Supplemental Material
Tracing human stem cell lineage during development using DNA methylation

Authors: Lucas A. Salas\(^1\)\(^\dagger\), John K. Wiencke\(^2\)\(^\ast\)\(^\dagger\), Devin C. Koestler \(^3\), Ze Zhang\(^4\), Brock C. Christensen\(^1\)\(^,\)\(^5\)\(^\dagger\)\(^\dagger\), Karl T. Kelsey\(^4\)\(^\dagger\)\(^\dagger\).

Affiliations:
\(^1\)Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Lebanon NH.
\(^2\)Department of Neurological Surgery, Institute for Human Genetics, University of California San Francisco, San Francisco CA.
\(^3\)Department of Biostatistics, University of Kansas Medical Center, Kansas City KS.
\(^4\)Departments of Epidemiology and Pathology and Laboratory Medicine, Brown University, Providence RI.
\(^5\)Departments of Molecular and Systems Biology, and Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon NH.

Correspondence: John.Wiencke@ucsf.edu.
\(^\dagger\) Contributed equally
\(^\dagger\)\(^\dagger\) Contributed equally

Table of contents
Supplemental Fig. S1. Pipeline for discovery of Fetal Cell Origin (FCO) methylation signature ................................................................. 2
Supplemental Table S1. Data sources and citations ................................................................................................................................. 3
Supplemental Fig. S2. Selection of invariant loci for the fetal cell origin-FCO signature .................................................................................. 5
Supplemental Fig. S3. Synthetic Mixture experiment ............................................................................................................................. 6
Supplemental Fig. S4. Estimated Fetal Cell Origin (FCO) in embryonic stem cells (ESC) and induced pluripotent stem cells (iPSC) through different number of cell culture passages (cell subcultures) using loess smoothing ................................................. 7
Supplemental Table S2. Fetal Cell Origin (FCO) signature deconvolution in pluripotent, fetal progenitors and adult CD34\(^\ast\) stem/progenitor cells .......................................................................................................................... 8
Supplemental Table S3. MSigDB pathways test for enrichment with DMRs contained in lineage invariant developmentally sensitive loci (N= 1218) .................................................................................................................. 9
Supplemental Table S4. Functional annotation using ENCODE data of the loci included in the FCO methylation signature ............. 11
Supplemental Table S5. Transcription factors with DMRs contained in lineage invariant developmentally sensitive loci (N= 1218) . . 12
Supplemental Table S6. Progenitor Cell Biology Consortium (PCBC) pathways test for enrichment using ToppGene with DMRs contained in lineage invariant developmentally sensitive loci (N= 1218) ........................................................................ 13
Supplemental Table S7. Age specific estimated FCO methylation fractions in blood leukocytes from birth to old age ..................... 15
Supplemental Methods S1. Stability of the FCO calculations ...................................................................................................................... 16

Supplemental Methods S1 Figure 1. Absolute difference between FCO estimated with one of the CpG probe lost versus the full set of 27 CpGs ......................................................................................................................... 17

Supplemental Methods S1 Figure 2. Root Mean Square Error increase per CpG lost ............................................................................. 18

Supplemental Methods S2. Synthetic mixture statistical validation ........................................................................................................ 19

Supplemental Methods S3. Maternal contamination sensitivity analyses ........................................................................................................... 20

Supplemental Methods S3 Figure 1. Evaluation of potential maternal contamination in the discovery datasets ............................. 21

Supplemental Methods S3 Figure 2. Evaluation of potential maternal contamination in the validation datasets ............................ 22

Supplemental Methods S3 Figure 3. Evaluation of potential maternal contamination in the five independent datasets compared to the FCO estimation ................................................................................................. 23

Supplemental File S1. List of 1218 candidate loci detected and the selected candidates (see Excel file) ........................................ 24

Supplemental Material References ......................................................................................................................................................... 25
1. Discovery datasets
Cell-specific methylation data from Bcells, CD4T cells, CD8T cells, NK cells, Granulocytes, and Monocytes

| Umbilical Cord-Blood (UCB) | Adult Whole Blood (AWB) |
|---------------------------|-------------------------|
| (26 subjects, 151 samples) | (6 subjects, 36 samples) |
| Bakulski et al. (2016)    | Reinius et al. (2012)   |

2. Identify library of fetal cell origin markers
The following 3-step filtering process was employed

a) Across the six cell types, compare DNA methylation between UCB and AWB samples.

- 1,255 CpGs were identified as differentially methylated (Q<0.05) in all six cell types.

b) Filter to CpGs with consistent directional difference in methylation across all cell types where |Δβ| ≥ 0.1

- 1218 CpGs

c) Filter to CpGs with minimal residual cell-specific effects using principal components analysis

- 27 CpGs

3. Estimate proportion of cells exhibiting the FCO signature
Using the final library of 27 CpGs, M, our estimate of the fraction of cells carrying the FCO signature, w, for a given sample, Y, was based on the constrained projection quadratic programming (CP/QP) approach of Houseman et al. (2012). Specifically:

\[
\text{arg min}_w \| Y - wM^T \|^2
\]

4. Replication and Statistical Validation
Three orthogonal approaches were used to assess the reliability and validity of our FCO signature.

- Replication using DNA methylation data on purified leukocyte cell types
  - GSE68456 (12 newborns, 45 samples)
  - GSE30870 (1 newborn and 1 adult sample)
  - GSE59069 (100 subjects, 199 samples)

- Classification (AUROC) UCB and AWB DNA methylation profiles
  - GSE80310, GSE74738, GSE54399, GSE79056, GSE62924 (123 newborns)
  - GSE74738, GSE54399 (34 adult subjects)

- Synthetic cell mixtures with varying proportions of UCB and AWB DNA methylation profiles
  - GSE66459 (22 newborns)
  - GSE43976 (52 adult subjects)

Supplemental Fig. S1. Pipeline for discovery of Fetal Cell Origin (FCO) methylation signature
### Supplemental Table S1. Data sources and citations

#### Discovery and validation datasets

| Discovery datasets | Repository | Lymphocytes | Myeloid cells |
|--------------------|------------|-------------|---------------|
|                    |            | B cell CD19+ | CD4T CD4+ | CD8 T CD8+ | NK CD56+ | Granulocytes recovery | Mono CD4+ | Subjects | Datasets |
| Umbilical cord blood | FlowSorted.CordBlood.450K (Bakulski et al. 2016) | 15 | 15 | 14 | 14 | 12 | 15 | 7 | 8 | 15 | 39.9(1.0) weeks |
| Peripheral blood   | FlowSorted.CordBloodNorway.450K (Gervin et al. 2016) | 11 | 11 | 11 | 11 | 11 | 11 | 6 | 5 | 11 | 39.3(1.2) weeks |
| Replication datasets | GSE35069 (Reinus et al. 2012) | 6 | 6 | 6 | 6 | 6 | 6 | 0 | 6 | 6 | 38 (13.6) years |

| Umbilical cord blood | GSE68456 (de Goede et al. 2015) | 7 | 7 | 6 | 6 | 7 | 12 | 7 | 5 | 12 | Term newborns |
| Peripheral blood   | GSE30870 (Heyn et al. 2012) | 1 | 0 | 0 | 0 | NA | NA | 1 | Term newborn |
|                    | GSE59065 (Tserel et al. 2015) | 0 | 99 | 100 | 0 | 0 | 0 | 52 | 48 | 100 | 52.6(23.7) years |
|                    | GSE30870 (Heyn et al. 2012) | 0 | 1 | 0 | 0 | NA | NA | 1 | 103 years |

