Oncogenic *Gata1* causes stage-specific megakaryocyte differentiation delay

Gaëtan Juban,[1a*] Nathalie Sakakini,[1,2b*] Hedia Chagraoui,[1] David Cruz Hernandez,[1,2] Qian Cheng,[3] Kelly Soady,[1,2] Bilyana Stoilova,[1,2] Catherine Garnett,[1,2] Dominic Waithe,[3] Georg Otto,[1c] Jessica Doondeea,[1,2] Batchimeg Usukhbayar,[1,2] Elena Karkoulia,[4] Maria Alexiou,[5d] John Strouboulis,[4e] Edward Morrissey,[5] Irene Roberts,[1,2,6] Catherine Porcher[1#] and Paresh Vyas[1,2,7#]

[1]MRC Molecular Hematology Unit WIMM, University of Oxford, Oxford, UK; [2]Haematology Theme Oxford Biomedical Research Center, Oxford, UK; [3]Center for Computational Biology WIMM, University of Oxford, Oxford, UK; [4]Institute of Molecular Biology and Biotechnology, Foundation of Research & Technology-Hellas, Heraklion, Crete Greece; [5]Biomedical Sciences Research Center "Alexander Fleming", Vari, Greece; [6]Department of Pediatrics University of Oxford, Oxford, UK and [7]Department of Hematology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

**GJ and NS contributed equally as co-first authors.*
#CP and PV contributed euqally as co-senior authors.

[a]Current address: Institut NeuroMyoGène, Université Claude Bernard Lyon 1, CNRS UMR 5310, INSERM U1217, Université Lyon, Lyon, France.

[b]Current address: Cambridge Stem Cell Institute, Cambridge, UK.

[c]Current address: Genetics and Genomic Medicine, University College London Institute of Child Health, London, UK.

[d]Current address: Department of Dentistry, University of Alberta, Edmonton, Alberta, Canada.

[e]Current address: Rayne Institute School of Cancer & Pharmaceutical Sciences, King's College London, London, UK

 doi:10.3324/haematol.2019.244541

Received: December 8, 2019.

Accepted: May 20, 2020.

Pre-published: June 11, 2020

Correspondence: *PARESH VYAS* - paresh.vyas@imm.ox.ac.uk

GAËTAN JUBAN - gaetan.juban@univ-lyon1.fr

Supplemental Information

Supplemental Methods

Targeting constructs and targeting approach: Two murine *Gata1* gene DNA fragments surrounding exon 2 (2.5 kb fragment upstream and 2.1 kb fragment downstream, **Supplemental Figure 1A**) were generated by PCR using 129/Ola strain DNA. For bioG1 targeting construct, the avi tag[1] was introduced in frame at exon 2 translational start codon via the *Nco I* restriction site. For bioG1s targeting construct, sequence including exon 2, intron 2 and the beginning of exon 3 (up to the translational start at amino acid 84) was excised and replaced with a sequence coding an avi-TEV-FLAG peptide, resulting in a fusion N-terminal GATA1s protein isoform with the avi-TEV-FLAG peptide. Targeting fragments were inserted either side of a Phosphoglycerate Kinase -neomycin (pgk-neo) cassette flanked by loxP sites. A negative selection cassette expressing the thymidine kinase gene regulated by MC1 promoter was included. IB10 ES cell line (subclone of E14) expressing a *BirA* transgene from the ROSA26 locus[2] was electroporated with targeting constructs and selected using G418 (0.3 mg/mL) and Ganciclovir (0.3 mM). Resistant colonies were characterized by PCR and Southern blotting. Normal karyotype ES clones (more than 80% of cells) were transiently electroporated with *Pgk-Cre* plasmid to delete the Neo cassette. Deletion was confirmed by PCR.

Growth and differentiation of murine ESCs: ESC lines were grown at 37°C in humidified 5% CO_2 incubator and maintained for up to 15 passages on gelatin-coated wells in KO-DMEM (Life Technologies, Warrington UK) supplemented with leukemia-inhibitory factor (LIF), 15% fetal calf serum (FCS), 2 mM L-glutamine and 100 μM β-mercaptoethanol. The protocol to differentiate megakaryocytes from ESC was modified from a previously published report[3]. Briefly, one day prior to embryoid body (EB) differentiation, cells were passaged into IMDM (Life Technologies, UK) with LIF, FCS and 0.3 mM monothioglycecerol (MTG). $2x10^4$ to $3x10^4$ ESCs were plated into 100-mm Sterilin dishes in IMDM with 5% PFHMII, 15% FCS, 2 mM L-glutamine, 1% Penicillin-Streptomycin, 0.4 mM MTG, 300 μg/ml human transferrin and 50 μg/ml ascorbic acid. After 6 days in culture, EBs were dissociated with 0.25%

Trypsin/EDTA. Cells were incubated with PE-conjugated anti-mouse CD41. Antibody-labelled cells were isolated with anti-PE magnetic microbeads (Miltenyi, Bisley UK) using LS columns on a MACS Separator and stained with APC-conjugated anti-mouse CD117 (kit). List of the antibodies used for the flow cytometry analyses is shown in **Supplemental Table 1**. FACS purified $CD41^{+}kit^{+}$ cells were seeded onto OP9 cells with 20 ng/ml TPO (Peprotech, London UK). After 3 days, cells were seeded onto fresh OP9 cells with 10 ng/ml TPO, 5 ng/ml IL6 and IL11 (Peprotech, UK). Cells were maintained under these conditions for up to 6 days.

Mice: Animal studies were conducted in accordance with the UK Home Office regulations. Targeting constructs were electroporated into mouse E14Tg2a ES cells maintained in Glasgow's MEM (ThermoFisher Scientific). After selection by G418 (0.3 mg/mL) and Gancyclovir (0.3 mM), resistant colonies were isolated and characterized by PCR and Southern blotting. ES clones with a normal karyotype (more than 80% of cells showing normal numbers of chromosomes) were injected into C57Bl/6 blastocysts and introduced into pseudo-pregnant C57Bl/6 females. After germ line transmission was obtained, heterozygous females were crossed to a male expressing Cre recombinase under the control of Gata1 promoter region[4] to excise the Neo resistance gene in germ cells. The Gata1-cre transgene was then bred out and mice were crossed with the $Rosa26^{BirA}$ strain[2]. Mice were maintained as homozygous and genotyped by PCR.

Embryos were staged based on somite pairs and morphological criteria. Embryos were removed from the uterine sac and kept on ice in DPBS (Life technologies UK) containing 10% FCS and 1% Penicillin-Streptomycin (Life Technologies, UK). Isolated yolk sacs were dissociated for 20 minutes at 37°C in PBS with 0.1% collagenase.

Flow cytometry: FACS analyses were performed either on LSR Fortessa X20 or FACS Canto II (BD, UK). FACS sorts were performed on FACS Aria Fusion (BD). Cells were labelled for 20 min at 4°C with Hoechst 33258 (Invitrogen, UK) and rat anti-mouse antibodies directed against: kit-APC, CD41-PE, Ter119-PECy7, Mac1-APCCy7, Gr1-FITC, and CD42b-FITC. List of the antibodies used for the flow cytometry analyses is shown in

Supplemental Table 1. Unstained, single stained and Fluorescence Minus One (FMO) controls were used to determine the background and fix the compensation in each channel. Data were analysed by FlowJo v10.1 software.

