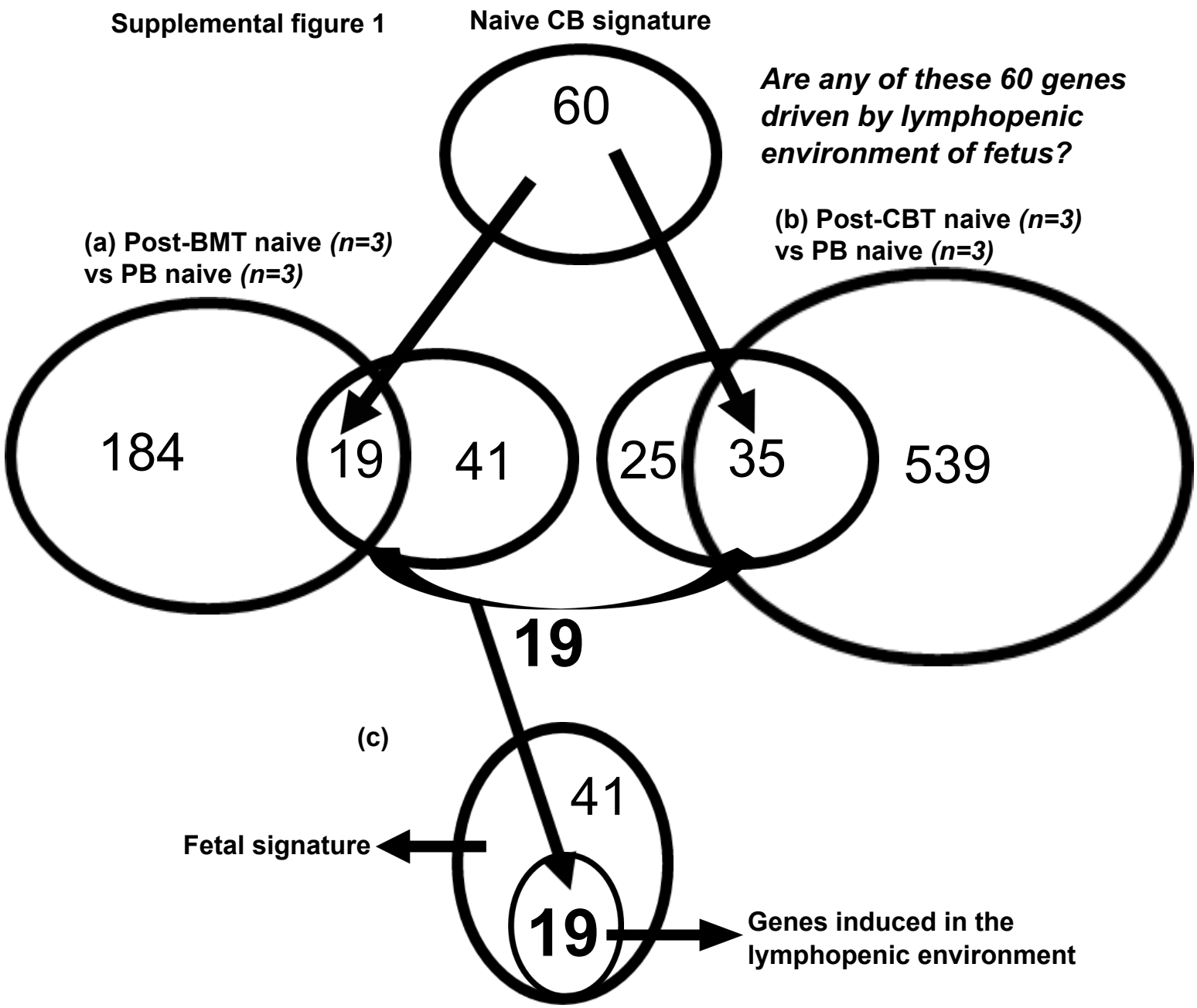
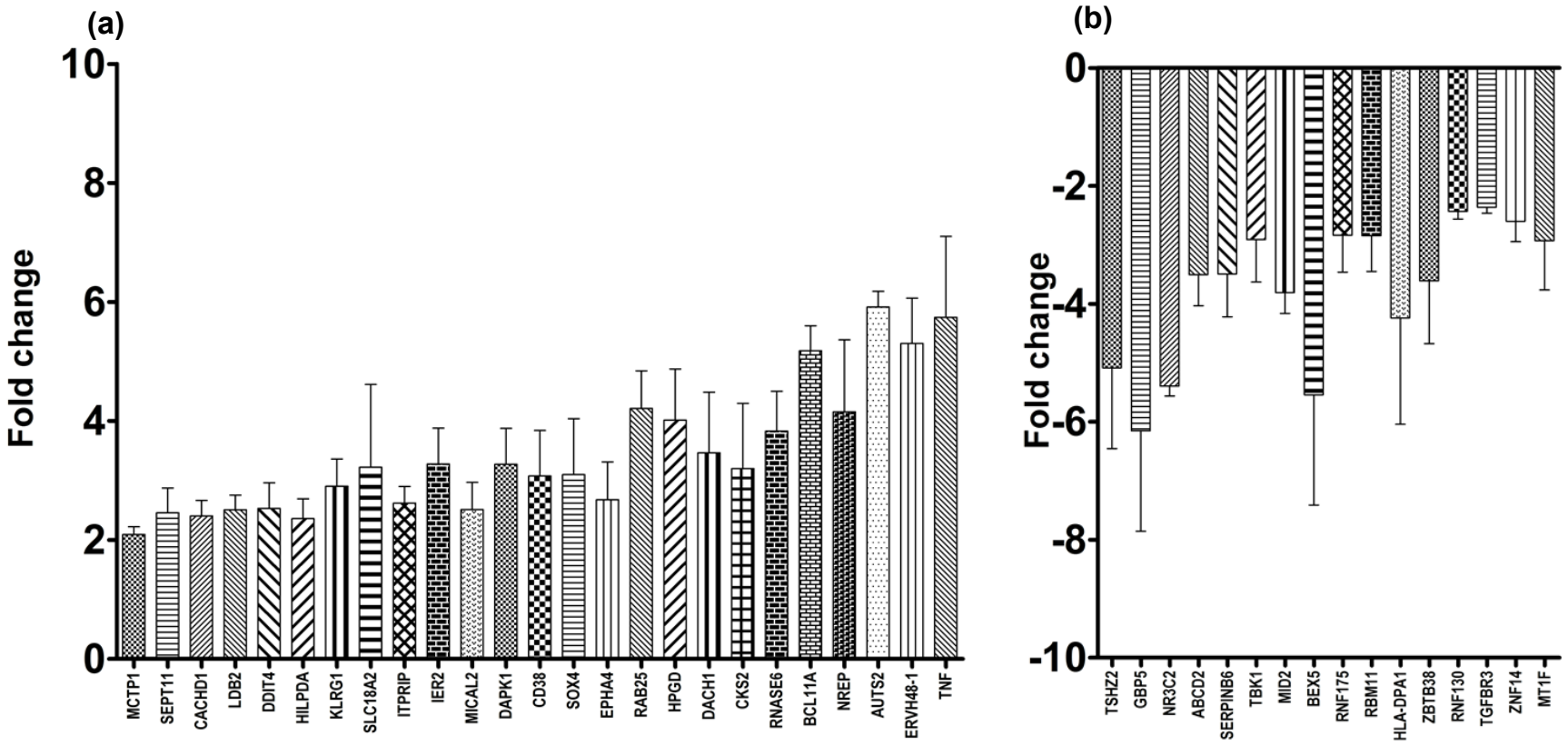