### AUROC datasets

| AUROC datasets | Repository | Whole blood | Females | Males | Total | Age mean(SD) |
|---------------|------------|-------------|---------|-------|-------|--------------|
| Umbilical cord blood | GSE80310 (Knight et al. 2016) | 24 | 13 | 11 | 24 | Term (38.1-42.9 weeks) newborns |
|                 | GSE74738 (Hanna et al. 2016) | 1 | 0 | 0 | 1 | Pooled sample (Unknown gestational age) |
|                 | GSE54399 (Montoya-Williams et al. 2017) | 24 | 10 | 14 | 24 | Term newborns, with unknown health conditions rural war area |
|                 | GSE79056 (Knight et al. 2016) | 36 | 19 | 17 | 36 | 14 preterm (24.1-34 weeks), 22 term (39-40.9 weeks) newborns |
|                 | GSE62924 (Rojas et al. 2015) | 38 | 22 | 16 | 38 | 39 (1.4) weeks |
| Peripheral blood | GSE74738 (Hanna et al. 2016) | 10 | 10 | 0 | 10 | 29.0 (9.7) years (healthy women) |
|                 | GSE54399 (Montoya-Williams et al. 2017) | 24 | 24 | 0 | 24 | 32.8 (7.4) years (unknown health conditions rural war area) |

### Synthetic mixtures datasets

| Synthetic mixtures datasets | Repository | Whole blood | Females | Males | Total | Age mean(SD) |
|-----------------------------|------------|-------------|---------|-------|-------|--------------|
| Umbilical cord blood | GSE66659 (Fernando et al. 2015) | 22 | 11 | 11 | 22 | 11 Term (38-41 weeks) and 11 preterm newborns (26-36 weeks) |
| Peripheral blood | GSE43976 (Marabita et al. 2013) | 52 | 52 | 0 | 52 | 42.2(8.4) years (healthy women) |

### Embryonic stem cells, induced Pluripotent stem cells and hematopoietic cell progenitors**

| Embryonic stem cells, induced Pluripotent stem cells and hematopoietic cell progenitors** | Repository | ESC | iPSC | CD34+ fetal | CD34+ Adult | MPP | L-MPP | CMP | GMP | MEP | Erythroid adult | PMC | PMN | Females | Males | Total | Age |
|-------------------------------------------------------------------------------------|------------|-----|------|-------------|-------------|-----|-------|-----|-----|-----|---------------|-----|-----|---------|-------|-------|-----|
| GSE31848 (Nazor et al. 2012)  | 19 | 29 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 42 | 12 | 54 | NA |
| GSE40799 (Weidner et al. 2013) | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | NA | NA | 3 | 12 |
| GSE56491 (Lassard et al. 2015) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | NA | NA | 12 | 12 |
| GSE56491 (Lassard et al. 2015) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | NA | NA | 12 | 12 |
| GSE50797 (Rönneblad et al. 2014) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | NA | NA | 12 | 12 |
| GSE63409 (Jung et al. 2015) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | NA | NA | 12 | 12 |

### Somatic tissues

| Somatic tissues | Repository | Adrenal | Brain | Heart | Liver | Lung | Muscle | Pancreas | Spleen | Stomach | Females | Males | Total | Age |
|-----------------|------------|---------|-------|-------|-------|------|--------|----------|--------|---------|---------|-------|-------|-------|-----|
| Fetal           | GSE61379 (Bonder et al. 2014) | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | NA | NA | 14 | 8-21 weeks |
|                 | GSE31848 (Nazor et al. 2012) | 3 | 4 | 4 | 5 | 0 | 0 | 3 | 5 | 4* | 2* | 6* | 14, 15, 18, and 20 weeks |
|                 | GSE56515 (Sliker et al. 2015) | 9 | 0 | 0 | 0 | 0 | 0 | 9 | 8 | 0 | NA | NA | 10* | 9,18 and 22 weeks |
|                 | GSE58885 (Spiers et al. 2015) | 0 | 179 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 79 | 100 | 179 | 3-26 weeks |
|                 | GSE61797 (Bonder et al. 2014) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 48 | 48 | 96 | 26.8 (10.5) years |
| Adult           | GSE31848 (Nazor et al. 2012) | 2 | 1 | 1 | 0 | 2 | 2 | 2 | 2 | 1 | 2* | 1* | 3* | 48.0 (8.5) years |
|                 | GSE48472 (Sliker et al. 2013) | 0 | 0 | 0 | 5 | 0 | 6 | 4 | 3 | 0 | NA | NA | 6* | 52.5 (7.5) years |
|                 | GSE41826 (Guinovart et al. 2013) | 0 | 29 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 14 | 29 | 33.3 (17.2) years |

### Aging datasets

| Aging datasets | Repository | Whole blood | Mononuclear cells | Females | Males | Total | Age |
|----------------|------------|-------------|-------------------|---------|-------|-------|-----|
| Permanent repository | 15 | 0 | 8 | 7 | 15 | 38.9 (1.3) weeks |
| FlowSorted.CordBlood.450K (Bakulski et al. 2016) | 11 | 0 | 6 | 5 | 11 | 39.3 (1.2) weeks |

### Umbilical cord blood

| Umbilical cord blood | Whole blood | Mononuclear cells | Females | Males | Total | Age |
|----------------------|-------------|-------------------|---------|-------|-------|-----|
| GSE30870 (Heyn et al. 2012) | 0 | 19 | NA | NA | 19 | 38.7 (1.9) weeks |
| GSE83334 (Urdingui et al. 2016) | 15 | 0 | 9 | 6 | 15 | 38.9 (1.4) weeks |
### Aging datasets

**Permanent repository**

| Dataset | Whole blood | Mononuclear cells | Females | Males | Total | Age |
|---------|-------------|--------------------|---------|-------|-------|-----|
| GSE62219 (Acevedo et al. 2015) | 60 | 0 | 60 | 0 | 60 | 2.3 (1.7) years |
| GSE36054 (Alisch et al. 2012) | 134 | 0 | 55 | 79 | 134 | 4.6 (4.1) years |
| GSE40279 (Hannum et al. 2013) | 656 | 0 | 338 | 318 | 656 | 64.0 (14.7) years |
| GSE35069 (Reinius et al. 2012) | 6 | 6 | 0 | 6* | 6* | 38 (13.6) years |
| GSE30870 (Heyn et al. 2012) | 0 | 19 | NA | NA | 19 | 92.6 (3.7) years |
| GSE35069 (Urdinguio et al. 2016) | 97 | 0 | 49 | 48 | 97 | 52.7 (23.7) years |

*Several samples were drawn from the same subject

**Peripheral blood**

| Dataset | Whole blood | Mononuclear cells | Females | Males | Total | Age |
|---------|-------------|--------------------|---------|-------|-------|-----|
| GSE83334 | 15 | 0 | 9 | 6 | 15 | 5 years |

