

**Cell-type specific DNA methylation
in neonatal cord tissue and cord blood:
a 850K-reference panel and comparison of cell-types**

Supplementary Material

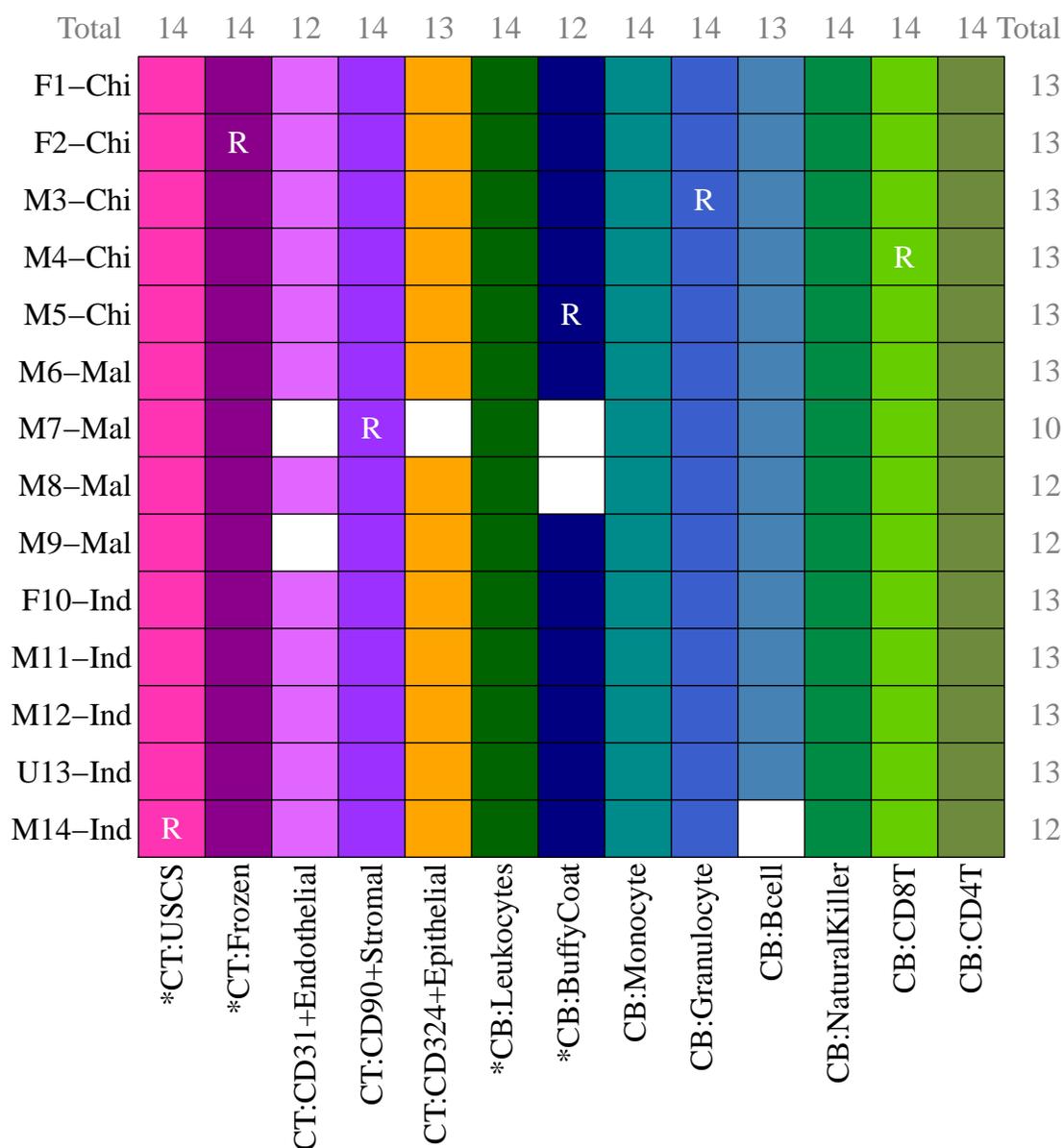
Version 9

Last updated: 23 August 2018

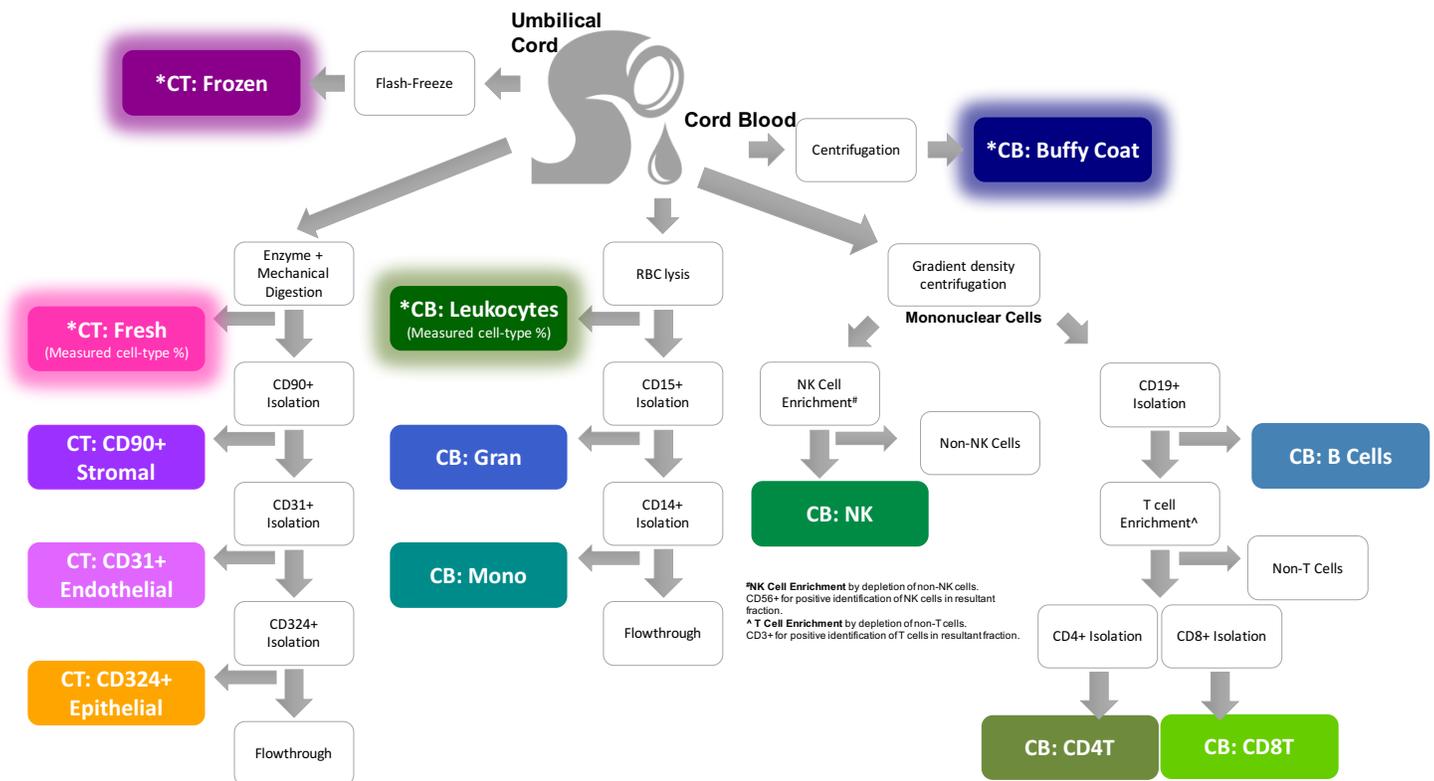
Contents

Supplementary A: Summary of QC and descriptive overview	2
Supplementary B: Benchmark reference panel in of cord tissue (sensitivity analysis)	10
Supplementary C: Benchmark reference panel in cord blood (sensitivity analysis)	20
Supplementary D: Comparison of cord tissue vs. cord blood	22
Supplementary E: Pathway enrichment analysis for cell-type specific CpGs	26

Supplementary Figure A1: **Epigenome-wide DNA methylation profiling of 3 isolated cell-type populations from infant cord tissue (CT), 6 isolated cell-type populations from cord blood (CB) and 4 non-isolated infant tissues (unsorted single cell suspension (USCS) CT, frozen CT, CB unsorted leukocytes, CB buffy coat) from 14 infants, using the Infinium MethylationEPIC BeadChip.** Color in heatmap represents all samples which passed quality control, where excluded samples are shown in white. Each row represents each infant. Infants are labeled according to infant sex and ethnicity. For example, “F2-Chi” indicates that the second infant is female (F) and from Chinese (“Chi”) ethnic group; “M6-Mal” indicates that the sixth infant is male (M) and from Malay (“Mal”) ethnic group; “U” represents unknown. Each column represents each cell-type/tissue. Five populations from CT are shown in purple/pink/orange and prefixed with “CT” in the labels; 8 populations from CB are shown in blue/green and prefixed with “CB” in the labels. Non-isolated infant tissues are labeled with an asterisk. Samples are labeled as “Cell-Type:Infant”. Technical replicates are indicated with a “R”.



Supplementary Figure A2: **Summary of isolated cell-types and non-isolated tissues from infant cord tissue (CT) and cord blood (CB) examined in present study.** Five populations from CT are shown in purple/pink/orange and prefixed with “CT” in the labels; 8 populations from CB are shown in blue/green and prefixed with “CB” in the labels. Non-isolated infant tissues are labeled with an asterisk. From CT, the following 3 populations were isolated sequentially: CD90+ stromal cells, CD31+ endothelial cells, CD324+ epithelial cells. Note that the CD90+ stromal cell population may comprise of sub-populations of cell-types including MSCs, fibroblast cells and smooth muscle cells. From CB, the following 6 populations were sorted: CD15+ granulocytes, CD14+ monocytes, B cells, CD8+ T-cells, CD4+ T-cells, CD56+ Natural Killer (NK) cells. Besides the 9 isolated cell-types (3 from CT and 6 from CB), we also examined 4 non-isolated heterogeneous tissues: unsorted single cell suspension (USCS) CT, frozen CT, CB unsorted leukocytes and CB buffy coat. Cell-type composition was measured in USCS CT and CB unsorted leukocytes.



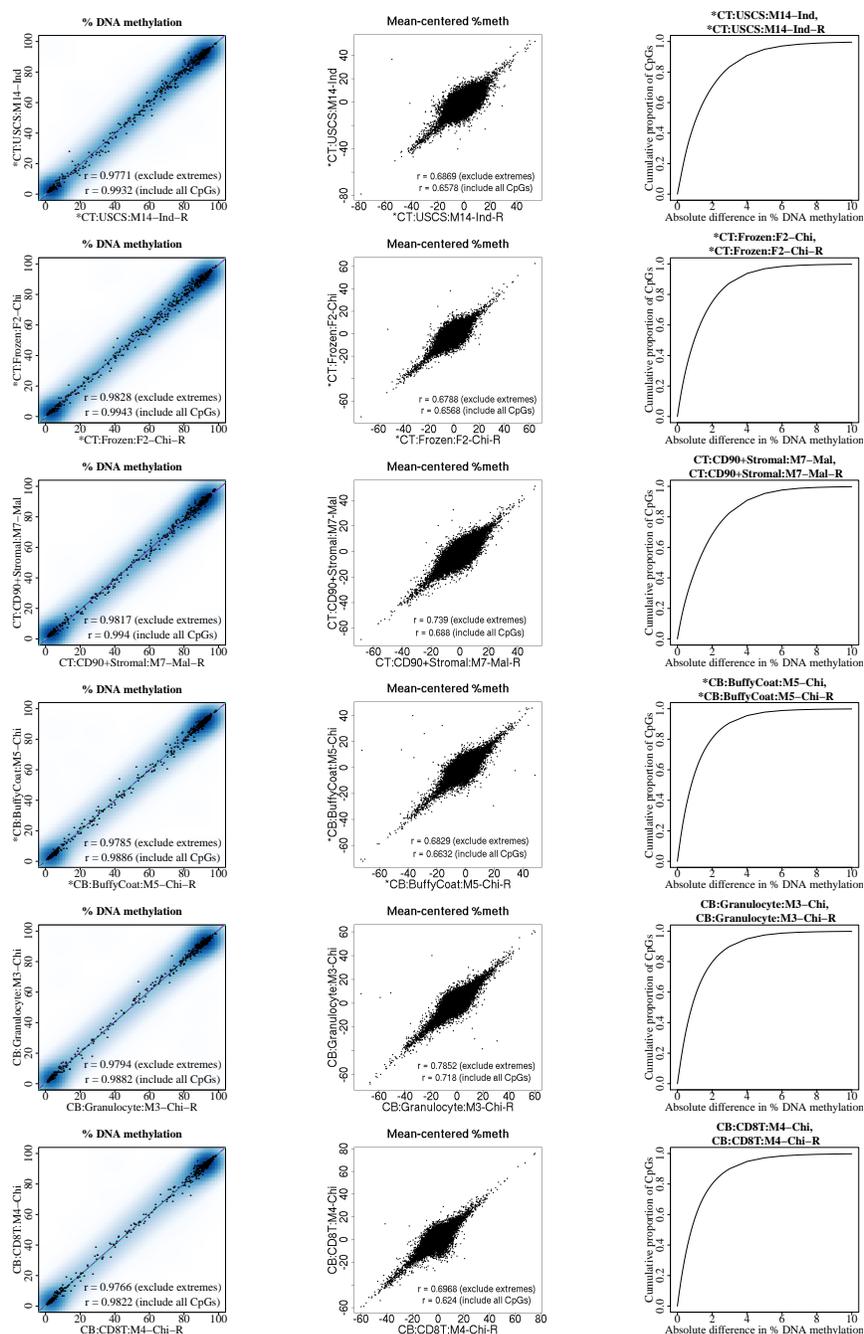
Supplementary Table A1: **Summary statistics on the 14 infant samples used in the current study.**

