

Supplementary Information Guide

Population Snapshots Predict Early Hematopoietic and Erythroid Hierarchies

Betsabeh Khoramian Tusi^{1*}, Samuel Wolock^{2*}, Caleb Weinreb^{2*}, Yung Hwang¹, Daniel Hidalgo¹, Rapolas Zilionis², Ari Waisman³, Jun Huh⁴, Allon M. Klein^{2#} and Merav Socolovsky^{1#}

¹Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA

²Department of Systems Biology, Harvard Medical School, Boston, MA

³Institute for Molecular Medicine, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany

⁴Division of Immunology, Department of Microbiology and Immunobiology and Evergrande Center for Immunological Diseases, Harvard Medical School and Brigham and Women's Hospital, Boston, Massachusetts 02115, USA

*Contributed equally to this work

#Corresponding authors

Contents:

1. Supplementary Data: PBA code

This file contains the input data files and code for running Population Balance Analysis on the Bone Marrow and Fetal Liver data sets.

2. Supplementary Figure 1: gel source data for Extended Data Figures 5b, 9d, 10h

3. Supplementary Tables 1 to 7

Supplementary Table 1 Lineage-specific marker genes

List of marker genes for mature cells of each lineage and multipotent progenitors. These genes were used for visualizing cell types on SPRING plots and for identifying the most mature cells for Population Balance Analysis. Marker gene lists are sample-specific due to sample-to-sample differences in the detected lineages and the maturity of these lineages.

Supplementary Table 2 Genes correlating with branching computationally predicted fate probabilities in hematopoietic progenitor cells.

The table shows the process by which genes were selected, as well as their identity. Column 1 (Cell subset) shows the PBA gates used, and for each cell subset, Column 2 (Lineage bias score) indicates the score to which the genes identified were correlated. Columns 3-5 give the Gene symbol, its Correlation to the lineage bias score in the cell subset, and the P-value of the correlation. For example, "P(Er)*P(Ba) > 0.07" identifies cells computationally predicted to reside close to the Erythroid-Basophil branch point, as measured by the product of the PBA scores P(Er) and P(Ba). Within this subset genes were tested for correlation against the PBA score "P(Er)-P(Ba)". Thus genes with positive correlation to this score were Erythroid biased and genes with negative correlation were Basophil biased.

Supplementary Table 3 Dynamic changes in gene expression along the erythroid trajectory

List of significantly varying genes in the basal BM dataset. For each gene, maximal and minimal expression levels, and the corresponding locations along the MPP to erythroid progression, are shown. See methods section for calculation of FDR-corrected p value.

Supplementary Table 4 Gene set enrichment analysis (GSEA) of dynamic gene clusters

Genes assigned to each dynamic cluster (sheet 1), and GSEA results for signaling pathways and Gene Ontology terms (sheet 2) and transcription factor targets (sheet 3). Dynamic gene clusters are numbered as in **Extended Data Fig. 5e**. In the GSEA tables (sheets 2 and 3), enriched gene set titles (column B), corresponding corrected p value (column C) and gene hits (column D) are shown.

Supplementary Table 5 Genes differentially expressed in stress

Genes differentially expressed (DE) in each of the MPP, EBMP, EEP, CEP, and ETD stages are shown, in two separate sheets for eBM and FL, respectively. See methods section for calculation of the DE score (column C).

