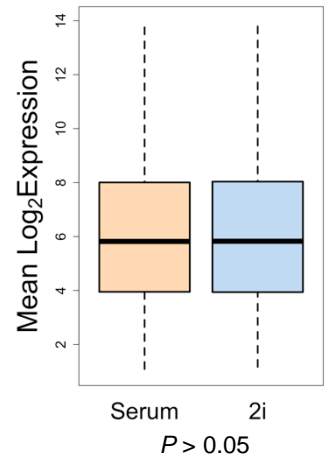
*Corresponding author

Condition		Cell type		Sex	
Serum	2i	ESC	EGC	Female	Male
9	13	10	12	10	12

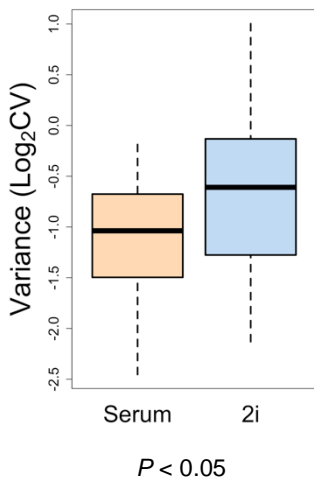
Supplementary Table 1: Sample numbers for each condition, related to Figure 1 and 2

Adpgk	Dclre1c	Itfg2	Plcg1	Rrbp1	Topors
Anapc5	Fbxo38	Jkamp	Plk1	Sel1l	Ttk
Aprt	Ganab	Lman1	Plk3	Srpr	Upf2
Axin1	Gca	Morf4l1	Pmel	Strn3	Utf1
Bloc1s1	Gja1	Nt5dc2	Pot1a	Stt3b	Xab2
Cacul1	Gtf2f1	P4hb	Rcn3	Surf4	Xbp1
Crebbp	Hyou1	Pdia5	Rfc4	Tiam1	Ccne1
Cyb5a	Inpp5a	Pdia6	Rpn2	Tmx1	

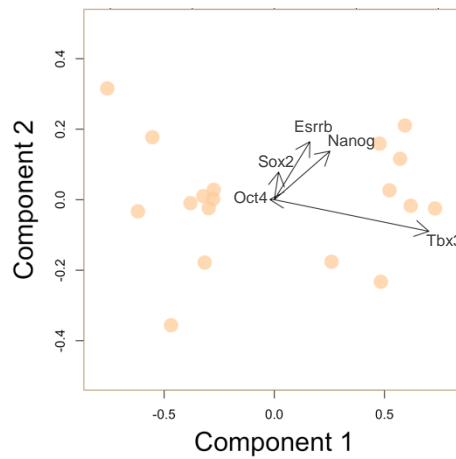
Supplementary Table 2: Genes tested by population and single-cell qPCR