**ESC:** undifferentiated embryonic stem cells, iPSC: undifferentiated induced pluripotent stem cells, CD34+ fetal: stem/progenitor cells from fresh umbilical cord blood, erythroid fetal and adult: CD34+ cells from fetal liver and bone marrow respectively differentiated ex-vivo to erythroid cells (transferrin receptor-CD71+, and glycoporphin-CD235α+), CD34+ adult: CD34+CD38-CD90-CDRA, adult bone marrow progenitors samples: MPP-multipotent progenitors CD34+CD38-CD90-CD45RA+, L-MPP-lymphoid primed multipotent progenitors CD34+CD38-CD90-CD45RA+, CMP- common myeloid progenitors CD34+CD38-CD123+CD45RA+, GMP-granulocyte/macrophage progenitors CD34+CD38-CD123+CD45RA+, MEP-megakaryocyte-erythroid progenitors CD34+CD38-CD123+CD45RA+, CD34+ myeloid progenitors: CMP- common myeloid progenitors CD34+CD38-CD123+CD110-CD45RA+, and GMP-granulocyte/macrophage progenitors CD34+CD38-CD123+CD110-CD45RA+, CD34 immature myeloid progenitors: PMC-promyelocyte/myelocyte CD34+CD117+CD33+CD13+CD11b+, PMN - metamyelocyte/band-myelocyte CD34+CD117+CD33+CD13+CD11b+. |
Supplemental Fig. S2. Selection of invariant loci for the fetal cell origin-FCO signature.
Panel A. Candidate loci (1,218 CpG) showed a high variability between umbilical cord blood and adult peripheral blood purified cells (principal component 1, x axis). Albeit small relative to the UCB/APB effect, there was a statistically significant cell type effect present among these 1,218 CpGs (principal components-PC 2 and 3, y axis upper panel and P heatmap in the lower panel in bold the significant variables). Panel B, the reduced library (27 CpGs), showed strong separation of UCB and APB samples (principal component 1, x axis), however the residual variability from cell type was attenuated (principal component 2, y axis upper panel, P heatmap lower panel). Abbreviations: mAge: DNA methylation age (Horvarth)
Supplemental Fig. S3. Synthetic Mixture experiment.
Panel A. When generating artificial synthetic mixtures, a high agreement was observed with a concordance correlation coefficient, CCC=0.97 ($P<0.05$). Panel B as we had samples from umbilical cord blood of preterms (<37 weeks of gestational age) and term newborns (≥37 weeks of gestation), we generated mixtures using these two different subgroups. The CCC for the mixtures using Preterm samples was slightly higher CCC=0.97 vs Term newborns CCC=0.96. Although there were differences with the largest proportions of cord blood mixtures, overall there were no statistically significant differences.
Supplemental Fig. S4. Estimated Fetal Cell Origin (FCO) in embryonic stem cells (ESC) and induced pluripotent stem cells (iPSC) through different number of cell culture passages (cell subcultures) using loess smoothing.

Notes: For the graph one observation with 105 passages was excluded. Number of passages ranging from 5 to 57 passages.
Supplemental Table S2. Fetal Cell Origin (FCO) signature deconvolution in pluripotent, fetal progenitors and adult CD34+ stem/progenitor cells.

| Fetal/embryonic | Cell Type       | n | mean (SD) |
|-----------------|-----------------|---|-----------|
| Fetal/embryonic | ESC             | 25 | 75.1 (9)  |
|                 | iPSC            | 29 | 81 (1.9)  |
|                 | CD34+ fetal     | 3  | 81.8 (2.3)|
|                 | Erythroid fetal | 12 | 63.6 (3.3)|
|                 | CD34+ adult     | 5  | 12.1 (6.7)|
| Adult           | MPP             | 5  | 2.6 (3.8) |
| progenitors     | L-MPP           | 5  | 4.3 (4.5) |
| (bone marrow)   | CMP             | 8  | 4.4 (3.7) |
|                 | GMP             | 8  | 4.8 (6.4) |
|                 | MEP             | 5  | 4.2 (4.5) |
|                 | Erythroid adult | 12 | 2.8 (3.8) |
|                 | PMC             | 3  | 2.7 (4.7) |
|                 | PMN             | 3  | 2.1 (3.7) |

Estimated mean (SD) FCO methylation fractions for embryonic/fetal cells are 75.9% (8.5) and 4.4% (5.1) for adult progenitors (bone marrow), \( P= 1.81 \times 10^{-86} \).

Abbreviations: Embryonic stem cells (ESC), Induced Pluripotent Stem cells (iPSC), CD34+ fetal (fresh cord blood cells expressing CD34+), Erythroid fetal (fetal liver CD34+ cells, differentiated \textit{ex vivo} to express transferrin receptor and glycophorin), CD34+ adult (bone marrow expressing CD34+ CD38- CD90+ CD45RA+), Multipotent progenitors (MPP), Lymphoid primed multipotent progenitors (L-MPP), Common myeloid progenitors (CMP), Granulocyte/macrophage progenitors (GMP), Megakaryocyte-erythroid progenitors (MEP), Erythroid adult (adult bone marrow CD34+ cells, differentiated \textit{ex vivo} to express transferrin receptor and glycophorin), Promyelocyte/myelocyte (PMC), metamyelocyte/band-myelocyte (PMN).
Supplemental Table S3. MSigDB pathways test for enrichment with DMRs contained in lineage invariant developmentally sensitive loci (N= 1218).