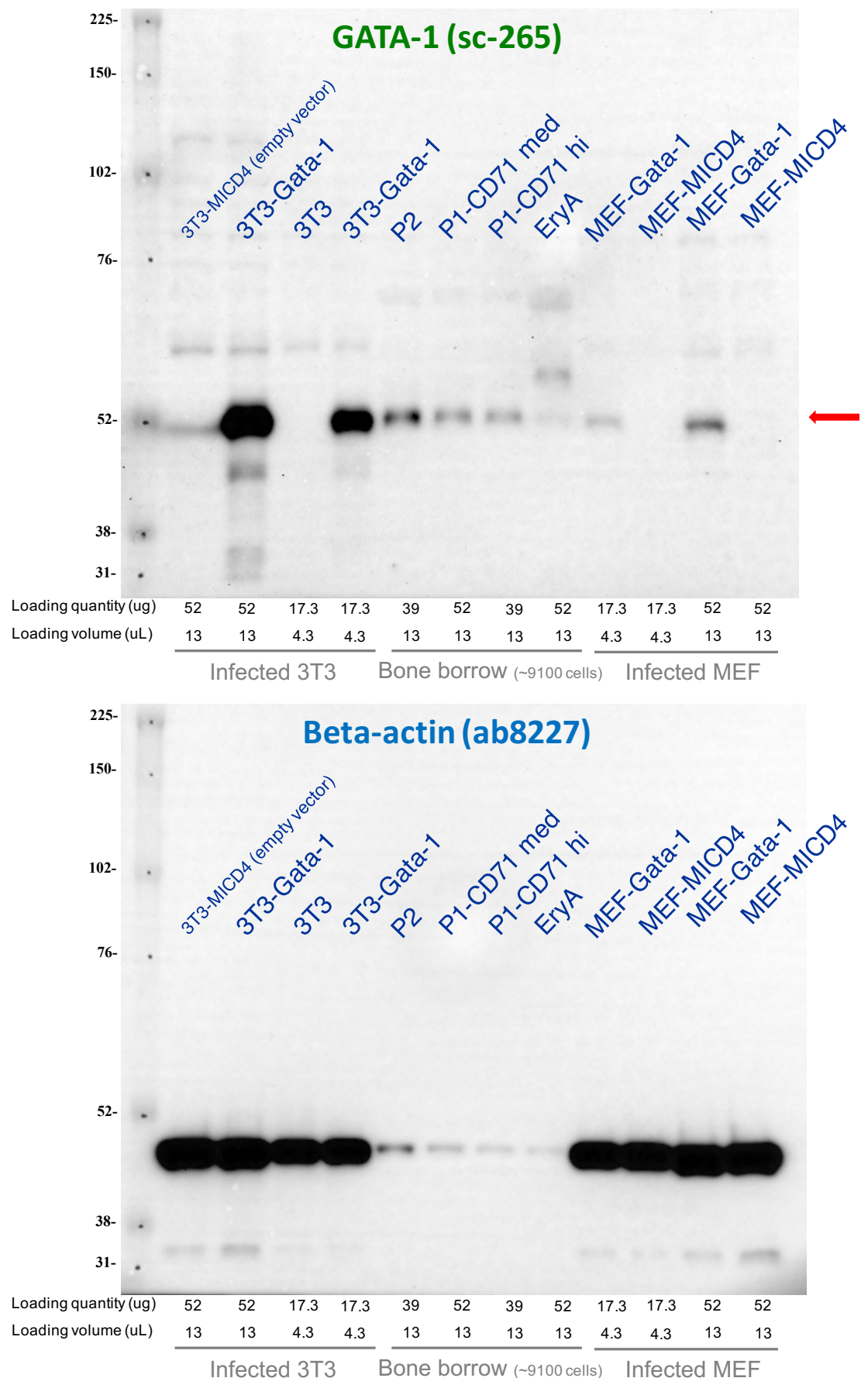
Supplementary Table 6 Genes correlated with progression through the CEP stage

For each gene expressed in the CEP stage, this table lists the “Linearity score”, a measure designed to identify genes that steadily increase or decrease with progression along the differentiation ordering (see Methods section “Identifying genes that change steadily in the CEP stage”). Large positive scores indicate genes that are strongly up-regulated in a linear fashion throughout the CEP stage; large negative scores indicate genes that are down-regulated linearly; and scores close to 0 indicate genes whose expression changes little or in a non-linear fashion.