A**B****C****D**

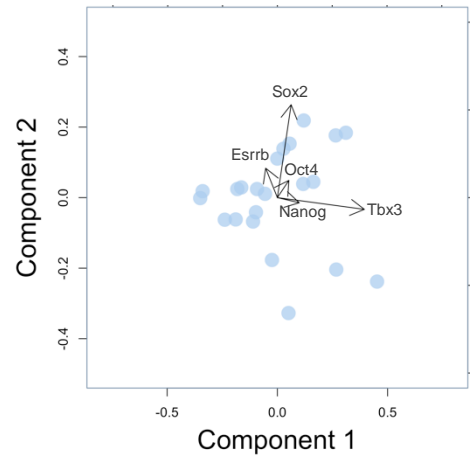
Population qPCR

**E**

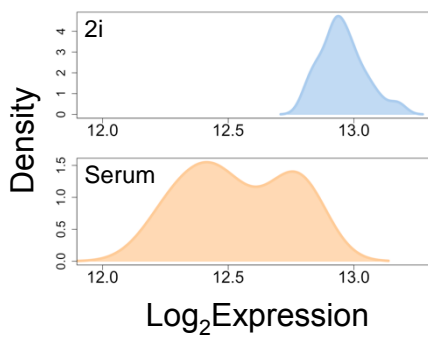
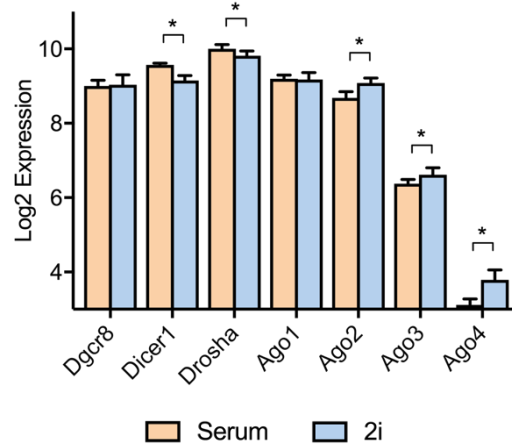
Pluripotency - serum