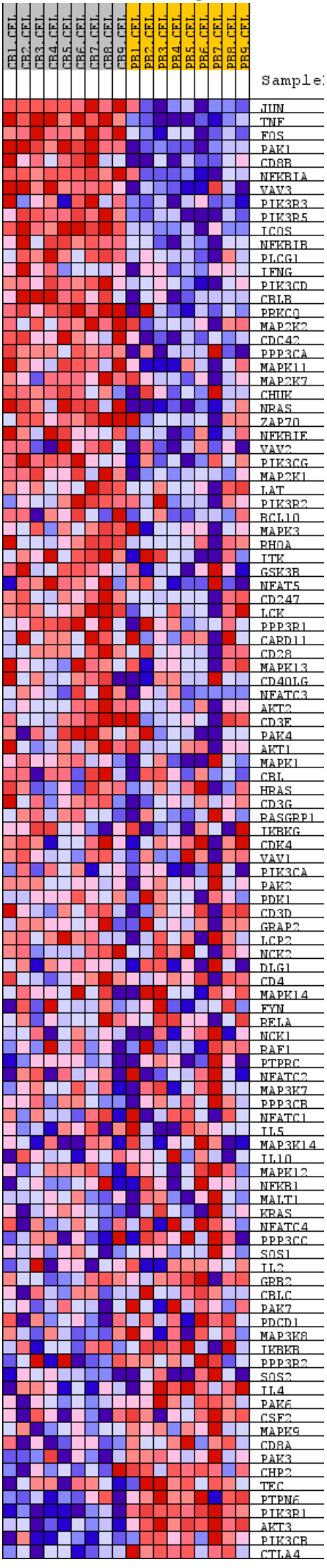
Supplemental figure 1 (a) Nineteen of the 60 genes that represent naive CB CD4+ T cell signature were induced in the reconstituting naive CD4+ T cells after BMT and (b) all the 19 genes remained differentially expressed in the reconstituing naive CD4+ T cells after CBT. (c) Thus, these 19 genes represent the signature induced in the lymphopenic environment and remaining 41 genes are likely to represent fetal signature.