| ID     | MSigDB Pathways                          | Cell target of the pathway                     | K    | DM (cis) | ToppGene | GREAT | missMethyl |
|--------|----------------------------------------|-----------------------------------------------|------|----------|----------|-------|------------|
|        | Genes identified by ChIP on chip as targets of a Polycomb protein or Polycomb Repression Complex 2 (bound to protein and H3K27 tri-methylation (H3K27me3)) |                                |      |          |          |       |            |
| M9898  | BENPORATH_SUZ12_TARGETS                | Human embryonic stem cells                    | 1038 | 112      | 183      | 2.86×10^46 | 1.33×10^37 | 2.09×10^36 | 1.61×10^35 | <2.0×10^-6 | <2.0×10^-6 |
| M7617  | BENPORATH_EED_TARGETS                  | Human embryonic stem cells                    | 1062 | 105      | 184      | 6.79×10^36 | 1.58×10^32 | 2.06×10^32 | 1.80×10^34 | <2.0×10^-6 | <2.0×10^-6 |
| M8448  | BENPORATH_PRC2_TARGETS                 | Human embryonic stem cells                    | 652  | 83       | 138      | 3.49×10^36 | 1.08×10^32 | 2.59×10^36 | 4.19×10^46 | <2.0×10^-6 | <2.0×10^-6 |
|        | Genes with high-CpG-density promoters (HCP) bearing the H3K27 tri-methylation (H3K27me3) |                                |      |          |          |       |            |
| M10371 | BENPORATH_ES_WITH_H3K27ME3             | Human embryonic stem cells                    | 1118 | 122      | 210      | 1.48×10^46 | 1.38×10^42 | 2.18×10^36 | 7.51×10^47 | <2.0×10^-6 | <2.0×10^-6 |
| M1938  | MEISSNER_BRAIN_HCWP_WITH_H3K27ME3      | Brain                                          | 269  | 39       | 80       | 2.16×10^39 | 3.36×10^14 | 3.71×10^14 | 1.31×10^14 | 4.40×10^46 | 2.90×10^-12 | 2.74×10^-9 |
| M1967  | MIKKELSEN_IPS_WITH_HCWP_H3K27ME3       | MCV8.1 (induced pluripotent cells, iPS)       | 102  | 22       | 28       | 3.53×10^15 | 4.11×10^12 | 4.99×10^12 | 7.61×10^36 | 4.27×10^35 | 8.32×10^-10 | 6.55×10^-7 |
| M2009  | MIKKELSEN_NPC_HCWP_WITH_H3K27ME3       | Neural progenitor cells (NPC)                 | 341  | 39       | 78       | 8.50×10^16 | 1.13×10^11 | 2.38×10^15 | 2.12×10^15 | 1.02×10^16 | 1.97×10^-8 | 1.17×10^-5 |
| M1932  | MEISSNER_NPC_HCWP_WITH_H3K27ME3        | Neural precursor cells (NPC)                  | 79   | 12       | 22       | 4.13×10^7  | 1.60×10^4  | 3.50×10^14 | 8.53×10^14 | 2.61×10^12 | 3.07×10^-8 | 9.06×10^-3 |
| M1954  | MIKKELSEN_MCV6_HCWP_WITH_H3K27ME3      | MCV6 cells (embryonic fibroblasts trapped in a differentiated state) | 435  | 43       |          | 5.14×10^12 | 5.00×10^11 | N.S          |          |          | 1.96×10^-7 | 9.27×10^-5 |
| M2019  | MIKKELSEN_MEF_HCWP_WITH_H3K27ME3       | MEF cells (embryonic fibroblast)              | 590  | 48       |          | 6.86×10^16 | 6.66×10^9  | N.S          |          |          | 2×10^-6 | 8.47×10^-4 |
|        | Genes with high-CpG-density promoters (HCP) that have no H3K27 tri-methylation (H3K27me3) |                                |      |          |          |       |            |
| M1936  | MEISSNER_NPC_HCWP_WITH_H3_UNMETHYLATED HYLATED | Neural precursor cells (NPC)                  | 536  | 44       | 65       | 1.65×10^12 | 1.18×10^9  | 2.06×10^15 | 1.69×10^14 | 4.36×10^12 | 3.4×10^-8 | 1.79×10^-5 |
|        | Genes with high-CpG-density promoters (HCP) bearing histone H3 dimethylation at K4 (H3K4me2) and trimethylation at K27 (H3K27me3) |                                |      |          |          |       |            |
| M1941  | MEISSNER_BRAIN_HCWP_WITH_H3K4ME3_AND_H3K27ME3 | Brain                                          | 1069 | 83       |          | 5.42×10^18 | 5.26×10^17 | N.S          |          |          | 1.86×10^-8 | 1.17×10^-5 |
| M1949  | MEISSNER_NPC_HCWP_WITH_H3K4ME2_AND_H3K27ME3 | Neural precursor cells (NPC)                  | 349  | 34       |          | 3.85×10^9  | 3.74×10^8  | N.S          |          |          | 9.3×10^-6 | 3.38×10^-3 |
|        | Genes hypermethylated in tumor cells | HATADA_METHYLATED_IN_LUNG_CANCER_UP | Lung cancer cells | 390  | 32       |          | 4.05×10^6  | 3.93×10^5  | N.S          |          |          | 2.5×10^-5 | 7.97×10^-3 |
|        | Genes up-regulated in tumor cells | MARTENS_TRETINOIN_RESPONSE_UP | NB4 cells (acute promyelocytic leukemia, APL) | 857  | 50       |          | 1.17×10^5  | 1.14×10^4  | N.S          |          |          | 3.5×10^-6 | 1.36×10^-3 |
Note: the table summarizes only the significant pathways overlapping three different methods to test for enrichment: 1) ToppGene, hypergeometric distribution to test for enrichment, 2) GREAT, binomial test to test for enrichment cis-regulatory regions, and 3) missMethyl which allows adjusting for array bias. 
Abbreviations: ID (MSigDB internal identifier), K (number of genes contained in the gene set), DM (differentially methylated genes overlapping the CpG site), DM (cis) (cis-regulatory regions either overlapping the differentially methylated CpG site or 1 Mb around the site), P (unadjusted P-value), FDR (False discovery), FE (Fold enrichment), N.S (not significant association, FDR>0.05)
## Supplemental Table S4. Functional annotation using ENCODE data of the loci included in the FCO methylation signature

| Probe ID   | Human Embryonic Stem cell | Human umbilical vein endothelial cell | Transcription factor 1 | Transcription factor 2 |
|------------|----------------------------|--------------------------------------|------------------------|------------------------|
| cg10338787 | 3_Poised_Promoter         | 12_Repressed                         | EZH2                   | EZH2                   |
| cg22497969 | 13_Heterochromatin/low signal | 13_Heterochromatin/low signal       |                        |                        |
| cg11968804 | 3_Poised_Promoter         | 12_Repressed                         | Pol2                   | EZH2                   |
| cg10237252 | 6_Weak_Enhancer           | 12_Repressed                         |                        |                        |
| cg17310258 | 3_Poised_Promoter         | 12_Repressed                         |                        |                        |
| cg13485366 | 13_Heterochromatin/low signal | 13_Heterochromatin/low signal       |                        |                        |
| cg03455765 | 2_Weak_Promoter           | 12_Repressed                         | USF-1                  | Bach1                  |
| cg04193160 | 3_Poised_Promoter         | 12_Repressed                         | SIN3A                  |                        |
| cg27367526 | 2_Weak_Promoter           | 1_Active_Promoter                    |                        |                        |
| cg03384000 | 3_Poised_Promoter         | 1_Active_Promoter                    |                        |                        |
| cg15575683 | 3_Poised_Promoter         | 12_Repressed                         | YY1                    |                        |
| cg17471939 | 3_Poised_Promoter         | 13_Heterochromatin/low signal        |                        |                        |
| cg11199014 | 3_Poised_Promoter         | 3_Poised_Promoter                    | Pol2                   | RBBP5                  |
| cg13948430 | 3_Poised_Promoter         | 12_Repressed                         |                        |                        |
| cg01567783 | 3_Poised_Promoter         | 12_Repressed                         |                        |                        |
| cg01278041 | 2_Weak_Promoter           | 11_Weak_Transcribed                  | CHD1                   | TAF1                   |
| cg19005955 | 7_Weak_Enhancer           | 4_Strong_Enhancer                    |                        |                        |
| cg16154155 | 3_Poised_Promoter         | 12_Repressed                         | EZH2                   | EZH2                   |
| cg14652587 | 3_Poised_Promoter         | 12_Repressed                         |                        |                        |
| cg19659741 | 6_Weak_Enhancer           | 12_Repressed                         | SUZ12                  |                        |
| cg06705930 | 3_Poised_Promoter         | 12_Repressed                         |                        |                        |
| cg23009780 | 5_Strong_Enhancer         | 12_Repressed                         |                        |                        |
| cg22130008 | 3_Poised_Promoter         | 3_Poised_Promoter                    |                        |                        |
| cg05840541 | 13_Heterochromatin/low signal | 13_Heterochromatin/low signal       |                        |                        |
| cg06953130 | 2_Weak_Promoter           | 5_Strong_Enhancer                    |                        |                        |
| cg11194994 | 2_Weak_Promoter           | 4_Strong_Enhancer                    |                        |                        |
| cg14375747 | 6_Weak_Enhancer           | 12_Repressed                         | TBP                    |                        |
Supplemental Table S5. Transcription factors with DMRs contained in lineage invariant developmentally sensitive loci (N= 1218).

