

1
2
3
4
5
6
7 **The Myogenesis Program Drives Clonal Selection and Drug Resistance in**
8 **Rhabdomyosarcoma**
9
10
11
12

13 Anand G. Patel^{1,2*}, Xiang Chen^{3*}, Xin Huang^{3*}, Michael R. Clay⁴, Natalia Komorova⁵,
14 Matthew J. Krasin⁶, Alberto Pappo², Heather Tillman⁷, Brent A. Orr⁷, Justina McEvoy¹, Brittney
15 Gordon¹, Kaley Blankenship², Colleen Reilly¹, Xin Zhou³, Jackie L. Norrie¹, Asa Karlstrom¹,
16 Jiyang Yu³, Dominik Wodarz⁸, Elizabeth Stewart^{1,2} and Michael A. Dyer¹
17
18
19
20

21 Departments of ¹Developmental Neurobiology, ²Oncology, ³Computational Biology, ⁶Radiation
22 Oncology and ⁷Pathology at St. Jude Children's Research Hospital, Memphis, Tennessee 38105,
23 USA
24

25 ⁴Department of Pathology, University of Colorado School of Medicine, Aurora, Colorado 80045,
26 USA
27

28 ⁵Department of Mathematics, University of California, Irvine, California 92697, USA
29

30 ⁸Department of Population Health and Disease Prevention, Program in Public Health, Susan and
31 Henry Samueli College of Health Sciences, University of California, Irvine, California 92697
32 USA
33
34

35 *These authors contributed equally
36
37
38

39 Correspondence and requests for materials should be addressed to:
40 Michael A. Dyer
41 Department of Developmental Neurobiology, MS-323
42 St. Jude Children's Research Hospital
43 262 Danny Thomas Place, Memphis, TN 38105-3678, USA
44 Phone: (901) 595-2257
45 Fax: (901) 595-3143
46 E-mail: michael.dyer@stjude.org

1 Abstract

2 Rhabdomyosarcoma (RMS) is a pediatric cancer with features of skeletal muscle; patients with
3 unresectable or metastatic RMS fare poorly due to high rates of disease recurrence. Here, we use
4 single cell and single nucleus RNA-sequencing to show that RMS tumors recapitulate the
5 spectrum of embryonal myogenesis. Using matched patient samples from a clinical trial and
6 orthotopic patient-derived xenografts (O-PDXs), we show chemotherapy eliminates the most
7 proliferative component with features of myoblasts; after treatment, the quiescent immature
8 population with features of paraxial mesoderm expands to reconstitute the developmental
9 hierarchy of the original tumor. We discovered that this paraxial mesoderm population is
10 dependent on EGFR signaling and is sensitive to EGFR inhibitors. Taken together, this data
11 serves as a proof-of-concept that targeting each developmental state in RMS is an effective
12 strategy for improving outcomes by preventing disease recurrence.

1 **Introduction**

2 Rhabdomyosarcoma (RMS) is the most common pediatric soft tissue sarcoma and has molecular,
3 cellular and histopathologic features of developing skeletal muscle¹⁻³. The alveolar form of RMS
4 (ARMS) is more differentiated than the embryonal form (ERMS) and each subtype has distinct
5 genomic and epigenomic landscapes^{2,4,5}. For newly diagnosed RMS patients, the overall survival
6 rate is 70% using multiagent chemotherapy combined with radiation and/or surgical resection^{6,7}.
7 Unfortunately, a subset of patients experience disease recurrence after treatment completion; for
8 those patients, overall survival rate drops below 20%⁸. Genomic studies have shown that clonal
9 selection occurs with disease recurrence, but no recurrent genetic lesion has been identified that
10 contributes to survival of the rare clones of cells for RMS^{2,5,9}. This raises the possibility that
11 other, non-genetic mechanisms may contribute to drug resistance and disease recurrence in RMS.
12 To explore this possibility, we performed single cell (sc) and single nucleus (sn) RNA-seq of
13 RMS patient tumors and matched orthotopic patient-derived xenografts (O-PDXs). We also
14 performed lentiviral barcode labeling to trace the clonal expansion of individual tumor cells
15 during normal growth and in response to treatment. Taken together, these studies showed that
16 individual tumor cells transition through myogenesis and the underlying myogenic
17 developmental hierarchy contributes to clonal selection with treatment. We used the
18 developmental program in RMS to identify therapeutic vulnerabilities that could be exploited to
19 reduce disease recurrence. Overall, this study reveals a developmental hierarchy with RMS and
20 introduces a novel approach to treating pediatric cancers, wherein targeting specific
21 developmental states that are destined to persist during therapy can be used to improve treatment
22 efficacy.

23

1 **Results**

2 **RMS tumors have developmental heterogeneity**

3 Skeletal muscle develops from the mesodermal cells of the somites during embryogenesis and
4 undergoes stepwise differentiation, which is typified by the expression of myogenic regulatory
5 factors^{10,11} (MRFs; Fig. 1A,B). RMS tumors have features of skeletal muscle including
6 myofibers and heterogenous expression of proteins such as myogenin (MYOG)^{3,12}. To further
7 investigate the transcriptomic heterogeneity within RMS, we performed droplet-based single-cell
8 RNA-sequencing (scRNA-seq) (Extended Data Tables 1 and 2). We obtained fresh ERMS and
9 ARMS patient tumor tissue (Fig. 1C,D) following surgical resection and generated single-cell
10 suspensions (>90% viable cells) for 3'-directed scRNA-seq (10x Genomics). Inference of
11 somatic copy number alterations was used to distinguish malignant cells from non-malignant
12 cells^{13,14} (Extended Data Fig. 1).

13 Single-cell analysis showed there were distinct populations of cells expressing
14 transcription factors characteristic of paraxial mesoderm (*MEOX2*, *PAX3*), myoblasts (*MYF5*,
15 *MSC*) and myocytes (*MYOG*, *MEF2C*) (Fig. 1E,F and data not shown). The proportion of *MYOG*
16 expressing cells in the scRNA-seq data was consistent with the proportion measured by
17 immunohistochemical staining (IHC) (Fig. 1C-F). The ARMS sample had fewer tumor cells
18 expressing the early paraxial mesoderm MRF *MEOX2* (2.1%) than the ERMS sample (29.4%),
19 and more cells expressing the late myocyte MRF *MYOG* (75.2% versus 25.6%; Fig. 1E,F). RNA
20 velocity analysis, which leverages the simultaneous measurement of spliced and unspliced RNA
21 transcripts to generate a model of the future state of cells¹⁵, showed unidirectional transit of cells
22 from the paraxial mesoderm through myoblast to the myocyte state in the ERMS tumor (Fig.

