Supplemental Material

TITLE

Redefinition of the human mast cell transcriptome by deep sequencing

AUTHORS

Efthymios Motakis,[1,*] Sven Guhl,[2,*] Yuri Ishizu,[1] Masayoshi Itoh,[1,3] Hideya Kawaji,[1,3] Michiel de Hoon,[1] Timo Lassmann,[1] Piero Carninci,[1] Yoshihide Hayashizaki,[3] Torsten Zuberbier,[2] Alistair R R Forrest,[1¶] Magda Babina[2¶] and the FANTOM consortium[1]

*These authors contributed equally

[1]RIKEN Centre for Life Science Technologies, Division of Genomic Technologies, Yokohama, Kanagawa, 230-0045 Japan

[2]Department of Dermatology and Allergy, Charité Campus Mitte, Berlin, Germany

[3]RIKEN Preventive Medicine and Diagnosis Innovation Program

¶To whom correspondence should be addressed

Short title: The mast cell transcriptome

RUNNING TITLE

The complete transcriptome of human skin mast cells by deep-CAGE

Supplemental METHODS

RNA preparation for HelicosCAGE

RNA was first isolated from the 9 MC samples specified above using the QIAGEN miRNEasy kit (cat#: 217004). To remove contaminating heparin, which inhibits reverse transcriptase, RNA samples were treated with heparinase (Sigma H2519-50UN) using a modification of Gilchrist et al.[1] Briefly, a 50 µl reaction containing 10 µg total RNA, 1 unit heparinase, 40 units RNAsin (Promega N2111), 5 mM Tris pH 7.5, 1 mM CaCl2 was incubated at 22° C for 2 h. The sample was then cleaned using a minelute column (QIAGEN cat#: 74204), with the following modification to keep small RNAs. At step 2, 1.5 volumes of 100% ethanol were used. 5 µg of total RNA was then subjected to the Helicos CAGE protocol.[2]

Bioinformatics analysis

Data Extraction

We used *samtools*[3] to extract the raw data counts for each Transcription Start Site (TSS) from all FANTOM5 *.bam* files (Forrest et al., manuscript submitted January 2013). Approximately, 4 million CAGE tags for each library were aligned to the genome (Hg19). For each TSS, *s*, we estimated a mapping quality score Q_s, $Q_s = -10 \times log_{10} p_s^w$. Term p_s^w denoted the posterior probability that *s* was incorrectly mapped and was estimated by approximation.[4] Only TSSs with $p_s^w \leq 1\%$ were kept for further analysis, corresponding to $Q_s \geq 20$ (Forrest et al., manuscript submitted January 2013). The quality mapped TSSs were grouped into CTSS (clusters of TSS with common start site).

The CTSS data were the summarized counts from all TSS included in the cluster (summation per sample). Then, “robust” and “permissive” Decomposition-based Peaks (DPI) were estimated (Forrest et al., manuscript submitted January 2013). These DPI peaks were annotated based on known transcript 5’-ends within 500 bases. We extracted the CAGE tags for 9 Mast Cell samples from the “robust” DPI peaks. The data consisted of 3 triplicated conditions: ex vivo MCs, cultured MCs (expanded) and cultured MCs (expanded & stimulated).

The raw counts $K_{i,j}$ of each promoter $i, i = 1, \dots, N$ ($N = 184{,}827$), and sample $j, j = 1, \dots, J$ ($J = 9$), were used for pre-processing (normalization and filtering) and subsequently for the main bioinformatics/statistical analysis (differential expression, correlation, multivariate analysis). To obtain the raw count data $K_{g,j}$for each gene and other transcribed regions g in sample *j*, we summarized the promoter counts as $K_{g,j} = \sum_{p=1}^{P} K_{p,j}$, where *P* is gene's/transcribed region's *g* number of promoters. Similar to the above pre-processing and main analysis steps were followed. For simplicity, hereafter we will use the term *DPI* for promoters, *genes* for the RefSeq annotated genes and *regions* for the whole set of genes and other transcribed regions.

Data filtering

We applied a pre-processing, filtering step on the DPI and the region data, separately, in order to keep only expressed DPIs/regions for further analysis. Filtering consisted of viewing the data and identifying an arbitrary cut-off with both biological and mathematical significance: the DPIs (or/and regions) falling below the cut-off are not expressed (CAGE expression counts are very low) and, in this sense, they did not contribute biologically in differential expressions analysis. Removing those DPIs/regions we reduced the dimensionality of the study, estimated dispersions from the expressed candidates and minimized the false positives/negatives. The procedure is part of both the DESeq[5,6] and edgeR[7,8] algorithms that are widely used in RNA-seq differential expression.

The DPIs (and likewise the regions') filtering consisted of the following steps (to simplify the notation we use $\{i, g\}$ subscript to denote the processing of DPI or the region data, respectively): (i) In order to adjust for the unequal sample library sizes, we transformed $K_{\{i,g\},j}$ into tags-per-million (TPM). The TPM-transformed data were obtained as $tmp\left(K_{\{i,g\},j}\right) = \left(K_{\{i,g\},j} \times 1e+6\right) / \sum_{i=1}^{N} K_{\{i,g\},j}$; (ii) we transformed the original counts into $vst\left(K_{\{i,g\},j}\right)$, i.e. the variance stabilized (generalized log transformation) expressions, by the R function *varianceStabilizingTransformation*;[5,6] (iii) we filtered out the DPIs/regions with $tpm(K_{\{i,g\},j}) < t$ in at least J^* samples in all experimental conditions (ex vivo MCs, cultured MCs expanded and

cultured MCs expanded and stimulated). We varied *t* to take values $\{0.5,1,1.5,2\}$ (Supplemental Figure 1A, left); (iv) using $vst(K_{\{i,g\},j})$, we plotted the histograms of DPIs/regions intensities for each sample and associated each *t* level to $vst(K_{\{i,g\},j})$, i.e. we found which VST expression level corresponds to the TPM cut-off *t*, $\{t = 0.5,1,1.5,2\}$. In short, steps (i)-(iv) attempt to find the minimum TPM that removes most of the background noise (the left end of the $vst(K_{\{i,g\},j})$ densities).

In the promoter level, we ended up with *t* = 1 for at least $J^* \geq 2$ corresponding to $vst(K_{\{i,g\},j}) \approx$ 3.2 and leaving for further study 55,355 DPIs. In the region level, we ended up with *t* = 1 for at least $J^* \geq 2$ (Supplemental Figure 1A, left) corresponding to $vst(K_{\{i,g\},j}) \approx$ 3.5 (Supplemental Figure 1A, right) and leaving for further study 33,515 RefSeq genes and other transcribed regions.