| Transcription factor | Name                                      |
|----------------------|-------------------------------------------|
| **Zinc-coordinating DNA-binding domains** |                                           |
| KLF9                 | Kruppel Like Factor 9                     |
| ZBTB46               | Zinc Finger BTB Domain Containing 46      |
| PRDM10               | PR/SET Domain 10                          |
| PRDM16               | PR/SET Domain 12                          |
| **Helix-turn-helix domains** |                                          |
| **Homeo domain factors** |                                           |
| HOXA2                | Homeobox A2                               |
| HOXB7                | Homeobox B7                               |
| HOXB-AS3             | HOXB Cluster Antisense RNA 3              |
| LBX2                 | Ladybird Homeobox 2                       |
| VAX2                 | Ventral Anterior Homeobox 2               |
| ALX4                 | ALX Homeobox 4                            |
| PITX3                | Paired Like Homeodomain 3                 |
| LHX6                 | LIM Homeobox 6                            |
| SIX2                 | SIX homeobox 2                            |
| POU2F1 (Oct-1)       | POU Class 2 Factor 1                      |
| POU3F1 (Oct-6)       | POU Class 3 Homeobox 1                    |
| **Paired box factors** |                                           |
| PAX6                 | Homeodomain Paired box 6                  |
| PAX8                 | Homeodomain Paired box 8                  |
| FOXE3                | Forkhead binding E3                       |
| FOXD2                | Forkhead binding D2                       |
| FOXI2                | Forkhead binding I2                       |
| FOXL2                | Forkhead binding L2                       |
| FOXL2NB              | FOXL2 Neighbor                            |
| **Tryptophan cluster factors** |                                      |
| ETV4                 | ETS variant 4                             |
| **ARID**             |                                           |
| ARID3A               | AT-Rich Interaction Domain 3A             |
| **Other all-α-helical DNA-binding domains** |                               |
| SOX18                | SRY-Box 18                                |
| **Immunoglobulin fold** |                                         |
| TBX1                 | T-Box 1                                   |
| TBX4                 | T-Box 4                                   |
| **β-Hairpin exposed by an α/β-scaffold** |                                      |
| NF-1X                | Nuclear Factor 1 X                        |
Supplemental Table S6. Progenitor Cell Biology Consortium (PCBC) pathways test for enrichment using ToppGene with DMRs contained in lineage invariant developmentally sensitive loci (N= 1218).