1 1G,H). Non-malignant cells including monocytes, fibroblasts, lymphocytes and vascular
2 endothelial cells were readily identifiable in our scRNA-seq dataset (Extended Data Fig. 1).

3 The rarity of childhood cancers limits the ability to obtain fresh tissue samples for
4 scRNA-seq. To increase the number of evaluable tumors, we validated single-nucleus RNA-
5 sequencing (snRNA-seq) of frozen tumor tissue and adapted our computational pipeline to
6 accommodate data generated from snRNA-seq¹⁴. Specifically, we compared scRNA-seq from
7 fresh tumors (SJRHB030680_R1 and SJRHB031320_D1) (Fig. 1E,F) to snRNA-seq of matched
8 frozen tumor specimens (Extended Data Fig. 2). As shown previously for neuroblastoma¹⁴, we
9 were able to recover more cells of the tumor microenvironment (TME) by snRNA-seq compared
10 to data generated by scRNA-seq (Extended Data Fig. 2). To extend our single cell transcriptional
11 profiling, we performed snRNA-seq on 18 RMS tumors (12 ERMS and 6 ARMS) (Extended
12 Data Tables 1 and 2). In total, 122,731 nuclei were analyzed from the patient tumors. As for the
13 fresh tumors, copy number inference was used to distinguish malignant nuclei (111,474) from
14 the normal nuclei (11,257) in the TME. The malignant nuclei were integrated using Conos, an
15 approach that leverages inter-sample mappings to generate a unified graph for the identification
16 of communal cell clusters¹⁶ (Fig. 2A). Leiden clustering identified 7 clusters, which we
17 combined into 1 mesoderm, 4 myoblast and 2 myocyte signature groups based on expression of
18 MRFs (Fig. 2B). The 2 myocyte populations were distinguished by expression of genes involved
19 in muscle differentiation and function (Extended Data Table 3). The 4 myoblast populations
20 were distinguished by ribosomal genes ($p=4.3\times 10^{-40}$) and muscle differentiation genes
21 ($p=0.0005$) (Extended Data Table 3). We identified 954 differentially expressed genes, of which
22 945 were cluster-type specific (Extended Data Table 3). Extracellular matrix and cell adhesion
23 pathways were enriched in the paraxial mesoderm-like tumor cells, ribosome biosynthesis

1 pathways were enriched in the myoblast-like cells and pathways involved in muscle function
2 were enriched in the myocyte-like cells (Extended Data Table 3). While all the tumors had a
3 mixture of cells with mesoderm, myoblast, and myocyte signatures, ARMS tumors contained
4 significantly fewer cells with the mesodermal gene expression signature ($p=0.008$; unpaired t-
5 test) and were skewed towards the myocyte signature (Fig. 2C and Extended Data Fig. 3). One
6 ERMS tumor (SJRHB010928_R1) was notable in that it contained a majority (97%) of tumor
7 cells with the mesodermal signature (Extended Data Fig. 3A,B). This sample was collected
8 during treatment (Extended Data Table 1) suggesting that mesodermal cells are more resistant to
9 treatment than the other cell populations. The proliferating cells were significantly enriched in
10 the myoblast population ($p<0.0001$; one-way ANOVA with multiple comparisons) (Fig. 2E,F).
11 All data can be viewed using an interactive viewer at: [https://pecan.stjude.cloud/static/RMS-
12 scRNA-atlas-2020/](https://pecan.stjude.cloud/static/RMS-scRNA-atlas-2020/).

13 The same approach was used to cluster the non-malignant cells within the TME
14 (Extended Data Figure 3C-F). Comparing normal cell populations between ERMS and ARMS
15 showed that fibroblasts in ARMS were significantly enriched in pathways involved in
16 extracellular matrix synthesis and organization as well as cell adhesion. In addition, *SFRP2* and
17 *SFRP4* were significantly ($p<1\times10^{-90}$) enriched in fibroblasts from ARMS (45% and 56% of
18 cells, respectively) relative to ERMS (1% and 3%, respectively) (Extended Data Table 4). A
19 previous pan-cancer analysis showed that *SFRP2* and *SFRP4* represent a tightly regulated
20 transcriptional program in cancer stroma that correlates with poor prognosis, EMT and
21 angiogenesis across multiple cancers¹⁷. The *HLA-A,B,C,E* and *B2M* and *CD74* genes were
22 significantly upregulated in lymphocytes from ARMS and *HLA-DRA*, *DRB1* and *DPB1* were
23 significantly upregulated in monocytes from ERMS (Extended Data Table 4).

1 We next investigated the spatial heterogeneity of malignant subpopulations using single
2 and multiplex immunohistochemistry (IHC) on 12 patient tumor specimens. Consistent with our
3 transcriptomic findings, there was heterogenous expression of MEOX2, MYF5 and MYOG
4 protein (Fig. 2G). The proportion of immunopositive cells were correlated with the proportion of
5 each population from the sc/snRNA-seq (Fig. 2H). Double IHC showed that these proteins were
6 expressed in a mutually exclusive pattern consistent with the distinct clusters of mesoderm,
7 myoblast and myocyte populations in RMS tumors from sc/snRNA-seq (Fig. 2I,J and data not
8 shown).

9

10 **Developmental indexing of RMS using embryonic snRNA-seq data**

11 To extend our analysis of the developmental trajectory of RMS beyond MRFs, we analyzed our
12 RMS data within the context of early muscle development using a single-nucleus atlas of
13 organogenesis from mouse embryos at E9.5, E10.5, E11.5, E12.5, and E13.5¹⁸. We extracted
14 data from the skeletal muscle lineage and performed trajectory analysis on half of the data to
15 generate a training dataset (Fig. 3A-D). We then adapted latent cellular analysis (LCA)¹⁹ to
16 calculate the similarity in the latent cellular space between cells in the remaining half of the
17 skeletal muscle dataset to cells used for training; a normalized muscle developmental index was
18 then calculated for each nucleus within the validation dataset (Fig. 3E,F). The developmental
19 index increased with embryonic age as expected within the validation dataset (Fig. 3E,F).