	Count (%)	Mean (SD)
Ethnicity		
Chinese	5 (36%)	
Malay	4 (29%)	
Indian	5 (36%)	
Infant Sex		
Female	3 (21%)	
Male	10 (71%)	
Unknown	1 (7%)	
Vaginal Delivery		
Yes	9 (64%)	
No	5 (36%)	
Birth weight (g)		3372.4 (316.5)
Birth length (cm)		50.6 (2.6)
Maternal Age (years)		32.0 (3.5)
Gestational Age (weeks)		39.3 (0.8)

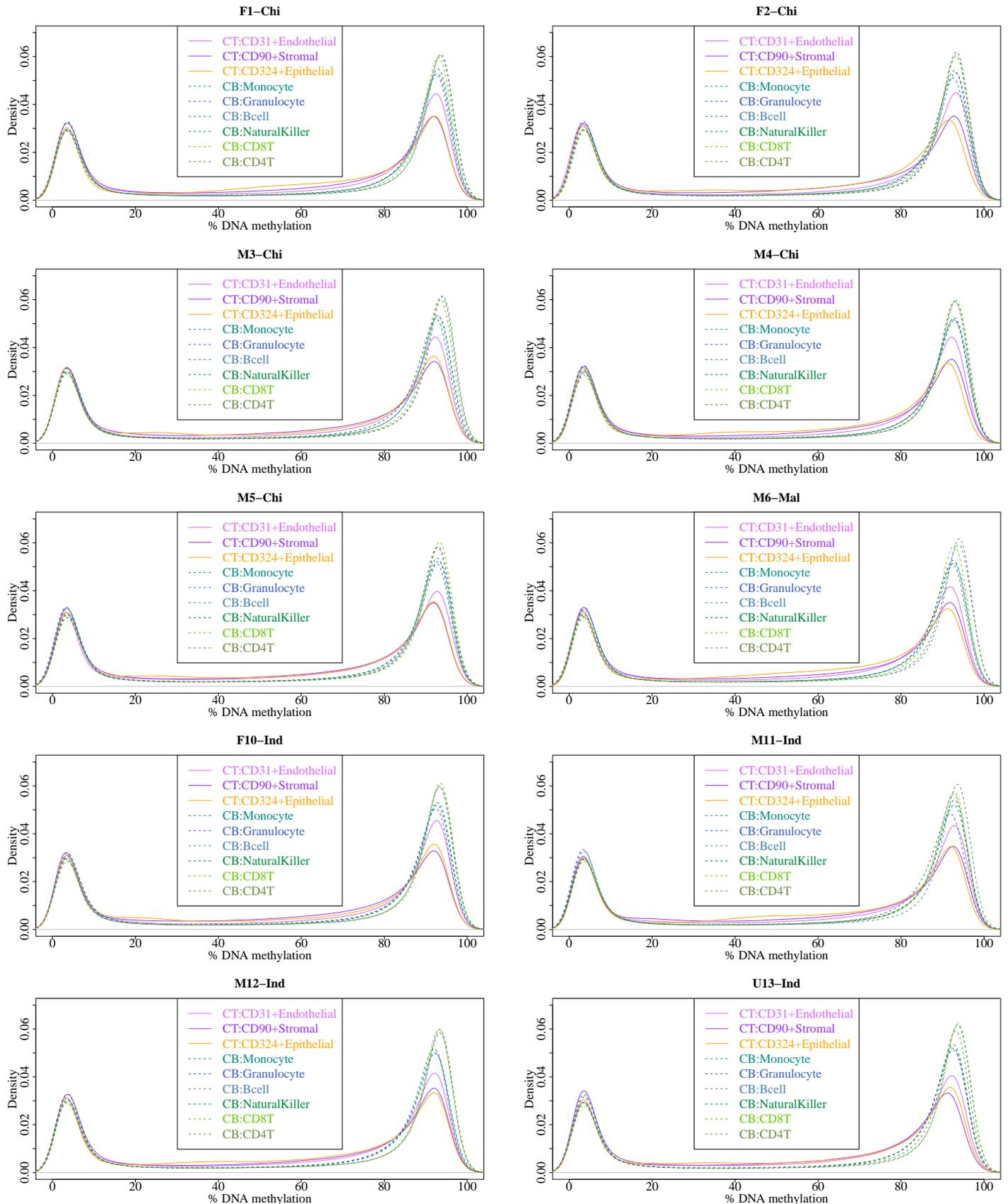
Supplementary Table A2: **Number of CpGs that passed each sequential step of quality control.**

	No. of CpGs
Initial	866,836
No SNP at CpG site or single-base extension	836,329
Detection p-value < 0.01 in all samples	666,140
On Chr 1-22	651,040
Not cross-hybridized	618,485
Final Passed QC	618,485

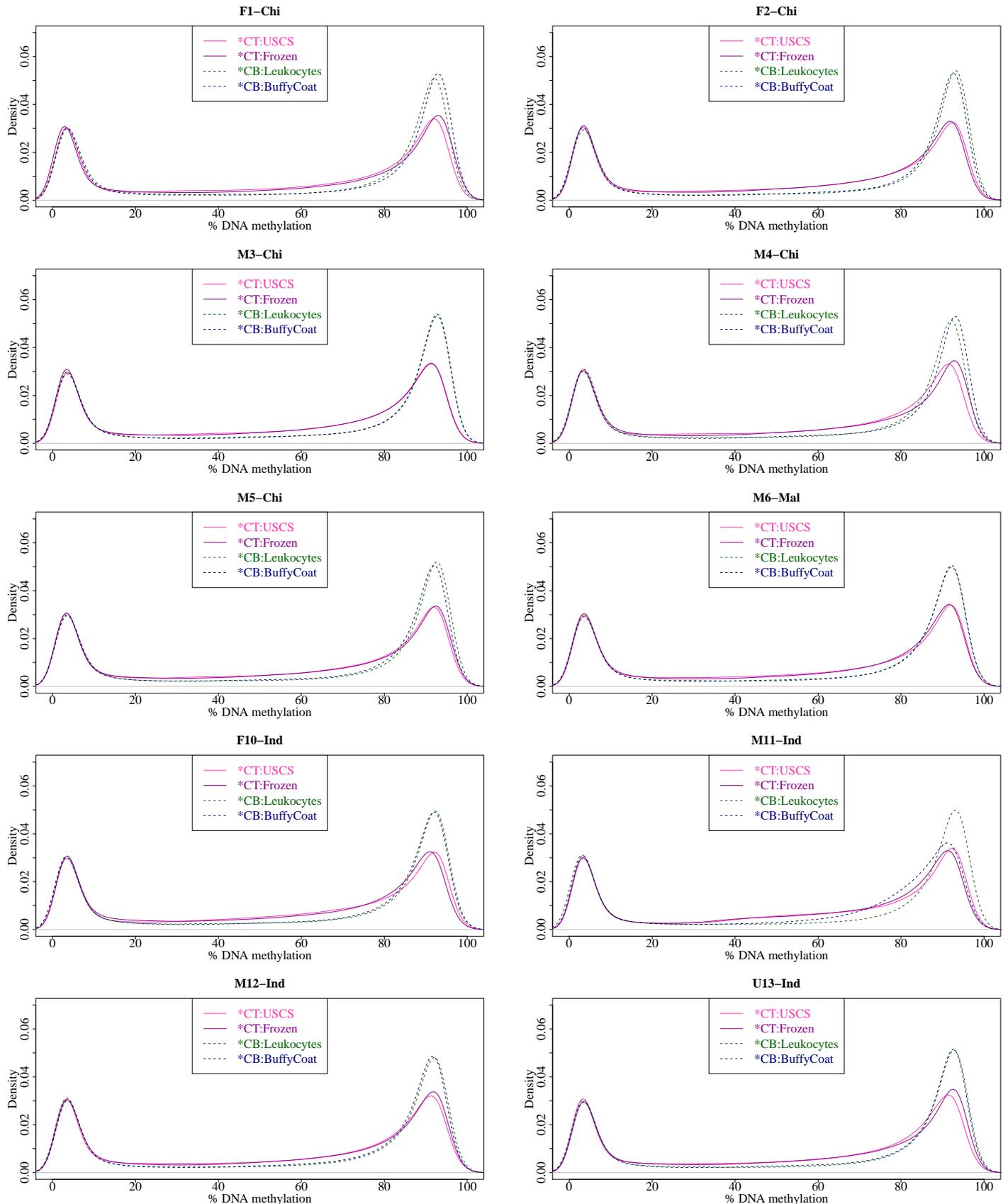
Supplementary Figure A3: **Concordance in DNA methylation values for 6 technical replicates.** Each row represents each sample and its corresponding technical replicate. Left panel gives spearman correlation and scatterplot with smoothed densities color representation of DNA methylation values, for replicate 1 (horizontal axis) and replicate 2 (vertical axis). Color represents density of CpG sites, with darker blue indicating higher density of CpG sites and lighter blue indicating lower density of CpG sites. Five hundred randomly selected CpG sites are shown as black points. Dotted green line gives $y = x$ line, solid purple line gives best-fit line; overlapping lines indicate high concordance between replicates. Two sets of spearman correlation were computed: (i) excluding extremes (CpGs where raw DNA methylation $< 20\%$ or $> 80\%$ in both samples) or (ii) using all CpGs. Middle panel gives mean-centered pearson correlation and scatterplot of mean-centered DNA methylation values, for replicate 1 (horizontal axis) and replicate 2 (vertical axis). Each point represents a single CpG site. Two sets of mean-centered pearson correlation were computed: (i) excluding extremes (CpGs where raw DNA methylation $< 20\%$ or $> 80\%$ in both samples) or (ii) using all CpGs. To compute the mean-centered correlation, for each CpG, the mean of the CpG across all samples of the specific cell-type/tissue was subtracted from the observed DNA methylation value, before computing the mean-centered correlation. Right panel gives absolute difference in DNA methylation values between technical replicates.



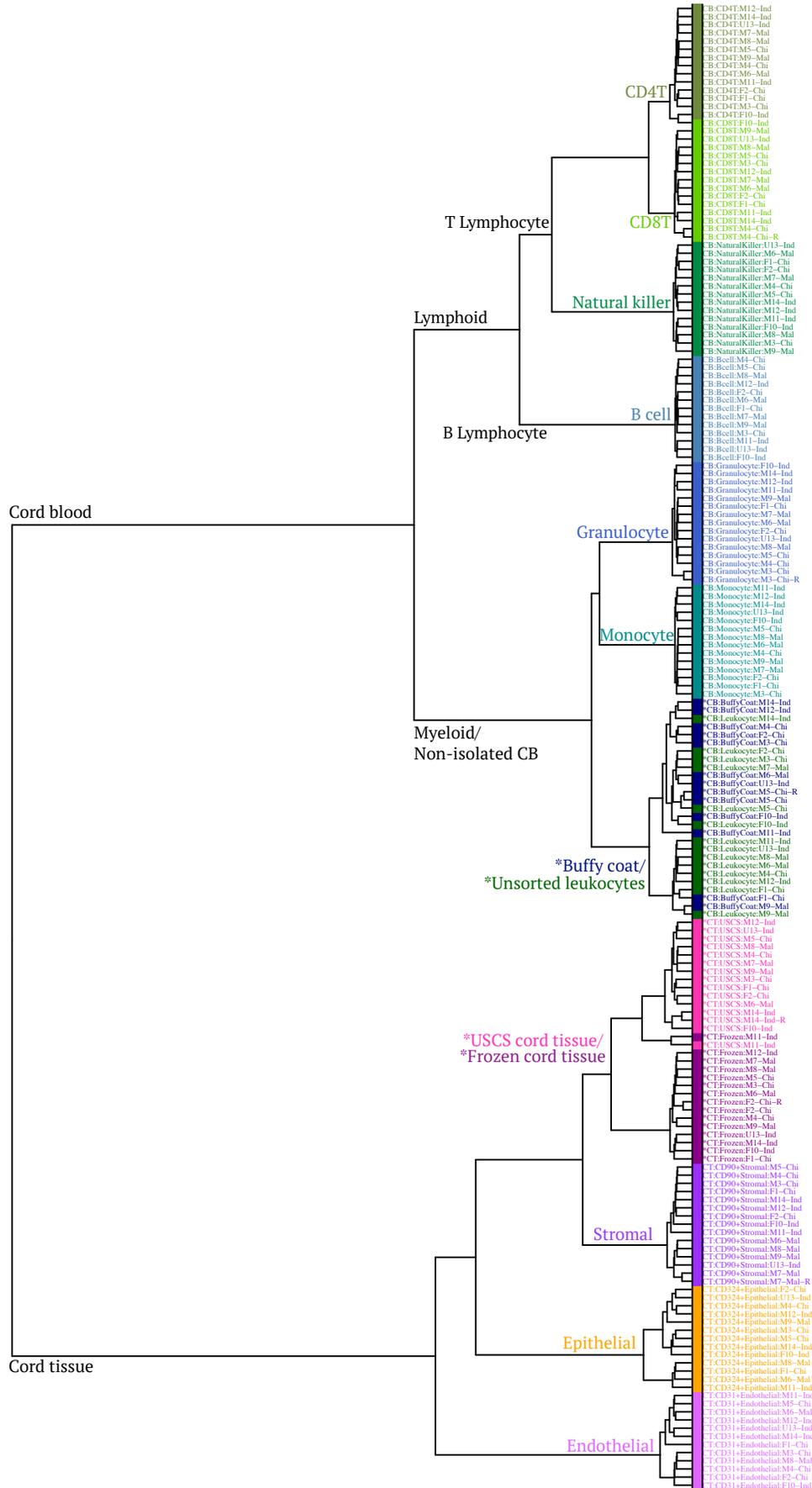
Supplementary Figure A4: Distribution of DNA methylation values for 9 isolated cell-type populations. Cord blood (dotted lines) had higher DNA methylation values compared to cord tissue (solid lines).