| PCBC Pathway |  | # Genes in Gene Set (K) | DM | P       | FDR      |
|---------------|---|------------------------|----|---------|----------|
| **Stem cells top expressed genes** |   |                        |    |         |          |
| Arv_EB-LF_2500_K2                  |   | 960                    | 59 | 3.21 x 10^{-10} | 1.04 x 10^{-8} |
| Arv_EB-LF_1000                     |   | 990                    | 58 | 2.73 x 10^{-9}  | 7.62 x 10^{-8} |
| Arv_EB-LF_1000_K4                  |   | 436                    | 33 | 2.67 x 10^{-8}  | 5.66 x 10^{-7} |
| Arv_EB-LF_500_K2                   |   | 256                    | 23 | 1.77 x 10^{-7}  | 3.11 x 10^{-6} |
| PCBC_SC_CD34+_1000                 |   | 987                    | 53 | 2.33 x 10^{-7}  | 3.77 x 10^{-6} |
| Arv_EB-LF_500                      |   | 499                    | 32 | 1.75 x 10^{-6}  | 2.45 x 10^{-5} |
| Arv_SC-LF_1000_K1                  |   | 679                    | 39 | 2.01 x 10^{-6}  | 2.74 x 10^{-5} |
| **Embryoid body vs Stem Cells**    |   |                        |    |         |          |
| PCBC_ratio_EB_vs_SC_1000            |   | 997                    | 86 | 8.85 x 10^{-24} | 5.43 x 10^{-21} |
| ratio_EB_vs_SC_2500_K3              |   | 1102                   | 79 | 4.62 x 10^{-17} | 9.46 x 10^{-15} |
| PCBC_ratio_EB_vs.SC_500             |   | 499                    | 47 | 1.01 x 10^{-14} | 1.03 x 10^{-12} |
| ratio_EB_vs_SC_1000_K5              |   | 418                    | 42 | 3.14 x 10^{-14} | 2.75 x 10^{-12} |
| ratio_EB_vs_SC_1000_K1              |   | 336                    | 29 | 1.09 x 10^{-8}  | 2.67 x 10^{-7}  |
| ratio_EB_vs_SC_500_K3               |   | 204                    | 22 | 1.26 x 10^{-8}  | 2.98 x 10^{-7}  |
| **Ectoderm vs Stem cell**          |   |                        |    |         |          |
| ratio_ECTO_vs.SC_2500_K3            |   | 854                    | 60 | 9.51 x 10^{-13} | 5.84 x 10^{-11} |
| ratio_ECTO_vs.SC_500_K1             |   | 283                    | 32 | 1.67 x 10^{-12} | 9.34 x 10^{-11} |
| ratio_ECTO_vs.SC_1000_K3            |   | 476                    | 42 | 2.47 x 10^{-12} | 1.26 x 10^{-10} |
| PCBC_ratio_ECTO_vs.SC_500           |   | 499                    | 42 | 1.14 x 10^{-11} | 5.01 x 10^{-10} |
| PCBC_ratio_ECTO_vs.SC_1000          |   | 994                    | 61 | 1.65 x 10^{-10} | 5.64 x 10^{-9}  |
| PCBC_ratio_ECTO_vs.SC_100           |   | 100                    | 14 | 2.32 x 10^{-7}  | 3.77 x 10^{-6}  |
| **Endoderm vs Stem cell**          |   |                        |    |         |          |
| PCBC_ratio_DE_vs.SC_500             |   | 499                    | 36 | 2.13 x 10^{-8}  | 4.66 x 10^{-7}  |
| ratio_DE_vs.SC_500_K5               |   | 300                    | 26 | 5.79 x 10^{-8}  | 1.15 x 10^{-6}  |
| ratio_DE_vs.SC_500_K1               |   | 377                    | 29 | 1.34 x 10^{-7}  | 2.50 x 10^{-6}  |
| ratio_DE_vs.SC_1000_K5              |   | 542                    | 36 | 1.68 x 10^{-7}  | 3.03 x 10^{-6}  |
| PCBC_ratio_DE_vs.SC_1000            |   | 998                    | 49 | 8.25 x 10^{-6}  | 1.01 x 10^{-4}  |
| ratio_DE_vs.SC_1000_K2              |   | 523                    | 31 | 1.24 x 10^{-5}  | 1.43 x 10^{-4}  |
| **Mesoderm vs Stem cell**          |   |                        |    |         |          |
| PCBC_ratio_MESO-5_vs.SC_500         |   | 499                    | 34 | 2.06 x 10^{-7}  | 3.51 x 10^{-6}  |
| PCBC_ratio_MESO-5_vs.SC_1000        |   | 994                    | 51 | 1.53 x 10^{-6}  | 2.24 x 10^{-5}  |
| ratio_MESO_vs.SC_500_K1             |   | 297                    | 22 | 8.01 x 10^{-6}  | 1.00 x 10^{-4}  |
| **Embryoid body top expressed genes** |   |                        |    |         |          |
| PCBC_EB_1000                        |   | 997                    | 81 | 9.22 x 10^{-21} | 2.83 x 10^{-18} |
| PCBC_EB_500                         |   | 499                    | 45 | 1.82 x 10^{-13} | 1.40 x 10^{-11} |
| **Embryoid body vs non-stem cells** |   |                        |    |         |          |
| PCBC_EB_blastocyst_1000             |   | 995                    | 74 | 7.21 x 10^{-17} | 1.11 x 10^{-14} |
| PCBC_EB_fibroblast_1000             |   | 992                    | 71 | 2.38 x 10^{-15} | 2.93 x 10^{-13} |
| PCBC_EB_fibroblast_500              |   | 499                    | 44 | 7.42 x 10^{-13} | 5.06 x 10^{-11} |
| PCBC Pathway                      | # Genes in Gene Set (K) | DM | P            | FDR            |
|-----------------------------------|-------------------------|----|--------------|----------------|
| PCBC_EB_blastocyst_500            | 498                     | 41 | 4.04 × 10⁻¹¹ | 1.55 × 10⁻⁹    |
| **Ectoderm top expressed genes** |                         |    |              |                |
| PCBC_ECTO_fibroblast_1000         | 996                     | 62 | 6.46 × 10⁻¹¹ | 2.33 × 10⁻⁹    |
| PCBC_ECTO_fibroblast_500          | 499                     | 39 | 5.61 × 10⁻¹⁰ | 1.72 × 10⁻⁸    |
| PCBC_ECTO_500                     | 498                     | 37 | 6.18 × 10⁻⁹  | 1.65 × 10⁻⁷    |
| PCBC_ECTO_1000                    | 997                     | 57 | 9.06 × 10⁻⁹  | 2.32 × 10⁻⁷    |
| PCBC_ECTO_blastocyst_1000         | 986                     | 56 | 1.55 × 10⁻⁸  | 3.53 × 10⁻⁷    |
| PCBC_ECTO_blastocyst_500          | 490                     | 34 | 1.34 × 10⁻⁷  | 2.50 × 10⁻⁶    |
| **Mesoderm top expressed genes** |                         |    |              |                |
| PCBC_MESO-5_blastocyst_1000       | 979                     | 52 | 4.26 × 10⁻⁷  | 6.71 × 10⁻⁶    |
| PCBC_MESO-5_fibroblast_1000       | 985                     | 50 | 2.64 × 10⁻⁶  | 3.53 × 10⁻⁵    |
| PCBC_MESO-5_500                   | 494                     | 30 | 1.08 × 10⁻⁵  | 1.29 × 10⁻⁴    |
| **Other differentiated cells**    |                         |    |              |                |
| JC_fibro_1000                     | 994                     | 64 | 7.28 × 10⁻¹² | 3.44 × 10⁻¹⁰   |
| geo_heart_1000_K5                 | 428                     | 38 | 2.36 × 10⁻¹¹ | 9.67 × 10⁻¹⁰   |
| JC_fibro_500                      | 497                     | 38 | 1.74 × 10⁻⁹  | 5.08 × 10⁻⁸    |
| PCBC_ctl_geo-heart_1000           | 997                     | 55 | 5.60 × 10⁻⁸  | 1.15 × 10⁻⁶    |
| JC_fibro_2500_K5                  | 826                     | 43 | 7.36 × 10⁻⁶  | 9.42 × 10⁻⁵    |
| JC_fibro_1000_K4                  | 177                     | 16 | 1.22 × 10⁻⁵  | 1.43 × 10⁻⁴    |
### Supplemental Table S7. Age specific estimated FCO methylation fractions in blood leukocytes from birth to old age

| Age group | N   | Min. | P10  | P25  | Median | Mean  | SD   | P75  | P90  | Max.  | P     |
|-----------|-----|------|------|------|--------|-------|------|------|------|-------|-------|
| Newborn   | 60  | 67.5 | 74.4 | 78.5 | 82.3   | 82.0  | 6.0  | 85.6 | 88.8 | 97.6  | Reference |
| <12mo     | 32  | 15.7 | 23.9 | 28.6 | 42.0   | 44.5  | 17.6 | 57.7 | 68.0 | 75.0  | 2.13 × 10^{-134} |
| 12-18mo   | 17  | 22.7 | 25.5 | 29.1 | 30.4   | 31.8  | 5.0  | 36.4 | 38.0 | 39.4  | 2.13 × 10^{-134} |
| 18-24mo   | 23  | 5.9  | 13.4 | 22.9 | 25.9   | 26.6  | 13.2 | 28.9 | 35.9 | 62.5  | 1.34 × 10^{-147} |
| 2-5yr     | 106 | 0    | 2.5  | 9.1  | 15.2   | 14.7  | 8.3  | 20.8 | 24.2 | 37.0  | 5.95 × 10^{-198} |
| 5-18yr    | 31  | 0    | 0    | 0    | 0.5    | 4.3   | 6.8  | 6.7  | 13.2 | 28.7  | <2.23 × 10^{-308} |
| 18-65yr   | 403 | 0    | 0    | 0    | 0      | 3.1   | 4.5  | 5.6  | 9.43 | 26.5  | <2.23 × 10^{-308} |
| >65yr     | 381 | 0    | 0    | 0    | 0      | 1.6   | 3.5  | 1.5  | 5.97 | 25.8  | <2.23 × 10^{-308} |

Notes: Minimum, maximum, percentile cutoff values (10, 25, 50, 75, 90), mean and standard deviations derived from population data combined from published methylation datasets: see Supplemental Table S1. Values < 0.1 were coded as 0. The reported P are based on linear model estimations adjusting for the age group using the newborns as the reference. We also used a linear mixed effect model adjusting for subject (for those measures with several samples), and Study as random effects, the P (using the Kenward Roger approximation for the degrees of freedom) were <2.23 × 10^{-308} for all the groups compared to the newborns.
Supplemental Methods S1. Stability of the FCO calculations

To establish the stability of the FCO signature, we evaluated the absolute difference in the FCO estimates when all the potential combinations of one to five CpGs were lost during the FCO estimations compared to the full set of 27 CpGs using the samples used for the AUROC analysis (umbilical cord blood GSE80310 (Knight et al. 2016), GSE74738 (Hanna et al. 2016), GSE54399 (Montoya-Williams et al. 2017), GSE79056 (Knight et al. 2016), GSE62924 (Rojas et al. 2015). Adult peripheral blood GSE74738 (Hanna et al. 2016), GSE54399 (Montoya-Williams et al. 2017). We also calculated the average root mean square error (RMSE) between the prediction using the 27 CpGs vs all the potential combinations when as few as one CpG and as many as five CpGs were excluded from the 27 FCO CpGs. Our results indicate that the 27 CpG sites is a minimum discriminatory set for a reliable FCO estimation.

