Single lineage transcriptome analysis reveals key regulatory pathways in primitive erythroid progenitors in the mouse embryo

Joan Isern 1,5,#, Zhiyong He 1,5, Stuart T. Fraser 1,5,6¶, Sonja Nowotschin 7, Anna Ferrer-Vaquer 7, Rebecca Moore 1,2, Anna-Katerina Hadjantonakis 7, Vincent Schulz 8, David Tuck 9, Patrick G. Gallagher 8 and Margaret H. Baron 1-6 *

Departments of Medicine 1, Developmental and Regenerative Biology 2, Oncological Sciences 3 and Gene and Cell Medicine 4, The Tisch Cancer Institute 5 and The Black Family Stem Cell Institute 6, Mount Sinai School of Medicine, New York, NY

7 Developmental Biology Program, Sloan-Kettering Institute, New York, NY

Departments of 8 Pediatrics and 9 Pathology, Yale U., New Haven, CT

* Corresponding author: Margaret H. Baron, MD PhD, Mount Sinai School of Medicine, Box 1079, 1468 Madison Avenue, Annenberg 24-04E, New York, NY 10029-6574; Phone: (212) 241-0825; FAX: (212) 849-2442; email: margaret.baron@mssm.edu

¶ Current address: Discipline of Physiology, School of Medical Sciences, University of Sydney, Camperdown NSW 2050, Australia

# Current address: Department of Cardiovascular Developmental Biology, Fundación Centro Nacional de Investigaciones Cardiovasculares (CNIC) Carlos III, Madrid 28029, Spain

Running Title: Embryonic erythroid development in the mouse

Scientific category: Red cells, Iron and Erythropoiesis
Abstract

Primitive erythroid (EryP) progenitors are the first cell type specified from mesoderm, late in gastrulation. We used a transgenic reporter to image and purify the earliest blood progenitors and their descendants from developing mouse embryos. EryP progenitors exhibited remarkable proliferative capacity in the yolk sac immediately before the onset of circulation, when these cells comprise nearly half of all cells of the embryo. Global expression profiles generated at 24 hour intervals from E7.5 through E12.5 revealed two abrupt changes in transcript diversity that coincided with the entry of EryP into the circulation and with their late maturation and enucleation, respectively. These changes were paralleled by expression of critical regulatory factors. Experiments designed to test predictions from these data demonstrated that the Wnt signaling pathway is active in EryP progenitors, whose numbers are regulated by TGF-β1 and hypoxia and which display an aerobic glycolytic profile. This is the first transcriptome assembled for a single hematopoietic lineage of the embryo over the course of its differentiation.
Introduction

The regulation of lineage commitment and differentiation of progenitor cells is a fundamental problem in developmental biology. In the postimplantation mammalian embryo, primitive erythroblasts (EryP) or red blood cells are the first cell type to be specified from nascent mesoderm, late in gastrulation \(^1\). EryP emerge in great numbers within the "blood islands" of the yolk sac (YS) and constitute the predominant circulating blood cell until a second wave of definitive, enucleated erythrocytes (EryD) is produced by the fetal liver \(^2-4\). EryP are crucial for the transition from rapidly growing embryo to fetus: failure in primitive erythropoiesis is uniformly associated with embryonic lethality. In addition to their function in oxygen delivery to cells within the embryo, EryP are thought to play a critical role in vascular remodeling during development \(^5,6\). The importance of this lineage is underscored by the fact that primitive erythropoiesis is conserved among vertebrate species \(^7\).

In the mouse, EryP progenitors are found in the YS between embryonic day (E) 7.25 and E9.0 \(^4\). Their numbers decrease abruptly within the next 12 hr and, by E9.5, when embryonic circulation has initiated, they can no longer be detected \(^4\). As EryP circulate, they continue to mature in a stepwise, essentially synchronous fashion \(^8\). These nucleated erythroblasts undergo a series of dramatic cellular and morphological changes, including upregulation of embryonic globin genes, expression of cell adhesion proteins, cytoskeletal reorganization, decreased cell proliferation, nuclear condensation, and, finally, from E12.5-E14.5, nuclear extrusion \(^8-11\). The enucleated EryP are rapidly outnumbered by adult type erythrocytes but remain in circulation through the end of gestation our unpublished data and ref. \(^8\). The transient appearance of EryP progenitors in the embryo and the synchronous, stepwise maturation of their progeny makes this lineage an attractive model for cell specification and terminal differentiation.
Early hematopoietic cells remain poorly characterized because of the relative inaccessibility and small size of the embryo, the transient appearance of their progenitors, and the lack of suitable cell surface markers for their isolation. Detailed chronological expression profiling will be essential for an understanding of the genetic networks that regulate the development and differentiation of the EryP lineage and for comparative analyses of embryonic versus adult erythropoiesis. We have developed a transgenic mouse system in which a nuclear green fluorescent protein (GFP) reporter is expressed specifically in EryP, allowing the tracking of these cells and their nuclei throughout gestation. Here, we show that expression of this transgene can be used to mark and prospectively isolate the earliest hematopoietic progenitors of the mouse embryo from their first appearance at ~E7.5. To identify the processes necessary for commitment, expansion, maturation and terminal differentiation of progenitors for this first hematopoietic lineage, a global transcriptional analysis was performed for successive stages of development from E7.5 through E12.5 and revealed not only well studied red blood cell genes but also some surprises. Experiments were designed to test predictions based on the expression profiles. We show that the Wnt pathway functions autonomously in EryP progenitors, the numbers of which are regulated by TGFβ1 and hypoxia. Interestingly, EryP progenitors express genes associated with aerobic glucose metabolism (the Warburg effect), a phenotype characteristic of cancer and other rapidly proliferating cells. This study is the first comprehensive genome-wide expression profiling undertaken for a single lineage from the gastrulating embryo and will provide a valuable resource for understanding early hematopoietic development.
Materials and Methods

Detailed experimental procedures are described in Supplementary Information. This study was approved by the Mount Sinai School of Medicine IACUC committee.

Mouse lines and embryo dissection

The \( \varepsilon \)-globin::H2B-EGFP transgenic mouse line has been described previously \(^{11}\). The Wnt reporter line TCF/Lef::H2B-GFP is described in ref.\(^{13}\). Embryos were dissected as described \(^{8,11}\).

Primary Microarray Data Acquisition and Analyses

The labeled cRNA samples were hybridized to Illumina Mouse WG-6 v1.1 Expression Bead Chip genome-wide arrays. The data files generated by the EryP array analyses have been submitted to Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24127) for use by other investigators at accession code GSE24127. The expression level cutoffs were set at 7.2 for E7.5 and E8.5 amplified samples and at 6.0 for E8.5-E12.5 samples (all log2 scale).