GATA1 Western blotting: This was performed as described[5] with modifications. Day 6 EB $CD41^+$ cells were lysed in hypotonic buffer containing 10 mM Hepes pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl (Sigma-Aldrich Poole UK) and a protease inhibitor cocktail (Roche Welwyn Garden City UK). Cell pellets were resuspended in hypertonic buffer (50 mM Hepes pH 7.9, 50 mM KCl, 300 mM NaCl, 0.1 mM EDTA, 10% Glycerol and a protease inhibitor cocktail). Sonicated lysates were incubated for 1h with 5 U/µl benzonase (Sigma-Aldrich). Protein concentration was determined using the Bradford Protein Assay (BioRad, Kidlington UK). 30 µg of nuclear extract were loaded onto 4 to 12% NUPAGE BisTris gradient SDS-PAGE gels (Invitrogen, Paisley UK) and transferred onto a PVDF membrane. After blocking overnight at 4°C in 5% milk, the blots were incubated with anti-GATA1 (sc-1234 Santa Cruz Biotechnology, M20, 1/200) and TBP (sc-204 Santa Cruz Biotechnology, 1/500), and secondary HRP-conjugated antibodies. Blots were revealed by chemiluminescence (GE Healthcare Life Science, UK).

RNA sequencing: P1 to P4 populations were FACS sorted from d8 cultures in quadruplicates and total RNA was isolated using RNeasy Plus Micro Kit (QIAGEN). cDNA synthesis was performed on 10 ng of RNA using SMARTer Low Input RNA Kit (Clontech) and libraries were generated using a Nextera XT DNA sample preparation kit and Index kit (Illumina). Library concentration was determined using a Qubit High-Sensitivity DNA kit (Invitrogen) and library size and quality were verified using an Agilent High-Sensitivity DNA chip on a Bioanalyser (Agilent Technologies). Indexed libraries were pooled to a final concentration of 10 nM and 50-bp single-end reads sequenced on a HiSeq 2000 apparatus (Illumina). Sequencing reads were mapped to the mm9 reference genome using TopHat v2.0.10[6]. Alignments were processed using Picard tools (http://picard.sourceforge.net/). Expression analysis was carried out using R version 3.1.1 (http://www.R-project.org). Sequencing reads were filtered for mapq ≥4 i.e. uniquely mapping reads. Lowly expressed

genes were filtered out by using a threshold of cpm > 5 in ≥ 2 samples, leaving 11365 genes for the analysis. The expression matrix was transformed to $\log_2$(cpm) scale, and PCA was carried out.

Gene expression analysis by dynamic arrays: Cells were FACS sorted into 9 µl of reaction buffer containing 5 µl of CellsDirect 2X reaction mix, 0.1 µl of RNase inhibitor, 2.5 µl of 0.2X Taqman gene expression assay mix, 1 µl of RT/Taq mix (all Life Technologies UK) and 0.4 µl of TE buffer. Reverse transcription and specific target amplification steps were performed as follows: 15 min at 50°C, 2 min at 95°C, 22 cycles of 15 sec at 95°C, 4 min at 60°C. cDNA diluted 1:5 in TE buffer was PCR-amplified with 48.48 dynamic array (Fluidigm, Cambourne UK). Each sample was analysed in technical duplicates or triplicates. Conditions with a quality score < 0.65 were considered as undetected. Biomark data were exported from the Fluidigm Data Collection Software and analysed using the $2^{-\Delta\Delta Ct}$ method. *GAPDH*, *TBP* and *HPRT* were used as housekeeping genes. For each gene, normalised ΔCt was calculated by subtracting the mean Ct of the three housekeeping genes to the gene of interest[7]. Downstream analyses were performed in R-3.3.3: pheatmap (heatmaps), hclust (hierarchical clustering with Euclidean distance measure and Ward.D2 agglomeration method), prcomp (PCA analysis using TRUE for the scale parameter), predict (for projection) and graphics (using the graphics package). List of genes analysed by Fluidigm is given in **Supplemental Table 2**.

Cell staining and microscopy: Cells were cytospun for 5 min at 400 rpm, air dried overnight and stained with May-Giemsa-Grunwald (MGG) (Sigma, Welwyn Garden City UK) or acetylcholine iodide (Sigma) to reveal acetylcholinesterase activity. Pictures were taken using an Olympus BX60 microscope with an Infinity 3S Luminera color camera.

Acetylcholinesterase staining quantitation: All images were analysed by Fiji/ImageJ macroscript available at (https://github.com/dwaithe)[8]. Images were colour inverted to create a negative image whereby the cellular staining was light against the background and the nuclear signals were particularly bright. Next, the image was smoothed lightly (sigma = 2.0) with a Gaussian kernel to reduce noise. The ImageJ ‘Find Maxima’ plugin was applied

(noise tolerance = 16) to locate each cell nucleus and the 'Maxima Within Tolerance' option used to export a binary image of cell nuclei. Output was processed by "Analyze Particles" plugin excluding any objects < 50 pixels in size, considered too small to be nuclei. From a mask image of the retained regions a Voronoi transformation was applied to demarcate the position of the nuclei. In this way, it was possible to threshold (Otsu) the entire cell and accurately associate each nucleus with its corresponding cytoplasmic staining. The original image was then colour deconvolved using the "H&E DAB" parameter set and the intensity in the resulting 'Colour_1' (blue) channel measured in the nuclear and cytoplasmic regions. Cytoplasmic and nuclear measurements were repeated for the 'Colour_2' (brown) deconvolved channel. After measurements, cellular masks were combined and selection inverted, and the background intensity measured in brown and blue channels. Data was then exported to Excel. For each picture in the brown channel the background intensity was subtracted to the cytoplasmic intensity. Beanplots were generated in R-3.3.3 using beanplot package.

Cell cycle: At day 9 of mESC differentiation, cell cycle was analysed with Click-iT EdU Alexa Fluor 488 Flow Cytometry Assay Kit (C10425; ThermoFisher Scientific, Hemel Hempstead UK). Cells were incubated for 1 h at 37°C with 10 μM EdU, then harvested and labelled with PE conjugated anti-kit and APC conjugated anti-CD41 antibodies for 20 minutes at 4°C. CD41/kit subpopulations were FACS-purified (FACS Aria II BD, UK). After fixation and permeabilization, Click-iT EdU Alexa Fluor 488 was detected and DNA was stained with SYTOX Advanced Dead Cell Stain Kit (S10349, ThermoFisher Scientific). Cells were analysed on a FACS Canto II (BD Biosciences).

Apoptosis: At day 8, cells were harvested and labelled with APC-conjugated anti-kit and PE-labelled anti-CD41 antibodies for 20 minutes at 4°C, washed in PBS with 10% FCS, resuspended in Binding Buffer (556454; BD Biosciences) and labelled with FITC conjugated Annexin V (A13199, ThermoFisher Scientific, UK). DNA was stained with SYTOX Advanced Dead Cell Stain Kit (S10349; ThermoFisher Scientific) at room temperature for 15 minutes. Cells were analysed on a FACS Canto II (BD Biosciences).