**F**

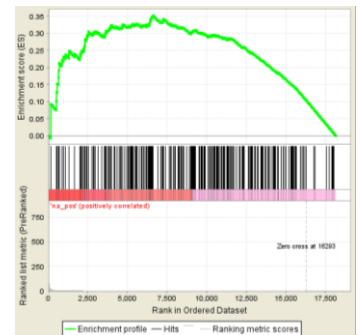
Pluripotency - 2i

**G**

Nanog – Population Level

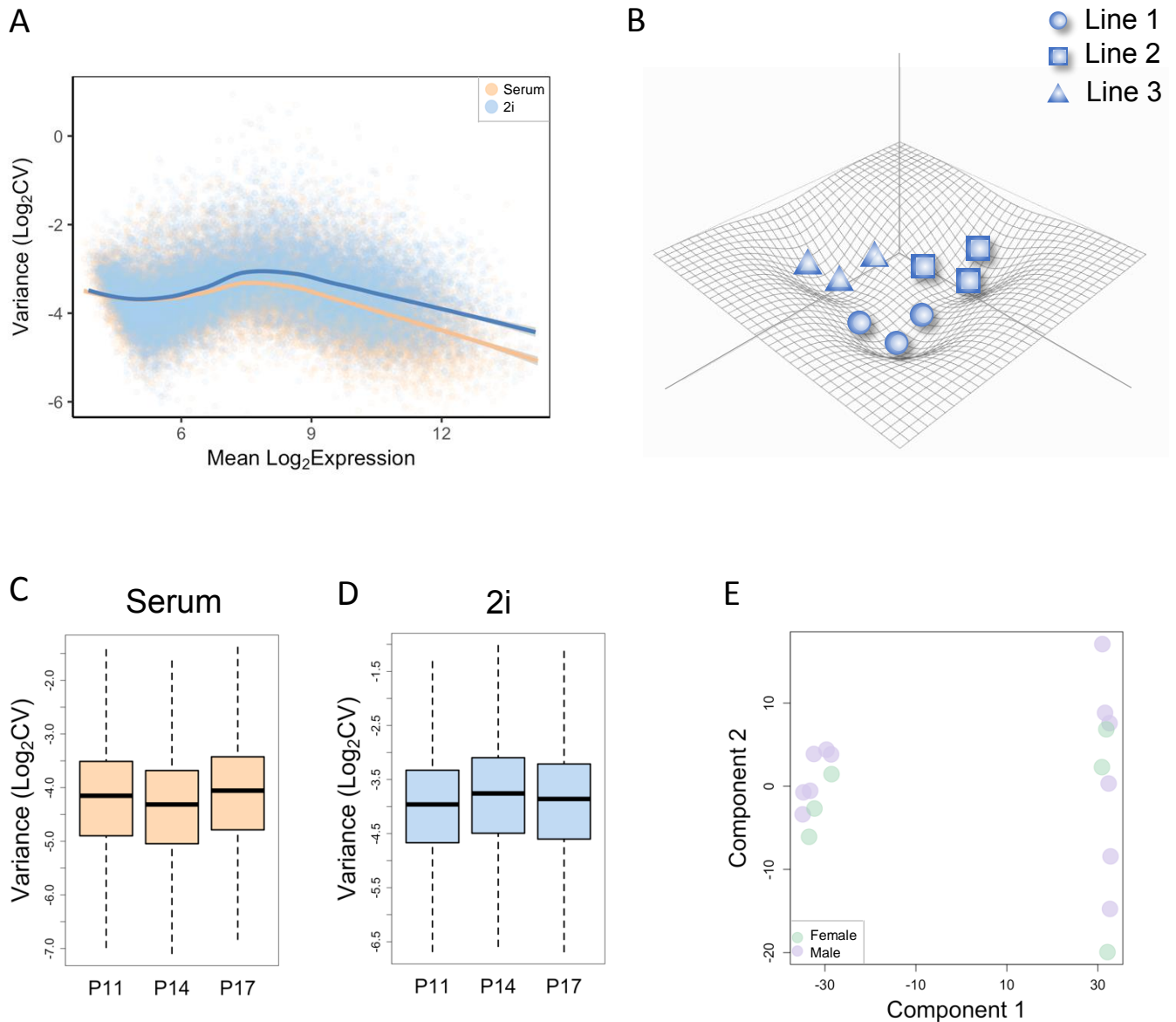
**H****I**

Pol II travelling ratio

 $FDR > 0.05$

Supplementary Figure 1: Transcriptional heterogeneity in 2i, related to Figure 1

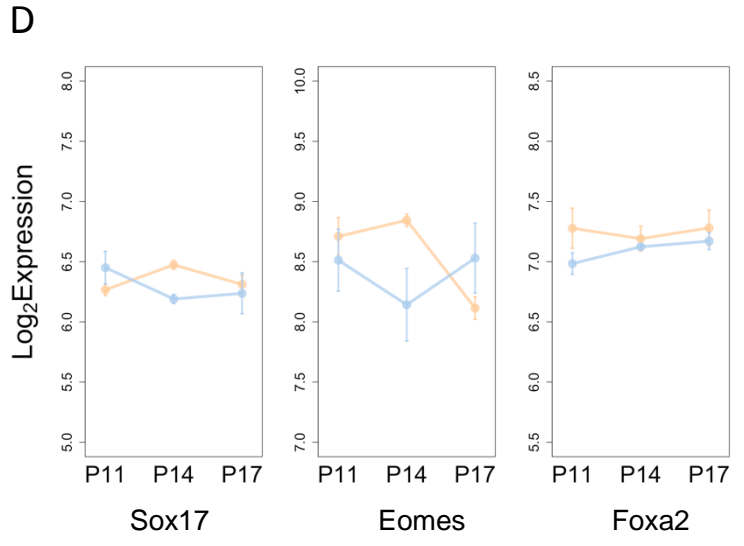
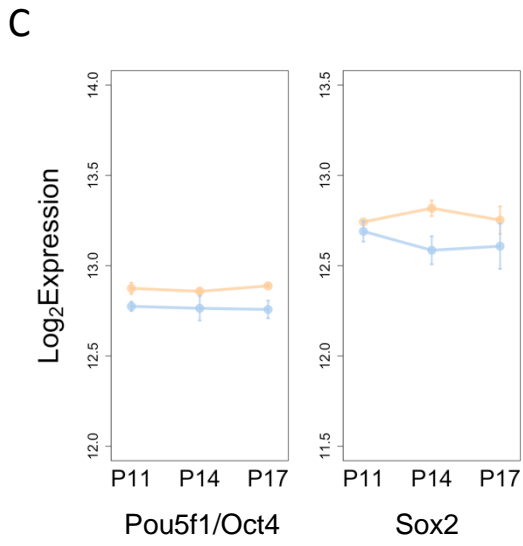
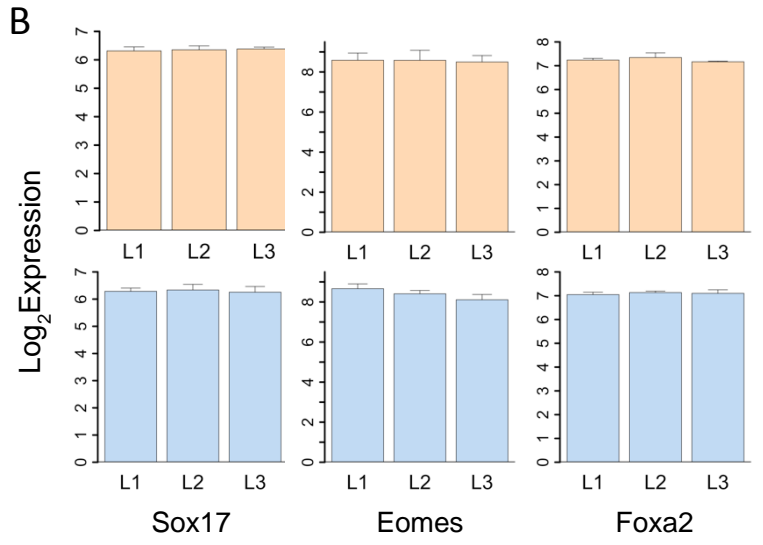
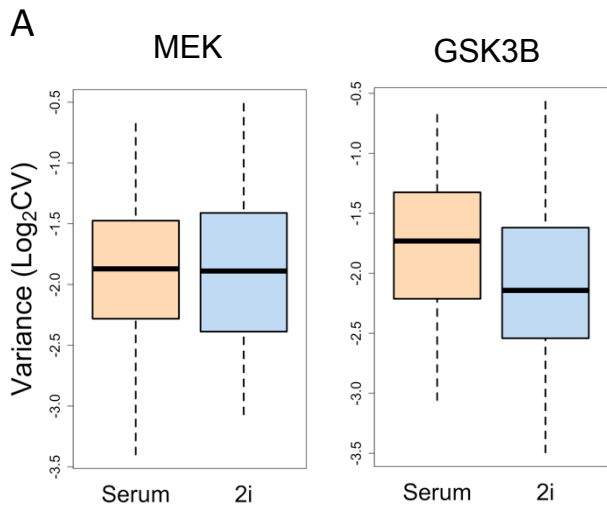
a, Schematic representation of variance measurement across populations. **b**, Variance (Log_2CV) for all genes in serum against all genes in 2i, showing skew towards 2i. **c**, Mean $\text{Log}_2\text{Expression}$ for all genes is not different between serum and 2i. **d**, qPCR shows independent validation of greater variance (Log_2CV) in 2i than serum. Candidate genes with significantly higher population heterogeneity in 2i ($FDR < 0.2$) were quantified ($n = 38$). **e**, MDS of pluripotency genes in serum, showing population samples separate across the first component which represents the major difference between samples. All serum population samples tested in the study were included ($n = 18$). **f**, In 2i ($n = 22$ samples), MDS of pluripotency genes shows samples clustering together, with lower variability than serum. **g**, Density distributions of *Nanog* showing heterogeneous expression in serum and more homogeneous expression in 2i for population data. **h**, Microarray gene expression for the miRNA biogenesis machinery in serum and 2i, * $FDR < 0.05$. **i**, GSEA showing RNA polymerase II travelling ratio is not significantly altered between serum and 2i for genes with greater 2i heterogeneity ($FDR > 0.05$).