20 Using this unsupervised developmental indexing approach, we confirmed that individual
21 RMS tumors have cellular heterogeneity that reflects normal myogenesis. For example, in
22 SJRHB030680_R1, an ERMS tumor, we identified a broad range of developmental indices
23 within the malignant components of the tumor (Fig. 3G). In contrast, in SJRHB031320_D1, an

1 ARMS tumor, the range of developmental indices was narrower and more skewed toward later
2 stages of myogenesis (Fig. 3H). Using our entire patient cohort of 18 tumors, we were able to
3 generalize these findings to RMS tumors - ERMS tumors had a wide diversity of developmental
4 indices while ARMS tumors narrowly centered with developmental indices from later stages of
5 murine myogenesis (Fig. 3I).

6

7 **O-PDXs and organoids recapitulate clonal heterogeneity in RMS**

8 We have previously established a panel of RMS O-PDXs and shared those models through the
9 Childhood Solid Tumor Network⁹. These O-PDXs encompass the clinical and molecular
10 diversity of RMS, and have previously undergone bulk genomic, transcriptomic, proteomic and
11 epigenomic analyses^{2,4,9}. We expanded our single-cell transcriptomic profiling to include the O-
12 PDXs that correspond to the 18 patient tumors profiled here (Extended Data Table 2 and
13 <https://pecan.stjude.cloud/static/RMS-scrna-atlas-2020/>). We performed the same analyses,
14 including developmental indexing (Fig. 3J). All 3 cell types (mesoderm, myoblast, and myocyte)
15 were preserved in the O-PDXs in the snRNA-seq and IHC analysis (Extended Data Fig. 4 and
16 data not shown). As expected, the O-PDXs lacked normal cells from the patient TME but
17 contained infiltration of murine monocytes (Extended Data Fig. 4). The patient tumor that was
18 collected during treatment and was enriched in cells with the mesodermal signature
19 (SJRHB010928_R1) re-established the developmental hierarchy in the O-PDX
20 (SJRHB010928_X1) (Extended Data Fig. 4).

21 To complement the O-PDXs, we also evaluated the transcriptomic heterogeneity of ex
22 vivo organoids derived from the O-PDXs (Supplemental Methods). Malignant cells within
23 organoids shared the cellular diversity seen in the originating patient tumor and O-PDX by single

1 cell transcriptional profiling (Extended Data Fig. 4). IHC for MEOX2, MYF5 and MYOG for
2 the organoids showed a similarity to their matched patient tumor and corresponding O-PDX
3 (Extended Data Fig. 4 and data not shown).

4

5 **RMS cells transition through developmental states**

6 RNA velocity analysis (see Fig. 1G,H) suggests that individual RMS tumor cells may transition
7 through developmental stages from mesoderm to myoblast and myocyte (Fig. 4A). Alternatively,
8 it is possible that there are distinct clones of cells that are restricted to their developmental stage
9 (Fig. 4B). To distinguish between these two possibilities, we used a lentiviral barcoding library
10 that incorporates a unique oligonucleotide barcode into the 3'-untranslated region of blue
11 fluorescent protein (BFP)^{20,21} (Fig. 4C,D). We infected 15 of the O-PDX models with the
12 barcode library and analyzed the barcode distribution in vivo by scRNA-seq. Following scRNA-
13 seq library generation, the barcode is retrievable by a separate PCR amplification step. In each of
14 the tumors that we analyzed, individual barcodes were found across all tumor cell types
15 (mesoderm, myoblast and myocytes) (Fig. 4E-G and Extended Data Table 5). Taken together,
16 these lineage tracing data, RNA-velocity analyses and genetic clonal analysis are consistent with
17 a model in which individual ERMS tumor cells can transition through developmental stages. The
18 same was true for ARMS tumors but the population of cells with paraxial mesoderm gene
19 expression signature was lower so some barcodes were found only in the myoblast and myocyte
20 population (Extended Data Table 5).

21

22 **Tumor cell heterogeneity reflects differential enhancer activity**

1 Several of the MRF genes that are turned on and off as cells transition through developmental
2 stages have core regulatory circuit super-enhancers⁴ (CRC-SEs) (Fig. 4H and Extended Data
3 Table 6). For example, *MEOX2* and *NFIX* (mesoderm), *PAX7* and *CREB5* (myoblast) and
4 *FOXO1* and *SOX6* (myocyte) each have CRC-SEs (Extended Data Table 6). To determine if the
5 chromatin accessibility of those CRC-SEs changes as individual cells transition through the
6 myogenic differentiation program, we performed droplet-based single-cell assay of transposase-
7 accessible chromatin sequencing (scATAC-seq) on 7 O-PDX tumors. We integrated scATAC-
8 seq and scRNA-seq profiles to investigate the chromatin accessibility of CRC-SEs for MRFs in
9 developmentally distinct subpopulations (Fig. 4H-J and Extended Data Table 6). Transferring
10 cell labels between scRNA-seq data and scATAC-seq data in SJRHB010927_X1 enabled us to
11 identify cell-type specific enhancer regions in *MYOD1*, *MSC*, *MEOX2* and several other
12 myogenic genes (Fig. 4H-J and Extended Data Table 6). These regions correspond to previously
13 reported core regulatory circuit domains for those genes^{4,22}. Analysis of all 7 O-PDX tumors
14 showed CRC-SEs that change in their chromatin accessibility in tumor cells with mesoderm
15 (*MEOX2*, *SMAD3*), myoblast (*CREB5*, *PAX7*), and myocyte (*MYOD1*, *FOXO1*) features
16 (Extended Data Table 6 and Extended Data Fig. 5). Collectively, these scATAC-seq studies
17 indicate that heterogeneity of developmental states within RMS tumors is reflected in chromatin
18 dynamics for myogenic CRC-SEs and genes.