Flow cytometry and cell sorting

Cell suspensions were sorted based on GFP fluorescence using an Influx (Cytopeia, Seattle, WA) or MoFlo (Dako-Cytomation, Glostrup, Denmark) cell sorter, with dead cells excluded by propidium iodine staining. For multicolor cell sorting, staining with 6-diamidino-2-phenylindole (DAPI) was used to exclude dead cells and surface fluorescence was analyzed using a LSR II flow cytometer (Becton Dickinson, Franklin Lakes, NJ). Data analysis was performed using FlowJo software (Tree Star Inc., Eugene, OR). Monoclonal antibodies used in this study are listed in Table S1.
Quantitative RT-PCR analysis and transcript verification

RNA from sorted cell populations was reverse transcribed to cDNA using Superscript III reverse transcriptase and oligo dT(20) as described. qRT-PCR was performed using either Taqman primers and probes or standard oligonucleotide primers and SYBR Green PCR master mix. Oligonucleotide primers are listed in Table S2.

Imaging

Images were acquired using a Zeiss axiocam MRc or MRm camera, analyzed with the Axiovision software package (Zeiss, Jena, Germany), and post-processed using Photoshop (Adobe, San Jose, California). Whole embryos were imaged on a Lumar V12 stereomicroscope (Zeiss, Jena, Germany) fitted with a NeoLumar S 1.5X objective. For time-lapse experiments, laser scanning confocal data were acquired using a Zeiss LSM510 META scan head fitted onto a Zeiss Axiovert 200M, with a plan-apochromat 10x/0.45 NA lens. For fluorophore excitation, a 488nm Argon laser line (3% power) was used. Giemsa-stained cytospun cells were imaged using an inverted Zeiss Axiovert 25 microscope (Zeiss, Jena, Germany) fitted with a 32X (LD-A-Plan/0.4 NA) objective.

Immunohistochemistry

Indirect fluorescent immunostaining was performed as described previously. Fluorescent images were acquired using a Zeiss AxioCam digital color camera mounted on a Zeiss AxioPlan microscope outfitted with a 63X objective.
Results

The *ε-globin::H2B-EGFP* transgene labels the first committed EryP progenitors in the developing mouse embryo

The identification and isolation of the first hematopoietic (EryP) progenitors has presented challenges owing to a lack of specific cell surface markers. While CD31 (PECAM1) and Tie-2 expression have been used to enrich for EryP progenitors from E7.5 embryos \(^{14}\), both of these proteins are also expressed on vascular endothelial and definitive hematopoietic progenitors. Similarly, low level expression of CD41 has been used to enrich for EryP progenitors \(^{15}\). CD41 is expressed on definitive hematopoietic but not on endothelial progenitors \(^{15,16}\). We have developed a transgenic mouse model in which expression of a histone H2B-GFP reporter is driven by human *epsilon-globin* regulatory elements specifically within the EryP lineage; this line has been validated by a variety of criteria and GFP shown to mark essentially all EryP but not EryD \(^{11,12}\). In addition, the same regulatory elements have been used by us for two related transgenic mouse lines \(^{8,17-20}\). While *globin* gene expression has traditionally been viewed as a late marker of differentiation, embryonic *globins* are in fact transcribed in the YS from as early as E7.5 \(^{17,18}\). We, therefore, asked whether the *ε-globin::H2B-EGFP* transgene is also expressed at this developmental stage and whether it might permit the identification and enrichment of EryP progenitors.

We first determined the time of onset of GFP fluorescence in *ε-globin::H2B-EGFP* embryos (Figure 1A). The first GFP(+) cells (EryP) were evident in the extraembryonic region of late streak (LS)/early bud (EB) embryos (~ E7.5) (Figure 1A). From neural stages onwards, all embryos displayed a ring of GFP(+) cells encircling the YS. Therefore, *ε-globin::H2B-EGFP*
transgene expression is activated toward the end of gastrulation, at the time when the first erythroid cells are specified from nascent mesoderm \(^1\). The emergence and expansion of GFP(+)/EryP were visualized using time-lapse confocal microscopy of cultured embryos (Figure 1B and Movie S1). The static and time-lapse images of GFP-expressing cells in the early postgastrulation embryo further confirms the utility of the \(\epsilon\)-globin::H2B-EGFP line as a model for primitive erythropoiesis.

EryP progenitors are present in the mouse embryo during a narrow developmental window: they first appear at E7.5 and are absent by E9.0 \(^{1,4}\). As bright GFP expression was easily detected as early as E7.5 in \(\epsilon\)-globin-H2B-EGFP transgenic embryos, we asked whether EryP progenitors could be identified by GFP expression. Whole E7.5 or E8.25 \(\epsilon\)-globin::H2B-EGFP embryos were dispersed and GFP(+) cells were isolated using FACS (Figure 1C) and were plated in clonogenic assays in methylcellulose. At E7.5 and E8.5, as many as 15% and 48%, respectively, of all cells in the embryo expressed GFP (Figure 1C). The sorted GFP(+) cells (Figure 1D) gave rise to EryP colonies (EryP-CFC, Figures 1D, S1A) containing green fluorescent erythroblasts. No EryP-CFC formed from the GFP(-) fractions. The number of EryP progenitors per embryo decreased significantly from E7.5 to E8.5 and were absent by E9.5 (Figure S1A; data not shown). GFP(+) cells from E8.5 embryos did not contain progenitors for definitive hematopoiesis (not shown). The endogenous \(\beta\)h1- and epsilon-globin genes were expressed only in the GFP(+) population, as early as E7.5 (Fig. 1E).

Expansion of the developing EryP lineage is rapid, as indicated by GFP expression (Figure 1A) and cell doubling times (Figure S1B). The highest proliferative activity was observed during the 24 hr period when EryP are first detected in the YS (E7.5) to shortly before the onset of
circulation (E8.5), with a >30-fold increase in cell numbers and a doubling time of ~4 hr (Figure S1B). By E12.5, when EryP proliferation has essentially ceased, the population has expanded by 30,000-fold (~14-15 cell divisions, Figure S1B). At present there is no unequivocal method to determine the percent of GFP(+) cells that are progenitors (give rise to colonies in clonogenic assays). In contrast with the large increase in GFP(+) cell number, progenitor numbers increase by only ~3 to 4-fold from E7.5 to E8.5 (our estimates from multiple experiments). These differences may indicate that not all of the GFP(+) cells found in the E7.5-E8.5 yolk sac are progenitors. In our best experiments, 25-30% (E7.5) or 10% (E8.5) of cells plated yield EryP colonies. However, it is possible that, during experimental manipulation (dissection, embryo dispersion to single cells, FACS sorting, and, finally, plating in methylcellulose), some cells are lost, damaged, or begin to differentiate prior to plating in methylcellulose.