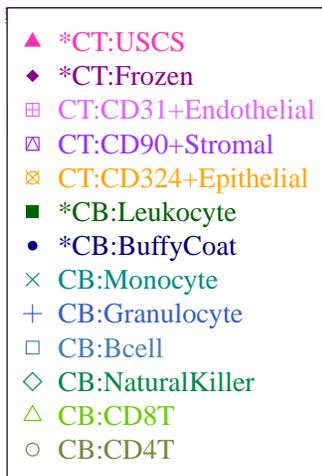
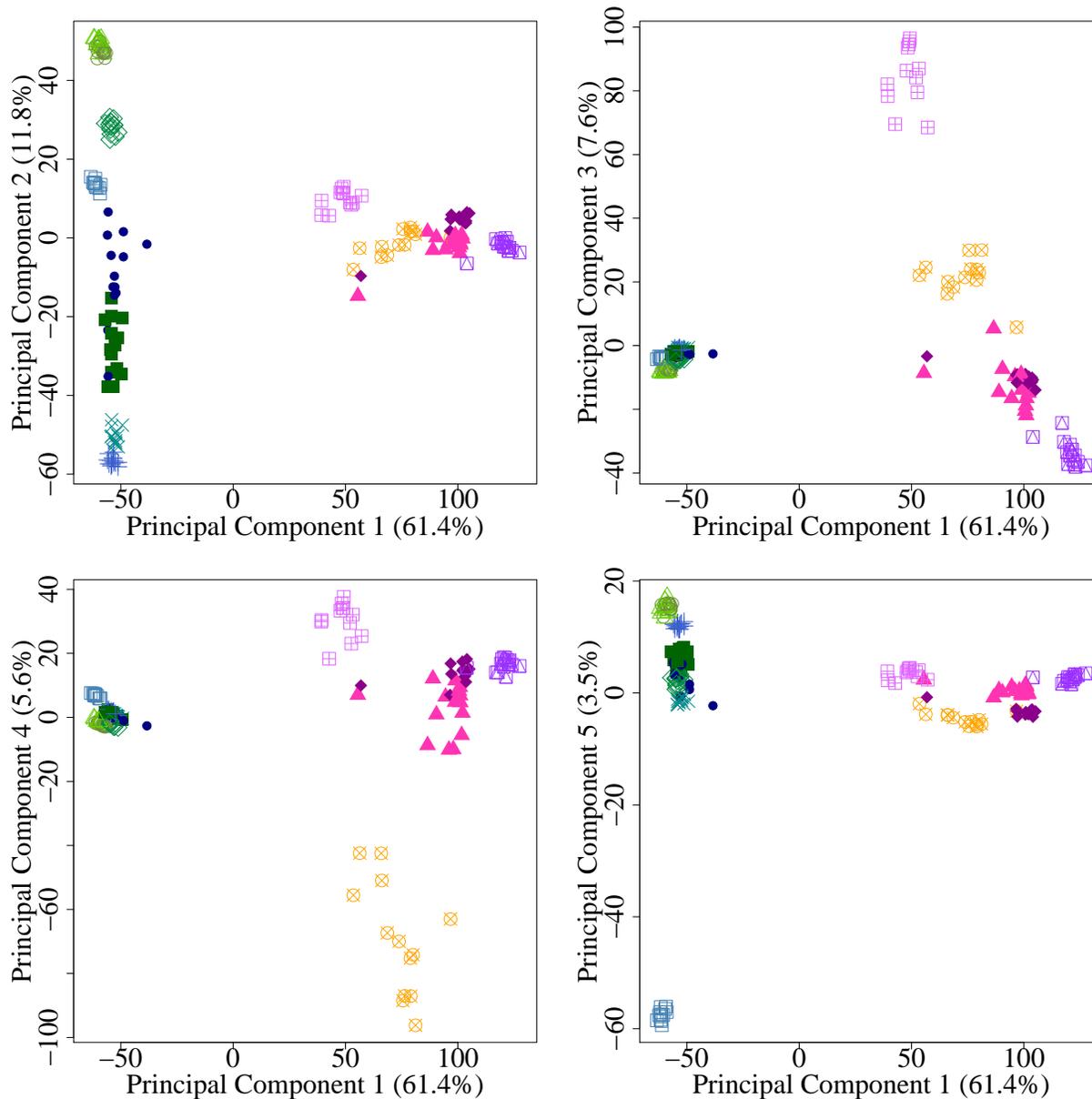
Supplementary Figure A5: Distribution of DNA methylation values for 4 non-isolated infant tissues. Cord blood (dotted lines) had higher DNA methylation values compared to cord tissue (solid lines).



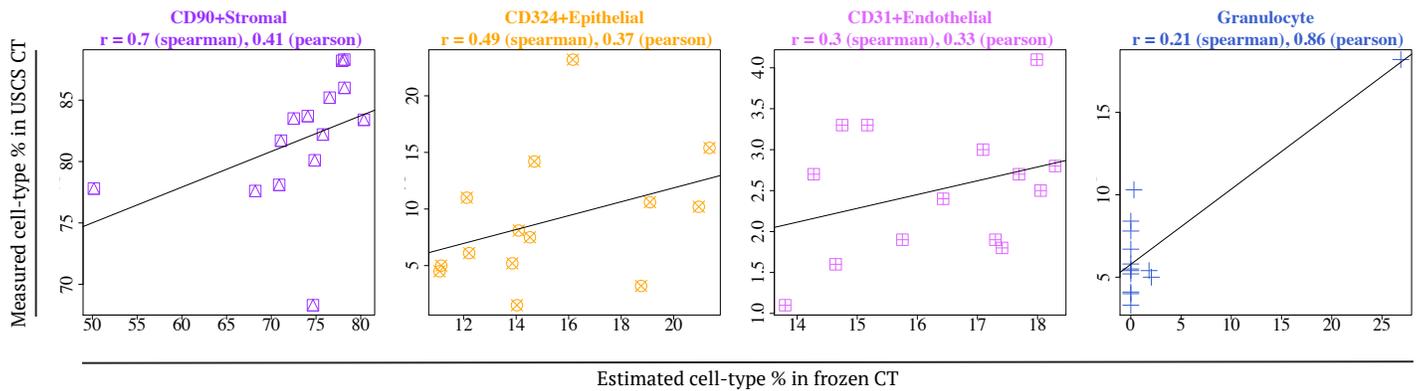
Supplementary Figure A6: Hierarchical clustering of 6 + 3 = 9 isolated cell-types that form cord tissue and blood reference panel, together with cord blood buffy coat, unsorted leukocytes and cord tissue (unsorted single cell suspension (USCS) and frozen). Technical replicates clustered together.



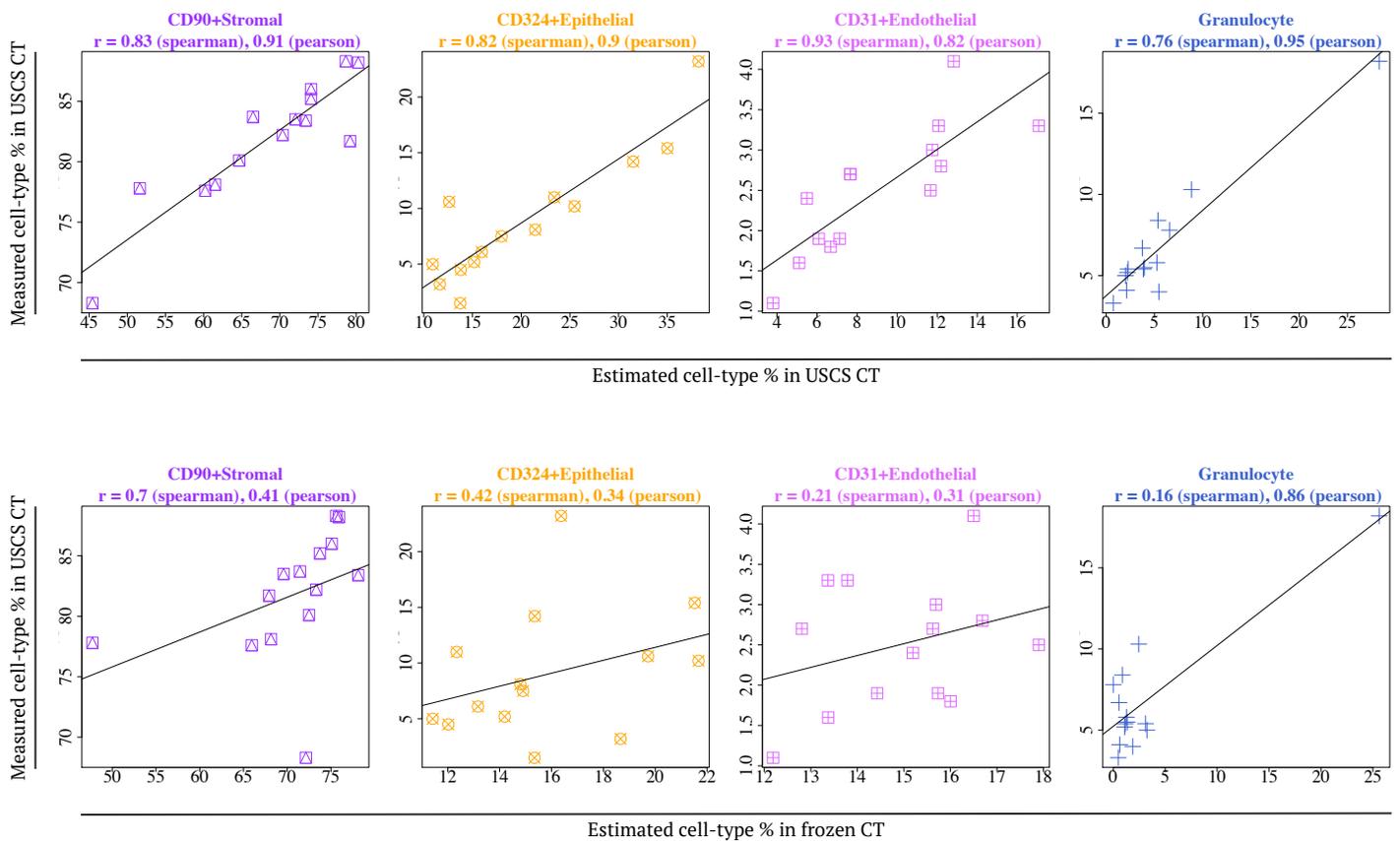
Supplementary Figure A7: Principal components analysis (PCA) of 6 + 3 = 9 isolated cell-types that form cord tissue and blood reference panel, together with cord blood buffy coat, unsorted leukocytes and cord tissue (unsorted single cell suspension (USCS) and frozen). Horizontal axis gives first PC. Vertical axis gives second to fifth PC.