Mathemathical Modelling: For the mathematical model we use a system of differential equations that describe the total number of cells for each population (P1, P2, P3, P4 and DN) and how they change over time.

$$\dot{x}_i = \sum_j A_{ij} x_j$$

Most of the A_{ij} are zero, as we only allow transitions described experimentally (see **Figure 2C** from main text). Also, aside from the cell growth, all the rates are balanced so that they represent movement between compartments. The above equation can be solved using the matrix exponential form:

$$X(t) = e^{At} X_0$$

We use a Bayesian approach to infer the model parameters from the data. The data used for fitting are the time series of measured proportions. We convert the solution of the ODEs to proportions $\pi_i(t) = x_i(t) / \sum x_j(t)$ and use a beta likelihood for the measured proportions p_i

$$p_{ik} \sim Beta(\pi_i(t_k)\sigma_i, (1 - \pi_i(t_k))\sigma_i)$$

where σ_i is a population specific parameter that controls the variability of the distribution. The proportions are sufficient to infer the transitions, however information on total cell numbers is not constrained. To constrain this part of the model we can use the total cell numbers that were experimentally measured separately. Rather than constraining the total number to a single value we specify that it comes from a normal distribution with mean the empirical mean and variance the empirical variance.

For the priors on the transition parameters we used a half-normal(0, 2).

The Bayesian model was coded using the probabilistic modelling language Stan.

Patient Samples: Parents gave written informed consent in accordance with the Declaration of Helsinki, and the study was approved by the Thames Valley Research Ethics Committee (06MRE12-10; NIHR portfolio no. 6362). Patient details are provided in **Supplemental Table 3.** Mononuclear cells from umbilical cord blood were enriched by ficoll gradient centrifugation and cryopreserved in fetal bovine serum supplemented with 10% DMSO. GATA1 mutation status was determined by targeted next generation sequencing as before[9].

Human Megakaryocyte Differentiation Assay: 20,000 CD34+ cells from each sample were sorted with the FACS Aria Fusion into 100 uL StemSpan™ SFEM II (Stem Cell Technologies, Cat No:09605) supplemented with human SCF (50 ng/ml, Peprotech), human TPO (100 ng/ml) (Peprotech), and 1% Penicillin-Streptomycin (Gibco, Cat No:15140122) and incubated at 37°C, 5% CO_2 for 15 days. Every 3 days, half of the culture was removed and replaced with fresh media. Every 3 days, half of the culture was analysed for megakaryocyte maturation by flow cytometry. Cells were washed and stained with antibodies against human CD117 (kit), CD235a, CD41 and CD42b (**Supplemental Table 1**). To discriminate dead cells, LIVE/DEAD™ Fixable Dead Cell Stain (Thermo Fisher, Hemel Hempstead UK) was used as recommended. Samples were run on a FACS Aria Fusion flow cytometer (Becton Dickinson, Oxford UK) and analysed with FlowJo-10.6.0 (Becton Dickinson, Oxford UK).

Supplemental Figure Legends

Supplemental Figure 1. Targeting of murine Gata1 locus and flow cytometric analyses. A) Top, murine *Gata1* (m*Gata1*) locus. Exons 1-6 are numbered (blue boxes). Below, targeting constructs used to generate the bio*GATA1* and bio*GATA1s* knock-in alleles with position of 5' and 3' homology arms (red boxes), the sequence encoding the AVI-TEV-FLAG tag (yellow box), the *neomycin* resistance gene (green arrow) and the *loxP* sequences for cre-recombinase.

B) Positions of the restriction enzymes sites (H, *Hind*III and B, *Bst*XI) with sizes (in base pairs) of the DNA restriction fragments and DNA probes (A and B) used for Southern blot screening in the wild type, modified bioGATA1 and bioGATA1s loci. Loci are depicted as in A.

C) Southern blot analysis of *Hind*III (top panel) and BstXI (bottom panel) digested ES cell DNA. Sizes of the digested fragments are shown.

D) Position of forward (fwd) and reverse (rev) primers in m*Gata1* loci used for PCR genotyping. Sizes of the amplified fragments are shown.

E) Agarose gel of PCR amplified DNA products, using fwd and rev primers shown in D, using template ES DNA as indicated.

F) Representative flow cytometry plots showing the serial gating strategy to analyse cultures at day 6 (d6) (top) and day 12 (d12) (below) based on morphology, isolation of single cells and Hoechst exclusion. On the right, Fluorescence Minus One (FMO) controls used to set positive gates for flow cytometry plots in **Figure 1E**.

G) Representative flow cytometry plots of kit and CD41 expression of d6 BirA (left), bioG1 (middle) and bioG1s (right) cultures of before (top) and after (below) CD41 bead enrichment. The box in the plots indicates the cell population sorted for d6-d12 cultures in the experiments.

H) Representative flow cytometry plots showing the serial gating strategy for flow cytometry plots in **Figure 1F**. Top, gates based on morphology, isolation of single cells and Hoechst exclusion. Below, single stained (left) and unstained (right) controls.

Supplemental Figure 2: FACS analysis of re-cultured FACS-purified P1-P4 populations from BirA, bioG1 and bioG1s cultures.

A) FACS-purified hemopoietic progenitors, $kit^{hi}CD41^{lo}$ (P1) from d6 EBs were cultured on OP9 stromal cells with cytokines for another 6 days. Representative flow cytometry plots of kit and CD41 expression from BirA (left), bioG1 (middle) and bioG1s (right) cultures. Aliquots

of culture were analysed daily for kit and CD41 and were monitored for 6 additional days (d7 to d12). The numbers indicated in the gates represents the mean percentage (±1SD) of parent population from 3 independent experiments. The data refer to **Figure 2D**.

B) Representative FACS plots of kit and CD41 expression. P1-P4 populations from either BirA or bioG1s at d8 were FACS-purified (left plot) and then re-cultured for another two days (d8+1, middle plot; d8+2, right plot). The numbers indicated in the gates represent the mean percentage (±1SD) of parent population from 3 (BirA) to 4 (bioG1s) independent experiments. The data refer to **Figure 2E**.

Supplemental Figure 3: Flow cytometric and transcriptomic analyses of P1 to P4 populations.

A) Representative flow cytometry plots showing serial gating strategy to isolate live-singlet cells based on morphology and Hoechst exclusion. Right, Fluorescence Minus One (FMO) controls used to set positive gates for **Figure 2D**.

B) Gating strategy to isolate P1 to P4 subpopulations from bioG1s and BirA clones (left) and post-sort purity check of sorted subpopulations from bioG1s clone (right). The data refer to **Figure 2E**.

C) PCA plots showing P1 to P4 populations from each genotype using 100 to 10,000 genes with the highest variance across all populations. Percentage variance for each PC is shown.

Supplemental Figure 4: Flow cytometric analyses of all populations for lineage markers and gene expression analysis.

A) Flow cytometry plots of Fluorescence Minus One (FMO) controls for kit, CD41, Ter119, Gr1 and Mac1 used to set gates for experiment in **Figure 3B**.

B) Flow cytometry plot of gating strategy to isolate live-singlet cells based on morphology and Hoechst exclusion for experiment in **Figure 3B**.

C) Representative flow cytometry plots of DN (kit^-CD41^-) cells (left) and expression of Ter119 (middle), Gr1 and Mac1 (right) in DN cells. The data refer to **Figure 3B**. Numbers in the

gates are mean percentages (±1SD) of the parent population from 3 independent experiments.