**Gene expression profiling of the EryP lineage throughout development**

The specific expression of GFP in maturing EryP and their progenitors allowed us to isolate these cells at distinct stages of differentiation and to define the EryP transcriptome during development. GFP(+) cells were FACS-sorted from ε-globin::H2B-EGFP embryos at 24 hr intervals during gestation, from E7.5 through E12.5 (Figure S2A). Total RNA was extracted from the purified GFP(+) cell populations (3 replicates per stage; Figure S2A legend), converted to labeled cRNA, and hybridized to the Illumina Mouse WG-6 beadchip platform. As the number of sorted cells obtained from pooled E7.5 embryos was very small (7,000-21,000 cells from >70 embryos per pool), it was necessary to amplify this RNA. RNA was also amplified from pooled E8.5 embryos, to permit direct comparison between E7.5 and E8.5 stages. Thus, two complementary EryP datasets were generated: one from amplified samples (E7.5 and E8.5, progenitor stages) and a second from samples that were not amplified (E8.5-E12.5, maturing
EryP). Pairwise comparison between samples revealed high sample-to-sample correlation (>98% concordance), with tight clustering between replicate samples from the same stage (Figures S2B, S2C).

**Dynamic gene expression patterns observed during EryP maturation**

The greatest numbers of increases and decreases in gene expression (>2-fold) were detected during two 24 hr windows, from E8.5-E9.5 (transition from YS to circulation) and from E11.5 to E12.5 (fetal liver stage) (Figure 2A), coinciding with key physiological changes in the embryo. Interestingly, we observed a nearly reciprocal expression of transmembrane protein genes in E8.5 versus E11.5 EryP (Figure S2D) suggestive of opposing functions that may reflect the need for these cells to respond to the distinct microenvironmental niches of the YS and circulation.

To identify functionally related groups of genes that cooperate in biological processes involved in erythroid cell maturation, hierarchical clustering of the gene expression data was performed. Genes clustered into seven major patterns (Figure 2B). We analyzed Gene Ontology (GO) terms for biological processes and found that each cluster was enriched for distinct functional classes of genes (Figure 2C). The GO terms that are overrepresented in clusters of genes that are upregulated during EryP maturation include heme biosynthesis, iron transport, macromolecule catabolism and autophagy (Figure 2C), all processes associated with erythroid maturation. Conversely, the GO terms overrepresented in clusters containing genes that are downregulated include DNA replication, ribosome and nucleolus biogenesis, and glucose metabolism, suggesting that early EryP are heavily engaged in these processes and that these processes decrease in significance as the cells mature. Cluster 7, in which genes are upregulated through E11.5 and are then rapidly downregulated, includes genes involved in DNA packaging and
chromatin assembly (Figure 2C). The peak at E11.5 may represent a “transition state” through which EryP pass as they enter the terminal stages of differentiation.

To validate the microarray results, we confirmed the temporal expression pattern of a battery of transcripts detected in the EryP dataset using real-time RT-PCR (Figure S3C) and/or FACS (Figure S3D). Representative qRT-PCR analyses of genes (Gata1 and Igf2) whose expression was found to increase or decrease, respectively, during development are shown in (Figure 2D). The transcription patterns identified by microarray (not shown) and qRT-PCR analyses (Figure S3C) were very similar. Additional validation of the microarray data was obtained using FACS analysis of cell surface proteins such as CD31, β1-integrin, endothelial cell-selective adhesion molecule (ESAM), and c-kit data (Figure S3D and see below); these were down-regulated on EryP, as predicted from the transcriptional profiling. Other genes and their translated proteins, such as Transferrin Receptor, TfR (CD71) and Glycophorin A, Gypa (Ter119), were upregulated (Figure S3C,D).

Molecular signatures of the earliest EryP progenitors

The earliest hematopoietic progenitors in the developing mouse embryo are thought to arise from a population of mesoderm that transiently expresses the receptor tyrosine kinase Flk-1 (Vascular Endothelial Growth Factor (VEGF) Receptor-2 or Kdr). Therefore, we asked whether mesodermal or endothelial genes are expressed in EryP progenitors. While Kdr mRNA was not detected in E7.5 EryP (Figure S3A), the protein was detected on the surface of a subset (~50%) of EryP/GFP(+) cells at this stage (Figure 3A), suggesting transcription of the Kdr gene in an earlier precursor. Cdh5, which encodes the endothelial adhesion protein VE-cadherin, was expressed by EryP at E7.5 and was downregulated by E8.5. VE-cadherin protein was expressed
on a subset of GFP(+) cells at E7.5 and, like Flk-1, was rapidly lost around the time of erythroid commitment (Figure 3A).

We identified a large group of genes that are expressed at high levels at E7.5 but are downregulated within the next 24 hr of development (Figure S3B). These transcriptional decreases are likely required for progenitor expansion and, perhaps, for the earliest steps in EryP differentiation. The sharpest drop was observed for Gata2, required for the proliferation and survival of hematopoietic progenitors \(^2^1\). Downregulation of other transcription factor genes (Runx1, its co-regulator Cbfb, Lmo2, and Etv2) was also detected.

Several Wnt/\(\beta\)-catenin pathway genes (encoding the transcription factor \(\beta\)-catenin, the Wnt receptor Fzd7 and the secreted activator Rspo3) were also expressed in EryP at E7.5 and were then downregulated (Figure S3B and data not shown). We investigated the potential involvement of canonical Wnt activity in EryP progenitors using an antibody against activated (stabilized) \(\beta\)-catenin. As shown in Figure 3C, EryP (green nuclei) stained with the antibody. That the Wnt pathway is active in these cells is further suggested by the bright fluorescence of an H2B-GFP transgenic Wnt/\(\beta\)-catenin reporter (\(TCF/Lef::H2B\)-GFP) \(^1^3\) in the blood islands of the YS (Figures 3D, 3E).

A number of collagen genes were expressed in E7.5 and E8.5 EryP (Figure S3B), suggesting that EryP progenitors and their early descendants secrete components of the extracellular matrix of the YS. Genes encoding cytoskeleton, receptor and cell communication proteins were also downregulated between E7.5 and E8.5, perhaps reflecting the imminent transition of EryP from the YS into the bloodstream.
The developmental window from E7.5 to E8.5 was also associated with substantial increases in gene expression. Most prominent among these were the embryonically expressed $\epsilon Y$, $\zeta$, and $\alpha$-globin genes ($Hbb$-$y$, $Hba$-$x$, and $Hba$-$a1$) and the anion transporter band 3/$Slc4a1$ (Figure 3B). Additional erythroid lineage-specific genes, including Glycophorin A and Alas synthase 2, were already expressed at significant levels at E7.5 and were further induced by E8.5 (Figure 3B). Others, including Eklf/Klf1, Tal1/Scl, Erythrocyte band 4.1 and Fog-1, were expressed at high levels at both E7.5 and E8.5 (Figure S3B, Table S3).

**Gene expression changes during the transition of EryP from the yolk sac to circulation stage of their development**

During the transition from progenitor to differentiating erythroblast (E8.5 to E9.5), expression of the transcriptional regulators Gata1 and Cited2 increased while expression of Lmo2, Hmga1 and Gatad2a decreased (Figure 4A). Among other classes of genes whose expression decreased from E8.5 to E9.5, most notable were those involved in cell signaling, glucose metabolism, and nitric oxide metabolism ($Ddah1$) (Figure 4A, Table S4B).