Within the 27 CpGs the loss of eight probes (cg01278041, cg05840541, cg11194994, cg11199014, cg13485366, cg14652587, cg17471939, cg22497969) had the biggest impact in the FCO calculations (RMSE>10). In contrast the loss of some other probes (e.g. cg01567783, absent in the EPIC array), only altered minimally the FCO estimates (RMSE:2.24). We suggest that the full set of probes will be used for the calculations but in the absence of specific probes the researcher should consider the increase in the estimation errors.
Supplemental Methods S1 Figure 1. Absolute difference between FCO estimated with one of the CpG probe lost versus the full set of 27 CpGs
Note: the y axis represent the difference in percentages
Supplemental Methods S1 Figure 2. Root Mean Square Error increase per CpG lost

Notes: In the x axis 0 corresponds to the reference including the 27 CpGs, 1, corresponds to 27 combinations losing one CpG, 2 to 351 combinations losing 2 CpGs, 3 to 2925 combinations losing 3 CpGs, 4 to 17550 combinations losing four CpGs, and 5 to 80730 combinations losing 5 CpGs.
Supplemental Methods S2. Synthetic mixture statistical validation

To establish the reliability of our fetal deconvolution methodology, we performed an additional experiment that involved first creating, and then deconvoluting synthetic mixtures of fetal UCB and adult peripheral blood DNA methylation profiles mixed in in predetermined proportions. To more precisely describe our approach, let $S^{CB}$ and $S^A$ represent $J \times 1$ vectors of methylation $\beta$-values for fetal UCB and adult peripheral blood (Fernando et al. 2015; Marabita et al. 2013), respectively, with $J$ denoting the number of CpG loci. The synthetic mixture, $M$, was generated as weighted linear combination of $S^{CB}$ and $S^A$, such that: $M = \pi S^{CB} + (1-\pi) S^A$ and $0 \leq \pi \leq 1$.

Assuming that $S^{CB}$ and $S^A$ represent the DNA methylation profile over “pure” populations of fetal and adult cells, respectively, $\pi$ represents the fraction of cells carrying the FCO signature within the synthetic mixture, $M$. Application of cell mixture deconvolution to $M$ using the FCO signature library allowed us to estimate the fraction of cells carrying the FCO signature, $\hat{\pi}$, which we compared to the “known” predetermined proportion, $\pi$.

To simulate synthetic mixtures we used two additional DNA methylation data sets: GSE66459 a fetal UCB ($n = 22$) data set (Fernando et al. 2015) and GSE43976 restricting to those samples of adult peripheral blood ($n = 52$) data set (Marabita et al. 2013). Importantly, neither of these data sets was used to identify or derive the FCO signature that forms the basis of deconvolution, and therefore represent truly independent data sets. Synthetic mixtures were generated by mixing randomly selected samples from both the fetal UCB and adult peripheral blood data sets, where the mixing parameter was selected to be $\pi = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. For each specification of $\pi$, $n = 10$ synthetic mixture were generated.
Supplemental Methods S3. Maternal contamination sensitivity analyses

During the peer review of this manuscript, one of the reviewers raised our attention to a recently published manuscript describing a marker for maternal blood contamination in cord blood samples (Morin et al. 2017). We thank the reviewer for this important observation; it was very helpful and stimulated a great deal of thinking for the authors. Clearly maternal blood is a potential issue for contaminating cord blood in our setting and we appreciate the reviewer bringing this to our attention.

Those researchers developed a signature of blood maternal contamination using 10 probes from the 450K array and validated their results using three pyrosequenced CpGs. Morin et al. used the Reinius et al. dataset (Reinius et al. 2012) as an adult comparison and whole umbilical cord blood samples to detect differences in a linear model without further adjustment by age. They found 2,250 CpGs as potential targets for the differences between adult peripheral blood and cord blood based on mixed samples, rather than purified cells. They used a random forest approach to select a subset of highly hypomethylated 10 CpGs in the cord blood, none of these CpGs were present within our FCO signature. From this set of 10 CpGs, they developed a semi-quantitative index, wherein if more than 5 CpGs out of 10 demonstrated greater than a 20% difference in methylation, then that sample would qualify as being suspicious of maternal contamination. Although their filtering was based on a strict statistical rule, declaration of contamination mostly involved a qualitative assessment.

In response to the reviewer’s concern, we assessed whether any potential maternal contamination had occurred in our datasets using the method from Morin et al. Only one donor sample
comprising all 6 isolated cells (indicated on the right side of the heatmap below) clustered slightly apart from the other samples (Supplemental Methods Figure 1). However, the DNA methylation age estimated for this sample (range: 0.82-2.95 years) was consistent with a UCB sample. We also clarified that the DNA methylation age margin of error reported by Horvath was >3.6 years (Horvath 2013). Thus, while the reviewer has raised a legitimate potential concern, we conclude there is no evidence of significant contamination in the discovery data set that we used. Nonetheless, we performed a sensitivity analysis eliminating all six cells from that sample and observed stable results. As the results were consistent, we only included the information of the sensitivity analyses in the Methods section and summarize this information for the reader here.

Supplemental Methods S3 Figure 1. Evaluation of potential maternal contamination in the discovery datasets
Notes: umbilical cord blood (UCB).
To further explore the idea of fetal contamination using the Morin makers we also explored our validation dataset and achieved the same results (Supplemental Methods Figure 2).

**Supplemental Methods S3 Figure 2.** Evaluation of potential maternal contamination in the validation datasets
Notes: umbilical cord blood (UCB), FCO estimated proportion (Fetal.proportion).

None of the samples were marked as suspicious using the Morin criteria.

Therefore, we do not believe the evidence supports maternal contamination as a factor influencing the validity or interpretation of our cord blood samples or any of the other fetal and adult data.

Morin et al. used five additional datasets that were not included in our first submission of the manuscript and they were included after the peer review. Using the 10 CpGs in Morin et al. we observed that one sample among the new data is clearly contaminated with maternal blood (Supplemental Methods Figure 3). The contaminated sample clusters with adult blood and has
FCO signature of 0%, as observed in the heatmap below. In addition, when calculating the DNA methylation age of this sample we estimated 44.5 years in the “cord blood sample” vs 45 years in the maternal blood pair. As not all Morin et al. CpGs were present in the GEO datasets accessed, we used a K-nearest neighbors imputation to predict the 10 CpGs in cases where data were missing. As this additional dataset (GSE54399) was used in the final manuscript we excluded this sample from the analyses. Taken together, these exercises give us confidence that we are able to detect maternal contamination using a combination of the Morin et al approach and the estimation of the DNA methylation age, should it exist, and that we can rule this factor out as playing a significant role in our final results.