Supplemental figure 2



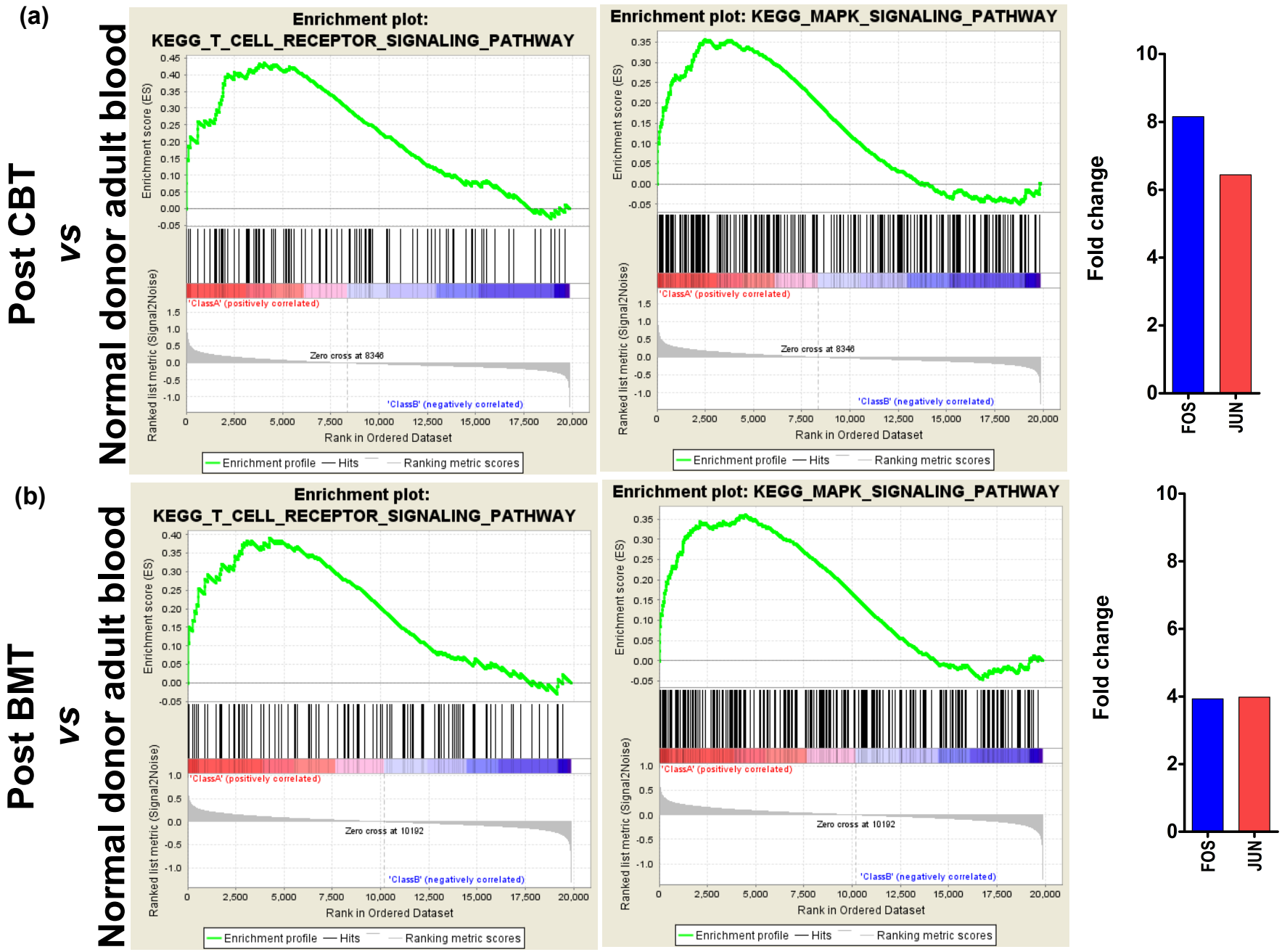
Supplemental figure 2 (a) and (b) Mean and standard deviation of upregulated and down regulated genes representing fetal signature.