At E7.5, EryP progenitors express Pecam1 and integrins $\alpha 5$, $\alpha 6$, $\beta 1$ and CD41 (Figures 4B, 4C). By E8.5, $\alpha V$, $\beta 2$- and $\beta 3$- integrins and CD44 were also detected (Figure 4C). From E8.5 to E9.5, expression of four adhesion molecule genes (Figure 4A) and their encoded proteins (Figures 4B, 4C, S3D) decreased. Most integrin protein and RNA expression was lost by E10.5 (not shown).

**Growth factor and cytokine pathways in EryP progenitors**

The growth factor requirements of EryP progenitor cells have not been well defined. Erythropoietin (Epo) is the sole cytokine known to stimulate the formation of EryP progenitors.
in vitro; EpoR expression was detected as early as E7.5 (Figure 5A). Profiling of E7.5 and E8.5 EryP allowed us to identify three growth factor signaling pathways that might regulate this lineage at the progenitor stage: c-kit/stem cell factor (SCF), Tie-2/Angiopoietin, and TGFβ1 receptor 1/TGFβ. At E7.5, EryP express mRNAs for receptors (c-kit, Tie-2, Acvr2b, Tgfbr1 and Tgfbr3) and downstream signaling components (e.g. Smad3 and Smad4) of these pathways (Figure 5A). With the exception of EpoR and Vegfa, their transcription begins to decrease by E8.5 (Figure 5A).

To determine whether the TGF-β signaling pathway can modulate the expansion of EryP progenitors, we tested recombinant forms of the three Tgfbr1 ligands TGFβ1, TGFβ2 and TGFβ3 in clonogenic assays. TGF-β1 showed a dose-dependent effect on progenitor activity: growth of EryP progenitors was inhibited at a low concentration (0.02 ng/mL) and stimulated at high concentrations (0.2 or 2 ng/mL) of this cytokine (Figure 5B). Neither TGFβ2 nor TGFβ3 had any effect on the numbers of EryP progenitors (not shown).

c-kit and Tie-2 proteins were detected on the surface of a significant subset of EryP at E7.5 and E8.5 but were nearly absent by E9.5 (Figure 5C). To determine whether c-kit marks EryP progenitors within the GFP(+) population (Figure 1D), GFP(+) c-kit(+) and GFP(+) c-kit(-) cells were sorted from E8.5 embryos and plated in clonogenic assays in methylcellulose (Figure 5D). The GFP(+) fraction that expressed c-kit gave rise to EryP colonies, while those cells lacking c-kit failed to do so (Figure 5D, right panel). Therefore, EryP progenitors are highly enriched within the c-kit(+) population.

While it has been reported that EryP progenitors are present within Tie-2(+) cell populations, these are heterogeneous. To determine the extent to which EryP progenitors are enriched in the
Tie-2(+) population of GFP(+) cells, EryP were FACS_sorted from E8.5 embryos and plated in methylcellulose. The Tie-2(+) fraction of GFP(+) cells showed a 4-fold enrichment of EryP progenitor activity (Figure 5E). The numbers of EryP progenitors were not consistently stimulated by addition of recombinant angiopoietin-1 (Ang-1), a Tie-2 ligand, or the c-kit ligand SCF, to the cultures (data not shown). It may be necessary to develop serum-free conditions for the growth of EryP progenitors in vitro to unmask possible Tie-2 and/or c-kit activity.

We next asked which cell types of the YS express genes encoding the cognate ligands for Tgfbr1 (Tgfβ1), c-kit (SCF), and Tie-2 (Ang-1). Cells of the three major YS lineages were FACS_sorted from E8.5 embryos on the basis of expression of Flk-1 (endothelial); GFP (from a Afp-GFP transgene 22; visceral endoderm, VE); and GFP from the ε-globin::H2B-EGFP transgene (EryP). Gene expression was analyzed using qRT-PCR (Figure 5F). YS endothelial cells expressed transcripts for all three ligands (Tgfβ1, SCF, and Ang-1), while only Ang-1 was expressed in VE. Therefore, EryP progenitor activity may be regulated by growth factors produced by endothelial and endodermal cells of the YS. Interestingly, Tgfβ1 was also transcribed in EryP (Figure 5F), suggesting the possibility of autocrine and/or paracrine signaling.

Changes in glycolytic gene expression during EryP maturation
Glucose metabolism was an overrepresented functional category in Cluster 5, in which genes are progressively downregulated (Figure 2B). To determine whether high glycolytic activity is a hallmark of EryP progenitors, we performed a qRT-PCR analysis (Figure S3) of genes predicted from the microarray (Figure 6A) to be transcribed in EryP and known to play pivotal roles in glucose metabolism. Hexokinase 2 (Hk2), Pyruvate kinase M2 (Pkm2), Lactate dehydrogenase a
(Ldha) and Phosphoglycerate kinase 1 (Pgk1) were all transcribed at high levels in E8.5 YS EryP, a profile characteristic of aerobic glycolysis (the "Warburg effect" 
23,24), but their expression decreased dramatically in cells that had entered the circulation (E9.5 and onwards; Figure 6A). In contrast, Pyruvate kinase, liver/red blood cell (Pklr) was not expressed in progenitor stage EryP but was rapidly upregulated in circulating EryP (Figure 6A).

**Oxygen tension is a regulator of EryP progenitor activity and gene expression**

The glucose metabolism-related genes (Figure 6A), as well as others identified in the transcriptome analysis (aldolase a, Aldoa; Vegfb; glucose transporter genes Slc2a3 and Slc2a1; Figure 6B) are known to be regulated by hypoxia24. Therefore, we next asked whether EryP development is sensitive to changes in oxygen tension and, in particular, whether hypoxia can regulate progenitor activity and gene expression.

EryP progenitors arise in the hypoxic environment of the pre-circulation YS, where oxygen availability is limited by diffusion. To determine whether the maintenance and/or expansion of EryP progenitors is regulated by oxygen levels, GFP(+) cells were sorted from E8.5 transgenic embryos, plated in methylcellulose and cultured under atmospheric (~21%) or low oxygen (2-5%) conditions. EryP progenitor numbers were increased by 2.5-fold in low oxygen (Figure 6C). The EryP colonies that formed in low oxygen were significantly larger than those formed at atmospheric oxygen (Figure 6D), suggesting that hypoxia is a growth signal for EryP progenitors. The increase in colony size was not due to cellular hypertrophy, as cells from colonies cultured under either low or atmospheric oxygen were similar in size (Figure 6E).

