M*Sc* Thesis

“Gene expression study of hematopoietic stem cells in an experimental model of Systemic Lupus Erythematosus (SLE)”

Maria Grigoriou

“Autoimmunity and Inflammation Lab”
A. Banos, G. Bertsias, DT. Boumpas

Joint Graduate Programme
“Molecular Biology and Biomedicine”
Department of Biology & Medicine
University of Crete & IMBB-FORTH
2015

Acknowledgements

Initially, I would like to thank Dr. Boumpas and Dr. Bertsias who gave me the opportunity to carry out this thesis. I express the warmest thanks for the guidance and for advice regarding my work. I could not forget to thank Aggelos Banos who. heard my various concerns without obligation and for his valuable assistance. Special thanks to Panos Verginis, Katerina Hatzioannou and Themis Alissafi who are always willing to help me with the experiment design and evaluation. Finally, I would like to thank Professor Athanassakis I. and Professor Papamatheakis J. for participating to my evaluation for obtaining this master degree.

Contents

Abbreviations

BM	Bone marrow
CDKs	Cdk inhibitors
CFU	Colony-forming unit
CLL	Chronic Lymphocytic Leukemia
DEGs	Differentially expressed genes
HSCs	Hematopoietic Stem Cells HSCs
HSPCs	Hematopoietic Stem/Progenitor Cells
IPA	Ingenuity pathway analysis
LSK	Lin- Sca-1+ c-Kit+
MACS	Magnetic cell sorting
PBMCs	Peripheral blood mononuclear cells
RNA-seq	RNA-sequencing
SLE	Systemic Lupus Erythematosus

Abstract

Hematopoietic Stem Cells (HSCs) represent a cell population that gives rise to every lineage of blood cells, including granulocytes, lymphocytes, monocytes and endothelial cells, all of which have been implicated in the pathogenesis of Systemic Lupus Erythematosus (SLE). Our working hypothesis is that the fundamental immune aberrations in SLE –genetic or epigenetic- may be easier to be traced back to the HSC population. Although HSCs are thought to exist in a dormant state within the bone marrow niche, recent evidence suggests that in lupus mice they may overproliferate due to both intrinsic and extrinsic effects. However, the molecular identity and the contribution of these alterations to lupus remain elusive. We found significantly increased frequencies as well as absolute numbers of HSCs in the BM of lupus NZBxNZW/F1 mice with established disease as compared to young NZBxNZW/F1 mice or to C57/BL6 mice. Bone marrow populations such as hematopoietic stem progenitors cells (HSPCs), lymphoid and myeloid lineages seem to differ in homogeneity depending upon either age or disease, indicating an alteration in HSC potential under inflammatory conditions. Accordingly, serum from F1 young mice promotes healthy HSCs to proliferation and skews their differentiation to myeloid lineage. Global gene expression analysis of HSCs originating from lupus mice revealed 547 differentially expressed (DEGs). These significantly DEGs contain crucial transcription factors for hematopoiesis, HSC function and homeostasis, as well as regulators for immune response during inflammation and autoimmune diseases. These data provide initial insights in the fundamental changes of HSC in lupus within the inflammatory milieu of the disease.

Introduction

Hematopoietic Stem Cells

Hematopoiesis is the lifelong process by which all the cells of the blood system are produced in a hierarchical manner from a small population of HSCs. HSCs give rise to progenitor cells that become gradually lineage restricted and ultimately differentiate into all lineages of mature blood cells (Figure 1). As HSCs continuously replenish cells that are lost or turned over, they must self-renew to maintain their number over the lifetime of the organism [1]. Homeostasis within the hematopoietic system depends on the replacement of the immune effector cells by hematopoietic precursors [2].