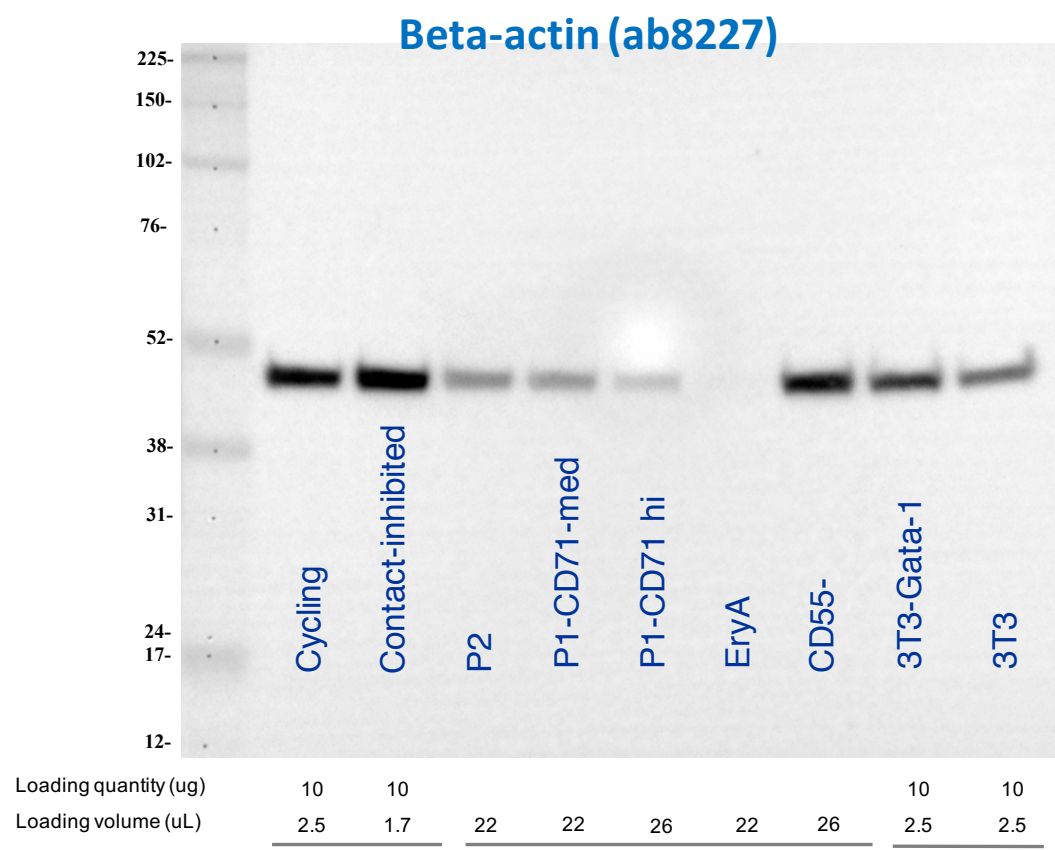
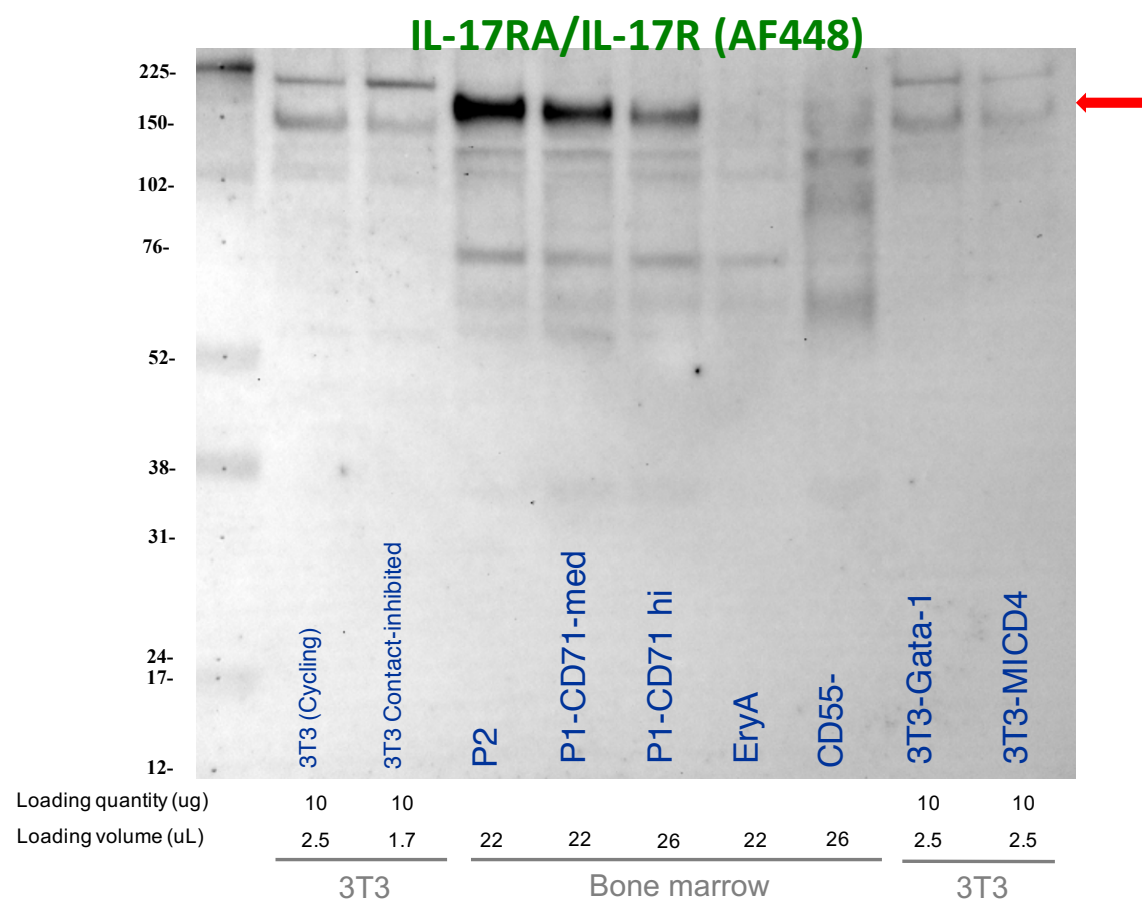
Supplementary Table 7 Modifications to the inDrops protocol

Detailed list of changes to the inDrops protocol by sample.