19

20 **The mesoderm-like RMS cells are drug resistant**

21 Current chemotherapeutic regimens for RMS include drugs that target proliferating cells. The
22 myoblast population has the highest proportion of dividing cells in the patient tumors, the O-
23 PDXs, and the ex vivo organoids (Fig. 5A,B and data not shown). In a pair of matched ERMS

1 samples obtained before and during treatment, SJRHB000026_R2 and SJRHB000026_R3
2 (Extended Data Fig. 3A), we noted that the post-treatment sample was skewed towards
3 mesoderm signature-expressing cells (28.6% post-treatment versus 3.4% pre-treatment) with a
4 concomitant reduction in cells expressing the myocyte signature (1.4% post-treatment versus
5 31.4% pre-treatment). Additionally, one ERMS patient tumor, SJRHB010928_R1, was obtained
6 during treatment with fewer than 5% viable cells by histology; in this sample, the majority
7 (96.8%) of remaining viable cells expressed the mesoderm signature (Extended Data Fig. 3).
8 Taken together, these data suggest that the myoblast population may be more sensitive to
9 chemotherapy and the mesoderm-like population is more likely to survive treatment. To
10 investigate this trend further, we evaluated matched formalin-fixed paraffin embedded (FFPE)
11 tissue from 11 patients obtained at diagnosis and mid-treatment on a single therapeutic clinical
12 trial, RMS13 (NCT01871766). We quantitated the number of cells in each sample expressing
13 MEOX2 and MYOG (Extended Data Table 7). There was a significant enrichment in MEOX2
14 immunopositive cells in the post-treatment tumors relative to the matched pre-treatment RMS
15 samples and a corresponding decrease in MYOG immunopositive cells (Fig. 5C).

16 To model clonal selection in the laboratory, we generated longitudinal samples from
17 repeat biopsy of O-PDXs treated with a standard drug combination used to treat patients with
18 RMS (vincristine (VCR) and irinotecan (IRN)) at clinically relevant doses and schedules^{4,9} (Fig.
19 5D). For each O-PDX (SJRHB000026_X1, SJRHB013758_X1, SJRHB011_X,
20 SJRHB013757_X1 and SJRHB013759_X14), biopsies were performed at multiple timepoints
21 (before treatment (day 0), day 3, day 7, day 14 and day 21 of the first course) when sufficient
22 tumor was present to sample (Fig. 5E,F). We also collected tumor biopsies after the tumors
23 recurred. A portion of each biopsy underwent formalin-fixation for IHC staining for MEOX2,

1 MYF5 and MYOG (Fig. 5G,H and data not shown). The remaining biopsy portion was utilized
2 for quantitative RT-PCR for 21 genes expressed in mesoderm, myoblast and myocyte-like RMS
3 tumor cells or snRNA-seq. In total, 250 biopsies were collected and 6,480 qRT-PCR reactions
4 were performed (Extended Data Table 8-13). As in patient samples, the myoblast and myocyte
5 populations were sensitive to treatment and the mesoderm tumor cells population was enriched
6 (Fig. 5I-K and Extended Data Table 8-13). Moreover, the normally quiescent mesoderm-like
7 cells re-entered the cell cycle to initiate myogenesis (Fig. 5K).

8 Taken together, our data suggest that ERMS tumor cells transition through distinct states
9 that represent progressive stages of myogenesis. These different states (paraxial mesoderm,
10 myoblast, myocyte) have distinct proliferation properties and differential sensitivity to
11 chemotherapy. To further refine our understanding of the cellular heterogeneity of ERMS
12 tumors, their developmental trajectory and clonal selection with treatment, we developed a
13 mathematical model that follows the fate of cells in both 3-dimensional space and time.
14 Experimentally determined barcode distribution in each compartment was used to develop the
15 model (Fig. 5L), and barcode diversity was tracked over time. We assumed that upon cell
16 division, cells maintain their barcodes and we included barcoded and non-barcoded cells to
17 reflect the *in vivo* experiments. The relative proportion of different division types (self-
18 renewing/differentiating) in the mesodermal compartment determines whether the tissue remains
19 in homeostasis and influences the degree of clonal diversity loss over time. To parameterize the
20 model, we used experimental data from 10 barcoded ERMS xenografts. The fraction of dividing
21 cells and distribution of cells across compartments was determined from the sc/snRNA-seq data.
22 Our ERMS model predicts a decrease in clonal diversity (as measured by barcode diversity) over
23 time and clonal selection with treatment for individual tumors (Fig. 5M-P). To test this

1 experimentally, we performed scRNA-seq on a barcoded ERMS tumor (SJRHB000026_X1)
2 after initial labeling and a subsequent passage in mice with and without clinically relevant
3 chemotherapy (vincristine+irinotecan). As predicted by the three-compartment model, there was
4 a decrease in clonal diversity over time and clonal selection with treatment (Fig. 5O-Q).
5 Additional iterations of modeling and comparison to in vivo barcode distribution data are
6 consistent with differential cytotoxicity across the cellular populations (mesoderm, myoblast,
7 myocyte). In particular, a subset of the mesoderm-like cells are proliferating thereby making
8 them sensitive to chemotherapy. Based on our model, partial elimination of the mesoderm-like
9 population is required to account for the clonal selection we observe in vivo in mice and in
10 patients.

11

12 **EGFR is a therapeutic vulnerability in paraxial mesoderm RMS cells**

13 Having shown that the paraxial mesoderm RMS cells are more quiescent and drug resistant than
14 the myoblast population, we set out to identify therapeutic vulnerabilities unique to this
15 population using a systems biology algorithm, NetBID (data-driven Network-based Bayesian
16 Inference of Drivers)^{23,24}. NetBID, which was originally developed for bulk -omics data, was
17 adapted to analyze snRNA-seq profiles of our panel of 18 RMS patient tumors. We first used the
18 SJARACNe algorithm²⁵ to reverse engineer cell type-specific interactomes for each of the 5
19 major cell types from the integrated snRNA-seq profiles (Fig. 6A). With a focus on signaling
20 drivers, we used the cell type-specific interactomes of 2,543 genes/proteins and inferred their
21 network activities in each nucleus using the NetBID algorithm. We then performed differential
22 activity analysis to identify cell type-specific therapeutic vulnerabilities in the RMS tumor cells
23 with the mesodermal signature. EGFR was significantly activated in the mesoderm population

1 compared to myoblasts ($p=4.4\times10^{-135}$) and myocytes ($p=1.8\times10^{-174}$) and the network was rewired
2 as cells transition through the developmental hierarchy (Fig. 6B). EGFR network activity was
3 also significantly higher in ERMS relative to ARMS ($p=5.4\times10^{-36}$) (Fig. 6C,D). These data are
4 consistent with previous integrated epigenetic/proteomic analyses for differential pathway
5 activity in ERMS and ARMS⁴. In addition, previous studies have shown heterogenous
6 expression of EGFR protein in FFPE samples of RMS²⁶⁻²⁸. To validate these data, we performed
7 IHC for EGFR alone and in combination with markers of each cell population. We used the 5B7
8 monoclonal antibody which has been previously shown to correlate with EGFR inhibitor
9 (EGFRi) responsiveness in lung cancer²⁹. There was co-localization of EGFR with MEOX2 in 2-
10 color IHC and EGFR was mutually exclusive with MYOG (Fig. 6E,F).