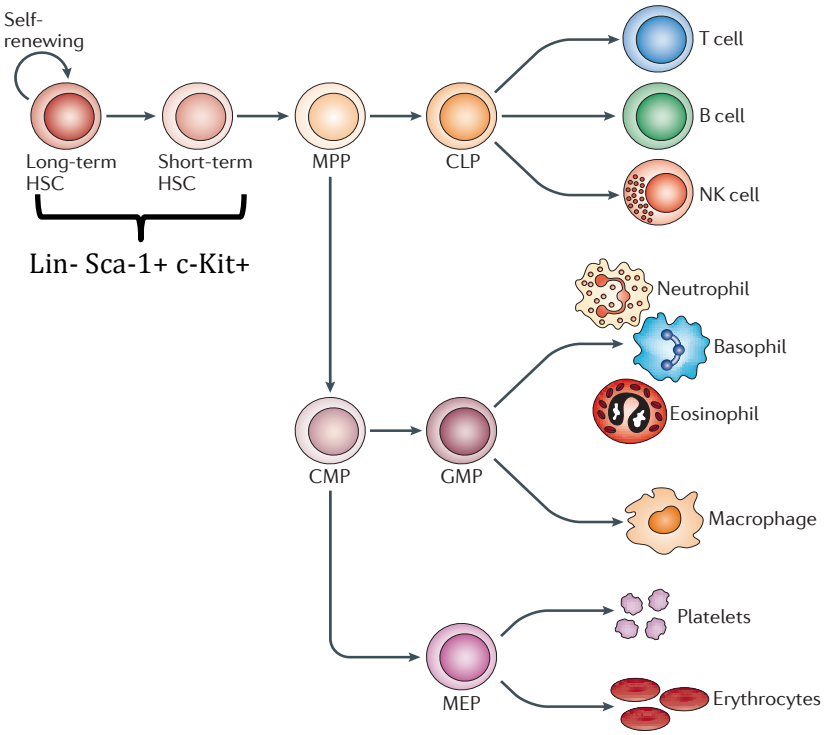


Figure 1. The Hematopoietic Tree[2]

LSK compartment

HSCs are the only cells within the hematopoietic system that possess the potential for both multi-potency (ability to differentiate into all functional blood cells) and self-renewal (ability to give rise to identical daughter HSCs without differentiation) [3]. The resultant population of enriched mouse HSCs have phenotype Lin^{-} $Sca\text{-}1^{+}$ $c\text{-}Kit^{+}$ and represent approximately 0.05% of the mouse adult BM cells [3]. Long-term HSCs are capable of nearly indefinite self-renewal and differentiation into the mature cells of blood lineages. These are the bone fide stem cells of hematopoiesis. They are operationally distinct from short-term HSCs which are capable of only limited self-renewal (but competent to differentiate into the mature cells of all blood lineages) and therefore cannot reconstitute hematopoiesis for the life of an organism [4].

Hematopoietic niche

HSCs reside primarily in the bone marrow within a specialized microenvironment called the HSC niche. The HSC niche provides soluble factors and cell-cell interactions that are crucial for regulating HSC proliferation, quiescence, self-renewal and differentiation. In the adult bone marrow there is a low frequency of HSCs, with two to five HSCs per 10^5 total bone marrow cells [4].

HSCs are first responders to both acute and chronic infection. Infection leads to changes in the HSC niche that contribute to changes in HSC biology. Pro-inflammatory cytokines that released during infection are crucially important to HSC regulation. These cytokines appear to be required for the maintenance of the appropriate number of HSCs and for the proliferation and differentiation of HSCs, both under homeostatic conditions and in response to stress [2]. Most of the cells in the bone marrow adult mice are quiescent (i.e. in a dormant or G_0 state). Thus, in the absence of any unusual perturbation, they are resistant to drugs or other treatments that specifically target dividing cells.