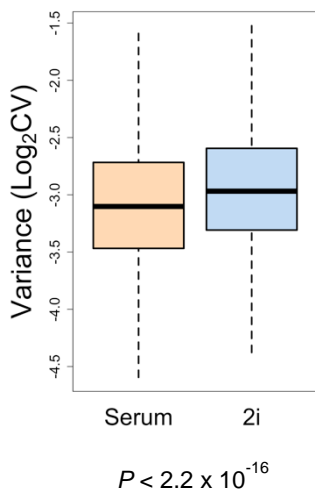
Supplementary Figure 2: Heterogeneity reflects dynamic gene expression, related to

Figure 3

a, Mean versus Log_2CV for all genes, showing greater 2i variance across expression levels for time-course experiments. **b**, Schematic landscape plot depicting clustering of samples by cell line (represented by shapes), which is expected to occur if distinct pluripotent states that are stable over time are captured during derivation in 2i, contrary to what is observed experimentally. **c**, Variance (Log_2CV) does not increase over time in serum or **d**, in 2i, but fluctuates with passage. **e**, MDS showing male and female PSCs do not cluster together indicating dynamic gene expression is a greater contributor to the transcriptome than sex.

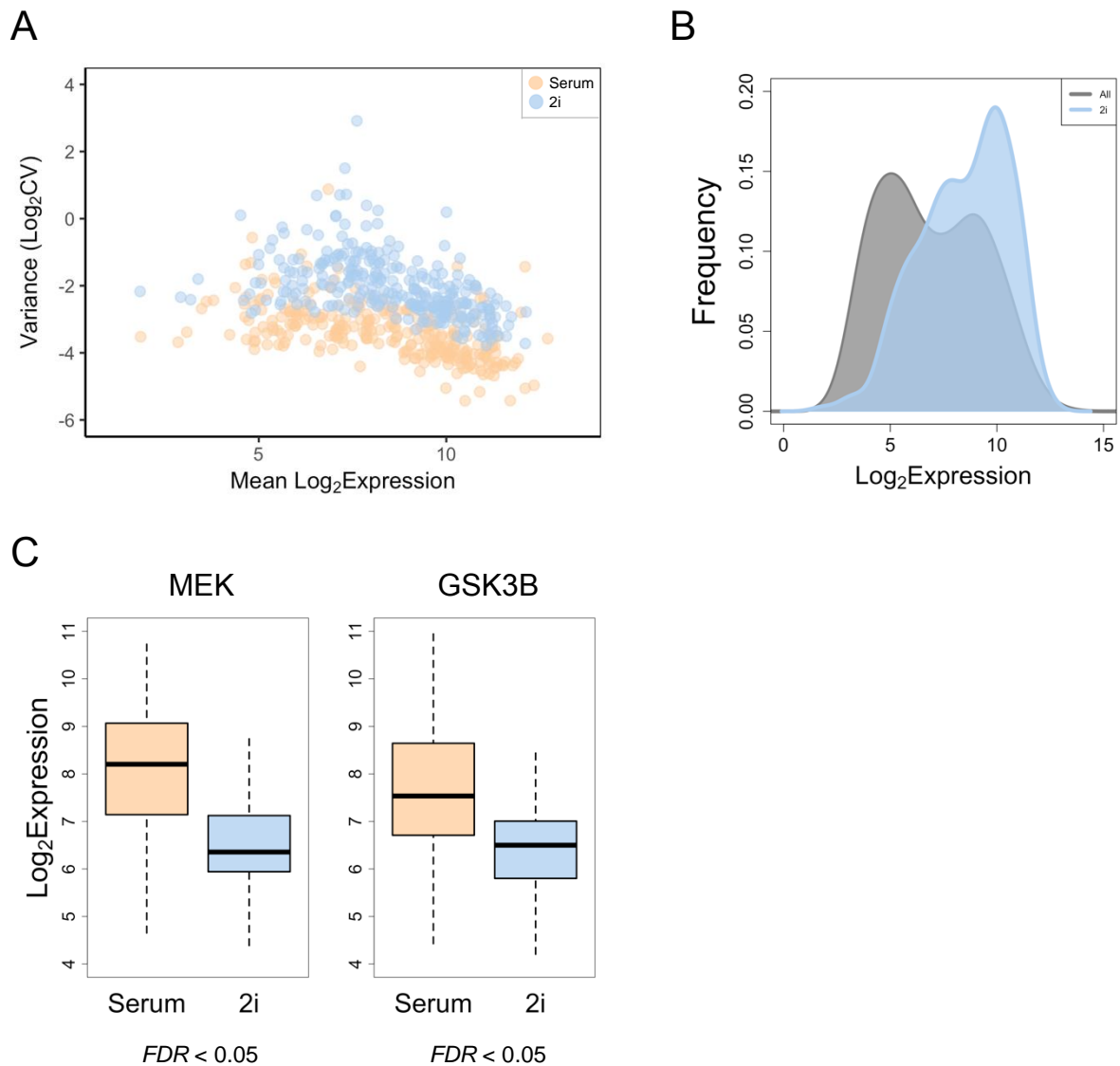


E Excluding cell subtype, sex and passage



Supplementary Figure 3: Heterogeneity reflects dynamic gene expression, related to Figure 3