The glucose metabolism genes Ldha1, Pkm2, Pgk1, and Hk2 and the endothelial growth factor gene Vegfb are expressed in EryP (Figure 6A, B) and are known to be responsive to oxygen
tension. A qRT-PCR analysis of cells from EryP colonies revealed that expression of these genes was higher in cells cultured under low than under atmospheric oxygen (Figure 6F). Thus, oxygen levels regulate gene expression in maturing EryP and/or EryP progenitors.
Discussion

Emergence of primitive hematopoietic progenitors in the late gastrulation stage embryo

During gastrulation, the embryo develops three germ layers, the descendant lineages of which must then be established. We have prospectively identified and isolated progenitors and maturing progeny for the earliest embryonic cell type (EryP) to be specified from mesoderm at the end of mammalian gastrulation and have systematically profiled their genome-wide transcriptomes throughout their development. In this report, we focus primarily on the progenitor and early circulation stages of the EryP lineage. This resource adds to the existing whole-embryo transcriptomes by providing information for the single most abundant lineage at several overlapping stages (E7.5-E9.5) of development.

A transcriptional roadmap of EryP throughout their development

EryP arise as a cohort that matures in a stepwise, essentially synchronous manner, a feature that, combined with the \(\varepsilon\)-\textit{globin::H2B-EGFP} transgenic mouse line, facilitated the isolation of these cells at distinct stages of their development and allowed us to generate a microarray database that provides a timeline of gene expression over 6 days of embryogenesis. In contrast with EryP, production of EryD within the fetal liver or bone marrow occurs continuously, so that isolation of these cells at discrete stages is not straightforward. To date, the transcriptional profiles reported for erythroid cells represented only a single stage or heterogeneous populations or were derived from cells differentiated in vitro.

Transcription in maturing EryP is characterized by two discrete waves that correlate with key developmental hallmarks. The first wave (E8.5~E9.5) coincides with the transition from YS erythropoiesis to the entry of EryP into the bloodstream, with concomitant loss of progenitor
activity. Once EryP are in circulation (E9.5 to E11.5), gene expression changes are more limited, until the second wave of transcriptional variation (E11.5~E12.5) that corresponds to a period of extensive morphological change, decreased cell division rate, cytoskeletal remodeling, and nuclear condensation and extrusion. These changes are reflected in the Gene Ontology functions that are enriched in developing EryP, including nuclear organization, DNA packaging and chromatin assembly.

As would be predicted from our previous validation of the \(\varepsilon\)-globin::histone H2B-GFP transgenic mouse line as a model for primitive erythropoiesis\(^{11,12}\), a large number of erythroid genes is expressed in the GFP(+) (but not the GFP-negative) population at high levels from E7.5 on. Many of the genes identified in this study are known or likely to serve important functions in the development of the definitive erythroid lineage. For example, genes encoding the well studied transcription factors Gata1, Gata2, Runx1, Eklf/Klf1, Lmo2, Ldb1, and Stat5 are expressed in both lineages. Less well known transcripts shared by the two lineages include the tetraspanin gene Penumbra\(^{31}\), Homeodomain interacting protein kinase 2 (Hipk2)\(^{32}\), a putative mitochondrial transporter gene Slc25a39 involved in the heme biosynthesis pathway\(^{33}\), and the antioxidant peroxiredoxin gene Prdx2\(^{34}\).

The EryP transcriptome includes some interesting surprises. For example, \(Ddah1\) (dimethylarginine dimethylaminohydrolase1) encodes an enzyme involved in metabolism of methylarginines, molecules that inhibit nitric oxide synthase (NOS)\(^{35}\) and regulate VEGF-mediated angiogenesis\(^{36}\). Other examples include the \textit{fragilis} gene family members Ifitm2 (\textit{fragilis}3) and Ifim3 (\textit{fragilis}1), which regulate migration of primordial germ cells\(^{37}\); \(Sox5\), which encodes an HMG domain transcription factor related to Sry\(^{38}\); and \textit{Muscleblind-1}, a
regulator of alternative splicing in muscle development\textsuperscript{39}. Future studies should help to clarify the functions of these genes in the EryP lineage.

**Regulation of hematopoietic progenitor activity**

An interesting feature of primitive erythropoiesis is the transient appearance of progenitors in the YS and abrupt loss of these cells around the onset of circulation. The molecular events underlying this transition include loss of membrane receptors for growth factors, orchestrated downregulation of a cohort of cell adhesion molecules, and massive alterations in metabolism. The growth factor receptors TGFβR1, c-kit and Tie-2 are expressed on YS EryP from E7.5 to E8.5 and are rapidly downregulated by E9.5. Genes encoding their ligands for (TGFβ1, SCF and Ang1) are expressed by one or more cell types of the YS, suggesting that local sources of these proteins are available in this microenvironment. While TGFβ1 has been reported to inhibit the growth of erythroid progenitors from adult bone marrow\textsuperscript{40}, its activity has not been evaluated for EryP. We observed a dose-dependent response of EryP progenitors to TGFβ1, with suppression of growth at a low concentration and enhancement at high concentrations, raising the possibility that TGFβ1 functions as a morphogen for these cells.

Wnt/β-catenin signaling is required for formation of the primitive streak\textsuperscript{41} and has been shown to regulate hematopoietic specification of mesoderm in differentiating embryonic stem (ES) cells in vitro\textsuperscript{42,43}. However, whether EryP cell fate is regulated by Wnt signaling in the embryo, and whether that pathway functions in EryP progenitors themselves or in a mesodermal precursor was unknown. A number of Wnt/β-catenin pathway genes are expressed in EryP from E7.5 embryos and are rapidly downregulated as the cells mature. In this report, we provide evidence
that the Wnt pathway is active and functions autonomously in EryP progenitors in the blood islands of the YS.

A variety of adhesion proteins are expressed on EryP progenitors but not on circulating EryP. These observations suggest that, within the YS, EryP may form tight associations with each other and/or with surrounding endothelial cells, as suggested by electron microscopy. Loss of adhesion proteins from the surface of EryP in the YS might facilitate their entry into the bloodstream. In the zebrafish, release of EryP into the circulation requires the activity of a metalloprotease. Interestingly, two metalloprotease genes, *Adam17* and *mmp2*, are expressed in yolk sac stage EryP (Figure S3B).

**EryP progenitors, hypoxia, and the Warburg effect**

During their maturation in successive microenvironments within the embryo, EryP must adapt to changes in oxygen and nutrient supplies. We found that the number of EryP progenitors was enhanced by low oxygen concentrations in culture. It is intriguing that the transcriptome of EryP progenitors revealed a signature generally associated with cancer and other rapidly proliferating cells that catabolize glucose aerobically and produce high levels of lactate in the cytosol, rather than using the more energy efficient mitochondrial pathway—the “Warburg effect”. EryP progenitors express high levels of *Pkm2* and *Hk2*, genes that play critical roles in the Warburg effect. The reciprocal expression of the pyruvate kinase genes *Pkm2* and *Pklr* in progenitors and maturing EryP, respectively, indicates that isoform switching is a feature of EryP differentiation and presumably reflects progressive adaptation of the rapidly dividing progenitors to new metabolic requirements. There may be a functional association between the development of the placenta and onset of circulation and the abrupt loss of progenitor activity as the cells enter
the bloodstream. Whether the aerobic glycolytic profile of EryP progenitors simply reflects the particular energy demands of these rapidly dividing cells or a more unique feature of primitive erythropoiesis is not yet clear.