D) Representative flow cytometry plots showing gates used to sort each population at d10 for experiments in **Figure 3A, C, F, G, H**.

E) Representative flow cytometry plots showing the post-sort purity of sorted populations. Genotypes and populations are indicated.

F) Table summarising mean (±1SD) percentage purity of each sorted population used for gene expression analysis shown in **Figure 2C, 2G-H**. Mean of 2 independent experiments.

Supplemental Figure 5: Megakaryocyte differentiation and maturation.

A) Photographs of May-Grunwald-Giemsa staining of purified P1, P2, P3 and P4 d10 cells from BirA (top), bioG1 (middle) and bioG1s (bottom). Scale bars, 10 μm. N = 3 independent experiments.

B) Representative flow cytometry plots showing Fluorescence Minus One (FMO) controls for experiment in **Figure 3E**.

C) Representative flow cytometry plots of gating strategy to isolate single cells based on morphology and Hoechst exclusion for experiment in **Figure 3E**.

D) Representative flow cytometry plots of CD42b expression in d12 DN and P1-P4 populations in BirA (top), bioG1 (middle) and bioG1s (bottom). Left, populations were first gated on kit and CD41 expression and then analysed for CD42b (right). In CD42b plots mean percentage (±1SD) of $CD42b^+$ cells (relative to parental) population from 4 independent experiments is indicated. Data used for **Figure 3E**.

E) Bar plot of mean (±1SD) percentage of $CD42b^+$ cells in DN, P1, P2, P3 and P4 populations for each genotype. N = 4 independent experiments. p-values were calculated using Student t-test.

F) Heatmap of mRNA expression of erythroid, myeloid and megakaryocytic genes in d10 P1-P4 populations. Genes are displayed horizontally. Duplicate biological samples were analysed (displayed vertically). Cell genotype is indicated below the heatmap. Data shown

here were used to generate the heatmap in **Figure 3G** and the hierarchical clustering in **Figure 3H**.

Supplemental Figure 6: *In vivo* validation of the GATA1s phenotype.

A) Representative flow cytometry plots showing Fluorescence Minus One (FMO) used to set gates in the experiment shown in **Figure 6A**.

B) Representative flow cytometry plots showing gating strategy followed to isolate live-singlet cells based on morphology and Hoechst exclusion for the experiment shown in **Figure 6A**.

C) Representative flow cytometry plots of kit and CD41 expression (top) and CD16/CD32 expression (below) in d6 EB derived cells from BirA (left), bioG1 (middle) and bioG1s (right). Numbers show the mean percentage (±1SD) of parent population from 3 independent experiments. Data used for **Figure 6A**.

D) Representative flow cytometry plots showing Fluorescence Minus One (FMO) used to set positive gates in the experiments shown in **Figure 6C, E**.

E) Representative flow cytometry plots showing the gating strategy followed to isolate live-singlet cells based on morphology and Hoechst exclusion in the experiments shown in **Figure 6C, E**.

F) Representative flow cytometry plots of kit and CD41 expression from bioG1 (top) and bioG1s (bottom) E10.5 yolk sac cells. Expression of Ter119, Gr1 and Mac1 was analysed as indicated. These data have been used to generate the box plots shown in **Figure 6C**.

G) Representative flow cytometry plots of CD42b expression from bioG1 (top) and bioG1s (bottom) E10.5 yolk sac cells from P1 to P4 populations. These data have been used to generate the box plots shown in **Figure 6E**.

H) Heatmap of mRNA expression of megakaryocytic, erythroid and myeloid genes in bioG1 (left) and bioG1s (right) E10.5 yolk sac DN, P1-P4 populations. Genes are displayed horizontally. Cells were purified from two independent litters for each genotype, in 2 independent experiments. Each condition was analysed as technical duplicates. Data shown here were used to generate the heatmap in **Figure 6F**.

Supplemental Figure 7: GATA1s megakaryocytic phenotype recapitulated in human patients.

A) Representative flow cytometry plots showing the gating strategy followed to isolate live-singlet cells based on their morphology and their negativity for Near IR Fixable Dead Cell staining. Cells from the erythroid lineage were excluded by plotting CD235A (y-axis) vs SSC-A (x-axis). $CD235A^-$ cells were used to analyse megakaryocytic maturation shown in **Figure 7A**.

B) Fluorescent Minus One (FMO) used to set initial gating for CD235A, kit, CD41 and CD42b markers for the data shown in **Figure 7A, B**.

C) Representative flow cytometry plots showing the level of expression of kit and CD41 (left), CD42b and CD41 (middle) in live/$CD235A^-$ cells. The plot on the right is an overlay showing the level of expression of kit and CD41 in the $CD41^+/CD42b^{-/low}$ (orange gate), $CD41^+/CD42b^+$ (blue gate) and $CD41^-/CD42b^-$ (red gating) populations to identify megakaryocytic maturation shown in **Figure 7C**.

Supplemental Tables

Supplemental Table 1: **Table of all antibodies used**

Supplemental Table 2: List of genes analysed by Fluidigm

Supplemental Table 3: Human patients details

References for Supplemental Data

1. de Boer E, Rodriguez P, Bonte E, et al. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc Natl Acad Sci U S A. 2003;100(13):7480-7485.

2. Driegen S, Ferreira R, van Zon A, et al. A generic tool for biotinylation of tagged proteins in transgenic mice. Transgenic Res. 2005;14(4):477-482.

3. Nishikii H, Eto K, Tamura N, et al. Metalloproteinase regulation improves in vitro generation of efficacious platelets from mouse embryonic stem cells. J Exp Med. 2008;205(8):1917-1927.

4. Mao X, Fujiwara Y, Orkin SH. Improved reporter strain for monitoring Cre recombinase-mediated DNA excisions in mice. Proc Natl Acad Sci U S A. 1999;96(9):5037-5042.

5. Hamlett I, Draper J, Strouboulis J, Iborra F, Porcher C, Vyas P. Characterization of megakaryocyte GATA1-interacting proteins: the corepressor ETO2 and GATA1 interact to regulate terminal megakaryocyte maturation. Blood. 2008;112(7):2738-2749.