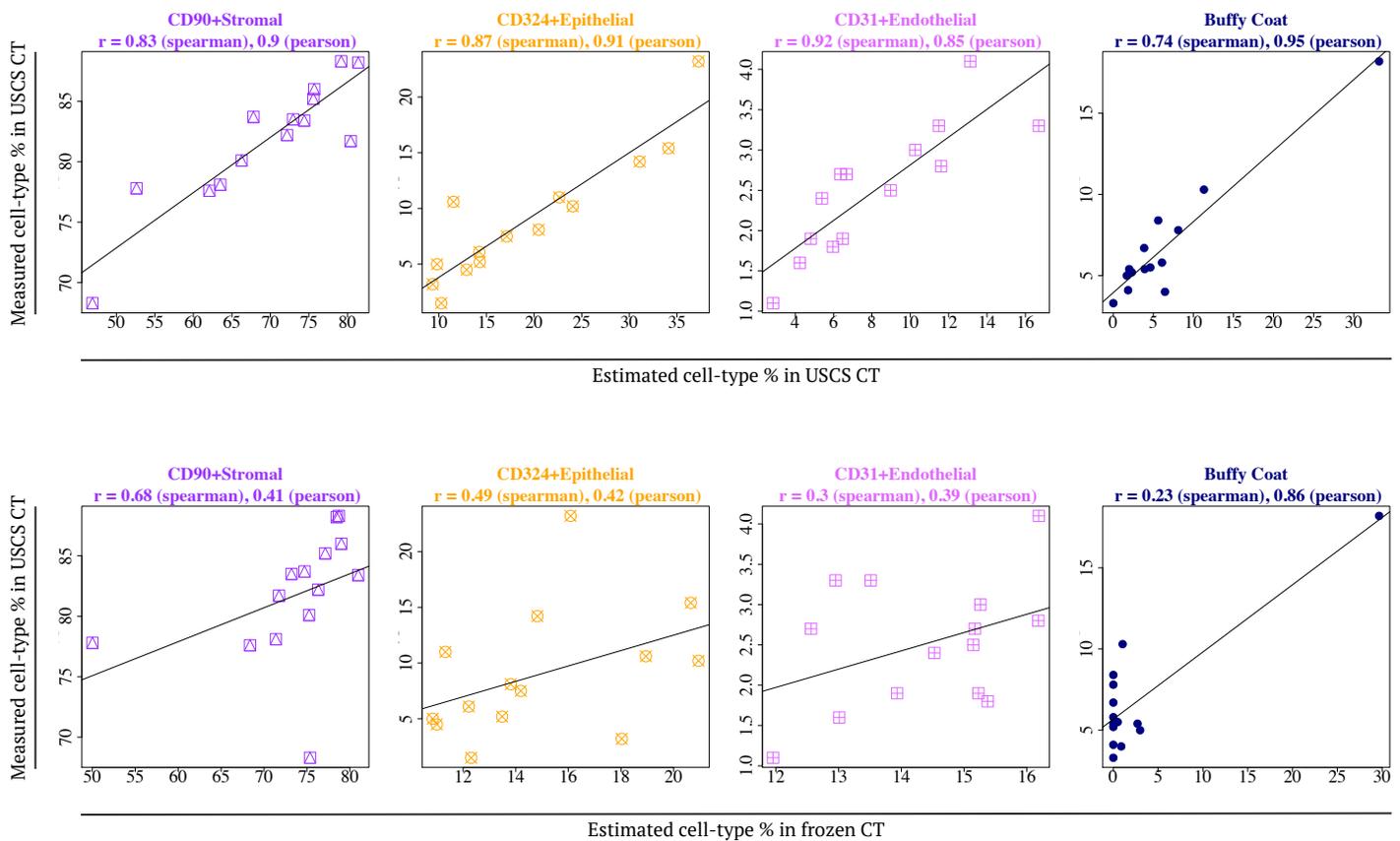
Supplementary Figure B1: **Benchmark reference panel: use of reference panel to capture cell-type composition in non-isolated frozen cord tissue (CT)**. Scatterplots of measured cell-type % in unsorted single cell suspension (USCS) infant CT (vertical axis) vs. estimated cell-type % in *frozen* infant CT (horizontal axis). Vertical axis gives raw/unscaled measured cell-type % for all cell-types except stromal cells, where the unaccounted measured cell-type % were added to the raw/unscaled measured stromal % so that the 4 measured cell-type % sum to 100%. As cellular proportions could not be measured in frozen CT, estimated cellular proportions in *frozen* CT were compared with cellular proportions in *USCS* CT; however cellular composition in *USCS* and frozen CT can differ. Cellular proportions were estimated using the reference panel in the current study following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 1000 cell-type informative CpGs, prioritized by both p-values and directionality of effect sizes (500 CpGs each), that best distinguished each cell-type from the remainder cell-types. Granulocytes isolated from cord blood (CB) were included in the reference panel to capture CB contamination in CT.



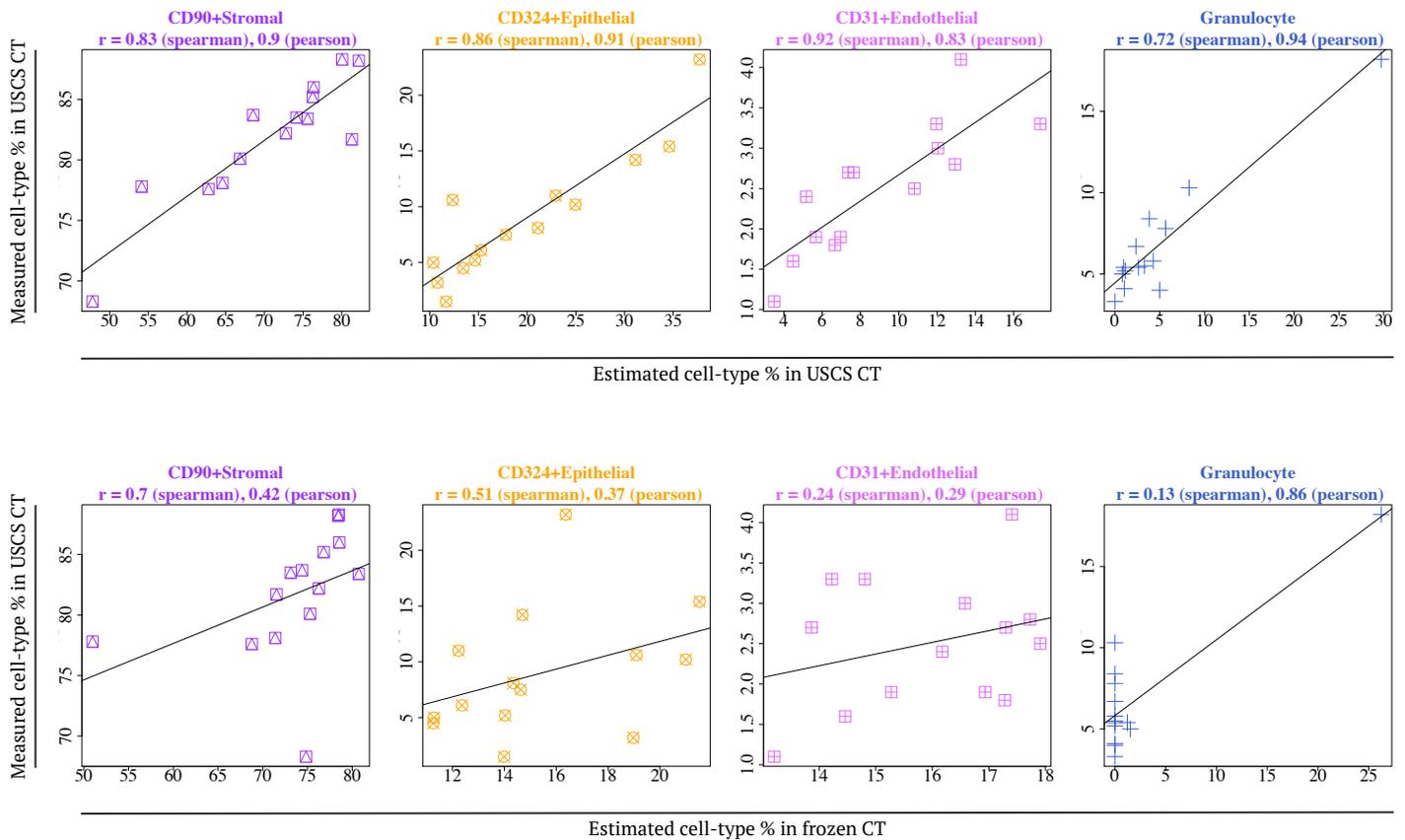
Supplementary Figure B2: **Benchmark reference panel - use of reference panel to capture cell-type composition in non-isolated unsorted single cell suspension (USCS) cord tissue (top panel) and non-isolated frozen cord tissue (bottom panel).** Scatterplots of measured cellular proportions in USCS infant CT (vertical axis) vs. estimated cellular proportions in *USCS* infant CT (horizontal axis, top panel) and *frozen* infant CT (horizontal axis, bottom panel), respectively. Vertical axis gives raw/unscaled measured cell-type % for all cell-types except stromal cells, where the unaccounted measured cell-type % were added to the raw/unscaled measured stromal % so that the 4 measured cell-type % sum to 100%. As cellular proportions could not be measured in frozen CT, estimated cellular proportions in *frozen* CT were compared with cellular proportions in *USCS* CT; however cellular composition in USCS and frozen CT can differ. Cellular proportions were estimated using the reference panel in the current study following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 1,000 cell-type informative CpGs, prioritized by only p-values, that best distinguished each cell-type from the remainder cell-types. Granulocytes isolated from cord blood (CB) was included in the reference panel to capture CB contamination in CT.



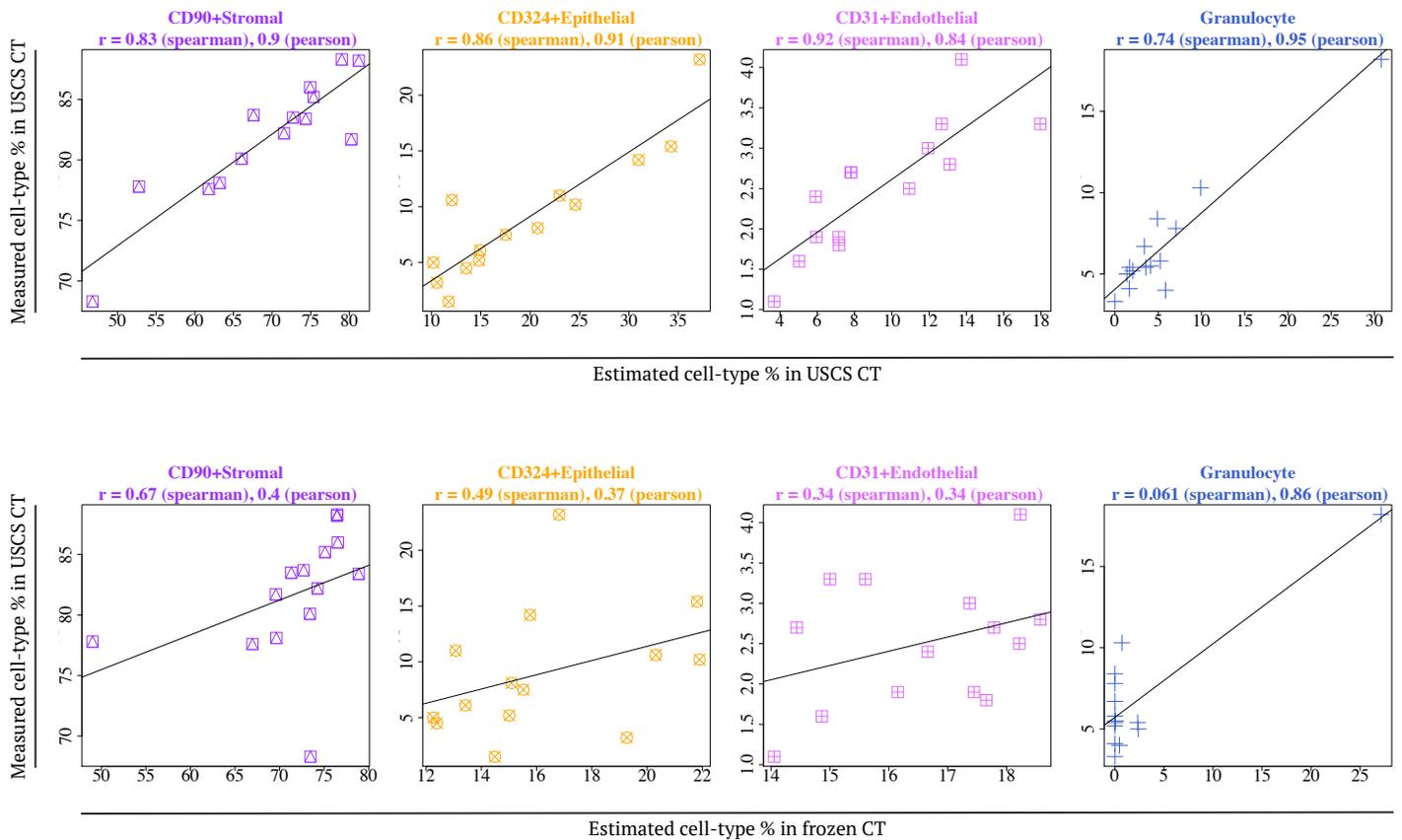
Supplementary Figure B3: **Benchmark reference panel - use of reference panel to capture cell-type composition in non-isolated unsorted single cell suspension (USCS) cord tissue (top panel) and non-isolated frozen cord tissue (bottom panel).** Scatterplots of measured cellular proportions in USCS infant CT (vertical axis) vs. estimated cellular proportions in *USCS* infant CT (horizontal axis, top panel) and *frozen* infant CT (horizontal axis, bottom panel), respectively. Vertical axis gives raw/unscaled measured cell-type % for all cell-types except stromal cells, where the unaccounted measured cell-type % were added to the raw/unscaled measured stromal % so that the 4 measured cell-type % sum to 100%. As cellular proportions could not be measured in frozen CT, estimated cellular proportions in *frozen* CT were compared with cellular proportions in *USCS* CT; however cellular composition in USCS and frozen CT can differ. Cellular proportions were estimated using the reference panel in the current study following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 1,000 cell-type informative CpGs, prioritized by both p-values and directionality of effect sizes (500 CpGs each), that best distinguished each cell-type from the remainder cell-types. Buffy coat from cord blood (CB) was included in the reference panel to capture CB contamination in CT.



