Supplementary information guide

Supplementary Tables:

Supplementary Table 1. **Inventory of tissues.** a) All tissues processed within the manuscript are listed with key biological and technical attributes. b) All VISIUM sections processed within the manuscript are listed with key technical attributes. c) Representative flow cytometry plots used for the sorting of CD31+ and/or CD34+ enriched cells in the AGM from the CS14 embryo. d) Representative flow cytometry plots used for the sorting of CD31+ and/or CD34+ enriched cells from the 8 weeks liver. e) Representative flow cytometry analysis of cells from human iPSC differentiated to hematovascular cells.

Supplementary Table 2. Genes enriched in nascent HSC in human AGM. a) Genes enriched in cluster 12 (HSC cluster) from all AGM cells (Fig. S1A). scRNAseq analysis of three AGM tissues (wk4.5-CS14, wk5-CS15a, wk5-CS15b) identifies one cluster (cluster 12) with a gene expression profile indicative of HSC. Pct1 is cluster 12 and pct2 all other clusters. The analysis was conducted with the FindAllMarkers Seurat function using default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25, p-value < 0.01, min.pct = 0.1). b) Differential gene expression analysis between HSC and other hematopoietic cells (ED Fig.1e). Genes enriched in HSC are shown. Pct1 is the HSC cluster and pct2 all other clusters. The analysis was conducted with the FindMarkers Seurat function using min.pct = 0.2, min.diff.pct = 0.2 and default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25, p-value < 0.01). c) Differential gene expression analysis between HSC and other hematopoietic cells (ED Fig.1e). Gene ontology analysis of genes enriched in HSC in (b), using Enrichr and selecting GO biological processes and human KEGG pathways, is shown. Fisher exact test, P-value < 0.01. d) Differential gene expression analysis between HSC and other hematopoietic cells (ED Fig.1e). Genes enriched in other hematopoietic cells are

shown. Pct1 are all other clusters, pct2 is the HSC cluster. The analysis was conducted with the FindMarkers Seurat function using min.pct = 0.2, min.diff.pct = 0.2 and default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25, p-value < 0.01). e) Gene ontology analysis of genes enriched in other hematopoietic cells in (d), using Enrichr and selecting GO biological processes and human KEGG pathways, is shown. Adjusted P-value < 0.05. f) HSC genes from the intersection of HSC-enriched genes compared to all other AGM clusters (in a) and to other hematopoietic cells (in b). Genes were selected for being expressed in less than 25% (pct <0.25) of the cells in the non-HSC group in each comparison.

Supplementary Table 3. Differential gene expression between SPINK2 cells from human CS14 AGM and Liver. a) Genes enriched in SPINK2+ cells from the CS14 AGM HSC cluster (ED Fig.2d-f). Direct comparison between SPINK2+ HSPC (selected for having >1 read of SPINK2) between wk4.5/CS14 AGM and liver from the same embryo. The analysis was conducted with the FindMarkers Seurat function using min.pct = 0.2, min.diff.pct = 0.2 and default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25, p-value < 0.01). b) Gene ontology analysis of genes enriched in SPINK2+ HSPC from CS14 AGM using Enrichr and selecting GO biological processes and human KEGG pathways, is shown. Fisher exact test, P-value < 0.05. c) Genes enriched in SPINK2+ cells from CS14 Liver (ED Fig.2 d-f). Direct comparison between SPINK2+ HSPC (selected for having >1 read of SPINK2) between wk4.5/CS14 AGM and liver from the same embryo. The analysis was conducted with the FindMarkers Seurat function using min.pct = 0.2, min.diff.pct = 0.2 and default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25, p-value < 0.01). d) Gene ontology analysis of genes enriched in SPINK2+ HPC from CS14 Liver using Enrichr and selecting GO biological processes and human KEGG pathways, is shown. Fisher exact test, P-value < 0.05.