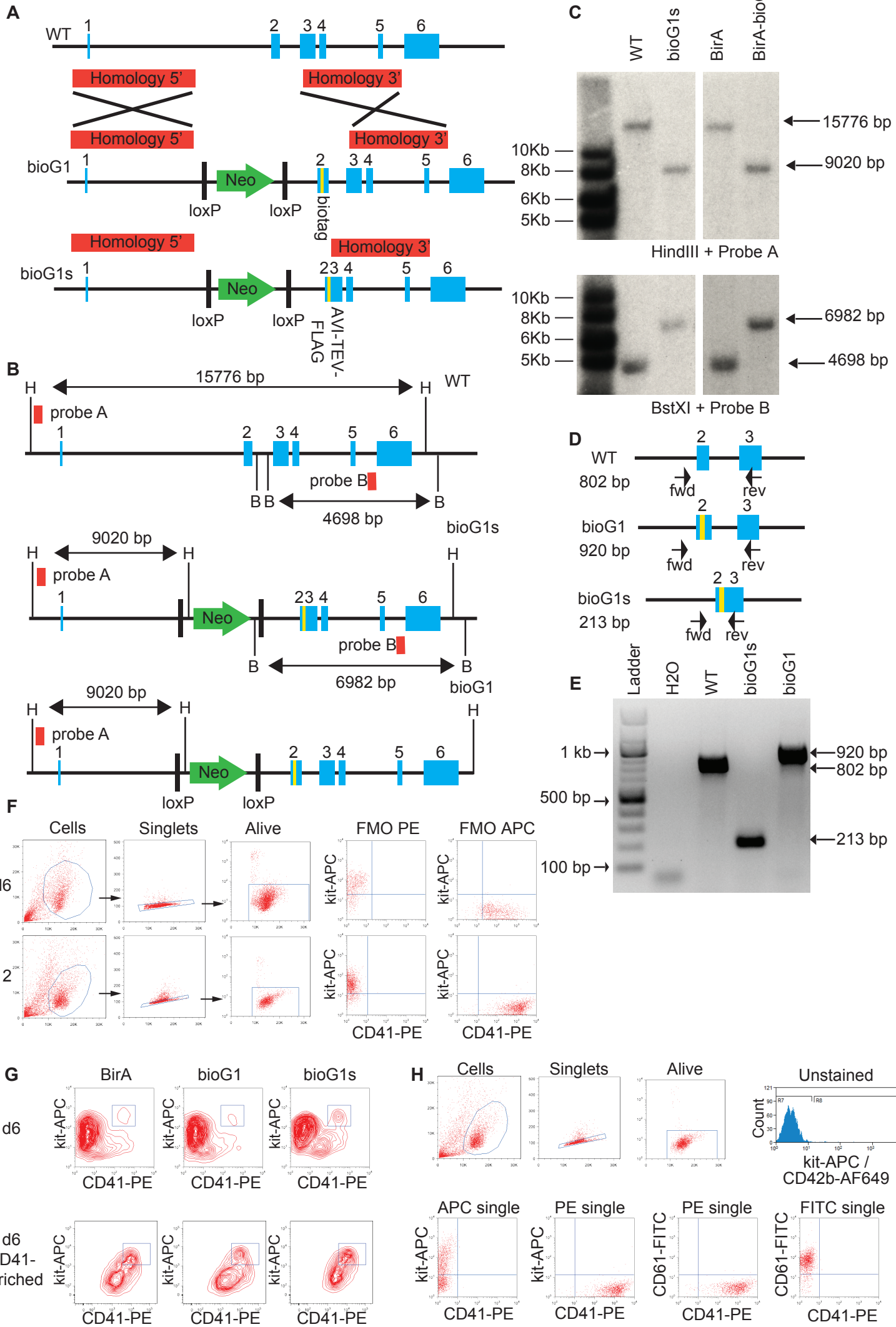
6. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions..Genome Biol. 2013; 4(4):R36.

7. Guo G, Luc S, Marco E, et al. Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. Cell Stem Cell. 2013;13(4):492-505.

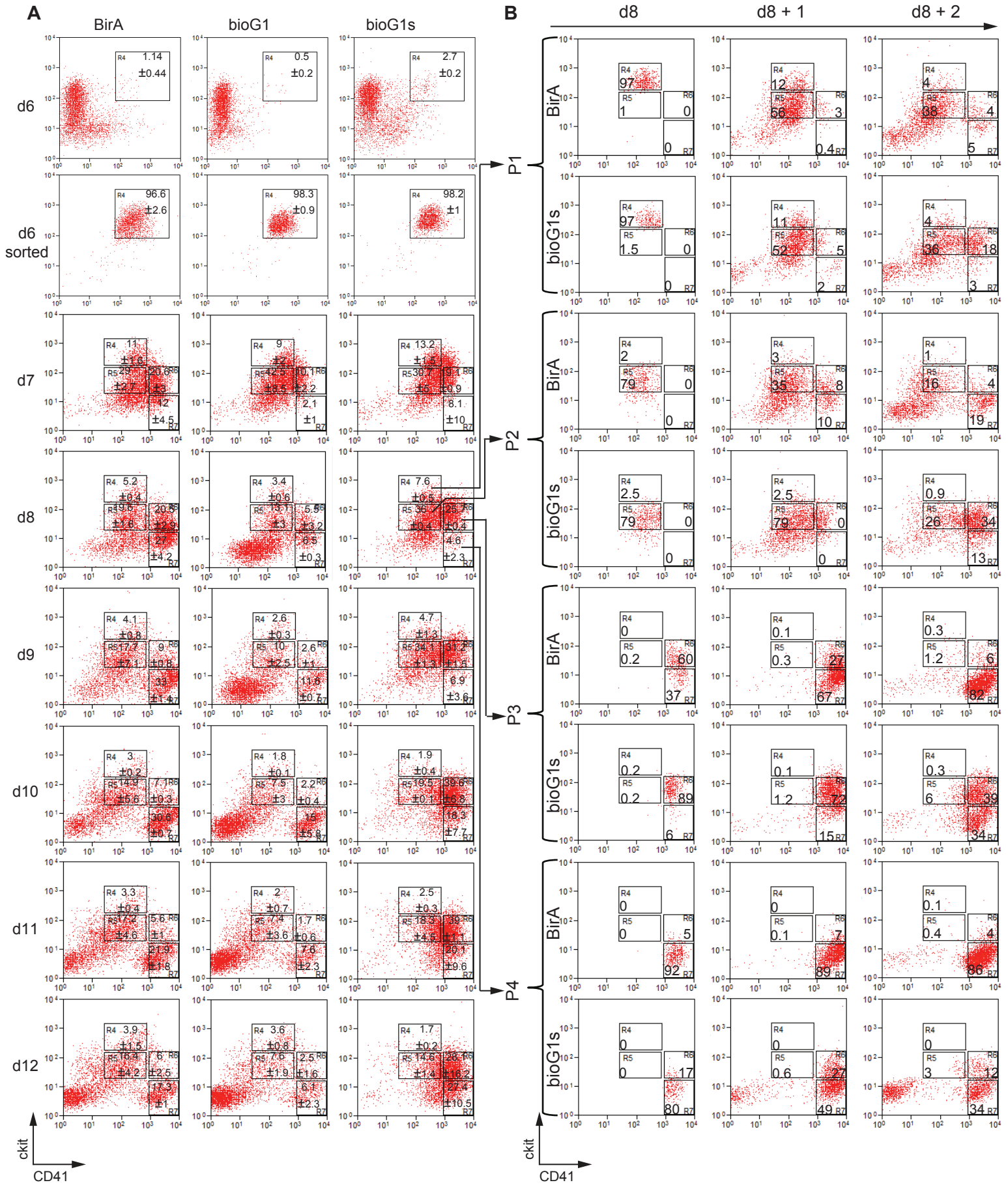
8. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676-682.

9. Roberts I, Alford K, Hall G, et al. GATA1-mutant clones are frequent and often unsuspected in babies with Down syndrome: identification of a population at risk of leukemia. Blood. 2013;122(24):3908-3917.