Extended Data
Figure 5b

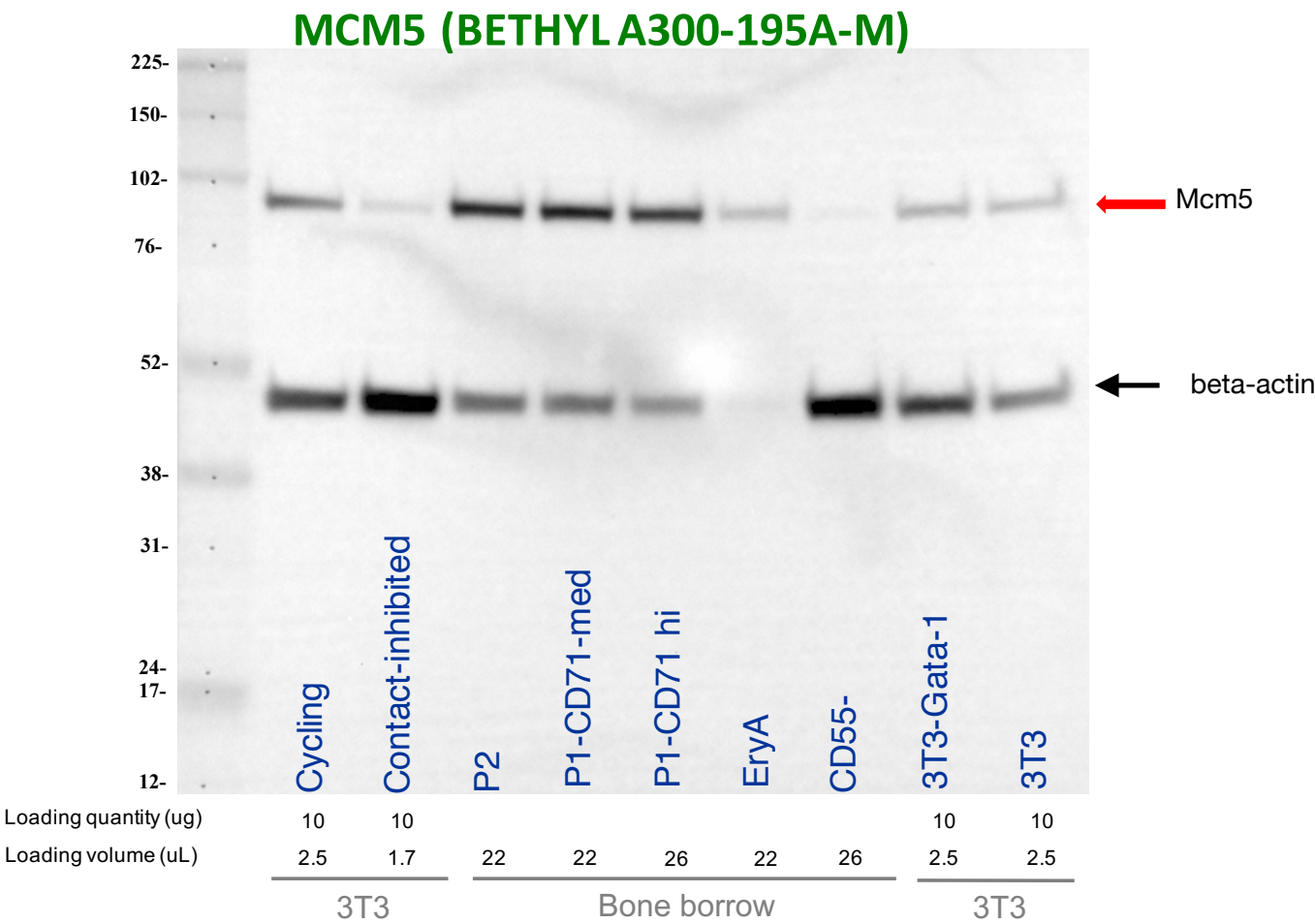


Extended
Data
Figure 9d

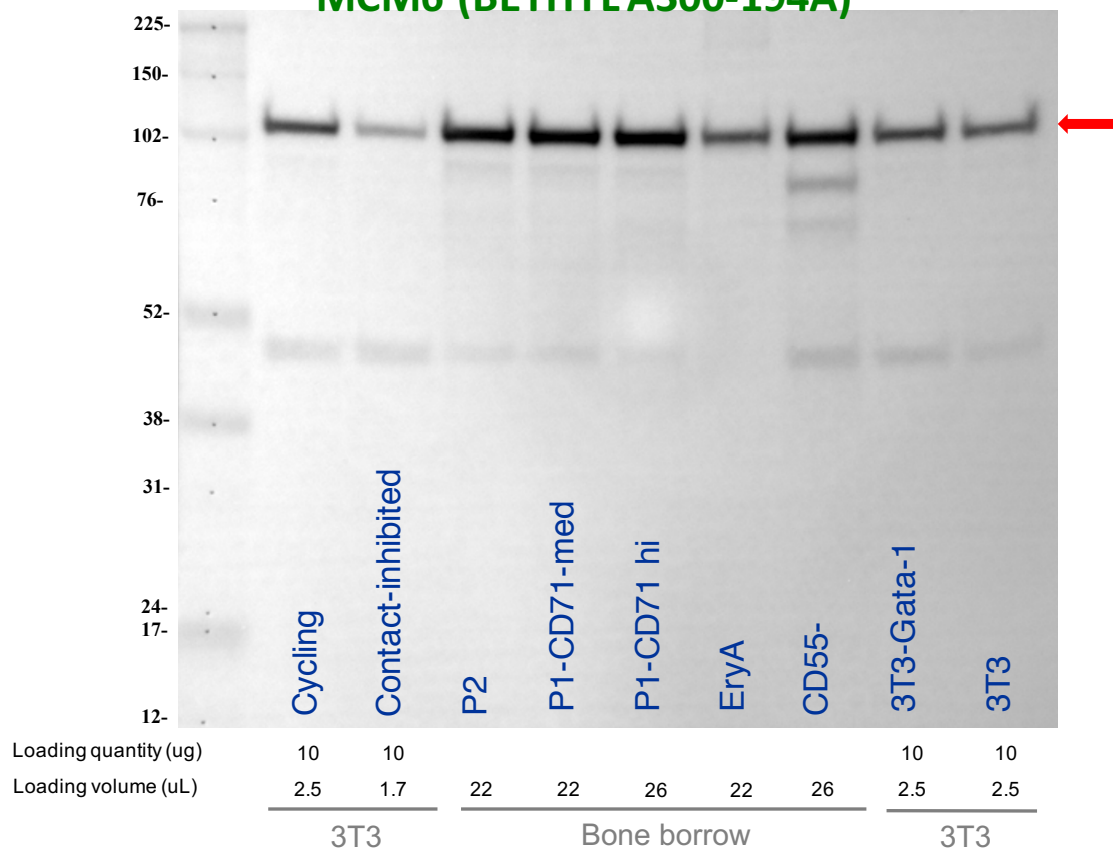


Extended Data Fig. 10h:

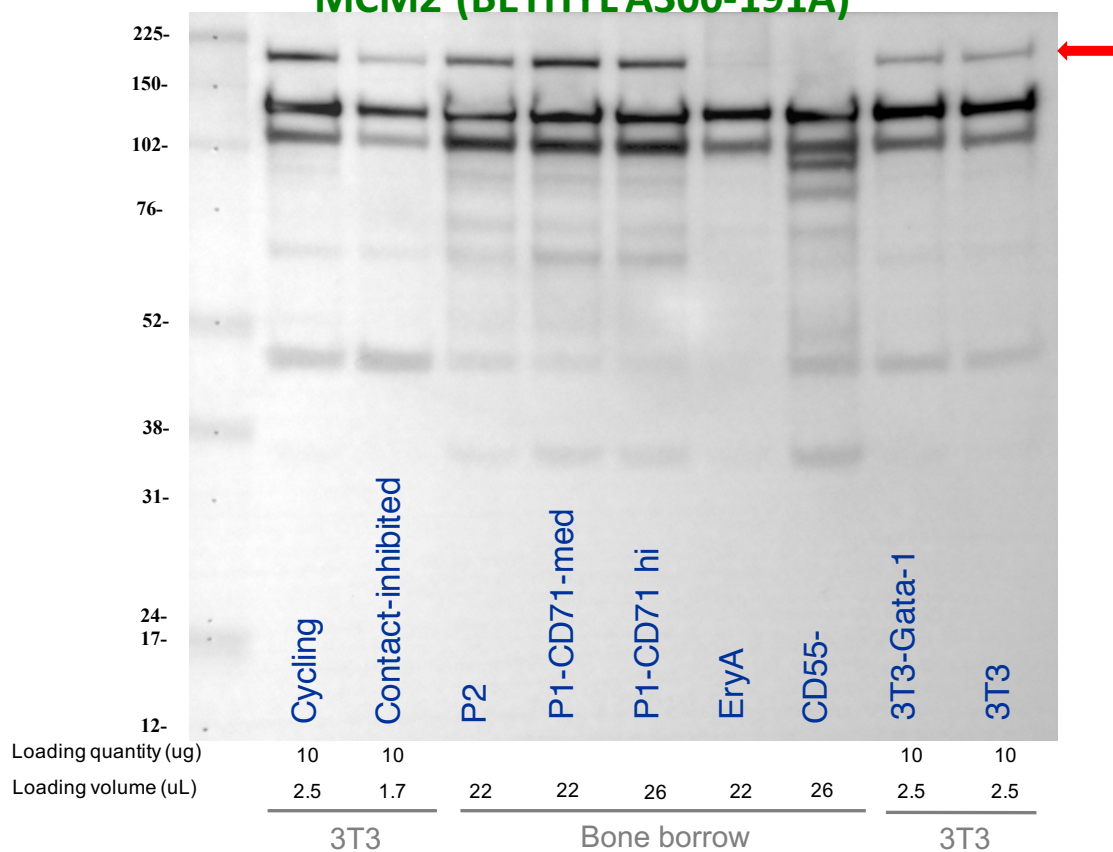
The same membrane as in Extended Data Fig. 9d was also probed for Mcm5, Mcm6, PCNA, Mcm2. (see above, for actin probe; the membrane was probed for Mcm5 without stripping the actin probe, which means that the actin bands are still visible following probing for Mcm5, below)