Supplementary Table 4. Gene expression changes during functional maturation of human HSC. a) Pseudotime analysis (Monocle) of HLF+ HSC from AGM and liver tissues that contain HSC during first and second trimester. Genes positively correlated (correlation coefficient >0.4) with pseudotime (upregulated during HSC maturation). Monocle statistic parameters using the default likelihood ratio test are reported. Dot plot for each gene is shown. b) Gene ontology analysis of genes positively correlated (correlation coefficient >0.4) with pseudotime (upregulated during HSC maturation, in (a)), using Enrichr and selecting human KEGG pathways, is shown. Fisher exact test, P-value < 0.05. c) Pseudotime analysis (Monocle) of HLF+ HSC from tissues containing HSC during first and second trimester. Genes negatively correlated (correlation coefficient <-0.4) with pseudotime (downregulated during HSC maturation). Monocle statistic parameters using the default likelihood ratio test are reported. d) Gene ontology analysis of genes negatively correlated (correlation coefficient <-0.4) with pseudotime (downregulated during HSC maturation). Fisher exact test, P-value < 0.05. e) Flow cytometry quantification of HSC maturation markers HLA-DR and CD133 (PROM1) in fetal liver and cord blood HSPC (CD43+CD45midCD34+CD38low/-CD90+ (top) or in fetal liver CD43+CD45midCD34+CD38low/-CD90+GPI-80+ HSC subset (bottom, Source data for Fig.2g.)) at different stages is shown.

Supplementary Table 5. **Differential gene expression analysis between cells at distinct stages of EHT.** a) Raw data table displaying genes that are significantly differentially expressed between HE (CDH5+RUNX1+PTPRC-) and all other clusters (see Fig3c, clusters 0-8 except HE). Pct.1 is HE and pct.2 is all other clusters. The analysis was conducted with the FindMarkers Seurat function using default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25). min.diff.pct was calculated as the difference between pct.1 and pct.2. a.1) Filtered data table containing most significantly (p_val_adj<0.01) enriched genes in HE (avg_logFC>0) compared to all other clusters (see Fig3c, clusters 0-8 except HE). Dot plot

displaying average and percentage expression of each gene in distinct stages of EHT is shown. b) Raw data table showing differentially expressed genes between cluster 2 and cluster 3. Pct.1 is cl.2 and pct.2 is cl.3. The analysis was conducted with the FindMarkers Seurat function using default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25). min.diff.pct was calculated as the difference between pct.1 and pct.2. b.1-b.2) Filtered data tables containing most significantly (p_val_adj<0.01 and min.pct.diff>0.3) enriched genes in cluster 2 (avg_logFC>0) compared to cluster 3 (b.1), or most significantly (p_val_adj<0.01 and min.pct.diff<-0.3) enriched genes in cluster 3 (avg_logFC<0) compared to cluster 2 (b.2) . Dot plot displaying average and percentage expression of each gene in distinct stages of EHT is shown. c) Raw data table showing differentially expressed genes between HE and cluster 3. Pct.1 is HE and pct.2 is cl.3. The analysis was conducted with the FindMarkers Seurat function using default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25). min.diff.pct was calculated as the difference between pct.1 and pct.2. c.1-c.2) Filtered data tables containing most significantly (p_val_adj<0.01 and min.pct.diff>0.3) enriched genes in HE (avg_logFC>0) compared to cluster 3 (c.1), or most significantly (p_val_adj<0.01 and min.pct.diff<-0.3) enriched genes in cluster 3 (avg_logFC<0) compared to HE (c.2). Dot plot displaying average and percentage expression of each gene in distinct stages of EHT is shown. d) Raw data table showing differentially expressed genes between HE and cluster 4. Pct.1 is HE and pct.2 is cl.4. The analysis was conducted with the FindMarkers Seurat function using default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25). min.diff.pct was calculated as the difference between pct.1 and pct.2. d.1-d.2) Filtered data tables containing most significantly (p_val_adj<0.01 and min.pct.diff>0.3) enriched genes in HE (avg_logFC>0) compared to cluster 4 (d.1), or most significantly (p_val_adj<0.01 and min.pct.diff<-0.3) enriched genes in cluster 4 (avg_logFC<0) compared to HE (d.2). Dot plot

displaying average and percentage expression of each gene in distinct stages of EHT is shown.