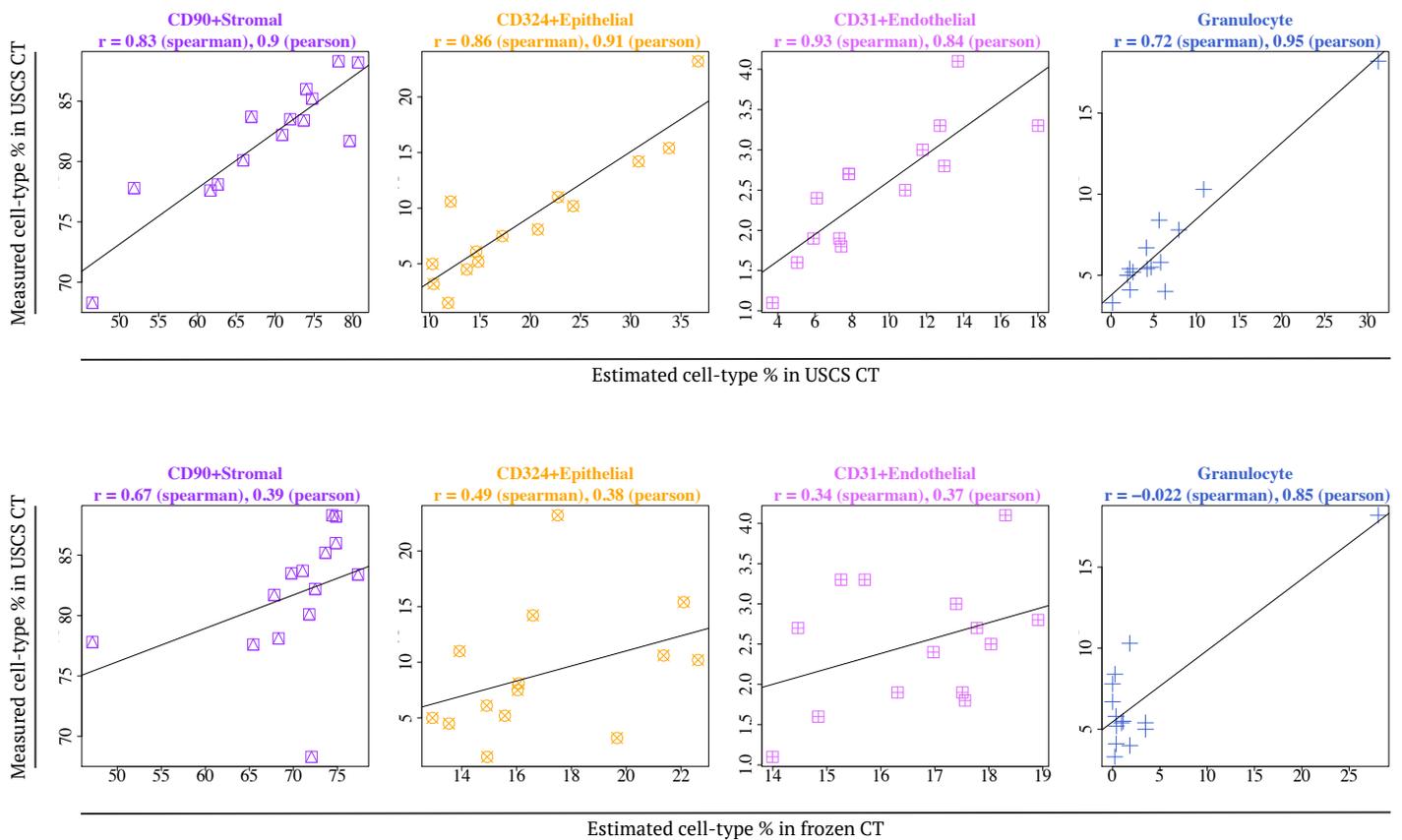
Supplementary Figure B4: **Benchmark reference panel - use of reference panel to capture cell-type composition in non-isolated unsorted single cell suspension (USCS) cord tissue (top panel) and non-isolated frozen cord tissue (bottom panel).** Scatterplots of measured cellular proportions in USCS infant CT (vertical axis) vs. estimated cellular proportions in *USCS* infant CT (horizontal axis, top panel) and *frozen* infant CT (horizontal axis, bottom panel), respectively. Vertical axis gives raw/unscaled measured cell-type % for all cell-types except stromal cells, where the unaccounted measured cell-type % were added to the raw/unscaled measured stromal % so that the 4 measured cell-type % sum to 100%. As cellular proportions could not be measured in frozen CT, estimated cellular proportions in *frozen* CT were compared with cellular proportions in *USCS* CT; however cellular composition in USCS and frozen CT can differ. Cellular proportions were estimated using the reference panel in the current study following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 500 cell-type informative CpGs, prioritized by both p-values and directionality of effect sizes (250 CpGs each), that best distinguished each cell-type from the remainder cell-types. Granulocytes isolated from cord blood (CB) was included in the reference panel to capture CB contamination in CT.



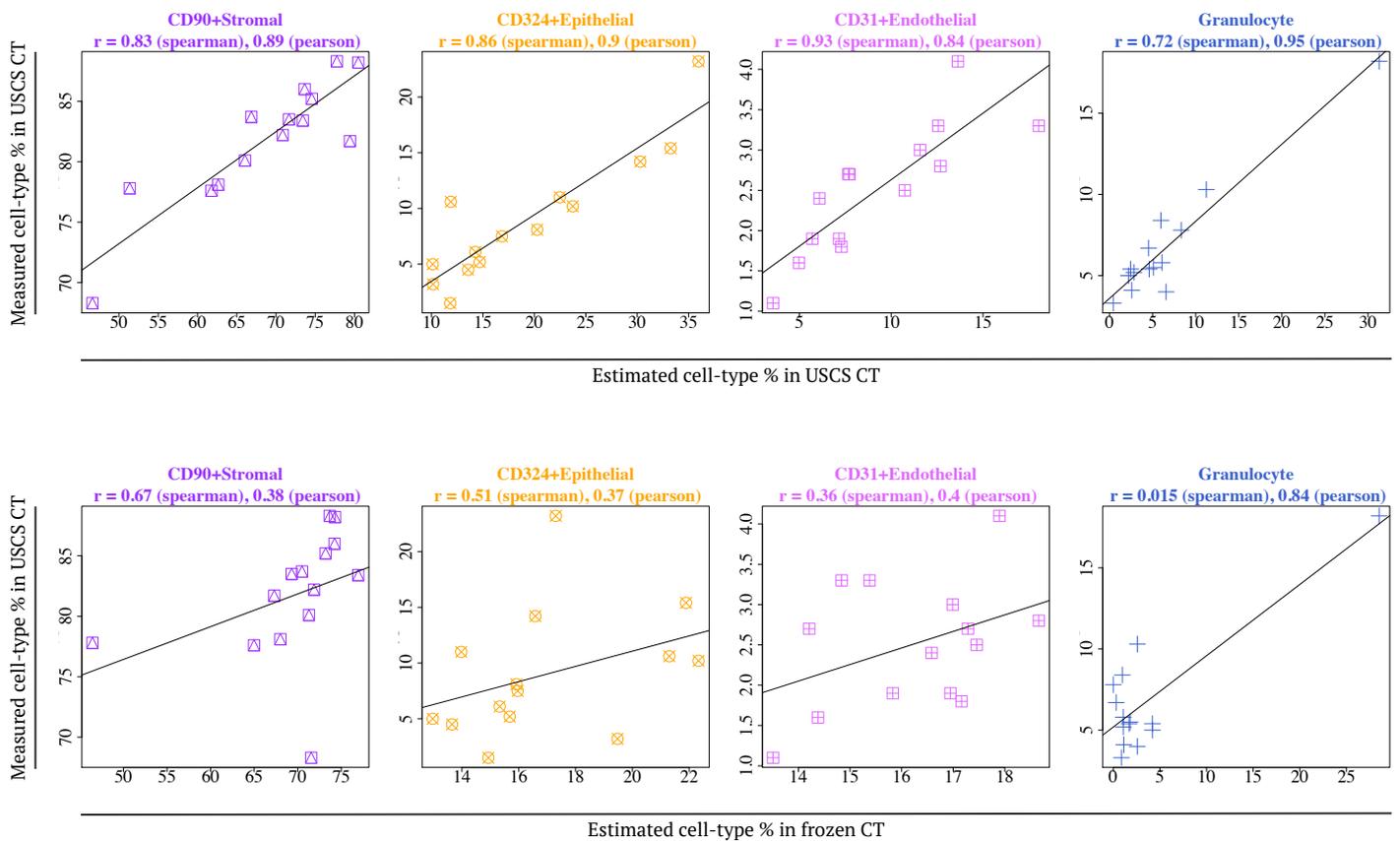
Supplementary Figure B5: **Benchmark reference panel - use of reference panel to capture cell-type composition in non-isolated unsorted single cell suspension (USCS) cord tissue (top panel) and non-isolated frozen cord tissue (bottom panel).** Scatterplots of measured cellular proportions in USCS infant CT (vertical axis) vs. estimated cellular proportions in *USCS* infant CT (horizontal axis, top panel) and *frozen* infant CT (horizontal axis, bottom panel), respectively. Vertical axis gives raw/unscaled measured cell-type % for all cell-types except stromal cells, where the unaccounted measured cell-type % were added to the raw/unscaled measured stromal % so that the 4 measured cell-type % sum to 100%. As cellular proportions could not be measured in frozen CT, estimated cellular proportions in *frozen* CT were compared with cellular proportions in *USCS* CT; however cellular composition in USCS and frozen CT can differ. Cellular proportions were estimated using the reference panel in the current study following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 2,000 cell-type informative CpGs, prioritized by both p-values and directionality of effect sizes (1,000 CpGs each), that best distinguished each cell-type from the remainder cell-types. Granulocytes isolated from cord blood (CB) was included in the reference panel to capture CB contamination in CT.



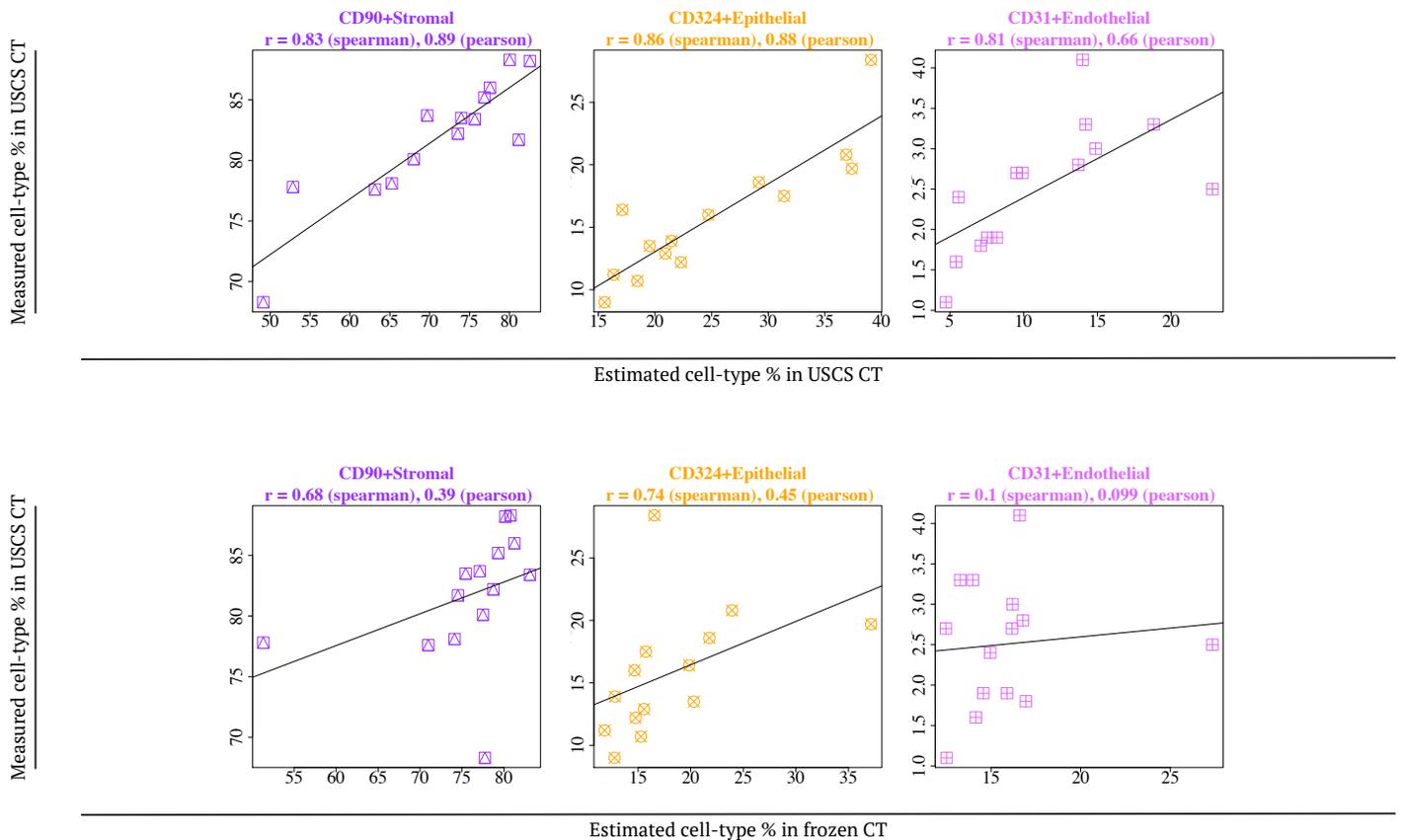
Supplementary Figure B6: **Benchmark reference panel - use of reference panel to capture cell-type composition in non-isolated unsorted single cell suspension (USCS) cord tissue (top panel) and non-isolated frozen cord tissue (bottom panel).** Scatterplots of measured cellular proportions in USCS infant CT (vertical axis) vs. estimated cellular proportions in *USCS* infant CT (horizontal axis, top panel) and *frozen* infant CT (horizontal axis, bottom panel), respectively. Vertical axis gives raw/unscaled measured cell-type % for all cell-types except stromal cells, where the unaccounted measured cell-type % were added to the raw/unscaled measured stromal % so that the 4 measured cell-type % sum to 100%. As cellular proportions could not be measured in frozen CT, estimated cellular proportions in *frozen* CT were compared with cellular proportions in *USCS* CT; however cellular composition in USCS and frozen CT can differ. Cellular proportions were estimated using the reference panel in the current study following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 5,000 cell-type informative CpGs, prioritized by both p-values and directionality of effect sizes (2,500 CpGs each), that best distinguished each cell-type from the remainder cell-types. Granulocytes isolated from cord blood (CB) was included in the reference panel to capture CB contamination in CT.