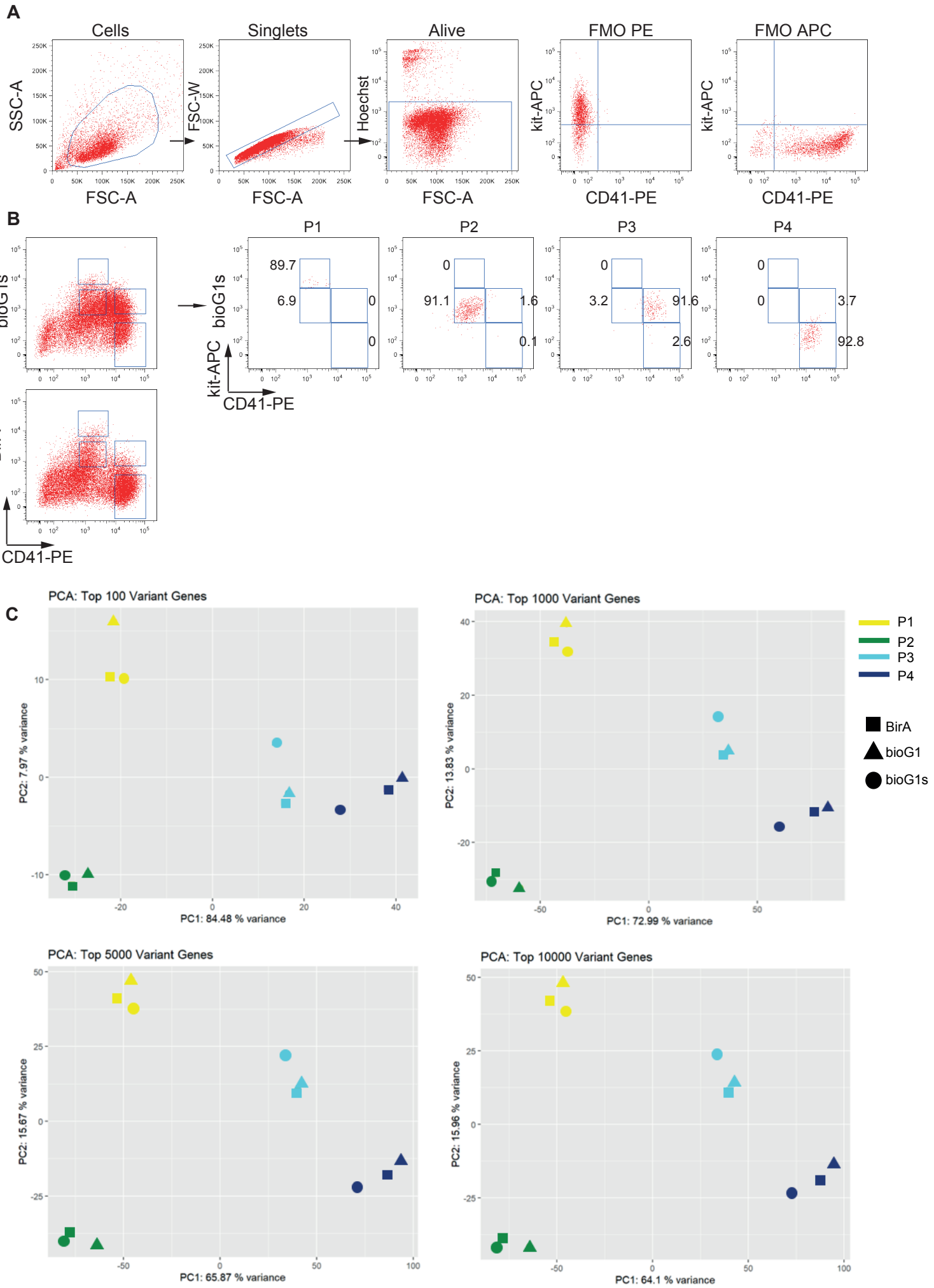
Supplemental Figure 1



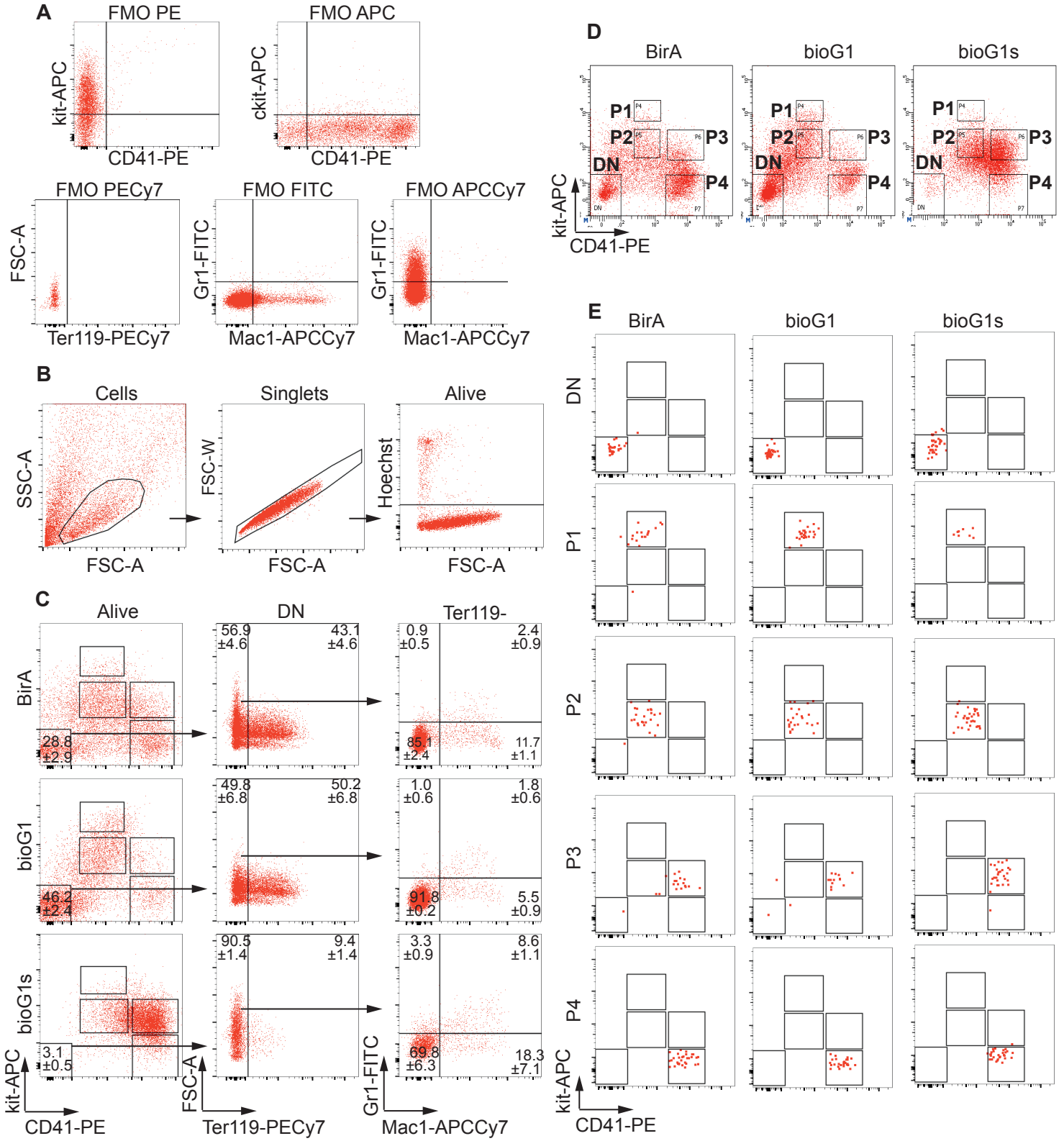
Supplemental Figure 2



Supplemental Figure 3



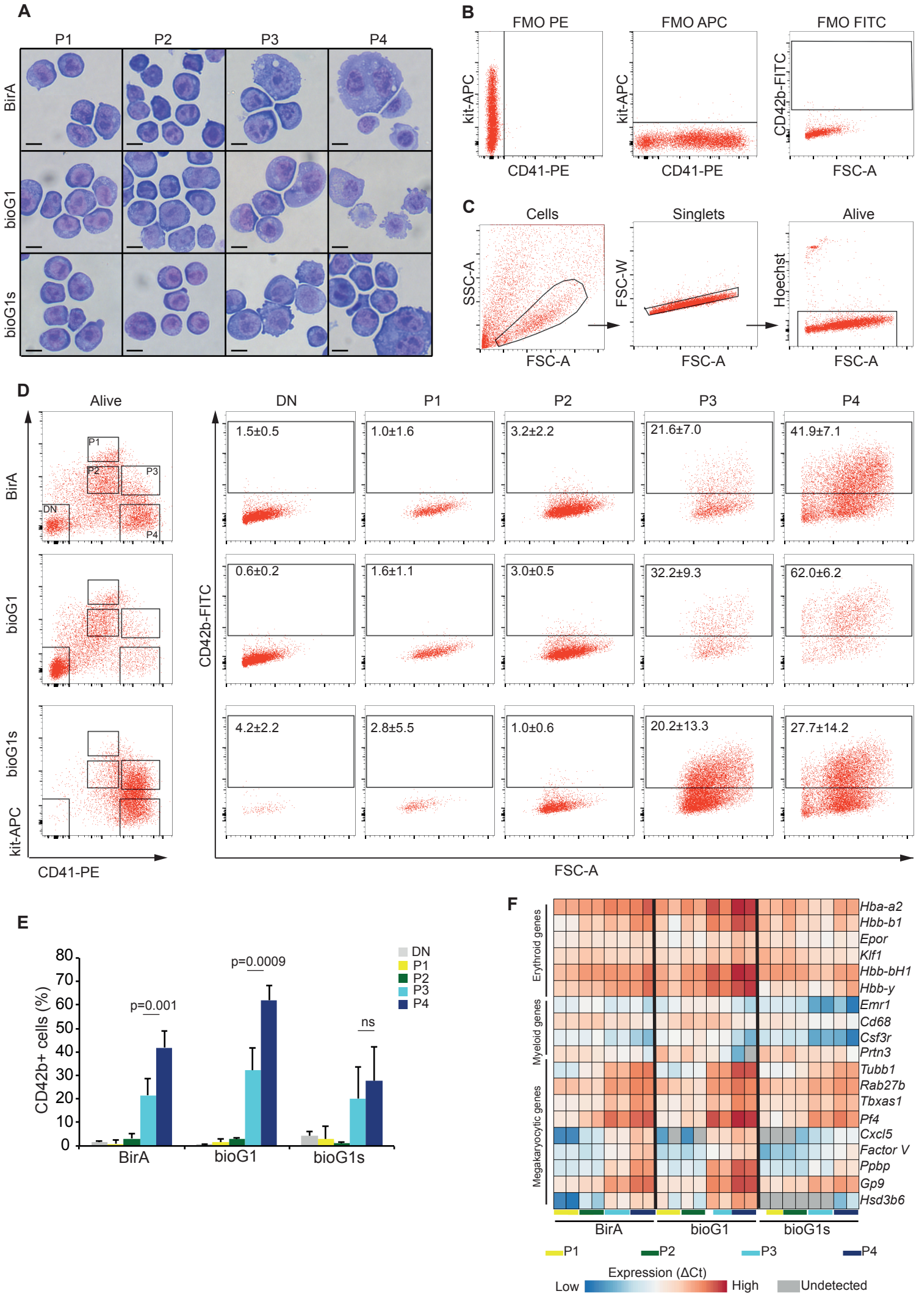
Supplemental Figure 4



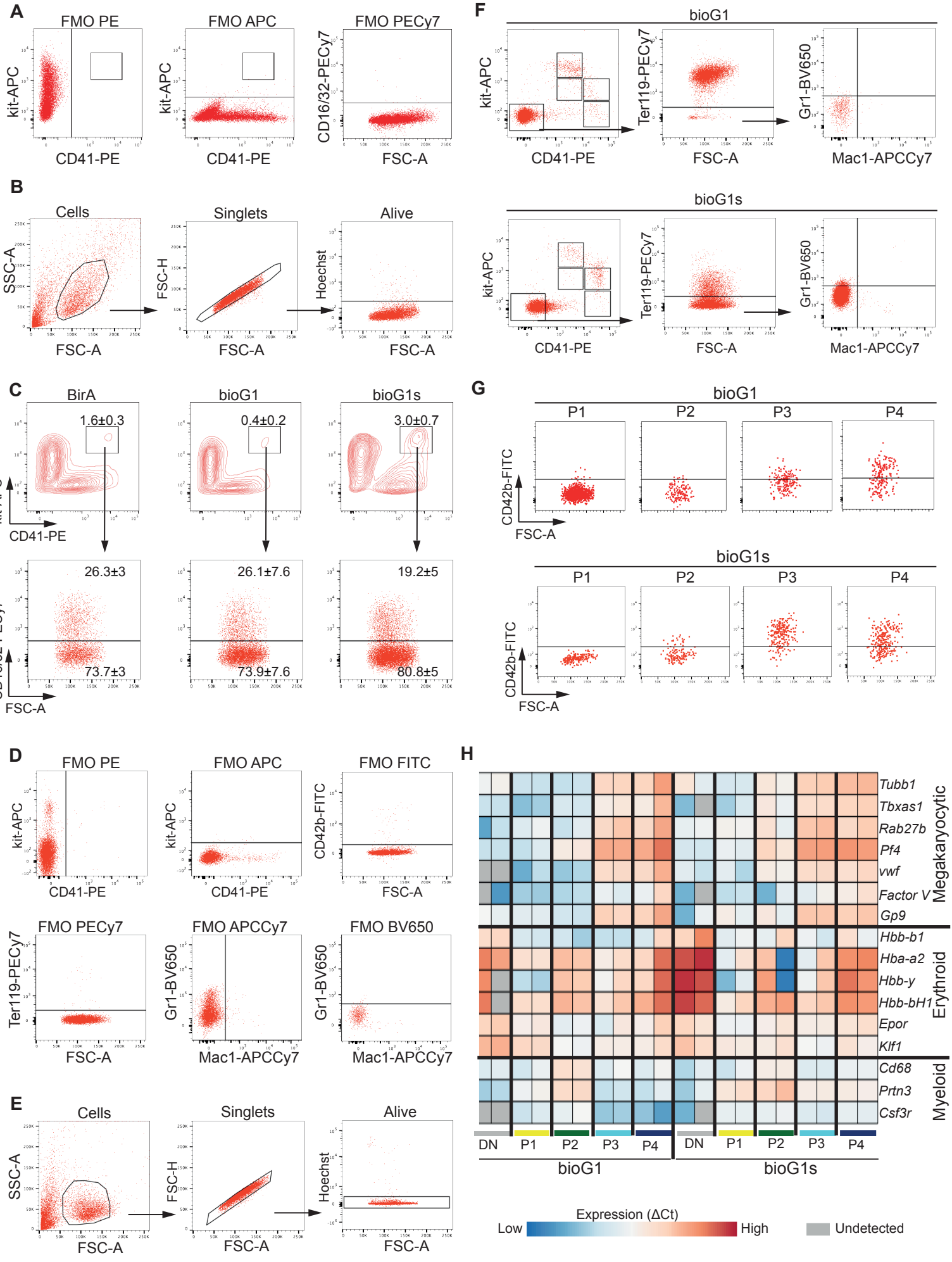
F

	DN	P1	P2	P3	P4
BirA	96.2±0.5	89.5±5.5	94.8±4.6	89.9±6.6	89.2±14.7
bioG1	99.4±0.5	89.1±3.4	97.6±2.6	87.5±6.5	88.2±14.7
bioG1s	90.0±14.1	95.4±5.1	97.7±2.0	90.2±5.9	87.8±12.5

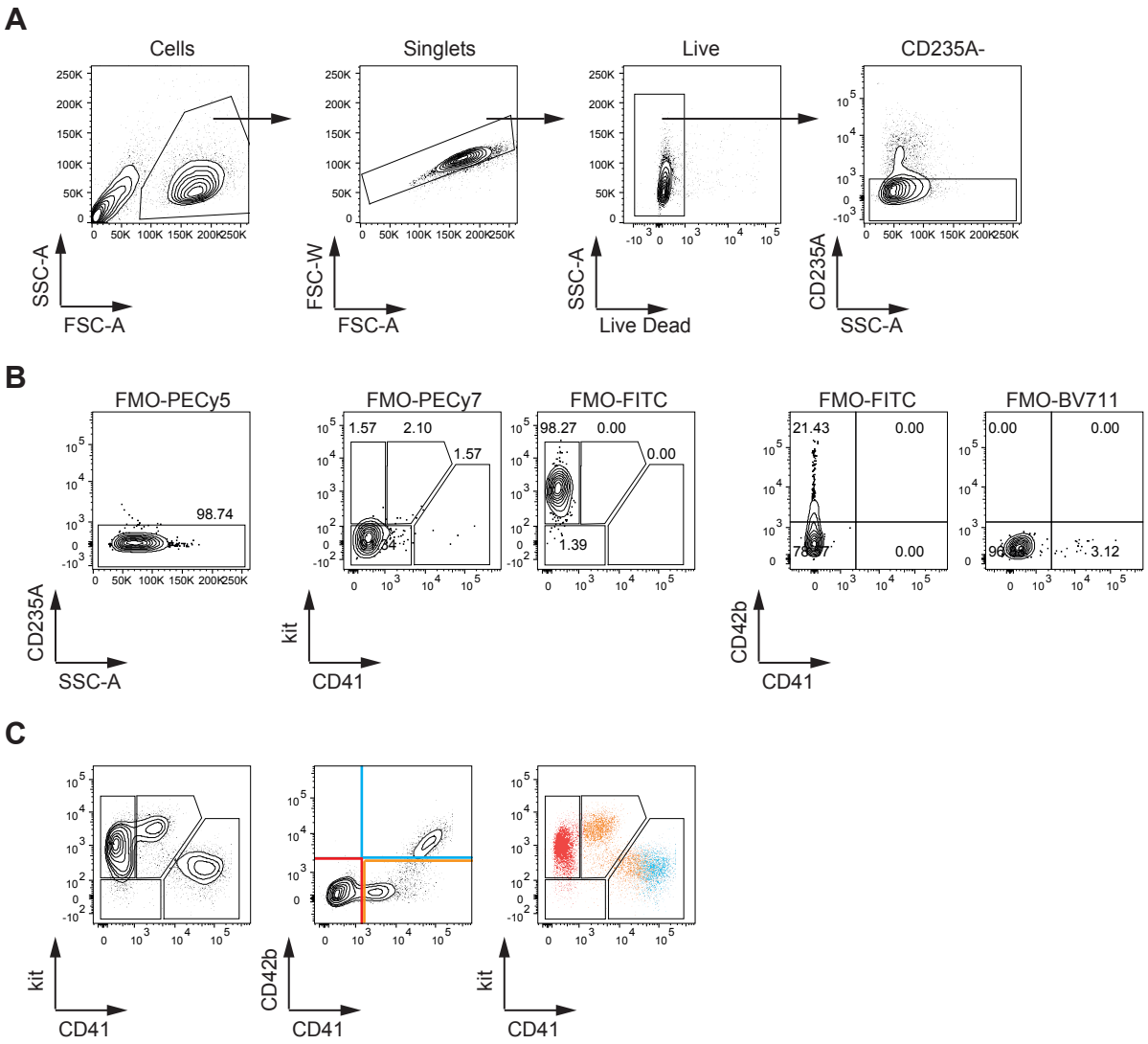
Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7