Principal Component Analysis and heatmaps

We ran the quality control algorithm of the Bioconductor R package *arrayQualityMetrics* (Bioconductor 2.11) to the filtered $vst(K_{g,j})$ data matrix[5] (see paragraph „Data Filtering“ for $vst(K_{g,j})$ description). Package *arrayQualityMetrics* checked for possible outlier samples through a series of visual, descriptive and statistical tests. Among them are heatmap / bi-clustering, density plots and Principal Components Anaysis (PCA). All heatmaps and PCA modelling of the main text were based on the filtered $vst(K)$ expressions using functions of the *arrayQualityMetrics* R package.

Correlation and multivariate analysis of the FANTOM5 blood cell samples

We estimated the correlation pattern of the 50 FANTOM5 blood cell samples (Forrest et al., manuscript submitted January 2013) by the function *cor.test* in R. We calculated all possible $0.5 \times 50 \times 49$ Pearson correlation coefficients and the respective p-values of the 50 samples from the filtered $vst(K_{g,j})$ data[5] (see paragraph „Data Filtering“ for $vst(K_{g,j})$ description). To

relax the distributional assumption of normality we also estimated the rank-based Kendal correlations.[9]

We ran PCA on the same 50 blood cell samples and selected by scree plot[9] (R function *screeplot*) the first 3 principal components (PC1-3), explaining approximately 82.3% of the total data variance (the cumulative variances explained by the 3 components were: 69%, 77.3% and 82.3%; the addition of the forth component added only 2% more). We plotted PC1 vs PC2 and PC1 vs PC3 (Figure 4A). PC1 separated the samples by the different compartments the cells were derived from (blood, skin, bone marrow) and treatments (cultured vs. *ex vivo*). It placed *ex-vivo* MCs and cultured MCs in different groups (Figure 4A left). PC2 mainly separated lymphoid from myeloid cells. PC3 showed the uniqueness of MCs compared to other blood samples as it clearly separated both ex-vivo and cultured MCs from the rest (Figure 4A right).

Model goodness of fit

We tested which of the competing models, Negative Binomial[10] vs Poisson-Tweedie[11], is more appropriate for analyzing the data. Poisson-Tweedie is a generalization of the Negative Binomial distribution[8], obtained as a mixture of overdispersed Poisson and Tweedie distributions. The respective algorithm of the *tweeDESeq* R package[12] was applied on the filtered DPI and region counts. The null hypothesis that *the data fit the Negative Binomial model* was tested with the likelihood ratio (LR) test[11] and the visually informative quantile-quantile (QQ) plot for model selection (Supplemental Figure 1B, left). The test statistic did not offer enough evidence to reject the null hypothesis (LR p-value = 0.376), which was also reflected in the QQ-plot of the deviance statistics.

Differential expression analysis

We employed two alternative models, DESeq[5] and edgeR[7], suggesting different parameterization and dispersion estimation algorithms, to identify sets of differentially

expressed DPIs and regions (from which we extracted the significant genes) between MCs treatment groups. Using as input the filtered counts of each dataset, we followed the experimental design and compared *ex vivo MC vs cultured MC expanded* and *cultured MC expanded vs cultured MC expanded and stimulated*. The common DESeq/edgeR hits of each comparison (as identified by the Benjamini-Hochberg (BH) adjusted p-values[13] at significance level 1%) were considered the most reliable targets. DESeq dispersions were estimated by Cox-Reid, pooled-CR parameterization. In edgeR we estimated the tagwise dispersions[8], whose fit we QQ-plotted in Supplemental Figure 1B (right).

The estimated log Fold Changes and BH adjusted p-values are given in Supplemental Tables 3a-c and 4a-c. The former set of tables gives the up-, down- and non-differentially regulated genes of *ex vivo MC vs cultured MC expanded*, respectively. Each gene's differentially expressed promoters are also depicted. Similar results for the *cultured MC expanded vs cultured MC expanded and stimulated* comparison are given in the latter set of tables.

Supplemental Table 8a-c shows the comparison between the ex-vivo MCs vs basophils (3 FANTOM5 CAGE samples from independent donors), conducted in the same way as above. The ex-vivo MC samples were the most similar to basophils according to the correlation and PCA analysis (Figure 4A).

Promoter hyperactivity

We estimated for each cell type $C = \{c = 1, \dots, 4; 1 = ex-vivo\ MCs, 2 = cultured\ (expanded)\ MCs, 3 = cultured\ (expanded\ and\ stimulated)\ MCs, 4 = basophils\}$ the ***tpm*** mean of replicates $\mu_{i,\{j \in C_c\}}$ and $\mu_{g^*,\{j \in C_c\}}$ where g^* denotes the genes and $j \in C_c$ the samples of each of the four cell types. Next, we defined the set $C' = \{c = 1, 2; 1 = all\ FANTOM5\ samples\ without\ MCs, 2 = all\ FANTOM5\ samples\ without\ basophils\}$. For each of the MCs types, we estimated the fold changes $FC_{i,C_c} = \mu_{i,\{j \in C_c\}} / \mu_{i,\{j \in C'_1\}}$ and $C_{g^*,C_c} = \mu_{g^*,\{j \in C_c\}} / \mu_{g^*,\{j \in C'_1\}}$. For basophils we estimated $FC_{i,C_4} = \mu_{i,\{j \in C_4\}} / \mu_{i,\{j \in C'_2\}}$ and $FC_{g^*,C_4} = \mu_{g^*,\{j \in C_4\}} / \mu_{g^*,\{j \in C'_2\}}$ in a similar fashion.

These fold changes indicated the promoters and genes, respectively, that were highly expressed in each cell type compared to the rest of the FANTOM5 samples. Using them, we found promoters i that are hyperactive in each MCs type and in basophils compared to their associated gene g^*, i.e. promoters with $FC_{i,C_c}/FC_{g^*,C_c'} > 50$. The analytical results are given in Supplemental Tables 5a-d.

Swapping promoters

An interesting finding of the MCs differential expression analysis is the identification of four differentially expressed and two non-differentially expressed genes having both up-regulated and down-regulated promoters in *ex vivo MC vs cultured MC* comparison. We present the gene names/locations, the promoter names/locations and the differential expression statistics (fold changes and BH P-values) at Supplemental Table 6.