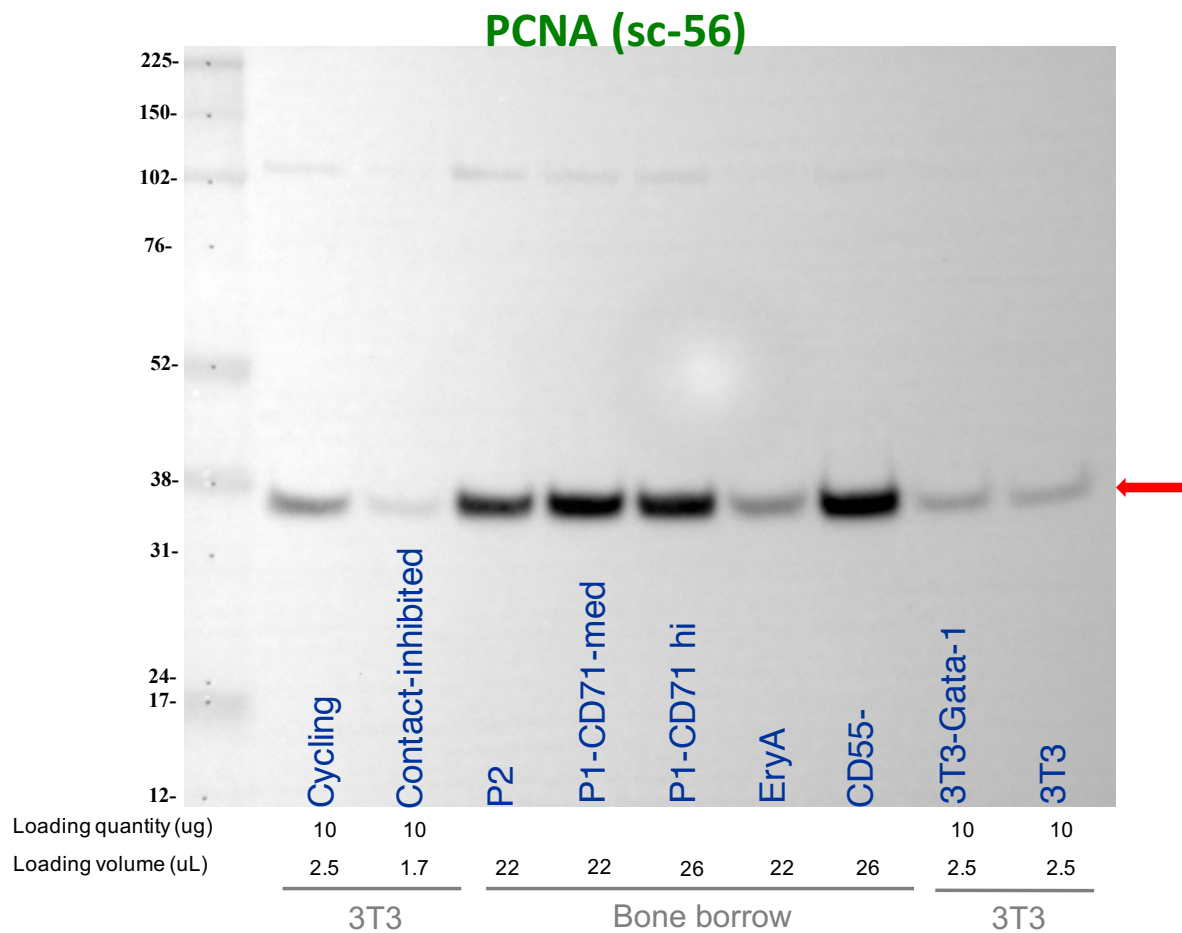


MCM6 (BETHYL A300-194A)



MCM2 (BETHYL A300-191A)