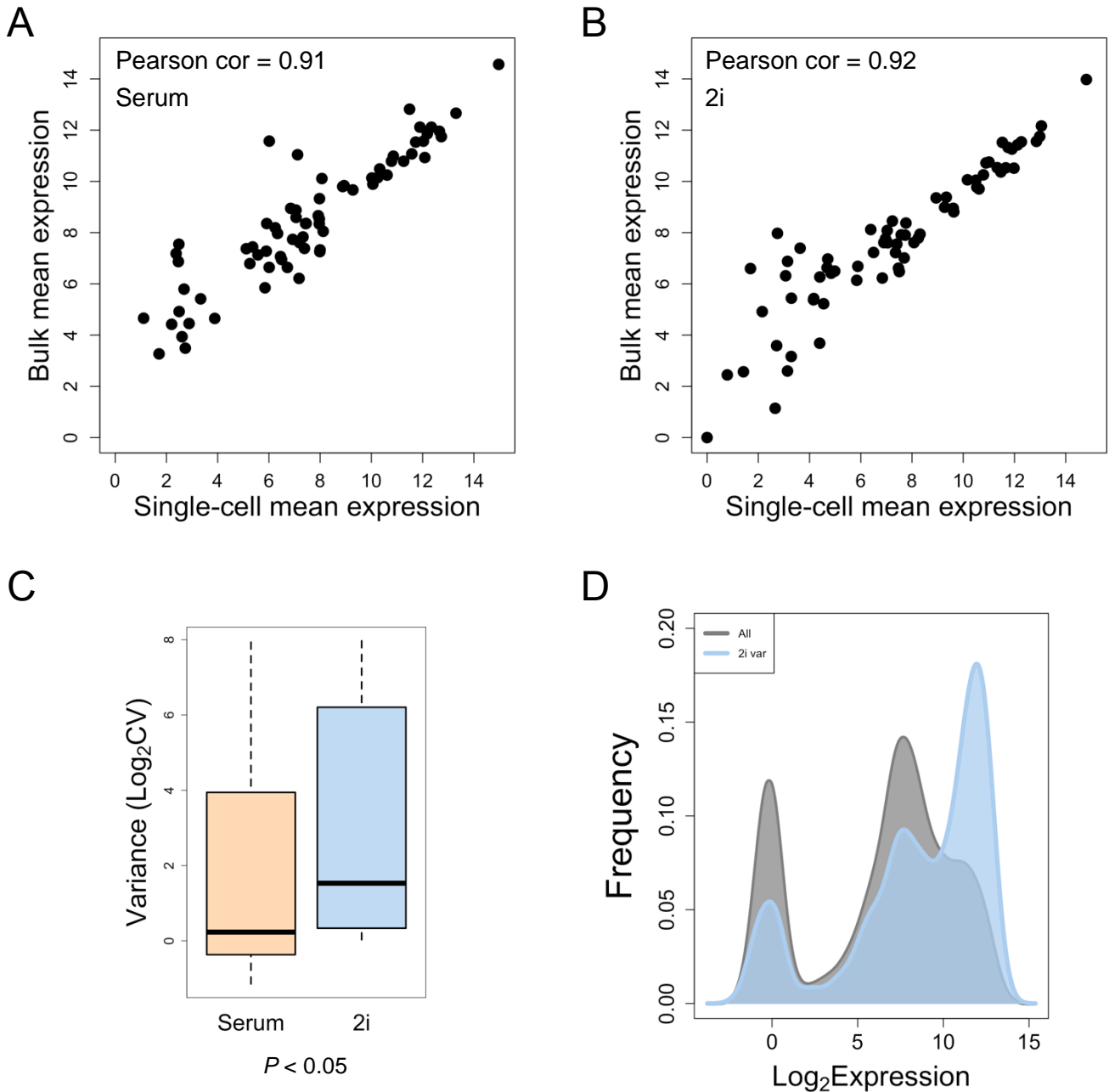
a, MEK ($n = 47$) and GSK3B ($n = 42$) pathways do not have increased heterogeneity in 2i showing heterogeneity is not due to experimental variation of inhibitor concentrations. **b**, Expression of developmental lineage markers does not significantly differ between cell lines indicating metastability does not bias individual cell lines towards a particular developmental lineage ($P > 0.05$, Kruskal-Wallis test). **c**, Pluripotency factor gene expression remains high over time. **d**, Expression of developmental regulators is not upregulated over time including primitive endoderm (*Sox17*), endoderm (*Eomes*) and mesoderm (*Foxa2*) lineage markers. **e**, Variance (Log_2CV) for all genes after regressing on sex, cell subtype and passage showing 2i heterogeneity remains; Wilcoxon signed rank test.