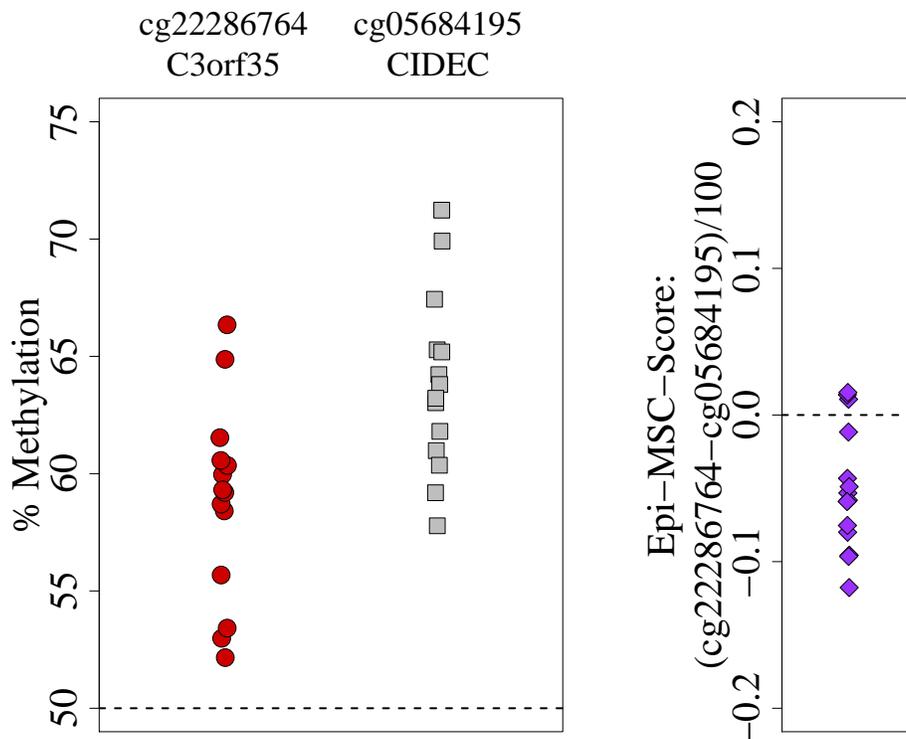
Supplementary Figure B7: **Benchmark reference panel - use of reference panel to capture cell-type composition in non-isolated unsorted single cell suspension (USCS) cord tissue (top panel) and non-isolated frozen cord tissue (bottom panel).** Scatterplots of measured cellular proportions in USCS infant CT (vertical axis) vs. estimated cellular proportions in *USCS* infant CT (horizontal axis, top panel) and *frozen* infant CT (horizontal axis, bottom panel), respectively. Vertical axis gives raw/unscaled measured cell-type % for all cell-types except stromal cells, where the unaccounted measured cell-type % were added to the raw/unscaled measured stromal % so that the 4 measured cell-type % sum to 100%. As cellular proportions could not be measured in frozen CT, estimated cellular proportions in *frozen* CT were compared with cellular proportions in *USCS* CT; however cellular composition in USCS and frozen CT can differ. Cellular proportions were estimated using the reference panel in the current study following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 10,000 cell-type informative CpGs, prioritized by both p-values and directionality of effect sizes (5,000 CpGs each), that best distinguished each cell-type from the remainder cell-types. Granulocytes isolated from cord blood (CB) was included in the reference panel to capture CB contamination in CT.



Supplementary Figure B8: **Benchmark reference panel - use of reference panel to capture cell-type composition in non-isolated unsorted single cell suspension (USCS) cord tissue (top panel) and non-isolated frozen cord tissue (bottom panel).** Scatterplots of measured cellular proportions in USCS infant CT (vertical axis) vs. estimated cellular proportions in *USCS* infant CT (horizontal axis, top panel) and *frozen* infant CT (horizontal axis, bottom panel), respectively. Vertical axis gives raw/unscaled measured cell-type % for all cell-types except stromal cells, where the unaccounted measured cell-type % were added to the raw/unscaled measured stromal % so that the 4 measured cell-type % sum to 100%. As cellular proportions could not be measured in frozen CT, estimated cellular proportions in *frozen* CT were compared with cellular proportions in *USCS* CT; however cellular composition in USCS and frozen CT can differ. Cellular proportions were estimated using the reference panel in the current study following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 1,000 cell-type informative CpGs, prioritized by both p-values and directionality of effect sizes (500 CpGs each), that best distinguished each cell-type from the remainder cell-types. When there is no cord blood (e.g. granulocytes) representation on the reference panel, the estimated cellular proportions for stromal and endothelial cells seem similar to before, and proportions previously attributed to “granulocytes” (blood) seem to be “attributed” to epithelial cells. The graph shows estimated cellular proportion for epithelial vs. measured epithelial composition + measured blood composition.



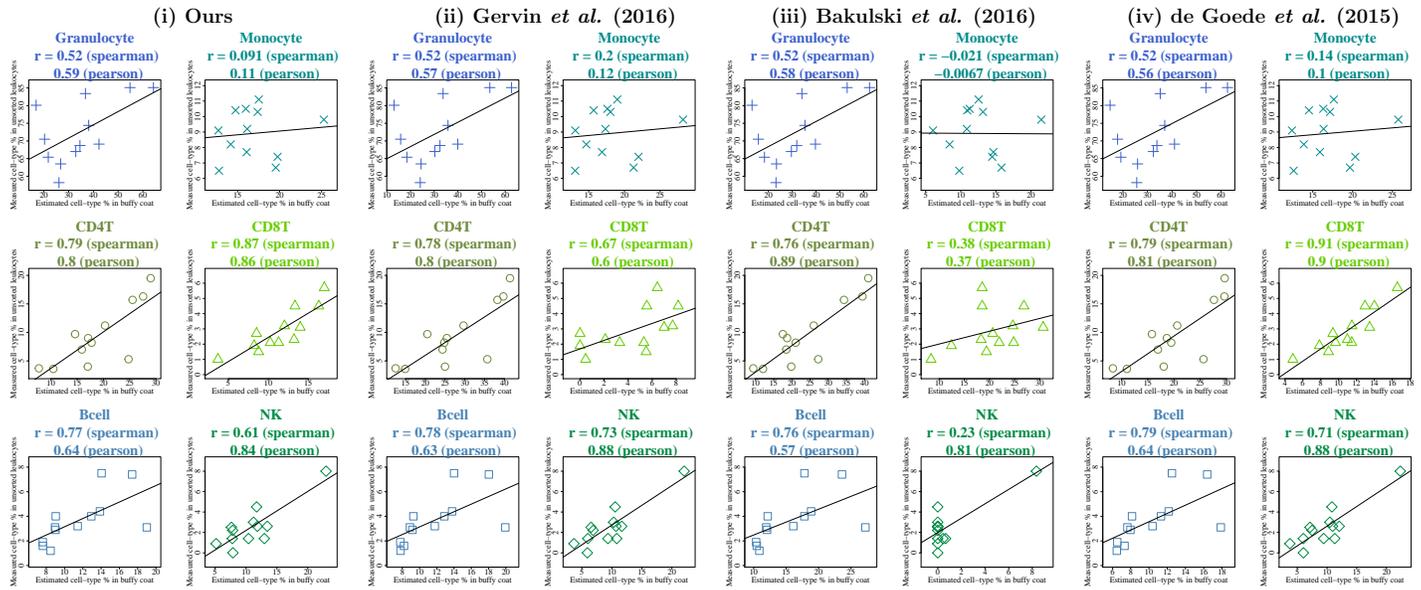
Supplementary Figure B9: **Methylation levels at cg22286764 (*C3orf35*), cg05684195 (*CIDEA*) and the resulting Epi-MSA score for 14 CD90+ stromal cells.** The Epi-MSA score is defined as the difference in methylation beta values between the 2 CpGs (where methylation beta values are represented on a 0-1 scale). A negative Epi-MSA score indicates fibroblast cells, while a positive score represents MSCs. When applied to CD90+ stromal cells isolated from CT, 11 of the 14 samples had a score below 0, while 3 samples had score greater than but close to 0, suggesting that proportion of fibroblasts exceeds the proportion of MSCs (slightly) in our CD90+ stromal cells pool. However, the Epi-MSA score was constructed from the methylomes of cultured fibroblasts and MSCs, and may not be applicable to non-cultured cell-types.



Supplementary Table B1: **Benchmark reference panel - use of reference panel to capture cell-type composition in non-isolated unsorted single cell suspension (USCS) (second column) and non-isolated frozen (third column) cord tissue (CT).** The numbers in the second column give normalized root-mean-square error (RMSE) comparing measured cell-type % in *USCS CT* vs. estimated cell-type % in *USCS CT*. The numbers in the third column give normalized RMSE comparing measured cell-type % in *USCS CT* vs. estimated cell-type % in *frozen CT*. The RMSE was normalized by the interquartile range of the measured cell-type composition in *USCS CT*. Raw/unscaled measured cell-type % were used for all cell-types except stromal cells, where the unaccounted measured cell-type % were added to the raw/unscaled measured stromal % so that the 4 measured cell-type % sum to 100%. As cellular proportions could not be measured in frozen CT, estimated cellular proportions in *frozen CT* were compared with cellular proportions in *USCS CT*; however cellular composition in USCS and frozen CT can differ. Cellular proportions were estimated using the reference panel in the current study following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 1000 cell-type informative CpGs, prioritized by both p-values and directionality of effect sizes (500 CpGs each), that best distinguished each cell-type from the remainder cell-types. Granulocytes isolated from cord blood (CB) were included in the reference panel to capture CB contamination in CT.

Cell-type	USCS CT	Frozen CT
CD90+Stromal	2.15	1.77
CD324+Epithelial	2.01	1.41
CD31+Endothelial	7.04	13.24
Granulocyte	1.72	2.50

Supplementary Figure C1: **Benchmark reference panel - use of 4 reference panels to capture cell-type composition in cord blood buffy coat.** Scatterplots of measured cell-type % in *unsorted leukocytes* (vertical axis) vs. estimated cell-type % in *buffy coat* (horizontal axis), using each of the 4 reference panels. Vertical axis gives raw/unscaled measured cell-type % for all cell-types except granulocytes, where the unaccounted measured cell-type % were added to the raw/unscaled measured granulocytes % so that the 6 measured cell-type % sum to 100%. Cellular composition was estimated using each reference panel following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 1,000 cell-type informative CpGs, prioritized by only p-values, that best distinguished each cell-type from the remainder cell-types. As cellular composition was not measured in buffy coat, estimated cellular composition in *buffy coat* was compared with cellular composition measured in *unsorted leukocytes*; however cellular composition in the 2 can differ.



Supplementary Table C1: **Benchmark reference panel - use of 4 reference panels to capture cell-type composition in cord blood unsorted leukocytes**. The numbers in the table gives normalized root-mean-square error (RMSE) comparing measured cell-type % in *unsorted leukocytes* vs. estimated cell-type % in *unsorted leukocytes*, using each of the 4 reference panels. The RMSE was normalized by the interquartile range of the measured cell-type composition. The raw/unscaled measured cell-type % was used for all cell-types except granulocytes, where the unaccounted measured cell-type % were added to the raw/unscaled measured granulocytes % so that the 6 measured cell-type % sum to 100%. Cellular composition was estimated using each reference panel following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 1,000 cell-type informative CpGs, prioritized by only p-values, that best distinguished each cell-type from the remainder cell-types.