Supplemental figure 3



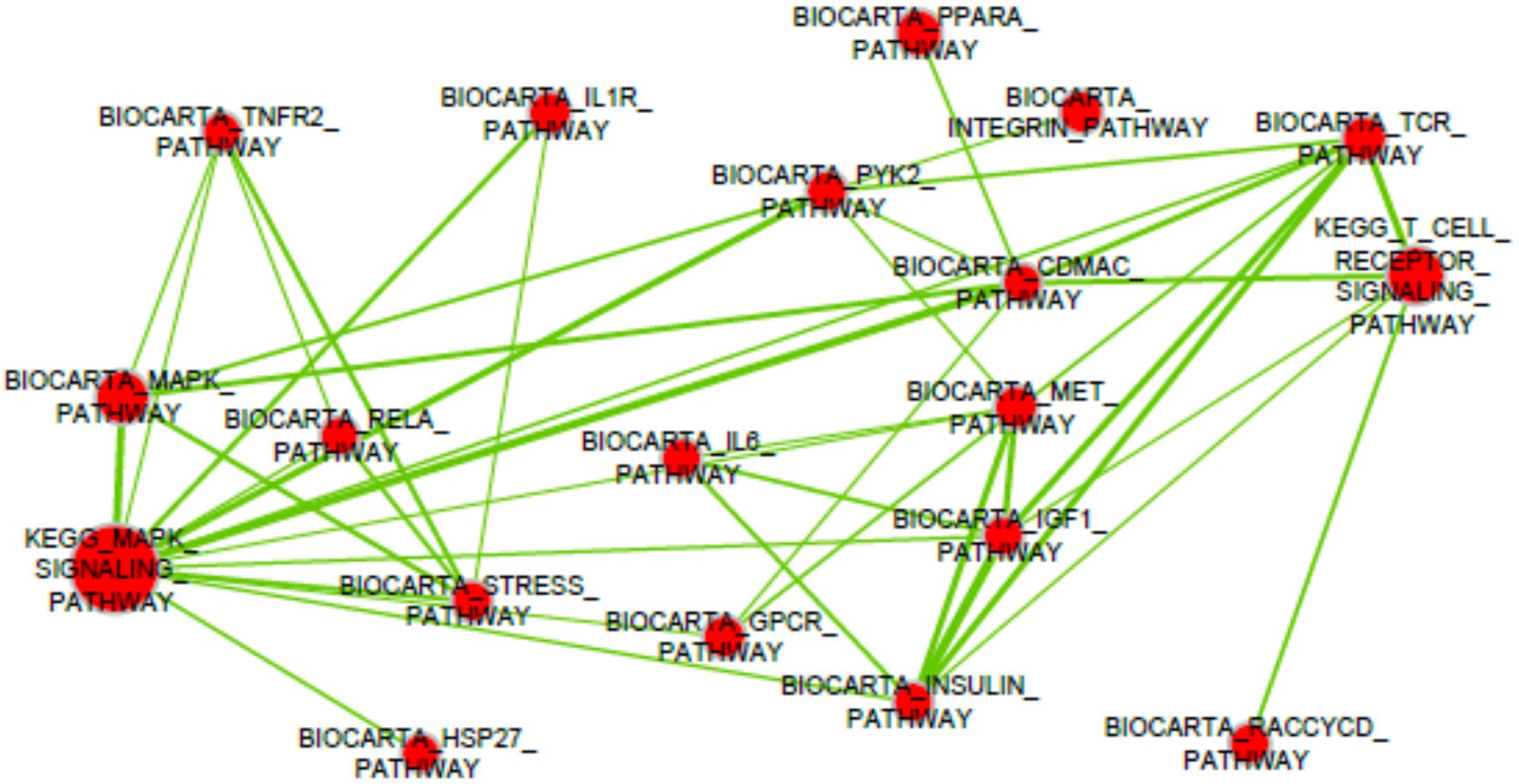
Supplemental figure 3 is a Blue-Pink O'gram of TCR signalling pathway derived after comparing transcription profiles of naive CD4+ T cells from the cord blood (*n*=*9*) and peripheral blood (*n*=*9*).