Supplementary Table 1

Sample	E	Ba	Meg	MPP	Ly	D	M	G	Eo	NK
bBM	Hbb-bt, Hba-a2, Hba-a1, Alas2, Bpgm	Lmo4, Ifitm1, Ly6e, Srgn	Pf4, Itga2b, Vwf, Pbx1, Mef2c	Hlf, Gcnt2	Cd79a, Igl1, Vpreb3, Vpreb1, Lef1	H2-Aa, Cd74, H2-Eb1, H2-Ab1, Cst3	Csf1r, Ly6c2, Ccr2	Lcn2, S100a8, Ltf, Lyz2, S100a9	NA	NA
eBM	Hbb-bt, Hba-a2, Hba-a1, Alas2, Bpgm	Lmo4, Ifitm1, Ly6e, Srgn	Pf4, Itga2b, Vwf, Pbx1, Mef2c	Hlf, Gcnt2	Cd79a, Igl1, Vpreb3, Vpreb1, Lef1	NA	Csf1r, Ly6c2, Ccr2	Lcn2, S100a8, Ltf, Lyz2, S100a9	NA	NA
FL	Hbb-bt, Hba-a2, Hba-a1, Alas2, Bpgm	Lmo4, Ifitm1, Ly6e, Srgn	Pf4, Itga2b, Vwf, Pbx1, Mef2c	Hlf, Gcnt2	Ccr9, Mzb1, Lsp1, Flt3, Ccr7, Lck	NA	Csf1r, Ly6c2, Ccr2	Elane, Mpo	NA	NA
P1-P5	Car2, Rhd, Tfrc	Cpa3, Lmo4, Srgn, Cd53	Pf4, Itga2b, Vwf, Pbx1, Mef2c	Hlf, Gcnt2	NA	NA	Csf1r, Ly6c2, Ccr2	Elane, Prtn3	Epx, Prg2	NA
Paul et al.	Alad, Hba-a2, Rhd, Cpox	Prss34, Alox5, Mcpt8, Lmo4	Pf4, Vwf, Pbx1	Hlf, Gcnt2	Dntt, Flt3	H2-Aa, Cd74, H2-Eb1, H2-Ab1, Cst3	Csf1r, Ly6c2, Ccr2, F13a1, Slpi	Mpo, Elane, Gstm1	NA	Ccl5, Shisa5, Lck, Ctsw, Nkg7
Nestorowa et al.	Hbb-bt, Hba-a2, Hba-a1, Alas2, Bpgm	Lmo4, Ifitm1, Ly6e, Srgn	Pf4, Itga2b, Vwf, Pbx1, Mef2c	Hlf, Gcnt2, Procr	Cd79a, Igl1, Vpreb3, Vpreb1, Lef1	NA	Csf1r, Ly6c2, Ccr2	Mpo, Elane	NA	NA

Supplementary Table 2: List of transcription factors, receptors and chromatin modifiers correlated with fate bias at key decision points.

Cell subset	Lineage bias score	Gene symbol	Correlation	P-value
P(Er)*P(Ba) > 0.07	P(Er)-P(Ba)	Lmo4	-0.58	1.00E-21
P(Er)*P(Ba) > 0.07	P(Er)-P(Ba)	Gata2	-0.44	1.00E-11
P(Er)*P(Ba) > 0.07	P(Er)-P(Ba)	Cebpa	-0.33	1.00E-06
P(Er)*P(Ba) > 0.07	P(Er)-P(Ba)	Klf1	0.33	1.00E-06
P(Er,Ba)*P(Mk) > 0.02	P(Er,Ba)-P(Mk)	Pbx1	-0.49	1.00E-33
P(Er,Ba)*P(Mk) > 0.02	P(Er,Ba)-P(Mk)	Mpl	-0.38	1.00E-19
P(Er,Ba)*P(Mk) > 0.02	P(Er,Ba)-P(Mk)	Mef2c	-0.27	1.00E-10
P(Er,Ba)*P(Mk) > 0.02	P(Er,Ba)-P(Mk)	Myb	0.28	1.00E-10
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Gfi1	-0.51	1.00E-80
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Cebpa	-0.42	1.00E-52
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Baz1a	-0.39	1.00E-44
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Cebpb	-0.36	1.00E-36
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Mybl2	-0.35	1.00E-35
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Spi1	-0.34	1.00E-33
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Irf8	-0.34	1.00E-32
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Cebpe	-0.33	1.00E-31
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Xbp1	-0.33	1.00E-31
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Cited4	-0.33	1.00E-30
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Hmgn2	-0.29	1.00E-24
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Ostf1	-0.28	1.00E-23
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Cbx4	-0.28	1.00E-22
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Atf4	-0.28	1.00E-22
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Csf1r	-0.28	1.00E-22
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Gtf2e2	-0.27	1.00E-20
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Arhgap17	-0.26	1.00E-20
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Mtf1	-0.26	1.00E-19
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Hmga1	-0.26	1.00E-18
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Mcm5	-0.26	1.00E-18
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Cited2	-0.25	1.00E-18
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Mef2d	-0.25	1.00E-18
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Lyl1	-0.25	1.00E-18
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Gata2	0.45	1.00E-59
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Gata1	0.38	1.00E-42
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Zfp1	0.36	1.00E-37
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Pbx1	0.31	1.00E-27
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Mpl	0.29	1.00E-23
P(Ba,Er,Mk)*P(G,D,Mo,Ly) > 0.07	P(Ba,Er,Mk)-P(G,D,Mo,Ly)	Ldb1	0.28	1.00E-22