Supplementary Table 6. Gene expression programs that evolve during EHT. a) Pseudotime analysis (Monocle) of EHT in the CS14-15 AGM based on endothelial and HSC clusters (clusters 0-5 in UMAP, see Fig.3a). plot_pseudotime_heatmap was used to generate the pseudotime heatmap (see Extended Data Fig.4g). The number of gene sets was established using the "num_clusters" argument within the monocle function plot_pseudotime_heatmap. After testing different number of clusters, 10 clusters were chosen as it allowed optimally separate the different modules of genes that co-vary across pseudotime (gene groups i-x). b-k) Gene ontology analysis each gene group (i-x) using Enrichr and selecting human KEGG pathways. Fisher exact test.

Supplementary Table 7. Cluster analysis of spatial transcriptomics data. Each sheet displays genes enriched in each cluster from the indicated sections. Spatial transcriptomics (Visium) analysis of seven sections from 5 weeks/CS15$_d$ embryo. The analysis was conducted with the FindAllMarkers Seurat function using Wilcoxon Rank Sum test with the parameters logfc.threshold = 0.01, p-value < 0.01, min.pct = 0.1.

Supplementary Table 8. **Differential gene expression analysis between early and HSC-forming hemogenic waves.** a) Raw data table showing differentially expressed genes between early wave HPCs (CS10 embryo and CS11 YS) and HSCs (CS13-CS17 AGM). Pct.1 is early wave HPCs and pct.2 is HSCs. The analysis was conducted with the FindMarkers Seurat function using default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25). a.1-a.2) Filtered data tables containing most significantly (p_val_adj<0.05) enriched genes in early HPCs (both CS10 embryo HPC and CS11 YS HPC) (avg_logFC>0) compared to HSCs (a.1), or most significantly (p_val_adj<0.05) enriched genes in HSCs (CS13-17 AGM) (avg_logFC<0) compared to early HPC (both CS10 embryo HPC and CS11 yolk sac

HPC) (a.2). Dot plot displaying average and percentage expression of each gene in all selected populations. a.3-a.4) Gene ontology analysis of genes enriched in early HPC (both CS10 embryo HPC and CS11 yolk sac HPC) (a.3), or in HSCs (CS13-17 AGM) (a.4) using Enrichr and selecting human KEGG pathways. Fisher exact test. b) Raw data table showing differentially expressed genes between CS10 embryo HPCs and HSCs (CS13-CS17 AGM). Pct.1 is CS10 embryo HPCs and pct.2 is HSCs. The analysis was conducted with the FindMarkers Seurat function using default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25). b.1-b.2) Filtered data tables containing most significantly (p_val_adj<0.05) enriched genes in CS10 embryo HPCs (avg_logFC>0) compared to HSCs (CS13-CS17 AGM) (b.1), or genes enriched in HSCs (CS13-CS17 AGM) (avg_logFC<0) compared to CS10 embryo HPCs (no gene appears significant) (b.2). Dot plot displaying average and percentage expression of each gene in all selected populations. b.3) Gene ontology analysis of genes enriched in CS10 embryo HPCs using Enrichr and selecting human KEGG pathways. Fisher exact test. c) Raw data table showing differentially expressed genes between CS11 yolk sac HPCs vs HSCs (CS13-CS17 AGM). Pct.1 is CS11 yolk sac HPCs and pct.2 is HSCs. The analysis was conducted with the FindMarkers Seurat function using default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25). c.1-c.2) Filtered data tables containing most significantly (p_val_adj<0.05) enriched genes in CS11 yolk sac HPCs (avg_logFC>0) compared to HSCs (CS13-CS17 AGM) (c.1), or most significantly (p_val_adj<0.05) enriched genes in HSCs (CS13-CS17 AGM) (avg_logFC<0) compared to CS11 yolk sac HPCs (c.2). Dot plot displaying average and percentage expression of each gene in all selected populations. C.3-c.4) Gene ontology analysis of genes enriched in CS11 yolk sac HPCs (c.3), or in HSCs (CS13-CS17 AGM) (c.4) using Enrichr and selecting human KEGG pathways. Fisher exact test. d) Overlap between genes enriched in early HPCs (CS10 embryo HPCs and CS11 yolk sac HPCs) (overlap between b.1 and c.1 lists). Gene ontology analysis