Supplemental figure 4



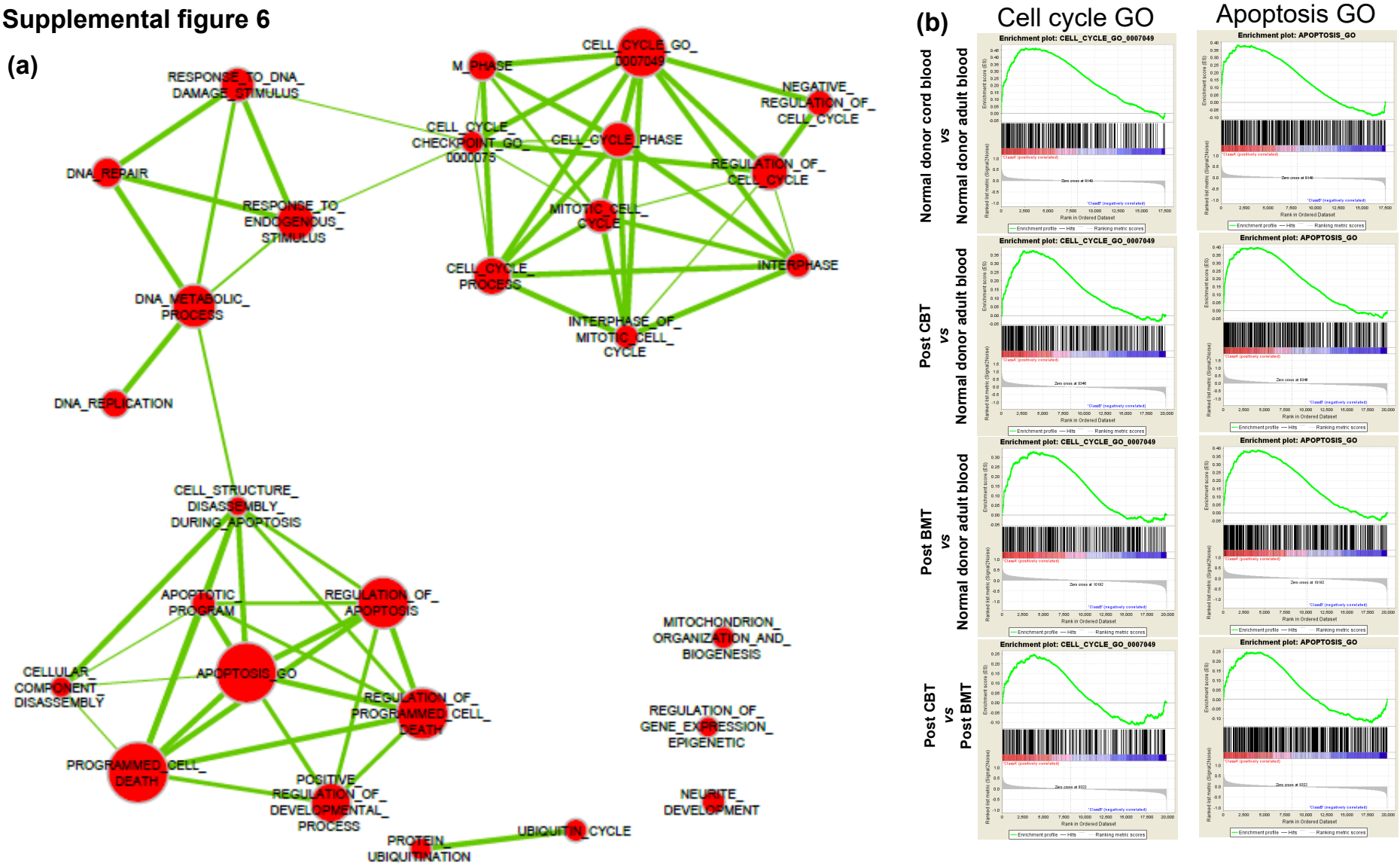
Supplemental figure 4 (a) and (b) Enrichment plots of TCR and MAPK signaling and the transcript values of two important transcription factors FOS and JUN (AP-1 complex) in the naive CD4+ T cells from post-transplant lymphopenic conditions i.e. during early T-cell reconsitution following CBT and BMT. FOS and JUN upregulation is expressed as mean.

Supplemental figure 5



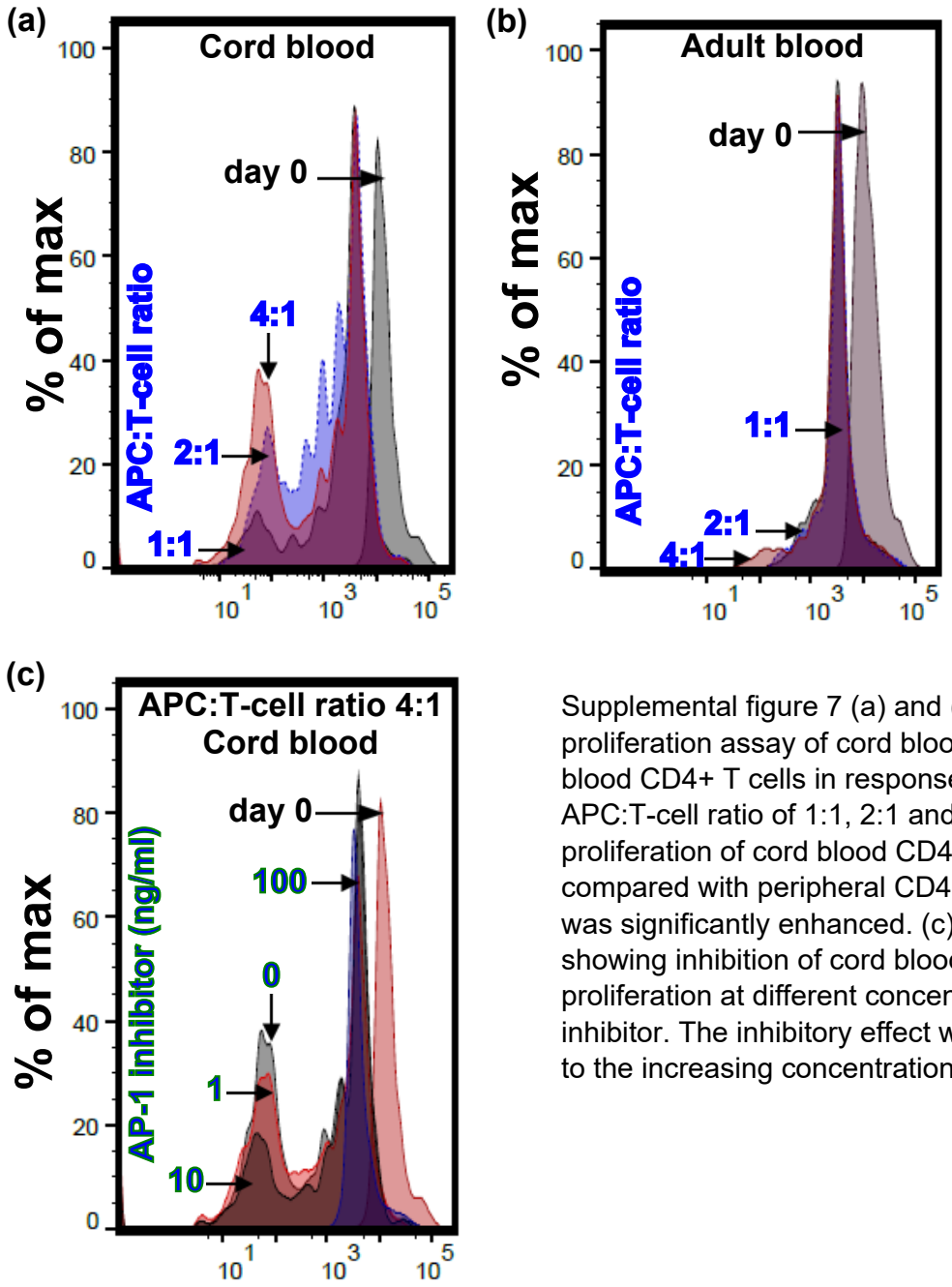
Supplemental figure 5 Enrichment map of upregulated pathways in the naive cord blood CD4+ T cells (n=9) compared with naive peripheral blood CD4+ T cells (n=9). The relationship of TCR and MAPK signalling with other upregulated pathways is shown.