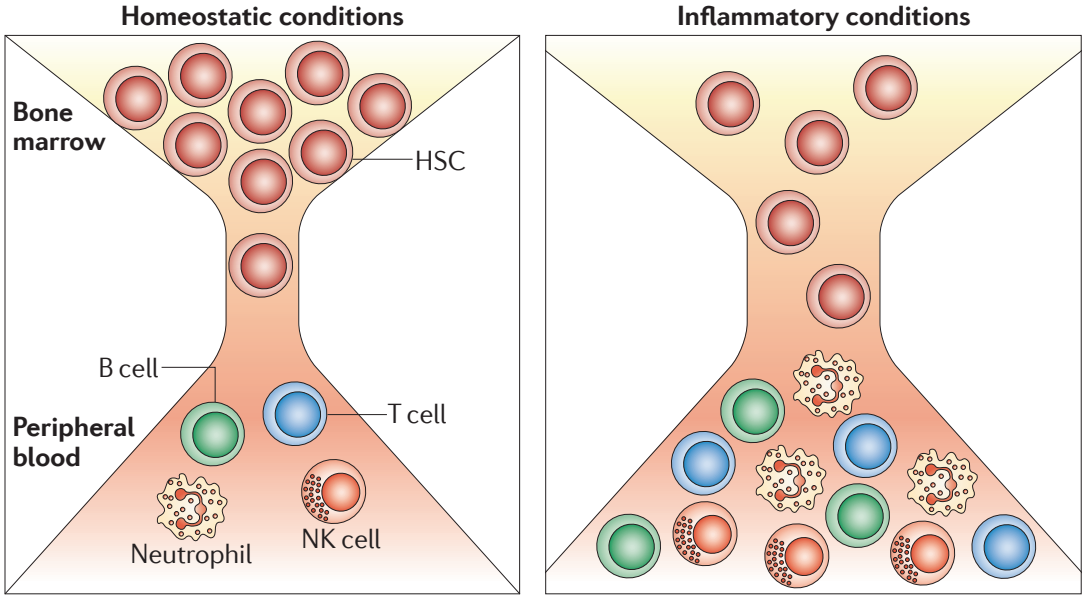


Figure 2. HSCs response under steady state and inflammatory conditions[2]

Inflammatory conditions increase proliferation and self-renewal among LSK compartment. HSCs divide as part of the primary immune response rather than simply to replace depleted progeny pool (Figure 2). They can also be "pulled" towards cell division following the depletion of committed progenitor populations from the bone marrow [212].

Hematopoietic Stem Cells and Cell cycle

Disruption of HSC quiescence in bone marrow leads to defects in HSC self-renewal and often results in HSC exhaustion [5], hence underscoring the critical importance of a constitutively low level of cell cycle activity for proper function of the blood system during adult life. DNA damage triggers signaling cascades that lead to cell cycle checkpoint activation, apoptosis or differentiation [4]. Cell

cycle regulators (Figure 3) mediate HSC continued maintenance in a quiescent or G_0 phase while allowing for their rapid entry into the cell cycle to respond to hematopoietic demand. These factors divide in (i) Cdks, which drive cell cycle progression; and (ii) CKIs which blunt progression through the cell cycle [6].

The activity of Cyclin D-Cdk4/6 complex that controls progression through G1 in response to mitogenic signals is likely a central determinant of HSC cell cycle activity. The Cyclin D family includes Cyclin D1 (Ccnd1), Cyclin D2 (Ccnd2) and Cyclin D3 (Ccnd3), which are all expressed, albeit at different levels in HSCs.