Motif analysis

We run Motif Activity Response Analysis (MARA[14]) to generate the motif activities, estimate differential activity p-values and extract motif importance scores for each C_c (see paragraph "Promoter hyperactivity" for the C_c definition) as in Forrest et al. (manuscript submitted January 2013). Denote by $K_{i,j}$ the expression level of promoter i in sample j, by $S_{i,m}$ the predicted number of functional sites for motif m in promoter i and $A_{m,j}$ the activity of motif m in sample j. Assuming Gaussian noise, we fit the model $K_{i,j} = \sum_m (S_{i,m} - \overline{S_m}) \times A_{m,j} + noise$ where $\overline{S_m}$ is the mean of $S_{.,m}$ across the promoters. The model's likelihood and $A_{m,j}$ parameter of interest are estimated by the Bayesian maximal posterior probability which is determined by the Singular Value Decomposition[14]. From this we calculate the standard error of $A_{m,j}$, $\sigma_{m,j}$, and subsequently the z-statistic $z_{m,j} = A_{m,j}/\sigma_{m,j}$.

To estimate differential activities among the C_c's, we exploit the Gaussian noise assumption and fit two linear weighted regression models: the full model $A_{m,j} = \alpha_m + \beta_m \times Group_j + noise_{m,j}$ and the reduced/nested $A_{m,j} = \alpha_m$.[14] Variable $Group$ takes values 0 or 1 depending

on which of the C_c the $A_{m,j}$ belongs to. For example to estimate differential activity between ex-vivo and cultured (expanded) MCs, *Group* takes 0's for the *j*'s belonging to ex-vivo and 1's otherwise. The weights of the regression are the estimated $\sigma_{m,j}$. We compare the two nested models with the Likelihood ratio test[7] and estimate the associated BH P-values. When the BH P-value is lower than 0.01, we conclude that the mean motif activity in the first C_c (e.g. ex-vivo MCs) are significantly different from the mean motif activity in the second C_c (e.g. cultured (expanded) MCs). These results are obtained for all comparisons of interest: ex-vivo MCs vs cultured (expanded) MCs, cultured (expanded) MCs vs cultured (expanded and stimulated) MCs, ex-vivo MCs vs basophils and each of the C_c's versus the rest of the FANTOM5 samples as before (paragraph "Promoter hyperactivity").

Finally, we calculate an overall importance of the motif by averaging the absolute values of $z_{m,j \in C_c}$.[14] Supplementary Table 7 depicts the differential activity p-values and the ranks of the motifs based on the importances (the most important motif is ranked as 1, the second most important as 2 and so on).

MC treatments

For functional studies involving BMPR1, MCs were treated with BMP2/BMP4 (both from R&D Systems, Wiesbaden, Germany) for the times and at the concentrations given in the Figure 1 legend prior to RNA extraction, stimulation, re-stimulation and histamine release experiments. For the single stimulation experiment, MCs were pre-treated with BMP4 and after 24 h assayed for histamine release. To assess the effect of BMP on MC recovery from refractoriness, MCs were stimulated with AER-37 for 2 h and then washed twice to remove unbound antibody. Cells were re-plated in fresh media with/out the addition of BMP4 (20 ng/ml) and SCF (100 ng/ml). After 48 h, MCs were stimulated by a second round of FcεRI-crosslinking and the released histamine quantified. For survival assessments, MCs were plated at $5x10^4$ cells per 100 µl in 96-well-plates in the presence of the mediators given in the figure.

Reverse transcription-quantitative PCR (RT-qPCR)

RT-qPCR was performed as described.[15,16] Briefly, total RNA was isolated using the RNeasy Total RNA Kit, digested with RNAse free DNAse (Qiagen, Hilden, Germany), and PCR carried out with the LC Fast Start DNA Master SYBR Green kit (Roche Applied Science). Primers for c-kit were 5'-CTATGCTCTCGCACCTTTCC and 5'- CAATGAAGTGCCCCTGAAGT; those for tryptase 5'-CACCTTGGCGCCTACACGGG and 5'-CACCTTGCACACCAGGGGCC. Other primer pairs were as described.[15,17] Values were normalized to the housekeeping gene β-actin.

Flow-cytometry

Flow-cytometric staining and analysis were performed according to established protocols.[17,18] In brief, MCs were blocked for 15 min at 4° C with human AB-serum (Biotest, Dreieich, Germany) and incubated with anti-OX40L-PE (clone Ik-1, Becton Dickinson) or anti-CD137-PE (clone 4B4, eBioscience). For negative control, cells were stained with the corresponding isotype control antibodies (clone MOPC-21, Becton Dickinson and clone eBMG2b, eBioscience).

Histamine release

Quantification of histamine release was performed exactly as described.[16-18] In brief, MCs were re-suspended in PAG-CM (at 1 x 10^5/ml), divided into aliquots and challenged at 20,000 cells/tube for 30 min at 37° C with the anti-IgER AER-37 (0.25 µg/ml) or kept in PAG–CM for spontaneous release. Supernatants were stored at 20° C until measurement. For total histamine content, mast cells were lysed in 1% perchloric acid for 30 min at 37° C centrifuged and the cell free supernatants stored at 20° C until measurement. Quantification of histamine content was performed by an automated fluorescence method, using an autoanalyzer (Borgwald Technik, Germany, Hamburg) referring to a 5-point histamine standard curve. All histamine determinations were performed in triplicate.

IL-31 ELISA

MCs were stimulated with AER-37 for 24 h in serum-free media and the concentration of IL-31 in the supernatants determined by ELISA (R&D Systems, Wiesbaden, Germany).

REFERENCES

1. Gilchrist M, MacDonald AJ, Neverova I, Ritchie B, Befus AD. Optimization of the isolation and effective use of mRNA from rat mast cells. *J Immunol Methods.* 1997;201(2):207-214.

2. Kanamori-Katayama M, Itoh M, Kawaji H, Lassmann T, Katayama S, Kojima M, Bertin N, Kaiho A, Ninomiya N, Daub CO, Carninci P, Forrest AR, Hayashizaki Y. Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome Res.* 2011;21(7):1150-1159.

3. samtools user's guide: Utilities for the Sequence Alignment/Map (SAM) format (http://samtools.sourceforge.net/samtools.shtml).

4. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads, calling variants using mapping quality scores. *Genome Res.* 2008;18(11):1851-1858.

5. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biology.* 2010;11(10):R106.

6. DESeq user's guide: Differential gene expression analysis based on the negative binomial distribution, Bioconductor 2.11; 2012.

7. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics.* 2008;9(2):321-332.

8. edgeR user's guide: Empirical analysis of digital gene expression data in R, Bioconductor 2.11; 2012.

9. Venables WN and Ripley BD. Modern Applied Statistics with S. Fourth Edition. Springer. New York: Academic; 2002.

10. Whitaker L. On the Poisson law of small numbers. *Biometrika.* 1914;10:36-71.

11. El-Shaarawi A, Zhu R, Joe H. Modelling species abundance using the Poisson-Tweedie family. *Environmetrics.* 2011;22:152-164.

12. tweeDEseq user's guide: Empirical analysis of digital gene expression data in R, Bioconductor 2.11; 2012.

13. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical, powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B.* 1995;57:289–300.

14. The FANTOM Consortium and RIKEN Omics Science Center. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature Genetics*. 2009;41(5):553-562

15. Babina M, Schülke Y, Kirchhof L, Guhl S, Franke R, Böhm S et al. The transcription factor profile of human mast cells in comparison with monocytes, granulocytes. *Cell Mol Life Sci.* 2005;62(2):214-226.

16. Babina M, Guhl S, Stärke A, Kirchhof L, Zuberbier T, Henz BM. Comparative cytokine profile of human skin mast cells from two compartments—strong resemblance with monocytes at baseline but induction of IL-5 by IL-4 priming. *J Leukoc Biol.* 2004;75(2):244-252.

17. Guhl S, Babina M, Neou A, Zuberbier T, Artuc M. Mast cell lines HMC-1, LAD2 in comparison with mature human skin mast cells - drastically reduced levels of tryptase, chymase in mast cell lines. *Exp Dermatol.* 2010;19(9):845-847.

18. Guhl S, Lee HH, Babina M, Henz BM, Zuberbier T. Evidence for a restricted rather than generalized stimulatory response of skin-derived human mast cells to substance P. *J Neuroimmunol.* 2005;163(1-2):92-101.

Supplemental Figure Legends

Supplemental Figure 1

DPI filtering and model goodness of fit

(A) DPI data filtering Left: Number of remaining regions at TPM cut-off $t = 1$; Right: Densities of VST-transformed region data by sample/condition. (B) Quantile-quantile plots for Goodness of fit. Left: fit of count data to the Negative Binomial model; Right: fit of count data with Tagwise dispersions to the edgeR model.

Supplemental Figure 2

Upregulation of T cell co-stimulatory receptors and donor-dependent IL-31 induction by FcεRI aggregation

(A) Cultured MCs were analyzed by flow-cytometry for ILA, OX40L expression 24 h following FcεRI crosslinking or with no stimulus Upper panel: mean values from 10 (ILA), and 5 (OX40L) independent assays * $p < 0.05$; *** $p < 0.001$. Lower panel: representative histograms (B). Quantification of IL-31 in the supernatant of MCs from 10 donors 24 h upon FcεRI aggregation. Note the great inter-individual variability in IL-31 production.

Supplemental Figure 3

Hierarchical clustering results for (A) GATA1/GATA2, (B) MITF, and (C) MRGPRX2

Higher-resolution heatmaps/hierarchical clustering of 50 blood FANTOM5 samples (as in Fig. 3) are provided.

Supplemental Figure 4

Promoter Swap in EXOC6B by cultured MCs

Comparison of preferential promoter activity of the EXOC6B gene between *ex-vivo* and cultured/expanded MCs. Note that cultured MCs acquire additional promoters barely expressed by other FANTOM5 samples.

Supplemental Table Legends

Supplemental Table 1

Number of differentially expressed DPIs (and RefSeq annotated genes) by method and MCs comparison

The table contains the number of differentially expressed DPIs (and RefSeq annotated genes) at BH adjusted P-value significance level = 0.01. Analytical results are given for each method (edgeR and DESEq; see Supplementary Methods) and each comparison of MC types. The actual genes, their annotation and the statistical estimates are provided in Supplementary Tables 3-4.

Supplemental Table 2

Genes overexpressed by MCs

Genes with highest expression in at least one of the 9 MC samples and enrichment by at least 10-fold compared to the mean of all 893 FANTOM5 samples. Expression levels (in tpm) are given individually for each MC preparation, followed by the mean of the 3 MC subsets (*ex vivo*, expanded, expanded+stimulated) and of the mean of all MCs. The mean of all FANTOM5 samples and sample with next best expression are also given for comparison. The (putative) gene function is specified whenever information was available.

Supplemental Table 3

Comparison between *ex vivo* and cultured MCs

Differentially expressed and non-differential genes of *ex vivo* and cultured MCs are given in separate worksheets (genes form the basis of Fig. 2A/B). Significantly differential promoters between the MC subsets are also specified.

Supplemental Table 4

Comparison between resting and stimulated MCs

Differentially expressed and non-differential genes of expanded and expanded+stimulated MCs are given in separate worksheets (genes form the basis of Fig. 2C/D). Significantly differential promoters between the MC subsets are also specified.

Supplemental Table 5

MC specific promoters and promoter versus gene expression analysis

Given are promoters with at least 50-fold higher activity in MCs than in non-MCs (FC=mean(MC)/mean(F5 w/o MC) >50). The fold change is likewise calculated for the level of the entire gene (i.e. all promoters combined). These two levels are compared to each other. At the top of the list are promoters with greatest selectivity in MCs if contrasted against their respective genes. At the bottom are those promoters for which full gene activity surpasses activity of the selected promoter (even though the promoter is still much more active in MCs than in non-MCs). *Ex vivo* MCs, expanded MCs, and expanded+stimulated MCs are given in separate worksheets. Basophils are also included for comparison.

Supplemental Table 6

Genes with different promoter sets active in *ex vivo* as compared to cultured MCs

Given is a list of genes, for which the two MC subsets utilize a different set of promoters. No genes of this kind were detected for expanded versus expanded+stimulated MCs. The promoter swap for the gene EXOC6B is illustrated in Suppl. Figure 4.

Supplemental Table 7

Motif activity in MCs

Motif activity was calculated by MARA (supplemental Methods) and is given for *ex vivo*, expanded, and expanded+stimulated MCs, as well as for basophils and non-MC blood samples (used for direct comparison). The motifs are ranked and sorted by “best motif ranking” for the *ex vivo* samples. Gene expression levels of the TFs binding to these motifs are also given for comparison.