Supplemental figure 6



Supplemental figure 6 (a) Enrichment map of biological process upregulated in the naive cord blood CD4+ T cells compared with naive peripheral blood CD4+ T cells. Cell cycle and apoptosis were the two dominant biological processes upregulated in the naive cord blood CD4+ T cells and (b) these biological processes were upregulated in all the lymphopenic conditions and were highly upregulated in naive CD4+ T cells after T-replete CBT compared with T-replete BMT.

Supplemental figure 7



Supplemental figure 7 (a) and (b) CFSE proliferation assay of cord blood and peripheral blood CD4+ T cells in response to increasing APC:T-cell ratio of 1:1, 2:1 and 4:1. The proliferation of cord blood CD4+ T cells (*n=3*) compared with peripheral CD4+ T cells (*n=3*) was significantly enhanced. (c) CFSE assay showing inhibition of cord blood CD4+ T-cell proliferation at different concentrations of AP-1 inhibitor. The inhibitory effect was proprotional to the increasing concentration of AP-1 inhibitor.

	Age at the time of transplant	HLA matching	Immunosuppression at the time of naïve CD4+ T-cell collection (2 months post-transplant)	CD4+ T-cell count on day 0 (per µL)	CD4+ T-cell count at the time of cell sorting (per µL)
BMT 1	9 years	10/10	Ciclosporin	0	180
BMT 2	1 year	10/10	Ciclosporin	0	390
BMT 3	5 years	10/10	Ciclosporin	0	440
CBT 1	1 year	8/10	Ciclosporin	0	320
CBT 2	4 months	7/10	Ciclosporin	0	600
CBT 3	9 months	10/10	Ciclosporin	0	210

Supplemental table 1 shows age, HLA matching, immunosuppression and CD4+ T-cell counts (day 0 and at the time-point when naïve CD4+ T cells were sorted) of bone marrow and cord blood transplant recipients that contributed to the microarray samples. (CBT=cord blood transplant, BMT=bone marrow transplant)

	Gene Name	Alternative name/function
High in cord blood T cells	FOS	Fos proto-oncogene, AP-1 transcription factor subunit(FOS)
	JUN	Jun proto-oncogene, AP-1 transcription factor subunit(JUN)
	SLC18A2	solute carrier family 18 member A2
	TNFAIP3	TNF alpha induced protein 3
	AUTS2	autism susceptibility candidate 2
	RGS1	regulator of G-protein signaling 1
	GZMA	granzyme A
	NFKBIA	NFKB inhibitor alpha(NFKBIA)
	IER2	immediate early response 2(IER2)
	TNF	tumor necrosis factor
High in adult T cells	Gene Name	Alternative name/function
	GBP5	guanylate binding protein 5(GBP5)
	TSHZ2	teashirt zinc finger homeobox 2
	ZNF204P	zinc finger protein 204, pseudogene
	SERPINB6	serpin family B member 6
	NR3C2	nuclear receptor subfamily 3 group C member 2

Supplemental table 2 List of selected genes found overexpressed in cord blood and adult blood T cells