Supplemental Table 1: Table of all antibodies used

Antigen	Conjugate	Supplier	Clone	Reactivity
CD11b (Mac1)	APCCy7	BioLegend	M1/70	Mouse, Human
CD16/32	PECy7	eBioscience	93	Mouse
CD41	PE	BioLegend	MWReg30	Mouse
CD41	APC	BioLegend	MWReg30	Mouse
CD41	BV421	BioLegend	MWReg30	Mouse
CD41	FITC	Biolegend	HIP8	Human
CD42b	BV711	BD Biosciences	HIP1	Human
CD42b	FITC	Emfret	Xia.G5	Mouse
CD61	PE	BioLegend	2C9.G2 (HMß3-1)	Mouse, Rat
CD117 (kit)	PE	BioLegend	2B8	Mouse
CD117 (kit)	APC	BioLegend	2B8	Mouse
CD117 (KIT)	PECy7	Biologend	14D2	Human
CD235A	PECy5	Biolegend	HIR2	Human
Gr1	FITC	eBioscience	RB6-8C5	Mouse
Gr1	BV650	BioLegend	RB6-8C5	Mouse
Ter 119	PECy7	BD Pharmingen	TER-119	Mouse

Supplemental Table 2: List of genes analysed by Fluidigm

Lineage	Gene name
Erythroid	EpoR
	Hba-a1
	Hbb-b1
	Hbb-bH1
	Hbb-y
	KLF1
Megakaryocytic	CXCL5
	Factor 5
	Gp9
	Hsd3b6
	Pf4
	Ppbp
	Rab27b
	Tbxas1
	Tubb1
	vwf
Myeloid	CD68
	Csf3r
	Emr1
	PRTN3
Housekeeping	GAPDH
	HPRT
	TBP

Supplemental Table 3: Human patients details

Sample	Extra chr. 21	GATA1s mut	VAF (% of MNC)
CB	-	-	/
T21	+	-	/
TMD1	+	exon2:c.G220C:p.V74L	10.3
TMD2	+	exon2:c.108_109del:p.S36fs	9.2