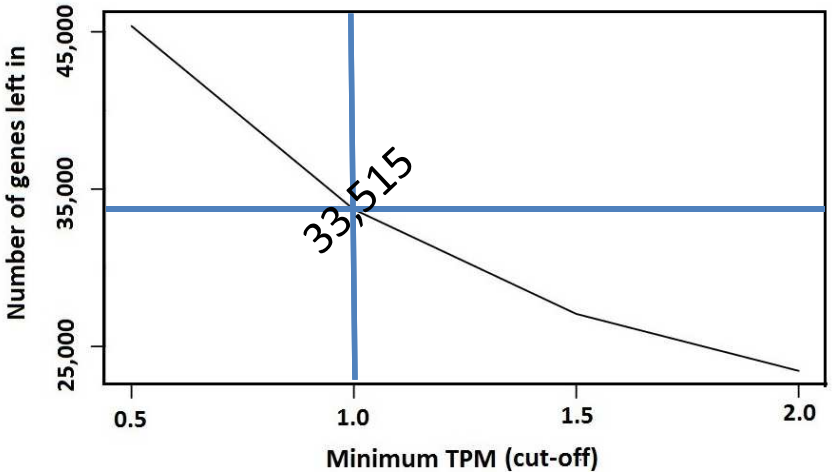
Supplemental Table 8

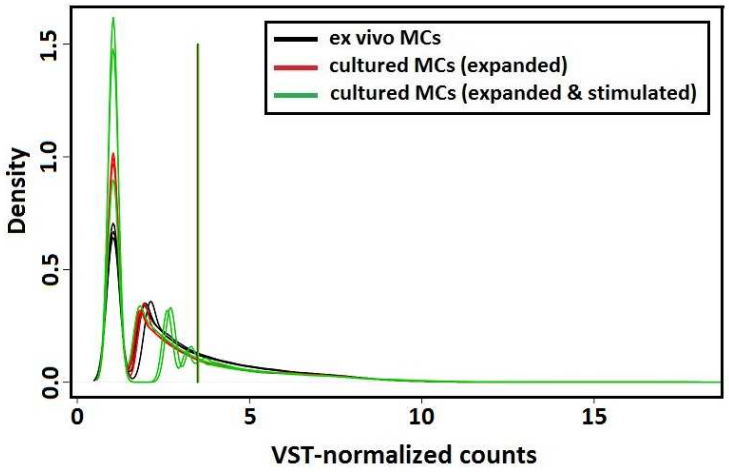
Comparison between basophils and MCs

Differentially expressed and non-differential genes of basophils and *ex vivo* MCs are given in separate worksheets (genes form the basis of Fig. 4C/D).