High in T cells from lymphopenic environment	Gene Name	Alternative name/function
	FOS	Fos proto-oncogene, AP-1 transcription factor subunit(FOS)
	JUN	Jun proto-oncogene, AP-1 transcription factor subunit(JUN)
	TNFAIP3	TNF alpha induced protein 3
	RGS1	regulator of G-protein signaling 1
	GZMA	granzyme A
	NFKBIA	NFKB inhibitor alpha(NFKBIA)
	PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1
	YARS	tyrosyl-tRNA synthetase
	CD69	CD69 molecule
	PDE6G	phosphodiesterase 6G
	PPP1R15A	protein phosphatase 1 regulatory subunit 15A
	YBX3	Y-box binding protein 3
	KLF6	Kruppel like factor 6
	ZNF462	zinc finger protein 462
	DACT1	dishevelled binding antagonist of beta catenin 1
	DUSP1	dual specificity phosphatase 1
	SGK1	serum/glucocorticoid regulated kinase 1
	SERPINF1	serpin family F member 1

Supplemental table 3 List of selected genes found overexpressed in the naïve CD4+ T cells from lymphopenic environment

NAME	SIZE	ES	NES	NOM p-val	FDR q-val
DNA_METABOLIC_PROCESS	247	0.5059515	2.1789281	0	0.002307
RESPONSE_TO_DNA_DAMAGE_STIMULUS	155	0.52454346	2.1087008	0	0.005804
CELL_CYCLE_CHECKPOINT_GO_0000075	48	0.6222186	2.054601	0	0.009687
INTERPHASE_OF_MITOTIC_CELL_CYCLE	62	0.5770853	2.0149944	0	0.01112
CELL_CYCLE_GO_0007049	309	0.4655104	2.0182862	0	0.013031
REGULATION_OF_CELL_CYCLE	180	0.47839648	1.9598776	0	0.018691
MITOTIC_CELL_CYCLE	151	0.4827955	1.9438541	0	0.019842
DNA_REPAIR	119	0.51513183	1.9633648	0	0.021051
INTERPHASE	68	0.54464364	1.9261076	0	0.022417
CELL_CYCLE_PHASE	166	0.47129884	1.9061838	0	0.026747
MITOCHONDRION_ORGANIZATION_AND_BIOGENESIS	47	0.5757308	1.8748549	0	0.028571
CELL_CYCLE_PROCESS	189	0.45616704	1.8823042	0	0.03004
NEGATIVE_REGULATION_OF_CELL_CYCLE	79	0.51832753	1.8762827	0	0.030403
RESPONSE_TO_ENDOGENOUS_STIMULUS	191	0.4572464	1.885719	0	0.031494
CELLULAR_COMPONENT_DISASSEMBLY	33	0.5860449	1.8021057	0	0.052743
APOPTOTIC_PROGRAM	59	0.5236585	1.8059676	0.001776199	0.054009
REGULATION_OF_GENE_EXPRESSION_EPIGENETIC	28	0.59163123	1.7378746	0.001956947	0.081819
UBIQUITIN_CYCLE	47	0.5175936	1.7416282	0.001855288	0.08269
APOPTOSIS_GO	415	0.38288313	1.6989125	0	0.086339
PROGRAMMED_CELL_DEATH	416	0.38030535	1.7039883	0	0.088013
DNA_REPLICATION	97	0.45436049	1.7076312	0.001680672	0.088296
REGULATION_OF_PROGRAMMED_CELL_DEATH	328	0.38511395	1.6878171	0	0.089241
REGULATION_OF_APOPTOSIS	327	0.38888615	1.7103146	0	0.089732
PROTEIN_UBIQUITINATION	39	0.5401635	1.7232333	0.003539823	0.090196
M_PHASE	110	0.44338146	1.7104169	0.001739131	0.093471
POSITIVE_REGULATION_OF_DEVELOPMENTAL_PROCESS	210	0.4172443	1.7120072	0	0.09613
NEURITE_DEVELOPMENT	50	0.5046813	1.6658067	0	0.096863
CELL_STRUCTURE_DISASSEMBLY_DURING_APOPTOSIS	18	0.6780949	1.7957022	0.003787879	0.052441

Supplemental table 4 shows significant biological processes in naïve CB CD4+ T cells after enrichment mapping (at p value < 0.005; fdr value < 0.1) of GSEA pathways

Supplemental methods

Preparation of RNA for microarray analysis

CD4+CD45RA+CCR7+ naïve T cells were sorted from blood samples of normal donor cord blood and peripheral blood and transplant recipients. Naïve CD4+ T cells from the transplant recipients without graft-versus-host disease and with 100% donor T-cell chimerism were sorted two months after transplant. The sorted naïve CD4+ T cells were checked for high purity (>98%), immediately resuspended in RNA lysis buffer (RNAeasy mini-kit) and stored at -80°C. RNA was isolated according to the manufacturer's protocol (RNAeasy mini-kit) and yield was determined on a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE). The samples that had an appropriate yield (typically >10 nanograms total RNA) were subsequently analyzed for RNA integrity using an Agilent Bioanalyzer (Agilent 2100, Agilent Technologies, Santa Clara, CA). RNA of high quality was converted to cDNA and amplified (NuGEN WT-Ovation, NuGEN, San Carlos, CA), purified to remove residual RNA (QIAquick PCR purification kit, Qiagen, Valencia, CA), fragmented (NuGEN WT-Ovation Kit), and labelled (FL-Ovation cDNA biotin module, NuGEN) for microarray analysis. The fragmented and labelled cDNA was hybridised to microarray chips (Affymerix HuGene ST 1.0 arrays). The hybridization of samples and data acquisition was performed by the Department of Genomics at University College London.