of genes in “common” list, using Enrichr and selecting human KEGG pathways. Fisher exact test. e) Raw data table showing differentially expressed genes between CS10 embryo HE and HSC-forming HE (CS13-CS17 AGM HE). Pct.1 is CS10 embryo HE and pct.2 is CS13-CS17 AGM HE. The analysis was conducted with the FindMarkers Seurat function using default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25). e.1-e.2) Filtered data tables containing most significantly (p_val_adj<0.05) enriched genes in CS10 embryo HE (avg_logFC>0) compared to HSC-forming HE (CS13-CS17 AGM HE) (e.1), or most significantly (p_val_adj<0.05) enriched genes in HSC-forming HE (CS13-CS17 AGM HE) (avg_logFC<0) compared to CS10 embryo HE (e.2). Dot plot displaying average and percentage expression of each gene in all selected populations. e.3-e.4) Gene ontology analysis of genes enriched in CS10 embryo HE (e.3), or in HSC-forming HE (CS13-CS17 AGM HE) (e.4) using Enrichr and selecting human KEGG pathways. Fisher exact test. f) Raw data table showing differentially expressed genes between CS11 yolk sac HE and HSC-forming HE (CS13-CS17 AGM HE). Pct.1 is CS11 yolk sac HE and pct.2 is CS13-CS17 AGM HE. The analysis was conducted with the FindMarkers Seurat function using default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25). f.1-f.2) Filtered data tables containing most significantly (p_val_adj<0.05) enriched genes in CS11 yolk sac HE (avg_logFC>0) compared to HSC-forming HE (CS13-CS17 AGM HE) (f.1), or most significantly (p_val_adj<0.05) enriched genes in HSC-forming HE (CS13-CS17 AGM HE) (avg_logFC<0) compared to CS11 yolk sac HE (f.2). Dot plot displaying average and percentage expression of each gene in all selected populations. f.3-f.4) Gene ontology analysis of genes enriched in CS11 yolk sac HE (f.3), or in HSC-forming HE (CS13-CS17 AGM HE) (f.4) using Enrichr and selecting human KEGG pathways. Fisher exact test. g) Raw data table showing differentially expressed genes between CS11 embryo HE and HSC-forming HE (CS13-CS17 AGM HE). Pct.1 is CS11 embryo HE and pct.2 is CS13-CS17

AGM HE. The analysis was conducted with the FindMarkers Seurat function using default parameters (Wilcoxon Rank Sum, logfc.threshold = 0.25). g.1-g.2) Filtered data tables containing most significantly (p_val_adj<0.05) enriched genes in CS11 embryo HE (avg_logFC>0) compared to HSC-forming HE (CS13-CS17 AGM HE) (g.1), or most significantly (p_val_adj<0.05) enriched genes in HSC-forming HE (CS13-CS17 AGM HE) (avg_logFC<0) compared to CS11 embryo HE (g.2). Dot plot displaying average and percentage expression of each gene in all selected populations. g.3-g.4) Gene ontology analysis of genes enriched in CS11 embryo HE (g.3), or in HSC-forming HE (CS13-CS17 AGM HE) (g.4) using Enrichr and selecting human KEGG pathways. Fisher exact test. h) Overlap between genes enriched in early HE (CS10 embryo HE, CS11 yolk sac HE and CS11 embryo HE) (overlap between e.1, f.1 and g.1 lists). Gene ontology analysis of genes in "common" list, using Enrichr and selecting human KEGG pathways. Fisher exact test. i) Overlap between genes enriched in HSC-forming HE (CS13-CS17 AGM HE) (overlap between e.2, f.2 and g.2 lists). Gene ontology analysis of genes in "common" list, using Enrichr and selecting human KEGG pathways. Fisher exact test.