Supplementary Figure 4: Signalling pathways are heterogeneous in 2i, related to

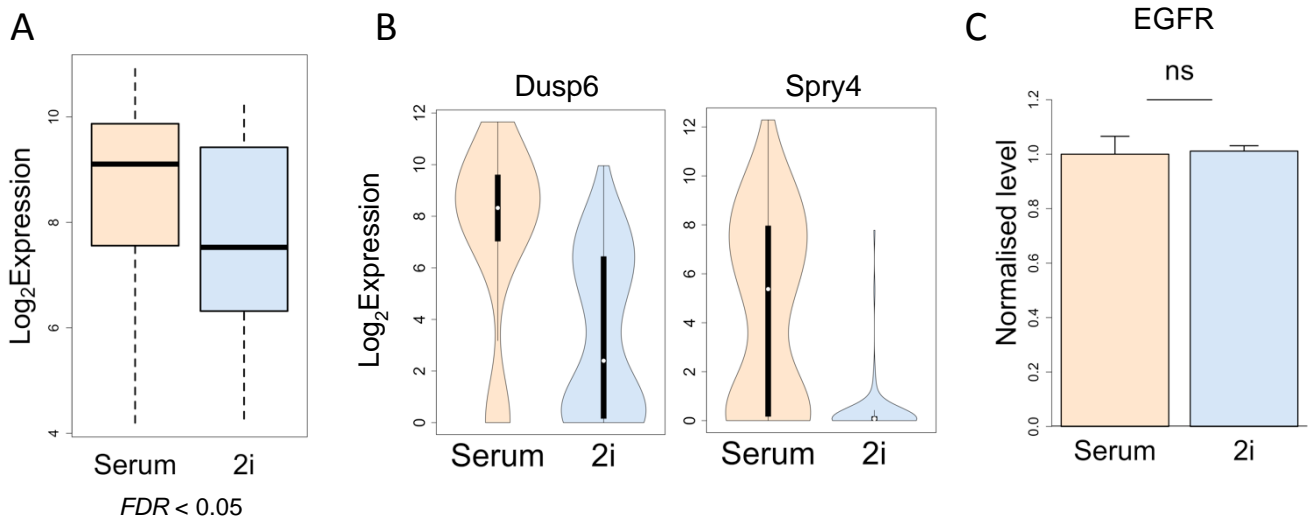
Figure 4

a, Mean versus Log₂CV for genes with significantly greater heterogeneity in 2i ($FDR < 0.05$, $n = 280$). **b**, Expression density distributions of genes with significantly greater heterogeneity in 2i than serum (blue) compared to all genes (gray). Non-expressed genes are not detected in 2i-specific heterogeneous genes and expression ranges from low to high. **c**, Expression levels of MEK ($n = 47$) and GSK3B ($n = 42$) target genes are reduced in 2i.



Supplementary Figure 5: Single-cell calcium signalling heterogeneity in 2i, related to Figure 5

a, Single-cell and bulk population mean expression are highly correlated in serum. **b**, Single-cell and bulk population mean expression are highly correlated in 2i. **c**, Variance of genes that have significantly increased heterogeneity in 2i at the single-cell level; Wilcoxon signed rank test ($n = 12$). **d**, Expression density distributions of genes that have significantly greater heterogeneity in 2i than serum at the single cell level (blue) compared to all genes tested. 2i variable genes show a tendency towards higher expression levels.



Supplementary Figure 6: Negative feedback is reduced in 2i, related to Figure 6

a, Expression of negative feedback regulators is reduced in 2i ($n = 20$). **b**, Mean expression levels of the negative feedback regulators *Dusp6* and *Spry4* are reduced at the single-cell level. **c**, Total EGFR is unchanged between 2i and serum. Two technical replicates and two biological replicates.