Gene set enrichment analysis

The default metric of signal to noise ratio was used to rank the differentially expressed genes. The three datasets were pooled and combatted to identify the differentially expressed gene pathways in naïve CD4+ T cells from normal donor cord blood *vs* peripheral blood. When GEPs from the same experiment were compared, Combat function was not used. Thus, for identifying differentially expressed gene pathways in reconstituting naive CD4+ T cells

following CBT and BMT uncombatted GEP's were compared with naive CD4+ T cells from volunteer adult donors. Similarly, uncombatted GEP's of reconstituting naive CD4+ T cells following CBT were compared with those following BMT to identify the pathways that may endow enhanced reconstituting ability to CB CD4+ T cells.

The Biocarta and KEGG gene set pathways were used to identify the differentially expressed canonical pathways, and biological gene sets were used to identify differentially expressed biological processes. Cytoscape software was used to perform enrichment mapping of the pathways upregulated at p values < 0.005 and false discovery rate values < 0.1.

Microarray data analysis – Quality control and statistical analysis

Assessment of hybridization quality was performed using the Bioconductor package affyPLM.[1] The preprocessing of the Affymetrix dataset adhered to the following procedures: background correction, normalization using the quantile method and summarization of probe set values using the RMA method (Robust Multi-array Average). The latter process fitted a specified robust linear model to the probe level data.

The probe sets were mapped to genes using the Bioconductor package AnnotationDbi.[2] Where multiple probe sets representing a gene were available, the probe with maximum average value was chosen. Thus, we identified 17,522 genes common to the Affymerix HuGene ST 1.0 and Affymetrix Human Genome U133 Plus 2.0 platform.

The dataset of 17,522 genes was combatted to remove the batch effects using Combat function in R as previously described.[3] Three-dimensional principal component analysis was performed on the combatted datasets using DUDI.PCA function. Hierarchical clustering of gene expression samples on combatted dataset was performed using the ward agglomerative hierarchical method.

The differentially expressed genes were identified by comparing the uncombatted data from each individual experiment. The Bioconductor package limma - a "moderated t test" application was used and unpaired or paired analysis was performed to identify differentially expressed genes.[4] Significant genes were identified using a threshold of $p < 0.05$ and a fold-change ≥ 2. All analysis was performed using R (cran.r-project.org) and Bioconductor (www.bioconductor.org).

Proliferation assay

CD3+ T cells were positively selected from cord blood and peripheral blood mononuclear cells using human CD3+ microbeads from Milteyni (130-050-101). The cells were incubated at 37^{o}C with 10µM carboxy-fluorescein diacetate succinimidyl ester (CFSE; Invitrogen) in X-VIVO 10 for seven minutes. CFSE-labelled CD3+ T cells were washed three times in X-VIVO 10 containing 10% Human AB serum to remove the excess CFSE. The cells were then suspended in X-VIVO 10 media and incubated at the appropriate concentration in 96-well round bottom plates.

We studied the role of TCR signalling in cord blood CD4+ T cell proliferation by co-culturing CFSE-labelled cord blood and peripheral blood T cells with CD3-negative cells from the same donors (self antigen-presenting cell fraction (APC)) at an APC:T cell ratio of 1:1, 2:1 and 4:1 for seven days.

The role of upregulated transcription factor complex AP-1 was tested by inhibiting AP-1 with a small molecule inhibitor SR 11302 from R and D systems. CFSE-labelled cord blood CD3+ T cells were cocultured with CD3-negative cells from the same donor at an APC:T cell ratio of 4:1. The inhibition of proliferation with 1, 10 and 100 ng/ml of AP-1 inhibitor was assessed after seven days.

CD4+ T cells were labelled with BV-605 anti-human CD4 antibody (Biolegend) and a minimum of 1000 CD4+ events were acquired on LSR BD II flow cytometer. These events were analysed using Flowjo software and are shown as histogram in Figure 6.

References

1) Bolstad BM, Collin F, Brettschneider J, Simpson K, Cope L, Irizarry RA, and Speed TP. (2005) Quality Assessment of Affymetrix GeneChip Data in Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Gentleman R, Carey V, Huber W, Irizarry R, and Dudoit S. (Eds.), Springer, New York.
2) Herve Pages, Marc Carlson, Seth Falcon and Nianhua Li (). AnnotationDbi: Annotation Database Interface. R package version 1.16.19.
3) Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8: 118-127.
4) Smyth, G. K. (2005). Limma: linear models for microarray data. In: 'Bioinformatics and Computational Biology Solutions using R and Bioconductor'. R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds), Springer, New York, pages 397--420.