The rapidly dividing EryP progenitor cell population effectively constitutes a transient amplifying pool from which large numbers of primitive erythroblasts are generated. Using the \( \varepsilon \)-globin::H2B-EGFP transgenic mouse system to measure the numbers of EryP at E7.5 and E8.5, we found that these cells represent a huge fraction of the embryo at these stages. Indeed, by E8.5, EryP comprise nearly half of all cells and are by far the most abundant cell type in the embryo. Thus, the embryo has set aside enormous resources specifically for the development of this single lineage.
Acknowledgements

We thank Dr. Saghi Ghaffari for providing access to the hypoxia chamber incubator and the Mount Sinai Flow Cytometry Shared Resource Facility for assistance with cell sorting. We thank Drs. Julie Baker, Jane Little, Philippe Soriano, and Daniel Weinstein for insightful comments on the manuscript. Transgenic mice were produced by the Mount Sinai Mouse Genetics Shared Research Facility (National Institutes of Health/National Cancer Institute Grant R24 CA88302). This work was supported by grants from the National Institutes of Health to M.H.B. (RO1 HL62248, DK52191, and EB02209), to P.G.G. (RO1 HL65448 and DK62039 and P30 DK072442) and to A.-K. H. (RO1 HD052115 and DK084391) and by grants to M.H.B. from the Roche Foundation for Anemia Research (grant 9699367999, cycle X) and the New York State Department of Health (NYSTEM grant N08G-024).

Author contributions:  J.I., Z.H., S.T.F. and M.H.B. designed the experiments, analyzed the data and prepared the figures; J.I., Z.H., and S.T.F., performed the experiments; B.M. performed the β-catenin immunostaining; S.N. performed the 3D time-lapse imaging; A. F.-V. generated and imaged the TCF/Lef::H2B-GFP strain; A.-K. H. provided the TCF/Lef::H2B-GFP mouse line and live imaging expertise; V.S., D.T. and P.G.G. helped design the microarray experiments and analyze and prepare data for publication; J.I., S.T.F. and M.H.B. wrote the paper.

Conflict of interest statement: The authors declare no competing financial interests.
REFERENCES


FIGURE LEGENDS

Figure 1. GFP expression from the ε-globin::H2B-EGFP transgene marks the primitive erythroid lineage in the blood islands of the yolk sac and can be used to identify and isolate EryP progenitors

(A) GFP expression in ε-globin::H2B-EGFP primitive streak to early somite stage transgenic embryos. To visualize the emergence and expansion of GFP(+)/EryP cells, time-lapse movies of cultured ε-globin::H2B-EGFP embryos were acquired. GFP(+) cells appear in a narrow band of 3-5 cell diameters in the proximal YS during mid-to-late gastrulation (MS/LS stage). Scale bar, 500 μm. MS: midstreak; LS: late streak; EB: early bud; LHF: late headfold. ss, somite stage.

(B) Selected snapshots from time-lapse movie (Movie S1) of an ε-globin::H2B-EGFP embryo cultured in vitro under physiological conditions from LS to ESom stage. Early bud stage embryos were imaged over a period of 14 hr, from a time prior to transgene induction (~EB, t=0) through early somite (ESom) stages (t=860 min) (Figure 1B and Movie S1). Esom, early somite. Confocal images were acquired as sequential optical x-y sections taken at 4 μm z intervals. Images were taken at 20 min intervals (total imaging time: 14.5 hr).

(C) Flow cytometric histogram profiles of dispersed ε-globin::H2B-EGFP transgenic embryos reveals a clearly identifiable GFP(+) population. (D) Cells from whole E7.5 or E8.5 embryos were FACS-sorted to GFP (+) and GFP-negative populations. Left panel, Giemsa stained cytospun cells from FACS sort. Scale bar, 20 μm. EryP progenitor numbers were measured using a clonogenic assay. Virtually all progenitor activity was recovered in the GFP(+) population. Characteristic EryP colonies (right panels) showed red pigmentation (hemoglobin) and GFP fluorescence. Scale bar, 50 μm. (E) Real-time RT-PCR expression of endogenous embryonic εy- and βh1-globin genes in GFP(+) and GFP-negative FACS-sorted cells from E7.5 ε-globin::H2B-EGFP
transgenic embryos. Expression was normalized to *ubiquitin b (Ubb)*. (F) EryP numbers at YS stages of development. ss, somite stage.

**Figure 2. Global gene expression profiling of the primitive erythroid lineage**

Labeled cRNA samples were hybridized to Illumina Mouse WG-6 v1.1 Expression Bead Chip genome-wide arrays. Quality control of array data was performed using the Bioconductor lumi R package. The filtered genes were clustered into seven major patterns using the maSigPro algorithm. We were able to survey a genome wide probe set representing 46,630 murine transcripts, encompassing the emergence of EryP progenitors in the yolk sac through successive stages of erythroblast differentiation in the circulation. Many of these probes target less well annotated transcripts or transcript isoforms of known genes. There are 21,174 unique genes in the University of California, Santa Cruz (UCSC) mouse mm9 refseq protein coding gene database, and the Illumina mouse-6 v1.1 microarray used in this study contains probes for 18,970 or 89.6% of the well-annotated refseq mouse genes. Analyses were performed using all probes with Entrez ID annotations found with the lumiMouseAll.db version 1.6.1 annotation package.

(A) Changes (increased or decreased) in transcript numbers during consecutive stages of EryP development. The graph represents the total numbers of transcripts showing a change of greater than 2-fold (p<0.01). Dotted red line: increasing expression; dotted gray line, decreasing expression. Peaks in transcription variation were identified during the windows from E8.5-E9.5 (transition from YS to circulation stage, 273 transcripts) and E11.5 to E12.5 (fetal liver stage, 351 transcripts). (B) Plot representations of seven specific clusters of transcripts with similar temporal expression patterns. Clusters were sub-classified into three groups, representing genes that are progressively upregulated (clusters 1, 2 and 3, red lines in Figure 2B); downregulated (clusters 4, 5 and 6, green lines in Figure 2B); or upregulated through E11.5 and then
downregulated rapidly over the next 24 hr of development (gray line, cluster 7). The peaks in transcription variation indicated in Figure 2A are especially evident in clusters 1, 4, and 7 (relatively sharp increases or decreases in expression, E8.5-E9.5, corresponding to the transition from YS to circulation stage) and in clusters 3, 6 and 7 (abrupt increases or decreases in expression, E11.5-E12.5, corresponding to fetal liver stage, when EryP complete their maturation and enucleate). Each individual point (open circles) represents the mean gene expression of the cluster genes from one microarray experiment. Each line connects mean values for all replicates. (C) Overrepresented gene ontologies for the clusters shown in (A). (D) Expression of a representative gene that is upregulated (Gata1) and one that is downregulated (Igf2) in the EryP microarray dataset, analyzed using qRT-PCR. Expression levels normalized relative to ubiquitin b (Ubb).