$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Gfi1	-0.38	1.00E-20
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Xbp1	-0.36	1.00E-18
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Baz1a	-0.34	1.00E-15
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Cebpa	-0.33	1.00E-15
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Cebpb	-0.32	1.00E-14
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Cited4	-0.31	1.00E-13
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Cited2	-0.3	1.00E-12
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Hmgn2	-0.29	1.00E-11
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Cebpe	-0.28	1.00E-10
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Mcm5	-0.27	1.00E-10
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Ostf1	-0.26	1.00E-09
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Phf10	-0.26	1.00E-09
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Dazap2	-0.26	1.00E-09
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Hmga1	-0.25	1.00E-09
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Satb1	0.43	1.00E-25
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Sox4	0.41	1.00E-22
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Mta3	0.41	1.00E-22
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Foxp1	0.26	1.00E-09
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Ebf1	0.26	1.00E-09
$P(\text{Ly}) * P(\text{Mo}, \text{G}, \text{D}) > 0.07$	$P(\text{Ly}) - P(\text{Mo}, \text{G}, \text{D})$	Hlf	0.26	1.00E-09
$P(\text{Mo}) * P(\text{G}) > 0.07$	$P(\text{Mo}) - P(\text{G})$	Gfi1	-0.5	1.00E-21
$P(\text{Mo}) * P(\text{G}) > 0.07$	$P(\text{Mo}) - P(\text{G})$	Myb	-0.43	1.00E-15
$P(\text{Mo}) * P(\text{G}) > 0.07$	$P(\text{Mo}) - P(\text{G})$	Cebpe	-0.38	1.00E-12
$P(\text{Mo}) * P(\text{G}) > 0.07$	$P(\text{Mo}) - P(\text{G})$	Baz1a	-0.29	1.00E-07
$P(\text{Mo}) * P(\text{G}) > 0.07$	$P(\text{Mo}) - P(\text{G})$	Irf8	0.74	1.00E-57
$P(\text{Mo}) * P(\text{G}) > 0.07$	$P(\text{Mo}) - P(\text{G})$	Csf1r	0.44	1.00E-16
$P(\text{Mo}) * P(\text{G}) > 0.07$	$P(\text{Mo}) - P(\text{G})$	Pou2f2	0.3	1.00E-07
$P(\text{Mo}) * P(\text{G}) > 0.07$	$P(\text{Mo}) - P(\text{G})$	Klf4	0.26	1.00E-05
$P(\text{Mo}) * P(\text{G}) > 0.07$	$P(\text{Mo}) - P(\text{G})$	Aes	0.25	1.00E-05

Supplementary Table 7

Sample set	Sequence of RT primers on hydrogel beads	Exo/Hinfl treatment (step 137 in (1))	PE2-N6 primer sequence (step 151 in (1))	PCR primer sequences (steps 157 and 160 in (1))	Custom sequencing primers?	Sequencing platform
Batch 1 of Kit+ samples: Basal bone marrow (bBM), +Epo bone marrow (eBM), fetal liver (FL)	5'- CGATGACGTAATACGA CTCACTATAGGG ATACCACCATGGCTCT TTCCTACACGACGCT CTCCGATC[bc1,8- 11nt]GAGTGATTGCTTG TGACGCCTT[bc2,8nt]N NNNNNTTTTTTTTTT TTTTTTTV-3' (same as in (1, 2))	37 °C for 60 min heat inactivation at 80 °C for 10 min (same as in (2))	5'- TCGGCATTCCT GCTGAACCGCT CTTCCGATC TNNNNN-3' (same as in (1))	5'- AATGATACGGCGACCACC GAGATCTACACTCTTTCCC TACACGA-3' 5'- CAAGCAGAAGACGGCATA CGAGATCGGTCTCGGCAT TCCTGCTGAAC-3' (same as in (2))	No	HiSeq 2000
Batch 2 of Kit+ samples	Same as above and in (1, 2)	37 °C for 30 min (same as in (1))	Same as above and in (1)	5'- CAAGCAGAAGACGGCATA CGAGATXXXXXCTCTTTC CCTACACGA-3' 5'- AATGATACGGCGACCACC GAGATCTACACGGTCTCG GCATTCTGCTGAAC-3' (same as in (1))	Read1: 5'- GGCATTCCTGCTG AACCGCTCTT CCGATCT -3' Index read: 5'- AGATCGGAAGAG CGTCGTGTAG- GGAAAGAG -3' Read2: 5'- CTCTTTCCTACA CGACGCTCTT CCGATCT -3' (same as in (1))	NextSeq 500
FACS subsets P1, P1-CD71hi, P2, P3, P4, P5	5'- CGATGACGTAATACGA CTCACTATAGGGTGTG GGGTGCAG[bc1,8nt]GT CTCGTGGGCTCGGAG ATGTGTATAAGAGACA G[bc2,8nt]NNNNNNNTT TTTTTTTTTTTTTTTV- 3'	Same as above and in (1)	5'- TCGTGCGCAGC ATCAGATGTGT ATAAGAGACAG NNNNN-3'	5'- AATGATACGGCGACCACC GAGATCTACACXXXXXX TCGTGCGCAGCGTC-3'	No	NextSeq 500

1. Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, et al. Single-cell barcoding and sequencing using droplet microfluidics. Nat Protoc. 2017;12(1):44-73.
2. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;161(5):1187-201.