11 To determine if EGFR is a therapeutic vulnerability in RMS, we exposed 3D ERMS
12 organoids that contain all 3 cell populations (Extended Data Fig. 6) to two different EGFRi's
13 (gefitinib and afatinib) with increasing concentrations of SN-38, the active metabolite of
14 irinotecan (Supplemental Information). The EGFRi's alone had no effect on overall organoid
15 viability as measured with CellTiter-Glo 3D which is not surprising given the low percentage of
16 mesoderm-like cells in the organoids (Extended Data Fig. 6). However, when the proliferating
17 myoblast population was targeted with increasing concentrations of SN-38, the addition of
18 EGFRi's significantly reduced viability (Extended Data Fig. 6). In a representative ERMS O-
19 PDX (SJRHB013758_X1), there was a significant improvement in outcome when afatinib or
20 gefitinib was combined with IRN+VCR (Fig. 6H,I). For a second ERMS O-PDX with a low
21 percentage of mesoderm-like cells, SJRHB010927_X1, O-PDXs responded robustly to
22 IRN+VCR therapy alone, reinforcing the importance of including standard-of-care treatment in
23 preclinical testing (Extended Data Fig. 6).

1 **Discussion**

2 In conclusion, we have discovered that RMS tumor cells can transition through different stages
3 of myogenesis from an immature quiescent paraxial mesoderm state through a highly
4 proliferative myoblast stage and into a more differentiated myocyte state. Not only do cells
5 undergo changes in gene expression during these developmental transitions but super-enhancer
6 chromatin accessibility is also dynamic. While proliferating cells can be identified in tumor cell
7 population in patient tumors and O-PDXs, the most proliferative cells are in the myoblast stage.
8 It is therefore not surprising that broad spectrum chemotherapeutic regimens that target rapidly
9 dividing cells efficiently reduce tumor volume by killing the myoblast-like RMS tumor cells.
10 The immature paraxial mesoderm-like RMS tumor cells are more quiescent and can survive
11 therapy and then expand to repopulate the myogenic lineage found in the primary tumor. These
12 observations are consistent with decades of clinical research showing that different combinations
13 of broad- spectrum chemotherapy or intensification of existing regimens have failed to improve
14 outcomes for children with RMS^{6,7}. Those different treatment approaches are killing the rapidly
15 dividing myoblast-like cells and the resistant mesoderm-like cells survive and contribute to
16 disease recurrence. By focusing our investigation on the mesoderm-like cells, we identified a
17 dependence on EGFR that can be exploited with EGFRi in vivo. While there were only a small
18 number of mesoderm-like cells in ARMS tumors, we discovered a dramatic upregulation of
19 EGFR during treatment suggesting that EGFRi's may be useful for both types of
20 rhabdomyosarcoma. Our study shows that treatment for RMS and possibly other pediatric solid
21 tumors should focus on total clonal elimination rather than continuing to target just the most
22 proliferative cell population that makes up the bulk of the tumor. Such an approach may reduce
23 disease recurrence and improve survival and quality of life for children with solid tumors.

24

1 **Figure Legends**

2 **Figure 1: Single-cell RNA-sequencing (scRNA-seq) reveal a developmental hierarchy**

3 **within RMS. A-B,** During fetal myogenesis, mesodermal cells of the somite migrate to form

4 skeletal muscle throughout the body (A). During that migration, these cells undergo stepwise

5 differentiation typified by the transient expression of myogenic regulatory factors¹⁰ (B). **C-D,**

6 Photomicrographs of an embryonal RMS tumor, SJRHB030680_R1 (C) and an alveolar RMS

7 tumor, SHRHB031320_D1. *Left*, H&E staining. *Right*, Myogenin (MYOG)

8 immunohistochemistry (IHC) with 20X magnification. *inset*, 80X magnification. **E-F**, UMAP

9 visualization of 3,973 malignant cells from SJRHB030680_R1 (E) and 2,414 malignant cells

10 from SJRHB031320_D1 (F). Cells are colored based on expression of *MEOX2* (*left*), *MYF5*

11 (*center*), and *MYOG* (*right*). **G-H**, RNA velocity analysis of SJRHB030680_R1 (G) and

12 SJRHB031320_D1 (H). Abbreviations: ERMS, embryonal rhabdomyosarcoma; UMAP, uniform

13 manifold approximation and projection. Scale bars: C,D, 100 μ m.

15

16 **Figure 2: Identification of major cell clusters within patient RMS tumors using single-**

17 **nucleus RNA-sequencing. A,** LargeVis visualization of snRNA-seq of 111,474 nuclei from 18

18 integrated patient RMS tumors, colored based on sample. **B,** Heatmap showing expression of

19 myogenic regulatory factor expression across seven Leiden clusters. **C,** Boxplot showing the

20 percentage of malignant nuclei within each muscle developmental state for each tumor. The

21 center line demarcates the median value with rectangle showing interquartile range (IQR)

22 between the first and third quartiles. The vertical bars extending from the rectangles indicate

23 maximum and minimum values with the exception of outliers that exceed more than 3 times the