Figure 3. Molecular signatures of EryP progenitors

(A) Hemangioblastic marker expression by early EryP. FACS plots showing expression of Flk-1 and VE-Cadherin surface protein in GFP (+) cells from ε-globin::H2B-EGFP embryos at E7.5 and E8.5. (B) Selected genes that are activated during EryP progenitor expansion from E7.5 to E8.5. Transcripts listed showed a low adjusted p value and median expression level <9.5 (log2 scale) for triplicates at E7.5 and were upregulated ≥3.5-fold by E8.5. Samples were amplified prior to hybridization to the microarray. Expression cutoff, 7.2. (C) Expression of activated β-catenin (*β-cat) in EryP at E8.0. Cells from dispersed E8.0 ε-globin::H2B-EGFP embryos were cytocentrifuged onto slides and then immunostained. (D, E) Expression of a TCF/Lef::H2B-GFP transgenic reporter for the canonical Wnt signaling pathway 13 in a ~E7.75 embryo. (D) Anterior view of late headfold (LHF) stage TCF/Lef::H2B-GFP embryo counterstained with Hoechst to highlight nuclei. The image is a 3D reconstruction of a z-stack and was acquired
using a Zeiss LSM 510 microscope outfitted with a plan-apochromat 20X/0.75 NA lens. White
box indicates the blood islands of the YS. Scale bar, 100 µm. (E) High magnification view of
boxed region in panel D. Scale bar, 50 µm.

Figure 4. Changes in gene and protein expression during the transition of EryP from the
yolk sac to the embryonic circulation

(A) Gene identifiers of EryP progenitors. Transcripts expressed by E8.5 EryP are grouped into
functional categories. Fold change from E8.5 and E9.5 is shown. Positive value, upregulated;
negative value, downregulated. (B) FACS profiles of CD41 protein expression on ϵ-globin::H2B-EGFP embryos from E7.5-E9.5. Expression increases from E7.5 to E8.5 and then
declines by E9.5. Expression is undetectable at later stages (not shown). (C) FACS histograms
showing downregulation of adhesion molecules on ϵ-globin::H2B-EGFP EryP during the
transition from the yolk sac stage to the circulation.

Figure 5. Growth factor and cytokine pathways in primitive erythroid progenitors

(A) Changes in expression of genes encoding growth factor or cytokine receptors and
downstream signaling components in E7.5 and E8.5 EryP. Shown in this table are median
expression levels (log2) from the microarray and linear fold change in expression. (B) Effect of
TGF-β1 on formation of EryP progenitors. EryP were FACS sorted from whole E8.5 embryos
and plated in the methylcellulose colony assays in the presence of the indicated concentrations of
TGF-β1. Data represent the average of triplicate samples from 4 experiments; error bars
represent SEM. (C) Expression of c-kit and Tie-2 protein on EryP at E7.5, E8.5 and E9.5. (D)
c-kit marks EryP progenitors within the GFP(+) cell population from E8.5 embryos. Cells were
FACS sorted and plated in triplicate in methylcellulose progenitor assays. Colonies were scored
at day 5. One representative experiment out of three is shown; error bars represent SEM. (E) Tie-2 marks EryP progenitors within the GFP(+) cell population from E8.5 embryos. Cells were FACS sorted and plated in triplicate in methylcellulose progenitor assays. Colonies were scored at day 5. One representative experiment out of three is shown; error bars represent SEM. (F) Real-time RT-PCR analysis of mRNA expression of Tgf-β1, Ang-1, Scf and Epo in FACS-sorted EryP, VE, and endothelial cells from YS or in cells from whole embryos at E8.5. Expression levels are shown relative to Ubb.

**Figure 6. Hypoxia regulates EryP progenitor activity**

(A) Expression of genes involved in glucose metabolism during EryP maturation. Relative mRNA levels from the microarrays, expressed on a log2 scale. Hk2, hexokinase 2; Pkm2 pyruvate kinase M2; Ldha, lactate dehydrogenase alpha; Pgk1, phosphoglycerate kinase 1; Pklr, pyruvate kinase lr. Note the isoform switching from Pgk1 to Pklr. (B) Transcripts known to be induced by hypoxia are downregulated during EryP maturation. Absolute expression (log2 scale) and fold change in expression are shown for the period from E8.5 to E11.5. Expression cutoff, 6.0. (C) Hypoxia increases EryP progenitor numbers in culture. E8.5 EryP were FACS sorted, plated in methylcellulose, and incubated under atmospheric or low oxygen (5%) conditions. Total EryP colony numbers were scored at day 5. (D) Increase in EryP colony size under low oxygen conditions. Photographs of representative EryP colonies grown at atmospheric or low oxygen. Scale bar, 50 μm. The graph displays mean radius of EryP colonies (day 4) grown at atmospheric or low oxygen. The EryP colonies that formed in low oxygen were significantly larger than those formed at atmospheric oxygen (33 μm vs 23.3 μm mean colony radius, respectively). (E) Giemsa staining of cytocentrifuged cells from EryP colonies grown at atmospheric or low oxygen and harvested at day 4. Scale bar, 50 μm. (F) Expression of
hypoxia-regulated genes in EryP colonies grown at atmospheric or low oxygen. Colonies were harvested at day 4 and RNA prepared for qRT-PCR analysis. Expression levels are shown relative to Ubb. Cells grown under hypoxic conditions maintain higher level expression of these genes than do cells grown at atmospheric oxygen. These genes are normally downregulated as EryP progenitors mature (panel A).
Figure 3

(A) Flk-1 and VE-Cadherin expression in H2BGFP at E7.5 and E8.5.