A

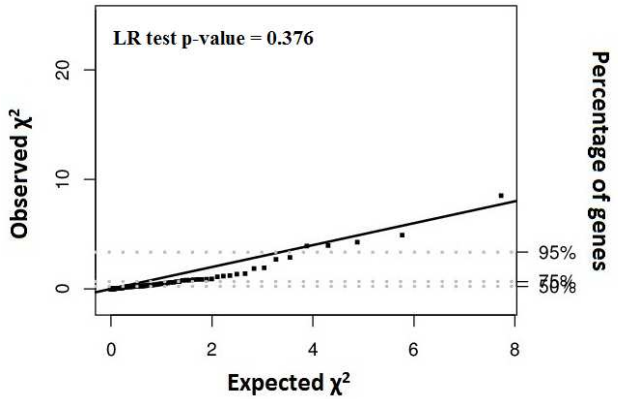
Supplemental Fig 1

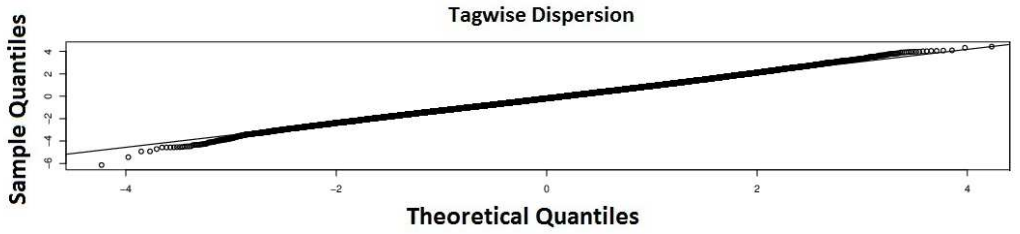




B

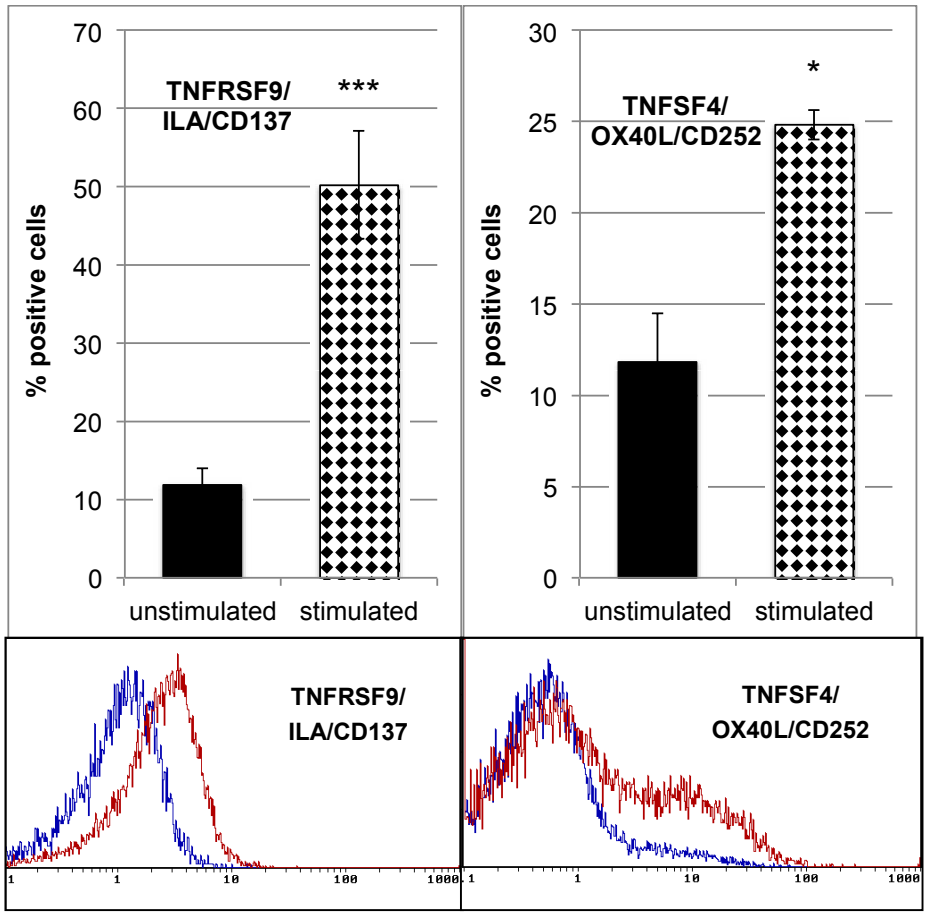
Negative Binomial edgeR model



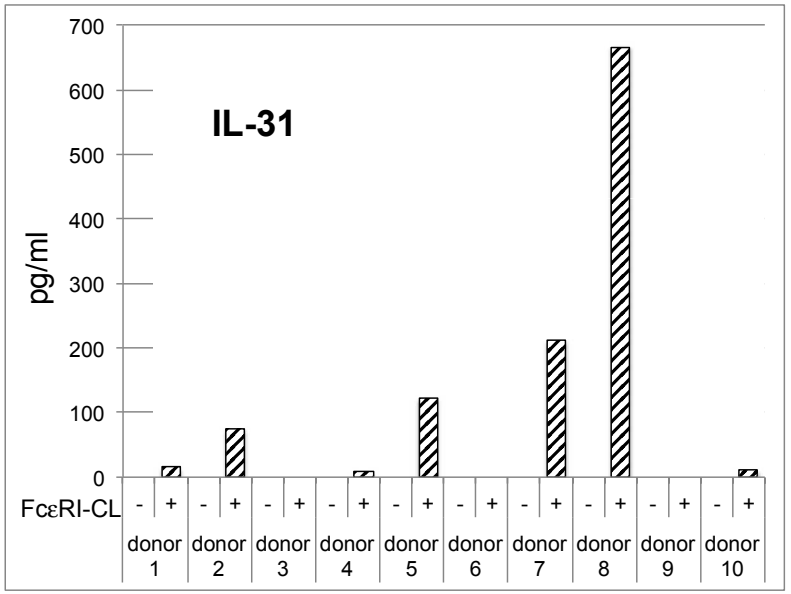


Supplemental Fig 2

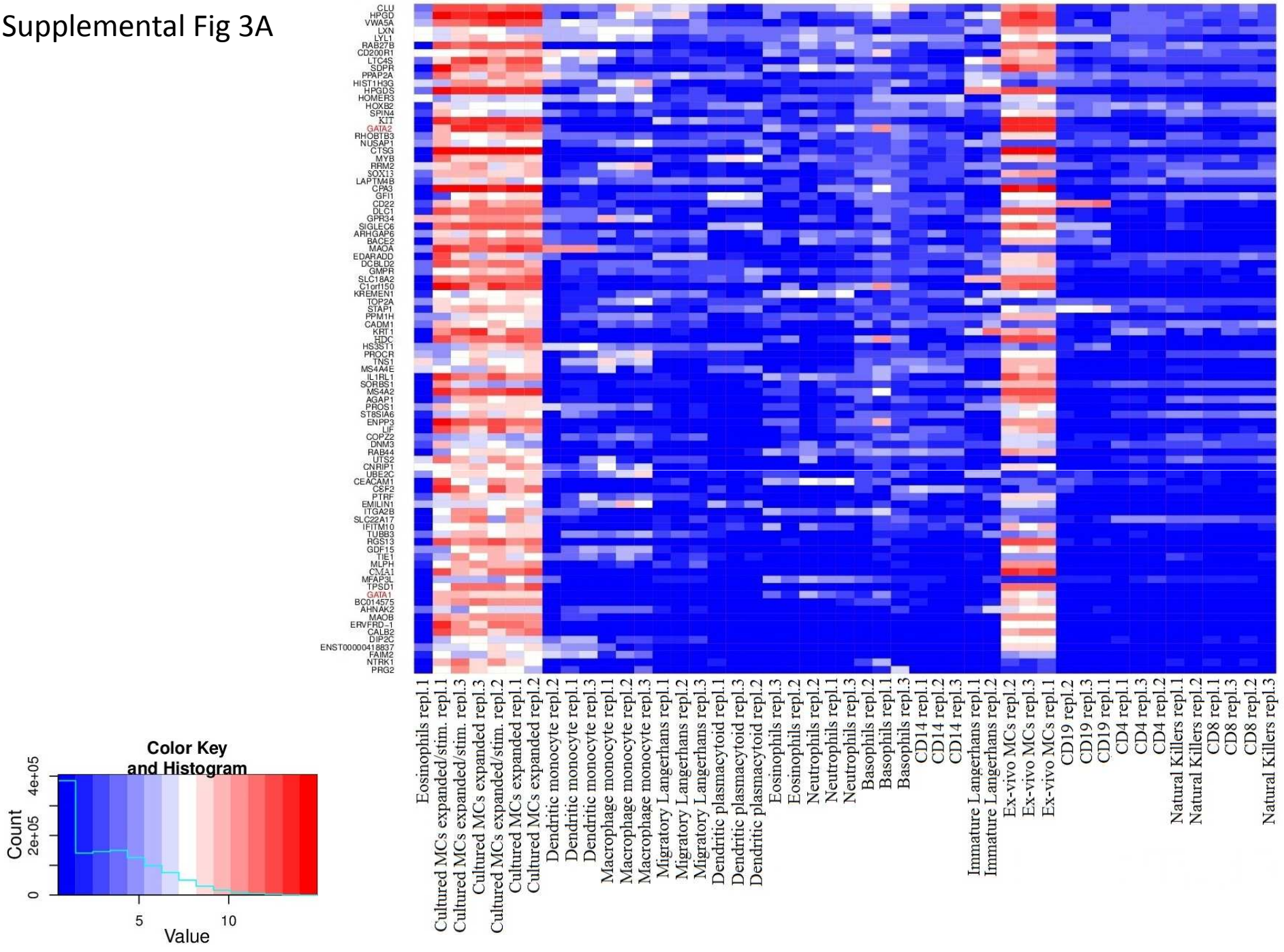
A



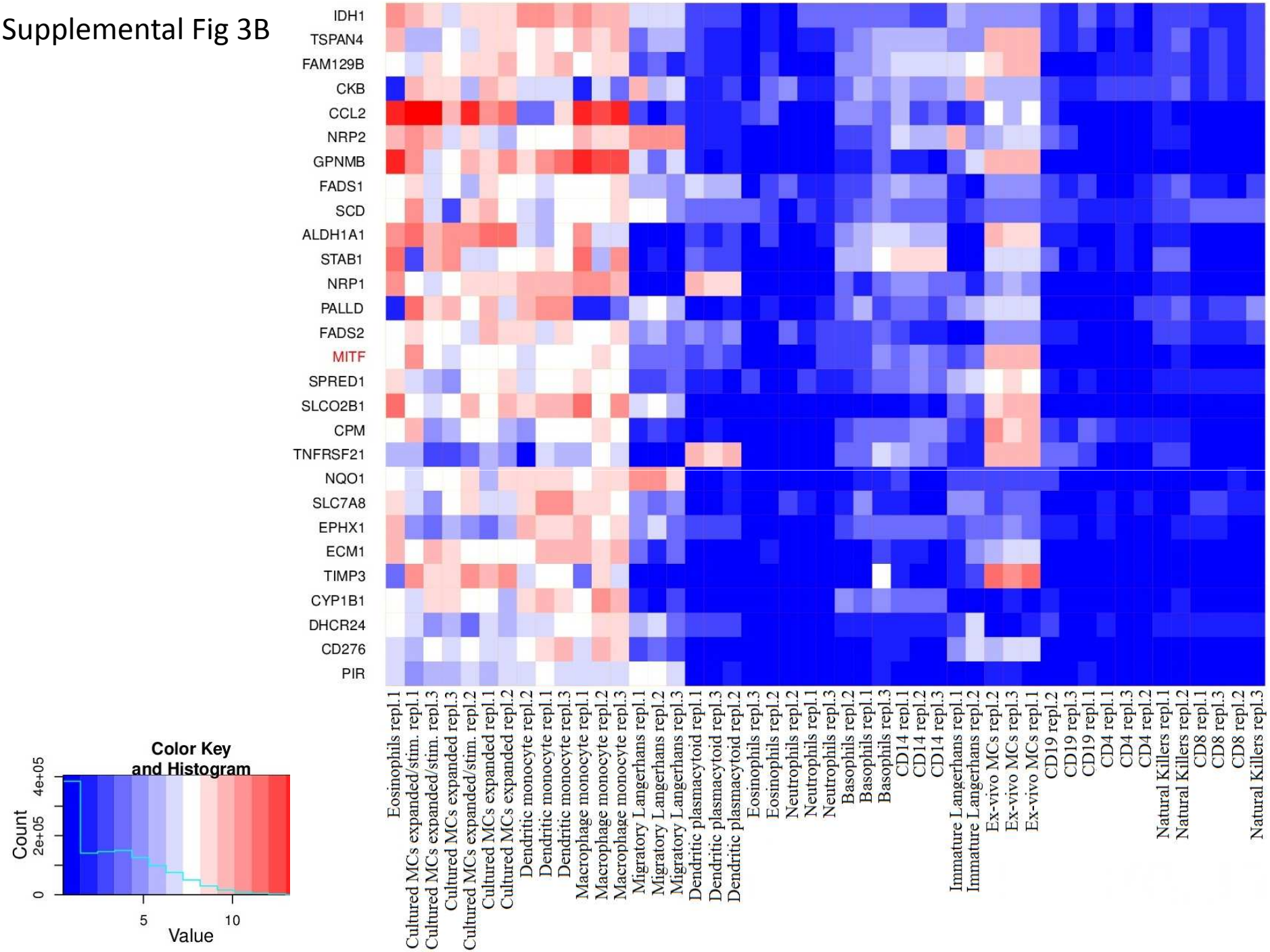
B



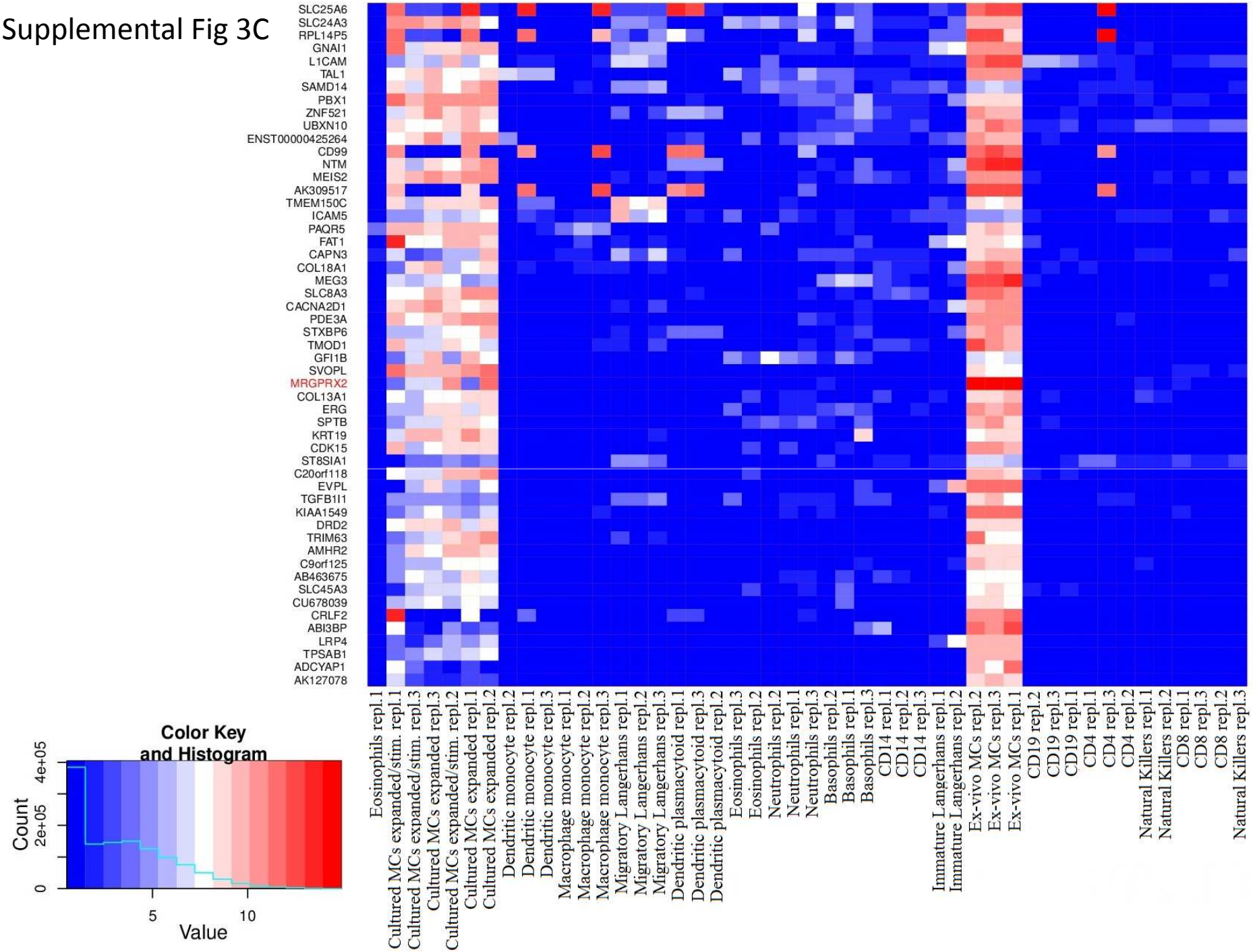
Supplemental Fig 3A