24 IQR. **D-E**, LargeVis visualization of Leiden clustering of snRNA-seq grouped based on

1 expression of mesoderm, myoblast, or myocyte myogenic regulatory factors (D) or colored by
2 predicted cell cycle phase (E). **F**, Plot of the proportion of proliferating cells (S/G2/M phase) in
3 each group, estimated using gene signatures associated with G1, S, and G2/M phases¹³. Circles
4 are ERMS and squares are ARMS. Center line and rectangle indicate the median and IQR as in
5 panel (C). Vertical bars indicate the maximum and minimum values with the exception of
6 outliers that exceed more than 3 times the IQR. **G**, Immunohistochemistry image of an ERMS
7 tumor, SJRHB013758_D2 stained with antibodies against MEOX2 (left), MYF5 (center) and
8 MYOG (right). **H**, Quantitation of the percentage of cells positive for MEOX2 (blue), MYF5
9 (green), or MYOG (red) immunohistochemical staining (x axis) compared to percentage of cells
10 within each developmental state as determined by snRNA-seq (y axis). **I-J**, Dual staining of
11 MEOX2 (purple) and MYOG (brown) within SJRHB013758 (I) with magnified view (J).
12 Abbreviations: ERMS, embryonal rhabdomyosarcoma; ARMS, alveolar rhabdomyosarcoma.
13 Scale bars: G, 10 μ m.
14

15 **Figure 3: Developmental indexing of patient RMS tumors and orthotopic patient-derived**
16 **xenografts. A**, UMAP plot of 1.5 million nuclei from the Mouse Organogenesis Cell Atlas¹⁸,
17 downsampled to 100,000 nuclei. Clusters are colored based on trajectory. **B**, UMAP plot of
18 576,560 nuclei from the mesenchymal trajectory with identification of the skeletal myogenesis
19 sub-trajectory. Nuclei are colored based on Leiden cluster. **C**, UMAP plot of 58,573 nuclei of the
20 skeletal muscle sub-trajectory with computational clustering that identifies nuclei from early
21 mesodermal progenitors, paraxial mesoderm, myoblasts, myocytes and myotubes. **D**, Heatmap of
22 aggregated transcription from each cluster demonstrating expression of myogenic regulatory
23 factors and additional mesodermal markers. **E**, Violin plot of projected developmental indices of

1 embryonic skeletal muscle data separated by mouse embryonic stage. **F**, UMAP plot of
2 developmental indices within the embryonic skeletal muscle sub-trajectory. **G-H**, Application of
3 developmental indices to an ERMS tumor, SJRHB030680_R1 (G) and an ARMS tumor,
4 SJRHB031320_D1 (H). **I-J**, Developmental indices of 18 patient RMS tumors (I) or 18 O-PDXs
5 (J). Abbreviations: ERMS, embryonal rhabdomyosarcoma; ARMS, alveolar rhabdomyosarcoma;
6 UMAP, uniform manifold approximation and projection.

7

8 **Figure 4: Developmental status in ERMS is plastic and associated with chromatin
9 accessibility at core regulatory superenhancer regions. A-B**, Two competing models of tumor
10 heterogeneity within RMS. In the first model, RMS cells transition across developmental states
11 (A); in the alternate model, genetically distinct clones are restricted to muscle developmental
12 states (B). **C**, Schematic of the lentiviral barcode plasmid.^{20,21} An 18-mer of random nucleotides
13 is incorporated into the 3'-untranslated region of a blue fluorescent protein (BFP) tag, enabling
14 barcode recovery from scRNA-seq libraries. **D**, Plot of frequency of individual barcodes for
15 subsequent passages of an individual ERMS O-PDX, SJRHB00026_X1. **E-F**, UMAP plot of an
16 ERMS O-PDX SJRHB013758_X2, colored based on developmental stage (E), or with 3 specific
17 barcodes highlighted (F). **G**, Quantitation of the developmental state diversity of all tumor cells
18 within SJRHB013758_X2, and from the 5 most prevalent barcoded clones. **H**, ChIP-seq and
19 chromHMM of *MYOD1* in an ERMS O-PDX, SJRHB10927_X1. Scales are indicated on the left,
20 and a previously identified CRC-SE⁴ is highlighted in blue. **I**, Comparison of H3K27
21 trimethylation in various pediatric O-PDXs. OS, osteosarcoma; EWS, Ewing sarcoma; LPS,
22 liposarcoma; HGS, high-grade sarcoma; NB, neuroblastoma. **J**, Single-cell ATAC-seq of
23 SJRHB010927_X1 at the *MYOD1* locus; cell identities were defined via gene activity estimation,

1 and dataset integration with scRNA-seq data³⁰. Abbreviations: ERMS, embryonal
2 rhabdomyosarcoma; ARMS, alveolar rhabdomyosarcoma; UMAP, uniform manifold
3 approximation and projection; RMS, rhabdomyosarcoma; OS, osteosarcoma; EWS, Ewing
4 sarcoma; LPS, liposarcoma; HGS, high grade sarcoma; NB, neuroblastoma.

5

6 **Figure 5. Chemotherapy treatment of ERMS selects for mesoderm developmental stages.**

7 **A-B**, Bar plots showing percentage of cells predicted to be dividing within each developmental
8 stage for patient tumors (a) and O-PDXs (b). **C**, Plots showing immunopositivity for MEOX2
9 (left) and MYOG (right) in patient samples from RMS13 obtained before treatment (“diagnosis”)
10 and during therapy (“mid-treatment”). **D**, Treatment schema for VI therapy of mice bearing RMS
11 O-PDXs. Needle biopsies were performed at days 0, 3, 7, 14, and 21 or when tumors were large
12 enough to sample. **E**, Photograph of needle biopsy of an orthotopically-injected xenograft. **F-H**,
13 Photograph of tissue obtained by a biopsied O-PDX (F), which was fixed and stained using H&E
14 (G) or MYOG (H). **I**, Plot showing longitudinal expression of *MEOX2* by qRT-PCR during
15 treatment. There is an increase in *MEOX2* during chemotherapy (days 7,14,21) but the
16 proportion resets to basal levels after 28 days. This was verified by IHC (lower panel). **J**,
17 Boxplot of all biopsies for ERMS tumor bearing mice for the untreated and treated samples. The
18 plot is an integration of expression of 6 genes (*MEOX2*, *PAX3*, *EGFR*, *CD44*, *DCN*, *POSTN*)
19 expressed as normalized relative fold. The center line demarcates the median value with
20 rectangle showing interquartile range (IQR) between the first and third quartiles. The vertical
21 bars extending from the rectangles indicate maximum and minimum values with the exception of
22 outliers that more than 3 times the IQR. **K**, Relative proportion of nuclei in each developmental
23 state for longitudinal biopsies of a single O-PDX, determined using snRNA-seq of biopsied