(B) Genes induced strongly during progenitor expansion.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Fold change E7.5 to E8.5</th>
<th>Fold change E7.5</th>
<th>Fold change E8.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slc4a1</td>
<td>Solute carrier family 4, member 1 (band 3)</td>
<td>35.8</td>
<td>7.6</td>
<td>12.8</td>
</tr>
<tr>
<td>Hba-x</td>
<td>Hemoglobin X, alpha-like embryonic chain</td>
<td>34.5</td>
<td>8.9</td>
<td>14.0</td>
</tr>
<tr>
<td>Hbb-y</td>
<td>Hemoglobin Y, beta-like embryonic chain (epsilon Y)</td>
<td>19.8</td>
<td>7.5</td>
<td>11.8</td>
</tr>
<tr>
<td>Hba-a1</td>
<td>Hemoglobin alpha, adult chain 1</td>
<td>18.3</td>
<td>8.9</td>
<td>13.1</td>
</tr>
<tr>
<td>Eraf</td>
<td>Erythroid associated factor</td>
<td>15.0</td>
<td>8.5</td>
<td>12.4</td>
</tr>
<tr>
<td>Hbb-b1</td>
<td>Hemoglobin, beta adult major chain</td>
<td>7.6</td>
<td>9.1</td>
<td>12.0</td>
</tr>
<tr>
<td>Mgst3</td>
<td>Microsomal glutathione S-transferase 3</td>
<td>7.4</td>
<td>7.8</td>
<td>11.2</td>
</tr>
<tr>
<td>Ipas</td>
<td>Hypoxia inducible factor 3, alpha subunit (Hif3a)</td>
<td>7.2</td>
<td>7.2</td>
<td>9.3</td>
</tr>
<tr>
<td>Alas2</td>
<td>Aminolevulinic acid synthase 2, erythroid</td>
<td>6.9</td>
<td>9.1</td>
<td>11.8</td>
</tr>
<tr>
<td>Gypa</td>
<td>Glycophorin A</td>
<td>6.9</td>
<td>8.2</td>
<td>10.9</td>
</tr>
<tr>
<td>Gtbp2</td>
<td>GTP binding protein 2</td>
<td>4.1</td>
<td>8.3</td>
<td>10.3</td>
</tr>
<tr>
<td>Trib3</td>
<td>Tribbles homolog 3</td>
<td>3.9</td>
<td>7.7</td>
<td>9.5</td>
</tr>
<tr>
<td>Tutt1</td>
<td>Tuttelin 1</td>
<td>3.6</td>
<td>8.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Kel</td>
<td>Kell blood group</td>
<td>3.5</td>
<td>7.3</td>
<td>9.1</td>
</tr>
</tbody>
</table>

(C) GFP expression and β-catenin localization.

(D) Transcription factor activity in E7.75 (LHF).

(E) TCF/Lef::H2B-GFP expression in Hoachst GFP.
Figure 4

A

Yolk sac to circulation transition

Transcriptional regulators

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Fold change E8.5 to E9.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gata1</td>
<td>GATA binding protein 1</td>
<td>1.2</td>
</tr>
<tr>
<td>Lmo2</td>
<td>LIM domain only 2</td>
<td>-2.4</td>
</tr>
<tr>
<td>Hmga1</td>
<td>High mobility group AT-hook 1</td>
<td>-4.1</td>
</tr>
<tr>
<td>Cited2</td>
<td>Cbp/p300-interacting transactivator, 2</td>
<td>1.3</td>
</tr>
<tr>
<td>Gata2a</td>
<td>Transcriptional repressor p66 alpha</td>
<td>-1.7</td>
</tr>
</tbody>
</table>

Signaling related molecules

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Fold change E8.5 to E9.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Igf2</td>
<td>Insulin-like growth factor II</td>
<td>-3.2</td>
</tr>
<tr>
<td>Igf2bp2</td>
<td>Insulin-like growth factor II mRNA binding protein 2</td>
<td>-3.7</td>
</tr>
<tr>
<td>Igf2bp3</td>
<td>Insulin-like growth factor II mRNA binding protein 3</td>
<td>-1.3</td>
</tr>
<tr>
<td>Mif</td>
<td>Macrophage migration inhibitory factor</td>
<td>-2.8</td>
</tr>
</tbody>
</table>

Surface molecules / Adhesion

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Fold change E8.5 to E9.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pecam1</td>
<td>Platelet/endothelial cell adhesion molecule 1 (CD31)</td>
<td>-1.8</td>
</tr>
<tr>
<td>Itga2b</td>
<td>Integrin alpha 2b (CD41)</td>
<td>-1.3</td>
</tr>
<tr>
<td>Itga6</td>
<td>Integrin alpha 6</td>
<td>-1.1</td>
</tr>
<tr>
<td>Itgb1</td>
<td>Integrin beta 1</td>
<td>-1.4</td>
</tr>
</tbody>
</table>

Glucose-related cell metabolism

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Fold change E8.5 to E9.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pkm2</td>
<td>Pyruvate kinase</td>
<td>-4.4</td>
</tr>
<tr>
<td>Enol1</td>
<td>Enolase alpha</td>
<td>-2.5</td>
</tr>
</tbody>
</table>

B

CD41

E7.5 | E8.5 | E9.5

C

% of expression in GFP(+) population

<table>
<thead>
<tr>
<th></th>
<th>E7.5</th>
<th>E8.5</th>
<th>E9.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>β1-integrin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α5-integrin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β3-integrin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α4-integrin</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5

A. EryP genes encoding receptors and downstream signaling components

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Expression levels</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>c-kit</td>
<td>Kit oncogene</td>
<td>E7.5: 8.0</td>
<td>E8.5: 7.7</td>
</tr>
<tr>
<td>Tek</td>
<td>Endothelial-specific receptor tyrosine kinase (Tie-2)</td>
<td>E7.5: 7.8</td>
<td>E8.5: 7.3</td>
</tr>
<tr>
<td>Tgfr1</td>
<td>Transforming growth factor, beta receptor I</td>
<td>E7.5: 9.0</td>
<td>E8.5: 8.3</td>
</tr>
<tr>
<td>Tgfr3</td>
<td>Transforming growth factor, beta receptor III</td>
<td>E7.5: 9.6</td>
<td>E8.5: 9.2</td>
</tr>
<tr>
<td>Acvr2b</td>
<td>Activin receptor II B</td>
<td>E7.5: 9.7</td>
<td>E8.5: 8.4</td>
</tr>
<tr>
<td>Smad3</td>
<td>MAD homolog 3 (Drosophila)</td>
<td>E7.5: 9.6</td>
<td>E8.5: 8.9</td>
</tr>
<tr>
<td>Smad4</td>
<td>MAD homolog 4 (Drosophila)</td>
<td>E7.5: 8.9</td>
<td>E8.5: 8.4</td>
</tr>
<tr>
<td>EpoR</td>
<td>Erythropoietin receptor</td>
<td>E7.5: 8.4</td>
<td>E8.5: 8.4</td>
</tr>
<tr>
<td>Vegfa</td>
<td>Vascular endothelial growth factor A</td>
<td>E7.5: 8.1</td>
<td>E8.5: 8.4</td>
</tr>
<tr>
<td>Vegfb</td>
<td>Vascular endothelial growth factor B</td>
<td>E7.5: 10.1</td>
<td>E8.5: 9.1</td>
</tr>
</tbody>
</table>

B. Relative colony number

C. c-kit and Tie-2 expression levels at E7.5, E8.5, and E9.5.

D. EryP colonies / 10,000 cells

E. Relative EryP colony number

F. Relative mRNA expression of Tgf-β1, Scf, and Ang1.
Single lineage transcriptome analysis reveals key regulatory pathways in primitive erythroid progenitors in the mouse embryo

Joan Isern, Zhiyong He, Stuart T. Fraser, Sonja Nowotschin, Anna Ferrer-Vaquer, Rebecca Moore, Anna-Katerina Hadjantonakis, Vincent Schulz, David Tuck, Patrick G. Gallagher and Margaret H. Baron