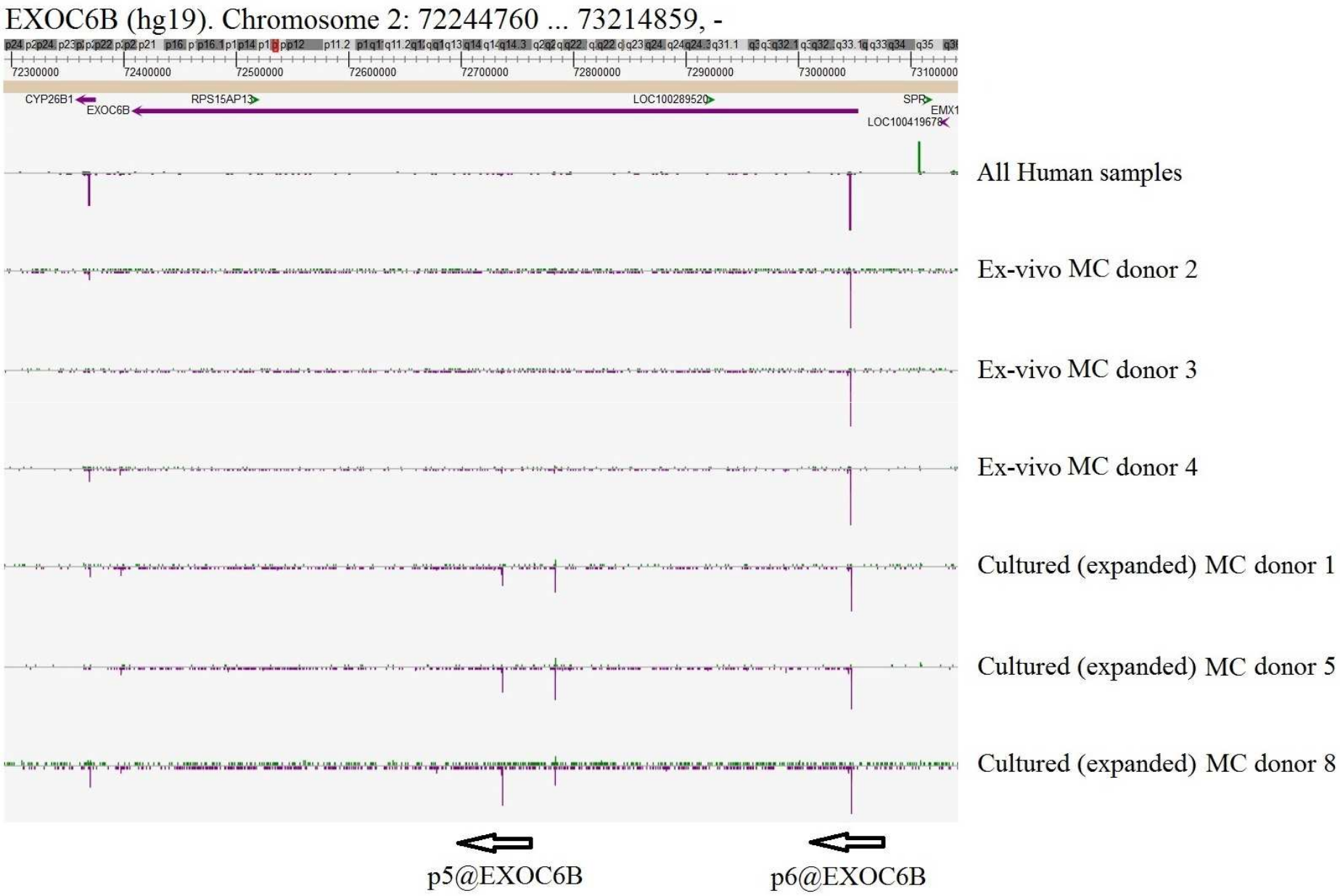
Supplemental Fig 3B



Supplemental Fig 3C



Supplemental Fig 4



Supplemental Table 1

Number of differentially expressed DPIs (and RefSeq annotated genes) by method/comparison

	Ex vivo MC vs cultured MC expanded		cultured MC expanded vs cultured MC expanded/stimulated	
	$\mu_{ev} > \mu_{exp}$ [1]	$\mu_{ev} < \mu_{exp}$ [1]	$\mu_{exp} > \mu_{exp\&stim}$ [1]	$\mu_{exp} < \mu_{exp\&stim}$ [1]
edgeR	7,369 (1,875)	1,392 (838)	104 (84)	638 (260)
DESeq	5,077 (1,322)	922 (583)	--- [2]	--- [2]
Common	4,980 (1,264)	846 (538)	104 (84)	638 (260)

[1] μ_{ev} is the mean of ex vivo samples, μ_{exp} is the mean of cultured (expanded) samples, $\mu_{exp\&stil}$ is the mean of cultured (expanded, stimulated) samples
[2] DESeq is not suitable for paired samples comparison