1 tissue. **L**, Diagram of the mathematical model of ERMS developmental heterogeneity. **M-N**,
2 Simulated average population size for an untreated ERMS tumor (M) or a treated ERMS tumor
3 (N) briefly exposed to an antiproliferative agent (gray bar). Average population size over 524
4 simulations are shown, standard error bars are too small to see. **O-P**, Simulated time course of
5 barcode dynamics for an ERMS tumor that was either untreated (O) or briefly treated (P;
6 duration of treatment in grey bar). Each curve represents a different barcoded lineage. One
7 realization of the stochastic dynamics is shown. Insets under each graph show spatial
8 distributions of bar codes (color coded) in myoblast cells at an early and late stage of tumor
9 growth (O) and pre- and post-therapy (P). **Q**, Temporal development of the average entropy
10 index (measure of barcode diversity) during barcoded ERMS tumor growth, either untreated or
11 briefly treated (grey bar). Average entropy values over 524 simulations \pm standard errors (dashed
12 lines) are shown. *Inset*, bar plot comparing the initial entropy index to the final entropy index of
13 untreated or treated tumors in experiments. Model parameters were: average value of L_{mes}
14 $=0.0035$ ($r_1=1.5$, $r_2=0.0001$), $L_{blast}=0.0045$, $P_{mes}=0.55$, $P_{blast}=0.49$, $D=0.035$, $\alpha_{mes}=0.0014$, α_{blast}
15 $=0.0035$. The parameter units are per minute. Abbreviations: ERMS, embryonal
16 rhabdomyosarcoma; ARMS, alveolar rhabdomyosarcoma; VCR, vincristine; IRN irinotecan.
17

18 **Figure 6. Mesoderm-like ERMS cells are uniquely vulnerable to EGFR blockade. A**,
19 Schematic workflow of NetBID algorithm to identify cell type-specific drivers from snRNA-seq
20 data. **B**, Volcano plot of differential activity analysis of signaling drivers in ERMSmesoderm vs.
21 other cell types. **C-D**, EGFR NetBID activity in different developmental states from snRNA-seq
22 data (C) and inferred from bulk RNA-seq of patient tumors (D). **E-F**, Dual IHC staining of
23 ERMS patient tumor, SJRHB030680_R1, combining EGFR (brown) with either MEOX2 (E) or

1 MYOG (F) in purple. **G**, Schedules of drugs used for preclinical study. **H**, Survival curves for
2 each treatment group for a ERMS tumor O-PDX (SJRHB01378_X1). **I**, Representative image of
3 bioluminescence for mice treated on the study, scale bar is photons/sec/cm² str. Scale bars: E,F,
4 10 µm. Abbreviations: VCR, vincristine; IRN, irinotecan; ERMS, embryonal
5 rhabdomyosarcoma; ARMS, alveolar rhabdomyosarcoma.

1 **Materials and Methods**

2 *See Supplementary Information.*

3

4 **Acknowledgements**

5

6 We thank A. McArthur for editing the manuscript. This work was supported by Cancer Center
7 Support (CA21765) and grants to M.A.D. from the National Institutes of Health (EY014867,
8 EY018599, and CA168875). The content is solely the responsibility of the authors and does not
9 necessarily represent the official views of the National Institutes of Health. Research was also
10 supported by American Lebanese Syrian Associated Charities. M.A.D. was also supported by the
11 Howard Hughes Medical Institute, Alex's Lemonade Stand, and Tully Family and Peterson
12 Foundations. A.G.P. was supported by Alex's Lemonade Stand, Hyundai Hope on Wheels, and
13 Damon Runyon Cancer Research Foundations.

14

15 **Author Contributions**

16

17 A.G.P., X.C. and M.A.D designed the study. A.G.P processed fresh and frozen tissue for single-
18 cell/nucleus RNA-seq and ATAC-seq. B.G., K.B., and E.S. assisted in patient sample and
19 xenograft tissue accrual for molecular analyses. A.G.P., X.C., X.H., and J.Y. performed
20 computational analysis of data. M.R.C., B.A.O., and H.T. supervised immunohistochemical
21 staining and pathology review. N.K. and D.W. performed mathematical modeling of tumors.
22 M.J.K. and A.P. wrote and supervised the RMS13 clinical trial. A.K. manages the Childhood
23 Solid Tumour Network. J.M. and K.B. assisted in molecular analysis. C.R. and X.Z. developed
24 the single-cell viewer portal. K.B. and E.S. performed the preclinical testing and tumor biopsies.
25 A.G.P. and J.L.N. developed the tumor organoid model.

26

1 **Competing Interest Statement**

2 The authors declare no competing financial interests.

3

4 **Supplementary Information**

5 **Supplementary Information** is available for this paper.

6

7 **Corresponding Author**

8 Correspondence and requests for materials should be addressed to Michael Dyer at

9 michael.dyer@stjude.org.

10

11 **Data and Source Code Availability**

12 NetBID: <https://jyyulab.github.io/NetBID>

13 All processed single-cell and single-nucleus RNA-sequencing data are publicly accessible via an
14 online data portal (<https://pecan.stjude.cloud/static/RMS-scrna-atlas-2020>).

15 All raw sequencing data will be uploaded and be publicly available at the time of publication
16 through GEO accession GSE174376.

17

18 **References**

19 1. Ries, L. *et al.* *Cancer Incidence and Survival among Children and Adolescents: United States SEER*
20 *Program 1975-1995*. (National Cancer Institute, SEER Program).

21 2. Chen, X. *et al.* Targeting oxidative stress in embryonal rhabdomyosarcoma. *Cancer Cell* **24**,
22 710–724 (2013).

23 3. Kahn, H. J. *et al.* Immunohistochemical and electron microscopic assessment of childhood
24 rhabdomyosarcoma. Increased frequency of diagnosis over routine histologic methods. *Cancer*
25 **51**, 1897–1903 (1983).