Cell-type	Current study	Gervin <i>at al.</i> (2016)	Bakulski <i>at al.</i> (2016)	de Goede <i>at al.</i> (2015)
Granulocyte	1.06	1.17	1.16	1.12
Monocyte	1.32	1.45	0.87	1.24
CD4T	0.30	0.87	0.55	0.33
CD8T	1.74	0.99	5.12	2.00
Bcell	2.83	3.01	4.81	2.17
NK	3.68	2.87	1.66	2.79

Supplementary Table C2: **Benchmark reference panel - use of 4 reference panels to capture cell-type composition in cord blood buffy coat**. The numbers in the table gives normalized root-mean-square error (RMSE) comparing measured cell-type % in *unsorted leukocytes* vs. estimated cell-type % in *buffy coat*, using each of the 4 reference panels. The RMSE was normalized by the interquartile range of the measured cell-type composition. The raw/unscaled measured cell-type % was used for all cell-types except granulocytes, where the unaccounted measured cell-type % were added to the raw/unscaled measured granulocytes % so that the 6 measured cell-type % sum to 100%. Cellular composition was estimated using each reference panel following the method described by Houseman *et al.* (2012), where pairwise t-tests were used to identify 1,000 cell-type informative CpGs, prioritized by only p-values, that best distinguished each cell-type from the remainder cell-types. As cellular composition was not measured in buffy coat, estimated cellular composition in *buffy coat* was compared with cellular composition measured in *unsorted leukocytes* ; however cellular composition in the 2 can differ.

Cell-type	Current study	Gervin <i>at al.</i> (2016)	Bakulski <i>at al.</i> (2016)	de Goede <i>at al.</i> (2015)
Granulocyte	2.76	2.94	2.96	2.85
Monocyte	3.21	3.73	2.00	3.26
CD4T	1.40	2.61	2.10	1.55
CD8T	6.03	1.85	12.47	5.63
Bcell	5.81	5.98	8.88	5.02
NK	6.81	5.83	1.69	6.03

Supplementary Table D1: **List of primary tissues/cells and primary Cultures/Embryonic Stem (ES) cell derived samples profiled using reduced representation bisulfite sequencing (RRBS) in the Epigenome Roadmap project. Samples with high missing rates were excluded.**

Primary Tissues/Cells		
Cell type/tissue group	EID	Epigenome name
Blood	E030	Primary neutrophils (from PB)
Blood	E031	Primary B cells from cord blood
Blood	E035	Primary haematopoietic stem cells (HSCs)
Blood	E050	Primary HSCs G-CSF-mobilized female
Blood	E051	Primary HSCs G-CSF-mobilized male
Brain	E068	Brain anterior caudate
Brain	E069	Brain cingulate gyrus
Brain	E072	Brain inferior temporal lobe
Brain	E073	Brain dorsolateral prefrontal cortex
Brain	E074	Brain substantia nigra
Digestive	E075	Colonic mucosa
Smooth Muscle	E076	Colon smooth muscle
Digestive	E077	Duodenum mucosa
Brain, fetal	E081	Fetal brain male
Heart, fetal	E083	Fetal heart
Kidney, fetal	E086	
Pancreas	E087	Pancreatic islets
Lung, fetal	E088	
Digestive	E101	Rectal mucosa donor 29
Digestive	E102	Rectal mucosa donor 31
Smooth Muscle	E103	Rectal smooth muscle
Muscle	E107	Skeletal muscle male
Muscle	E108	Skeletal muscle female
Digestive	E110	Stomach mucosa
Smooth Muscle	E111	Stomach smooth muscle

Primary Cultures/Embryonic Stem (ES) cell derived		
Cell type/tissue group	EID	Epigenome name
ES cell	E002	ES-WA7 cells
ES cell	E003	H1 cells
ES cell	E008	H9 cells
ES cell	E014	HUES48 cells
ES cell	E015	HUES6 cells
ES deriv	E009	H9 derived neuronal progenitor cultured cells
ES deriv	E010	H9 derived neuron cultured cells
ES deriv	E011	HUES64 derived CD184+ endoderm
iPSC	E018	iPS-15b cells
iPSC	E019	iPS-18 cells
iPSC	E020	iPS-20b cells
Mesench	E026	Bone marrow derived MSCs
Mesench	E049	Mesenchymal stem cell deriv. chondrocyte

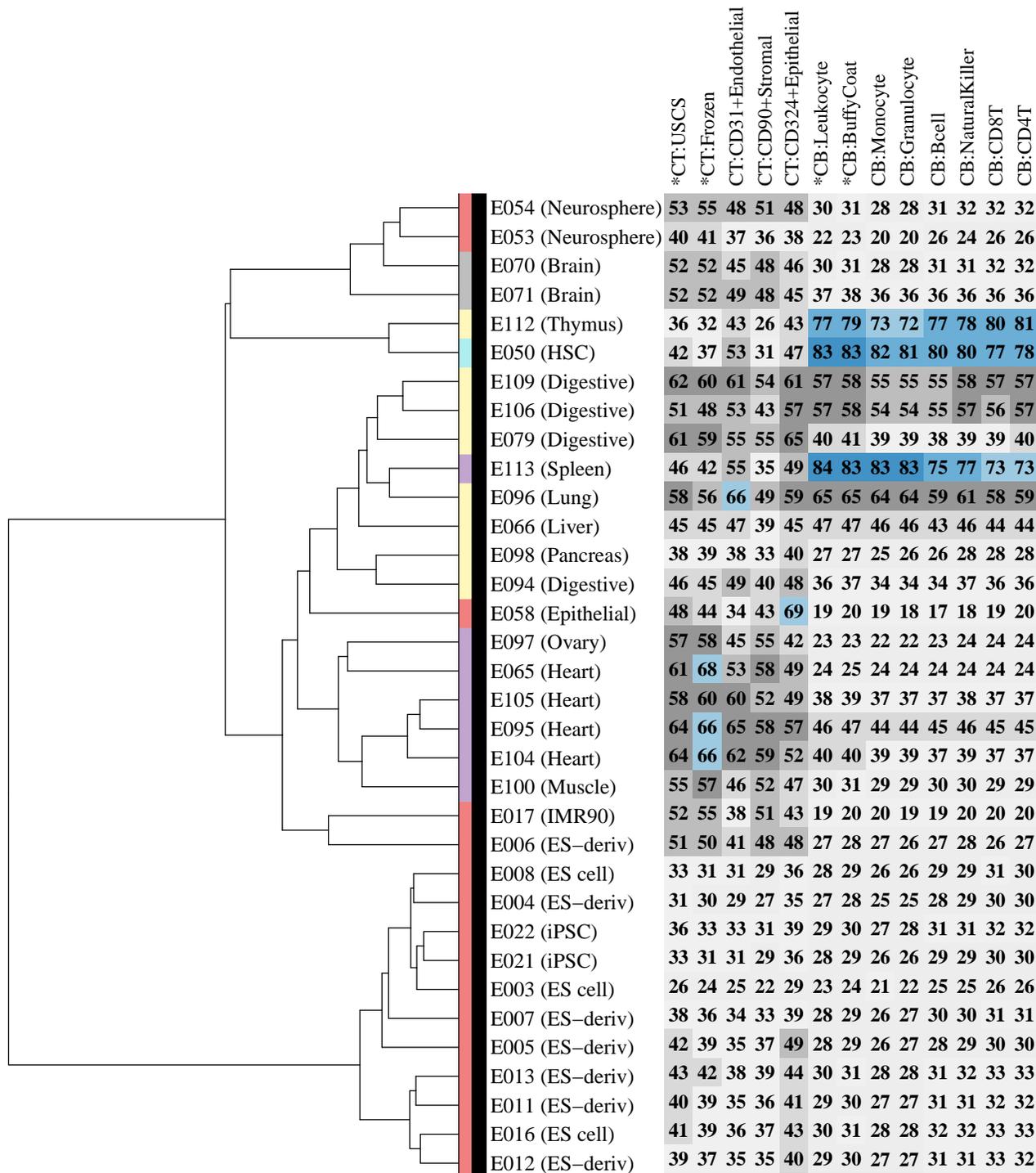
Supplementary Table D2: **List of primary tissues/cells and primary Cultures/Embryonic Stem (ES) cell derived samples profiled using whole genome bisulfite sequencing (WGBS) in the Epigenome Roadmap project. Samples with high missing rates were excluded.**

Cell type/tissue group	EID	Epigenome name
HSC	E050	Primary HSCs G-CSF mobilized female
Heart	E065	Aorta
Liver	E066	
Brain	E070	Brain germinal matrix
Brain	E071	Brain hippocampus middle
Digestive	E079	Oesophagus
Digestive	E094	Gastic
Heart	E095	Left ventricle
Lung	E096	
Ovary	E097	
Pancreas	E098	
Muscle	E100	Psoas muscle
Heart	E104	Right atrium
Heart	E105	Right ventricle
Digestive	E106	Sigmoid colon
Digestive	E109	Small intestine
Thymus	E112	
Spleen	E113	

Primary Cultures/Embryonic Stem (ES) cell derived

Cell type/tissue group	EID	Epigenome name
ES cell	E003	H1 cells
ES cell	E008	H9 cells
ES cell	E016	HUES64 cells
ES deriv	E004	H1 BMP4 derived mesendoderm
ES deriv	E005	H1 BMP4 derived trophoblast
ES deriv	E006	H1 derived mesenchymal stem cells
ES deriv	E007	H1 derived neuronal progenitor cultured cells
ES deriv	E011	HUES64 derived CD184+ endoderm
ES deriv	E012	HUES64 derived CD56+ ectoderm
ES deriv	E013	HUES64 derived CD56+ mesoderm
IMR90	E017	IMR90 fetal lung fibroblasts
iPSC	E021	iPS DF 6.9 cells
iPSC	E022	iPS DF 19.11 cells
Neurosphere	E053	Cortex derived neurospheres
Neurosphere	E054	Ganglion eminence derived neurospheres
Epithelial	E058	Foreskin keratinocyte

Supplementary Figure D1: **Comparison of Epigenome Roadmap samples with isolated cell-types from cord tissue (CT), cord blood (CB) and non-isolated infant CT/CB tissues.** Heatmap (number and color) represents spearman correlation between CT/CB samples and 34 Epigenome Roadmap samples profiled using whole genome bisulfite sequencing (WGBS). Each column represents each of the 9 isolated CT/CB cell-type populations or 4 non-isolated infant CT/CB tissues. Each row represents each sample from the Epigenome Roadmap project. Spearman correlation is represented on 0-100 scale, and color in heatmap changes from grey to blue as correlation increases. Dendrogram shows hierarchical clustering of Epigenome Roadmap samples. Color in dendrogram represents germinal origins of samples from the Epigenome Roadmap project. Primary tissues/cells of mesodermic (MSC-derived), mesodermic (HSC-derived), endodermic and ectodermic germinal origins are represented in light purple, light turquoise, light yellow and grey, respectively. Embryonic stem (ES) cell derived and primary culture samples are represented in light red.



Supplementary Table D3: Comparison of inter-individual variation in cell-types/tissues from cord tissue and cord blood - cell-types/tissues from cord tissue were more variable than those from cord blood. For each isolated cell-type or non-isolated tissue, the inter-quartile range (IQR) in % methylation was computed using 10 samples which had DNA methylation profiled across all 13 cell-types/tissues. First column gives the % methylation IQR value and the remainder columns give the proportions of CpGs with % methylation IQR exceeding the value in the first column for each cell-type/tissue.

%meth IQR	*CT:USCS	*CT:Frozen	C:T:CD31+Endothelial	C:T:CD90+Stromal	C:T:CD324+Epithelial	*CB:Leukocyte	*CB:BuFFyCoat	CB:Mono	CB:Gran	CB:Bcell	CB:NK	CB:CD8T	CB:CD4T
0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1	0.72	0.76	0.69	0.70	0.69	0.64	0.69	0.56	0.60	0.55	0.59	0.56	0.58
2	0.50	0.53	0.45	0.44	0.51	0.36	0.41	0.29	0.31	0.28	0.31	0.29	0.30
3	0.34	0.36	0.31	0.28	0.38	0.22	0.25	0.17	0.18	0.17	0.19	0.18	0.18
4	0.23	0.23	0.22	0.18	0.30	0.14	0.16	0.11	0.11	0.11	0.12	0.11	0.11
5	0.16	0.14	0.16	0.11	0.094	0.094	0.11	0.068	0.069	0.071	0.075	0.072	0.073
6	0.10	0.091	0.11	0.075	0.19	0.063	0.074	0.046	0.046	0.048	0.049	0.048	0.048
7	0.066	0.059	0.080	0.050	0.15	0.043	0.051	0.032	0.032	0.033	0.033	0.032	0.033
8	0.042	0.038	0.058	0.033	0.12	0.029	0.036	0.023	0.023	0.024	0.023	0.023	0.023
9	0.027	0.026	0.042	0.023	0.097	0.020	0.025	0.017	0.017	0.017	0.017	0.017	0.017
10	0.017	0.017	0.031	0.016	0.078	0.015	0.018	0.013	0.013	0.013	0.013	0.012	0.012

Supplementary Table E1: **CpGs variably methylated across isolated cell-types in cord tissue (CT) were enriched for intercellular matrix pathways.** F-tests were used to identify cell-type specific CpGs in CT. A KEGG enrichment analysis was performed for the top 1000 cell-type specific CpGs (smallest p-values), using an adapted G0seq procedure implemented in the missMethyl R package, that accounts for the non-random selection/number of probes of each gene represented on the array. We report pathways at a false discovery rate of 0.01 and with number of genes (N) in the KEGG pathway ≤ 150 .