Supplemental Methods S3 Figure 3. Evaluation of potential maternal contamination in the five independent datasets compared to the FCO estimation
Notes: umbilical cord blood (UCB), FCO estimated proportion (Fetal.proportion).
Supplemental File S1. List of 1218 candidate loci detected and the selected candidates (see Excel file)
Supplemental Material References

Acevedo N, Reinius LE, Vitezic M, Fortino V, Söderhäll C, Honkanen H, Veijola R, Simell O, Toppari J, Ilonen J, et al. 2015. Age-associated DNA methylation changes in immune genes, histone modifiers and chromatin remodeling factors within 5 years after birth in human blood leukocytes. *Clin Epigenetics* 7: 34.

Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, Warren ST. 2012. Age-associated DNA methylation in pediatric populations. *Genome Res* 22: 623–632.

Bakulski KM, Feinberg JI, Andrews S V., Yang J, Brown S, L. McKenney S, Witter F, Walston J, Feinberg AP, Fallin MD. 2016. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics* 11: 354–362.

Bonder MJ, Kasela S, Kals M, Tamm R, Lokk K, Barragan I, Buurman WA, Deelen P, Greve J, Ivanov M, et al. 2014. Genetic and epigenetic regulation of gene expression in fetal and adult human livers. *BMC Genomics* 15: 860.

de Goede OM, Razzaghian HR, Price EM, Jones MJ, Kobor MS, Robinson WP, Lavoie PM. 2015. Nucleated red blood cells impact DNA methylation and expression analyses of cord blood hematopoietic cells. *Clin Epigenetics* 7: 95.

Fernando F, Keijser R, Henneman P, van der Kevie-Kersemaekers A-MF, Mannens MM, van der Post JA, Afink GB, Ris-Stalpers C. 2015. The idiopathic preterm delivery methylation profile in umbilical cord blood DNA. *BMC Genomics* 16: 736.

Gervin K, Page CM, Aass HCD, Jansen MA, Fjeldstad HE, Andreassen BK, Duijts L, van Meurs JB, van Zelm MC, Jaddoe VW, et al. 2016. Cell type specific DNA methylation in cord blood: a 450K-reference data set and cell count-based validation of estimated cell type composition. *Epigenetics* 2294: 00–00.
Guintivano J, Aryee MJ, Kaminsky ZA. 2013. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. Epigenetics 8: 290–302.

Hanna CW, Peñaherrera MS, Saadeh H, Andrews S, McFadden DE, Kelsey G, Robinson WP. 2016. Pervasive polymorphic imprinted methylation in the human placenta. Genome Res 26: 756–67.

Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y, et al. 2013. Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Mol Cell 49: 359–367.

Heyn H, Li N, Ferreira HHJ, Moran S, Pisano DG, Gomez A, Diez J. 2012. Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci U S A 109: 10522–10527.

Horvath S. 2013. DNA methylation age of human tissues and cell types. Genome Biol 14: R115.

Jung N, Dai B, Gentles AJ, Majeti R, Feinberg AP. 2015. An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. Nat Commun 6: 8489.

Knight AK, Craig JM, Theda C, Bækvad-Hansen M, Bybjerg-Grauholm J, Hansen CS, Hollegaard M V, Hougaard DM, Mortensen PB, Weinsheimer SM, et al. 2016. An epigenetic clock for gestational age at birth based on blood methylation data. Genome Biol 17: 206.

Lessard S, Beaudoin M, Benkirane K, Lettre G. 2015. Comparison of DNA methylation profiles in human fetal and adult red blood cell progenitors. Genome Med 7: 1.

Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerström-Billai F, Jagodic M, Sundberg CJ, Ekström TJ, Teschendorff AE, Tegnér J, et al. 2013. An evaluation of analysis pipelines
for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics* **8**: 333–46.

Montoya-Williams D, Quinlan J, Clukay C, Rodney NC, Kertes DA, Mulligan CJ. 2017. Associations between maternal prenatal stress, methylation changes in IGF1 and IGF2, and birth weight. *J Dev Orig Health Dis* **9**: 1–8.

Morin AM, Gatev E, McEwen LM, MacIsaac JL, Lin DTS, Koen N, Czamara D, Rääkkönen K, Zar HJ, Koenen K, et al. 2017. Maternal blood contamination of collected cord blood can be identified using DNA methylation at three CpGs. *Clin Epigenetics* **9**: 75.

Nazor KL, Altun G, Lynch C, Tran H, Harness J V., Slavin I, Garitaonandia I, Müller FJ, Wang YC, Boscolo FS, et al. 2012. Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* **10**: 620–634.

Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén SE, Greco D, Söderhäll C, Scheynius A, Kere J. 2012. Differential DNA methylation in purified human blood cells: Implications for cell lineage and studies on disease susceptibility. *PLoS One* **7**.

Rojas D, Rager JE, Smeester L, Bailey KA, Drobná Z, Rubio-Andrade M, Stýblo M, García-Vargas G, Fry RC. 2015. Prenatal arsenic exposure and the epigenome: identifying sites of 5-methylcytosine alterations that predict functional changes in gene expression in newborn cord blood and subsequent birth outcomes. *Toxicol Sci* **143**: 97–106.

Röännerblad M, Andersson R, Olofsson T, Douagi I, Karimi M, Lehmann S, Hoof I, de Hoon M, Itoh M, Nagao-Sato S, et al. 2014. Analysis of the DNA methylome and transcriptome in granulopoiesis reveals timed changes and dynamic enhancer methylation. *Blood* **123**: e79-89.

Sliker RC, Bos SD, Goeman JJ, Bovée JV, Talens RP, van der Breggen R, Suchiman HED,
Lameijer E-W, Putter H, van den Akker EB, et al. 2013. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin* **6**: 26.

Slieker RC, Roost MS, van Iperen L, Suchiman HED, Tobi EW, Carlotti F, de Koning EJP, Slagboom PE, Heijmans BT, Chuva de Sousa Lopes SM. 2015. DNA Methylation Landscapes of Human Fetal Development. *PLoS Genet* **11**: e1005583.

Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CCY, O’Donovan MC, Bray NJ, Mill J. 2015. Methylomic trajectories across human fetal brain development. *Genome Res* **25**: 338–52.

Tserel L, Kolde R, Limbach M, Tretyakov K, Kasela S, Kisand K, Saare M, Vilo J, Metspalu A, Milani L, et al. 2015. Age-related profiling of DNA methylation in CD8+ T cells reveals changes in immune response and transcriptional regulator genes. *Sci Rep* **5**: 13107.

Urdinguio RG, Torró MI, Bayón GF, Álvarez-Pitti J, Fernández AF, Redon P, Fraga MF, Lurbe E. 2016. Longitudinal study of DNA methylation during the first 5 years of life. *J Transl Med* **14**: 160.

Weidner CI, Walenda T, Lin Q, Wöfler MM, Denecke B, Costa IG, Zenke M, Wagner W. 2013. Hematopoietic stem and progenitor cells acquire distinct DNA-hypermethylation during in vitro culture. *Sci Rep* **3**: 3372.