- 1 4. Stewart, E. *et al.* Identification of Therapeutic Targets in Rhabdomyosarcoma through
2 Integrated Genomic, Epigenomic, and Proteomic Analyses. *Cancer Cell* **34**, 411–426 (2018).
- 3 5. Shern, J. F. *et al.* Comprehensive Genomic Analysis of Rhabdomyosarcoma Reveals a
4 Landscape of Alterations Affecting a Common Genetic Axis in Fusion-Positive and Fusion-
5 Negative Tumors. *Cancer Discov* **4**, 216–231 (2014).
- 6 6. Pappo, A. S. & Dirksen, U. Rhabdomyosarcoma, Ewing Sarcoma, and Other Round Cell
7 Sarcomas. *J Clin Oncol* **36**, 168–179 (2017).
- 8 7. Arndt, C. A. S., Bisogno, G. & Koscielniak, E. Fifty Years of Rhabdomyosarcoma studies on
9 both sides of the Pond and Lessons Learned. *Cancer Treat Rev* **68**, 94–101 (2018).
- 10 8. Pappo, A. S. *et al.* Survival After Relapse in Children and Adolescents With
11 Rhabdomyosarcoma: A Report From the Intergroup Rhabdomyosarcoma Study Group. *J Clin
12 Oncol* **17**, 3487–3493 (1999).
- 13 9. Stewart, E. *et al.* Orthotopic patient-derived xenografts of paediatric solid tumours. *Nature*
14 **549**, 96–100 (2017).
- 15 10. Chal, J. & Pourquié, O. Making muscle: skeletal myogenesis *in vivo* and *in vitro*.
16 *Development* **144**, 2104–2122 (2017).
- 17 11. Buckingham, M. & Rigby, P. W. J. Gene Regulatory Networks and Transcriptional
18 Mechanisms that Control Myogenesis. *Dev Cell* **28**, 225–238 (2014).
- 19 12. Sebire, N. J. & Malone, M. Myogenin and MyoD1 expression in paediatric
20 rhabdomyosarcomas. *J Clin Pathol* **56**, 412–416 (2003).
- 21 13. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human
22 oligodendrogloma. *Nature* **539**, 309–313 (2016).
- 23 14. Slyper, M. *et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen
24 human tumors. *Nat Med* **26**, 792–802 (2020).
- 25 15. Manno, G. L. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
- 26 16. Barkas, N. *et al.* Joint analysis of heterogeneous single-cell RNA-seq dataset collections. *Nat
27 Methods* **16**, 695–698 (2019).
- 28 17. Vincent, K. M. & Postovit, L.-M. A pan-cancer analysis of secreted Frizzled-related proteins:
29 re-examining their proposed tumour suppressive function. *Sci Rep* **7**, 42719 (2017).
- 30 18. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature*
31 **566**, 496–502 (2019).

- 1 19. Cheng, C. *et al.* Latent cellular analysis robustly reveals subtle diversity in large-scale single-cell RNA-seq data. *Nucleic Acids Res* **16**, e143 (2019).
- 3 20. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867–1882 (2016).
- 5 21. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853–1866 (2016).
- 7 22. Gryder, B. E. *et al.* PAX3-FOXO1 Establishes Myogenic Super Enhancers and Confers BET Bromodomain Vulnerability. *Cancer Discov* **7**, 884–899 (2017).
- 9 23. Du, X. *et al.* Hippo/Mst signalling couples metabolic state and immune function of CD8 α + dendritic cells. *Nature* **558**, 141–145 (2018).
- 11 24. Wijaya, J. *et al.* An ABC Transporter Drives Medulloblastoma Pathogenesis by Regulating Sonic Hedgehog Signaling. *Cancer Res* **80**, 1524–1537 (2020).
- 13 25. Khatamian, A., Paull, E. O., Califano, A. & Yu, J. SJARACNe: a scalable software tool for gene network reverse engineering from big data. *Bioinformatics* **35**, 2165–2166 (2018).
- 15 26. Ganti, R. *et al.* Expression and genomic status of EGFR and ErbB-2 in alveolar and embryonal rhabdomyosarcoma. *Modern Pathol* **19**, 1213–1220 (2006).
- 17 27. Grass, B. *et al.* Immunohistochemical detection of EGFR, fibrillin-2, P-cadherin and AP2beta as biomarkers for rhabdomyosarcoma diagnostics. *Histopathology* **54**, 873–879 (2009).
- 19 28. Wachtel, M. *et al.* Subtype and Prognostic Classification of Rhabdomyosarcoma by Immunohistochemistry. *J Clin Oncol* **24**, 816–822 (2006).
- 21 29. Mascaux, C. *et al.* EGFR Protein Expression in Non–Small Cell Lung Cancer Predicts Response to an EGFR Tyrosine Kinase Inhibitor—A Novel Antibody for Immunohistochemistry or AQUA Technology. *Clin Cancer Res* **17**, 7796–7807 (2011).
- 24 30. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902 (2019).
- 25 31. Stewart, E. *et al.* The Childhood Solid Tumor Network: A new resource for the developmental biology and oncology research communities. *Dev Biol* **411**, 287–293 (2016).
- 27 32. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491–499 (2017).
- 29 33. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).

1 34. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq
2 data using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019).

3 35. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat
4 Biotechnol* **37**, 38–44 (2019).

5 36. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data
6 analysis. *Genome Biol* **19**, 15 (2018).

7 37. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to
8 transient cell states through dynamical modeling. *Nat Biotechnol* **1**–7 (2020)
9 doi:10.1038/s41587-020-0591-3.

10 38. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional
11 profibrotic macrophage. *Nat Immunol* **20**, 163–172 (2019).

12 39. Traag, V. A., Waltman, L. & Eck, N. J. van. From Louvain to Leiden: guaranteeing well-
13 connected communities. *Sci Rep* **9**, 5233 (2019).

14 40. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory
15 inference methods. *Nat Biotechnol* **37**, 547–554 (2019).

16 41. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell
17 transcriptomics. *BMC Genomics* **19**, 477 (2018).

18 42. Hanawa, H. *et al.* Comparison of Various Envelope Proteins for Their Ability to Pseudotype
19 Lentiviral Vectors and Transduce Primitive Hematopoietic Cells from Human Blood. *Mol Ther*
20 **5**, 242–251 (2002).

Figure 1




Figure 2




Figure 3




Figure 4




Figure 5




Figure 6