Identifier	Pathway	N	DE	P.DE	FDR
hsa04933	AGE-RAGE signaling pathway in diabetic complications	96	11	2.7e-09	2.1e-07
hsa00532	Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	19	6	2.7e-08	1.2e-06
hsa04072	Phospholipase D signaling pathway	140	12	9.9e-08	3.4e-06
hsa04270	Vascular smooth muscle contraction	118	10	2.2e-07	6.2e-06
hsa04012	ErbB signaling pathway	84	9	7.2e-07	1.5e-05
hsa04611	Platelet activation	121	9	3.5e-06	5.5e-05
hsa05146	Amoebiasis	94	8	4.2e-06	6.2e-05
hsa04066	HIF-1 signaling pathway	98	8	6.1e-06	8.3e-05
hsa01521	EGFR tyrosine kinase inhibitor resistance	79	8	6.8e-06	8.7e-05
hsa04666	Fc gamma R-mediated phagocytosis	91	8	7.1e-06	8.7e-05
hsa05323	Rheumatoid arthritis	86	6	7.2e-06	8.7e-05
hsa04750	Inflammatory mediator regulation of TRP channels	98	8	1.5e-05	1.7e-04
hsa04971	Gastric acid secretion	74	7	1.6e-05	1.7e-04
hsa05100	Bacterial invasion of epithelial cells	76	7	2.1e-05	2.2e-04
hsa04650	Natural killer cell mediated cytotoxicity	120	7	2.9e-05	3.0e-04
hsa05162	Measles	123	7	3.2e-05	3.2e-04
hsa04142	Lysosome	118	7	3.5e-05	3.3e-04
hsa05211	Renal cell carcinoma	64	6	4.8e-05	4.3e-04
hsa00240	Pyrimidine metabolism	100	6	5.7e-05	4.9e-04
hsa04725	Cholinergic synapse	110	8	5.9e-05	5.0e-04
hsa05231	Choline metabolism in cancer	98	7	9.1e-05	7.4e-04
hsa04068	FoxO signaling pathway	130	7	1.4e-04	1.1e-03
hsa04261	Adrenergic signaling in cardiomyocytes	144	8	1.6e-04	1.3e-03
hsa04722	Neurotrophin signaling pathway	116	7	1.9e-04	1.4e-03
hsa05120	Epithelial cell signaling in Helicobacter pylori infection	66	5	2.0e-04	1.5e-03
hsa04071	Sphingolipid signaling pathway	119	7	2.1e-04	1.5e-03
hsa05161	Hepatitis B	137	7	2.4e-04	1.6e-03
hsa04520	Adherens junction	71	6	2.9e-04	1.9e-03
hsa04911	Insulin secretion	82	6	3.5e-04	2.2e-03
hsa04912	GnRH signaling pathway	90	6	3.7e-04	2.2e-03
hsa04370	VEGF signaling pathway	61	5	4.3e-04	2.5e-03
hsa04310	Wnt signaling pathway	139	7	6.1e-04	3.5e-03
hsa05321	Inflammatory bowel disease (IBD)	61	4	6.3e-04	3.5e-03
hsa04150	mTOR signaling pathway	150	7	6.7e-04	3.7e-03
hsa00760	Nicotinate and nicotinamide metabolism	29	3	8.5e-04	4.5e-03
hsa05110	Vibrio cholerae infection	50	4	8.6e-04	4.5e-03
hsa05014	Amyotrophic lateral sclerosis (ALS)	51	4	1.0e-03	5.3e-03
hsa04970	Salivary secretion	84	5	1.0e-03	5.3e-03
hsa05214	Glioma	65	5	1.1e-03	5.5e-03
hsa00480	Glutathione metabolism	52	3	1.2e-03	5.9e-03
hsa05220	Chronic myeloid leukemia	71	5	1.3e-03	6.1e-03
hsa04330	Notch signaling pathway	48	4	1.5e-03	6.9e-03
hsa04961	Endocrine and other factor-regulated calcium reabsorption	46	4	1.5e-03	6.9e-03
hsa05142	Chagas disease (American trypanosomiasis)	102	5	1.7e-03	7.6e-03
hsa04550	Signaling pathways regulating pluripotency of stem cells	139	6	1.9e-03	8.3e-03
hsa04668	TNF signaling pathway	108	5	1.9e-03	8.3e-03
hsa04919	Thyroid hormone signaling pathway	114	6	1.9e-03	8.4e-03
hsa04540	Gap junction	87	5	2.0e-03	8.4e-03

Supplementary Table E2: CpGs variably methylated across isolated cell-types in cord blood (CB) were enriched for immune-related pathways. F-tests were used to identify cell-type specific CpGs in CB. A KEGG enrichment analysis was performed for the top 1000 cell-type specific CpGs (smallest p-values), using an adapted G0seq procedure implemented in the missMethyl R package, that accounts for the non-random selection/number of probes of each gene represented on the array. We report pathways at a false discovery rate of 0.01 and with number of genes (N) in the KEGG pathway ≤ 150 .

Identifier	Pathway	N	DE	P.DE	FDR
hsa04662	B cell receptor signaling pathway	70	14	1.5e-13	2.1e-11
hsa04380	Osteoclast differentiation	125	16	2.0e-13	2.1e-11
hsa04660	T cell receptor signaling pathway	101	14	1.7e-11	1.4e-09
hsa04666	Fc gamma R-mediated phagocytosis	91	11	3.5e-08	1.5e-06
hsa05220	Chronic myeloid leukemia	71	10	5.3e-08	2.1e-06
hsa05221	Acute myeloid leukemia	54	9	7.4e-08	2.3e-06
hsa04650	Natural killer cell mediated cytotoxicity	120	10	1.2e-07	3.0e-06
hsa04664	Fc epsilon RI signaling pathway	68	9	1.4e-07	3.0e-06
hsa04659	Th17 cell differentiation	102	10	1.5e-07	3.0e-06
hsa05212	Pancreatic cancer	63	9	2.1e-07	3.8e-06
hsa04071	Sphingolipid signaling pathway	119	11	3.2e-07	5.5e-06
hsa04210	Apoptosis	136	10	1.1e-06	1.6e-05
hsa04915	Estrogen signaling pathway	100	10	1.3e-06	1.7e-05
hsa05162	Measles	123	9	1.9e-06	2.4e-05
hsa04722	Neurotrophin signaling pathway	116	10	2.4e-06	2.8e-05
hsa04918	Thyroid hormone synthesis	72	8	2.5e-06	2.8e-05
hsa05215	Prostate cancer	86	9	2.7e-06	2.9e-05
hsa04145	Phagosome	144	9	3.0e-06	3.1e-05
hsa04066	HIF-1 signaling pathway	98	9	3.5e-06	3.5e-05
hsa04630	Jak-STAT signaling pathway	141	9	3.9e-06	3.8e-05
hsa04725	Cholinergic synapse	110	10	4.9e-06	4.6e-05
hsa04658	Th1 and Th2 cell differentiation	88	8	5.7e-06	5.3e-05
hsa05321	Inflammatory bowel disease (IBD)	61	6	1.0e-05	9.1e-05
hsa03460	Fanconi anemia pathway	49	6	1.2e-05	1.0e-04
hsa04068	FoxO signaling pathway	130	9	1.3e-05	1.1e-04
hsa04150	mTOR signaling pathway	150	10	1.5e-05	1.2e-04
hsa05211	Renal cell carcinoma	64	7	1.7e-05	1.4e-04
hsa05161	Hepatitis B	137	9	2.3e-05	1.8e-04
hsa04064	NF-kappa B signaling pathway	88	7	2.6e-05	2.0e-04
hsa04910	Insulin signaling pathway	134	9	3.0e-05	2.2e-04
hsa04070	Phosphatidylinositol signaling system	94	8	3.4e-05	2.4e-04
hsa04072	Phospholipase D signaling pathway	140	10	3.4e-05	2.4e-04
hsa03440	Homologous recombination	39	5	4.1e-05	2.9e-04
hsa04724	Glutamatergic synapse	113	9	4.4e-05	2.9e-04
hsa04320	Dorso-ventral axis formation	28	5	5.7e-05	3.7e-04
hsa05216	Thyroid cancer	29	5	6.1e-05	3.8e-04
hsa04750	Inflammatory mediator regulation of TRP channels	98	8	6.2e-05	3.8e-04
hsa05224	Breast cancer	143	9	8.0e-05	4.7e-04
hsa05142	Chagas disease (American trypanosomiasis)	102	7	9.6e-05	5.5e-04
hsa04540	Gap junction	87	7	1.0e-04	5.9e-04
hsa05146	Amoebiasis	94	7	1.3e-04	7.4e-04
hsa03013	RNA transport	149	7	1.6e-04	8.3e-04
hsa04912	GnRH signaling pathway	90	7	1.6e-04	8.3e-04
hsa04933	AGE-RAGE signaling pathway in diabetic complications	96	7	1.8e-04	8.7e-04
hsa04925	Aldosterone synthesis and secretion	81	7	1.8e-04	8.7e-04
hsa01521	EGFR tyrosine kinase inhibitor resistance	79	7	2.1e-04	9.9e-04
hsa04012	ErbB signaling pathway	84	7	2.2e-04	1.1e-03
hsa04720	Long-term potentiation	64	6	2.3e-04	1.1e-03
hsa05145	Toxoplasmosis	112	7	2.5e-04	1.1e-03
hsa05160	Hepatitis C	125	7	2.5e-04	1.1e-03
hsa04917	Prolactin signaling pathway	72	6	2.5e-04	1.1e-03
hsa05223	Non-small cell lung cancer	55	6	2.5e-04	1.1e-03
hsa00051	Fructose and mannose metabolism	32	4	2.7e-04	1.2e-03
hsa05231	Choline metabolism in cancer	98	7	3.5e-04	1.5e-03
hsa04310	Wnt signaling pathway	139	8	3.9e-04	1.6e-03
hsa01522	Endocrine resistance	96	7	4.1e-04	1.7e-03
hsa04146	Peroxisome	81	5	4.2e-04	1.7e-03
hsa00061	Fatty acid biosynthesis	12	3	4.5e-04	1.8e-03
hsa04930	Type II diabetes mellitus	47	5	5.5e-04	2.1e-03
hsa05120	Epithelial cell signaling in Helicobacter pylori infection	66	5	6.1e-04	2.3e-03
hsa04914	Progesterone-mediated oocyte maturation	85	6	6.3e-04	2.4e-03
hsa04922	Glucagon signaling pathway	96	6	6.9e-04	2.5e-03
hsa04924	Renin secretion	64	5	7.7e-04	2.8e-03
hsa04611	Platelet activation	121	7	7.8e-04	2.8e-03
hsa04668	TNF signaling pathway	108	6	8.4e-04	3.0e-03
hsa04973	Carbohydrate digestion and absorption	41	4	9.2e-04	3.3e-03
hsa04744	Phototransduction	27	3	9.6e-04	3.3e-03
hsa00600	Sphingolipid metabolism	46	4	9.8e-04	3.3e-03
hsa04114	Oocyte meiosis	111	6	1.1e-03	3.6e-03
hsa03430	Mismatch repair	22	3	1.1e-03	3.6e-03
hsa04620	Toll-like receptor signaling pathway	97	5	1.1e-03	3.6e-03
hsa04919	Thyroid hormone signaling pathway	114	7	1.1e-03	3.7e-03
hsa04920	Adipocytokine signaling pathway	67	5	1.2e-03	3.8e-03
hsa05213	Endometrial cancer	50	5	1.3e-03	4.2e-03
hsa00514	Other types of O-glycan biosynthesis	21	3	1.5e-03	4.8e-03
hsa04110	Cell cycle	121	6	1.5e-03	4.8e-03
hsa00562	Inositol phosphate metabolism	67	5	1.6e-03	4.8e-03
hsa05210	Colorectal cancer	61	5	1.6e-03	4.8e-03
hsa04211	Longevity regulating pathway	93	6	1.6e-03	4.9e-03
hsa05219	Bladder cancer	40	4	1.7e-03	4.9e-03
hsa00532	Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	19	3	1.7e-03	4.9e-03
hsa04931	Insulin resistance	106	6	1.7e-03	5.0e-03
hsa05214	Glioma	65	5	2.7e-03	7.6e-03
hsa04270	Vascular smooth muscle contraction	118	6	2.8e-03	8.1e-03
hsa04261	Adrenergic signaling in cardiomyocytes	144	7	2.9e-03	8.1e-03
hsa04972	Pancreatic secretion	92	5	2.9e-03	8.1e-03
hsa03320	PPAR signaling pathway	70	4	3.0e-03	8.3e-03
hsa04330	Notch signaling pathway	48	4	3.3e-03	9.1e-03
hsa01523	Antifolate resistance	30	3	3.4e-03	9.1e-03
hsa04640	Hematopoietic cell lineage	90	4	3.6e-03	9.6e-03
hsa05100	Bacterial invasion of epithelial cells	76	5	3.7e-03	9.8e-03