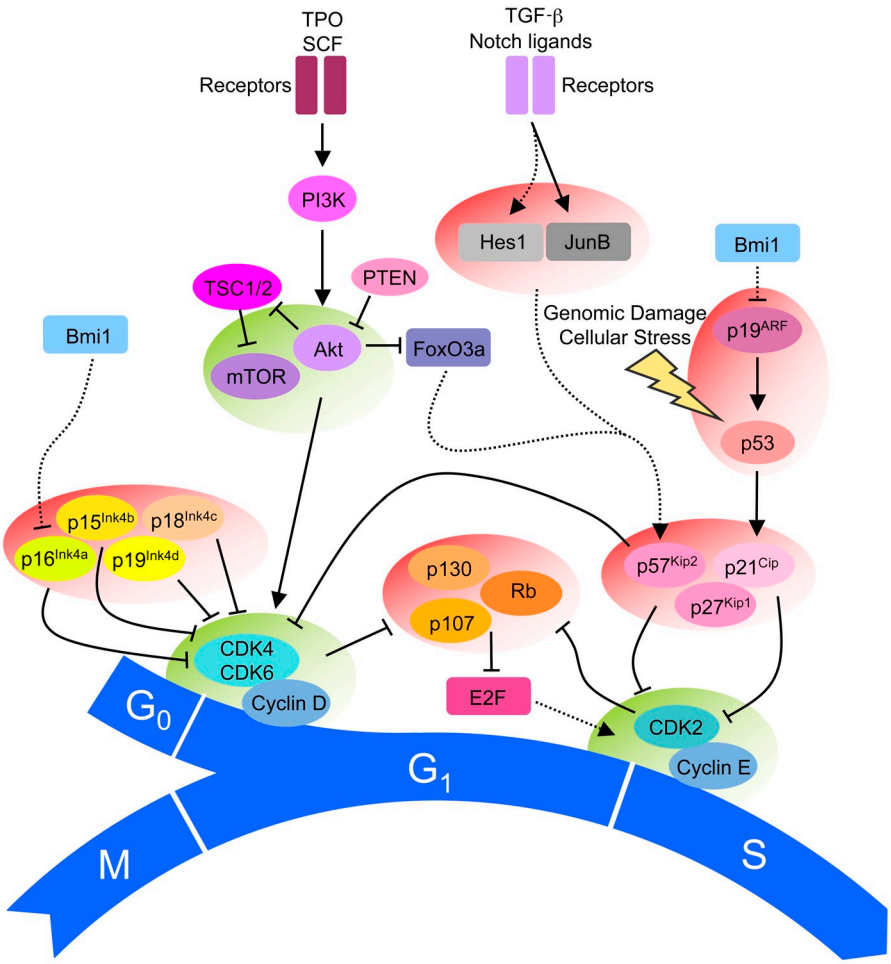


Figure 3. Cell cycle regulation[1]

The Ink4 family includes the CKIs $p15^{Ink4b}$, $p16^{Ink4a}$, $p18^{Ink4c}$ and $p19^{Ink4d}$ and a functionally distinct protein $p19^{ARF}$. The Ink4 proteins all function as antagonists of the Cyclin D-Cdk4/6 complex, thereby blocking phosphorylation of Rb family members and subsequent entry into S phase [1].
The CIP/KIP family includes the CKIs $p21^{Cip}$, $p27^{Kip1}$ and $p57^{Kip2}$, which also restrain entry into S phase by inhibiting the activity of the Cyclin E-Cdk2 complex. p21 is expressed at somewhat greater levels in adult HSCs relative to their differentiated progeny or to fetal HSCs. p27 appears to affect the cell cycle activity of more committed progenitor populations [1].

Systemic Lupus Erythematosus

Systemic autoimmune diseases result from interactions between genes and environmental triggers that build on overtime until clinical symptoms appear. A complex interplay between innate and adaptive immunity lies at the core of most of these diseases. This interplay is not static, since initial inflammatory cascades might change as organ damage accumulates. Furthermore, these diseases can be heterogeneous regarding the type of organs involved, clinical course and response to treatment [7].

SLE is a chronic autoimmune inflammatory disease that can affect the majority of organs and tissues. There are periods of remission and flares. At systemic level, both the adaptive and innate branches of the immune system contribute to the development of SLE. The two predominant cell types involved in the adaptive immune system B and T cells, are both essential for the development of lupus. B cells are pathogenic in SLE because of the autoantibodies and cytokines that they produce. T cells drive the systemic and intra-renal activation of B cells [8].

In the early stages of disease, dendritic and other myeloid cells, activate T cells, and produce key mediators such as B cell activating factor. This effect results in the activation of the adaptive immune system. Activation of the systemic immune system leads to the generation of effector T cells and autoantibodies that subsequently target organs such as kidneys. These processes together with numerous soluble mediators elicit chronic inflammation within glomerular and tubulointerstitial sites in the kidneys [8].

There is a large need in both diagnosis and treatment of SLE. The heterogenous and flaring nature of the disease makes assessment of drug efficacy in clinical trials difficult.

RNA-seq expression

RNA sequencing refers to techniques used to determine the sequence of RNA molecules. It includes high-throughput sequencing of cDNA molecules obtained by reverse transcription from RNA, and next-generation sequencing technologies to sequence the RNA molecules within a biological sample in an effort to determine the primary sequence and relative abundance of each RNA molecule. RNA-seq measures global transcriptional profiles, quantitative measurement of mRNA and non-coding RNAs and can detect splicing variants [7]. In addition, paired-end RNA-seq can be particularly advantageous for fusion identification because of the increased physical coverage it offers [9].

The transcriptome is the complete set of transcripts in a cell and their quantity is specific to developmental stage or each condition. Understanding the transcriptome is essential for interpreting the functional elements of the genome and revealing the molecular constituents of cells and tissues and also for understanding the development and disease. In addition, other issues that RNA-seq is answering are the determination of the transcriptional structure of genes in terms of their start sites, 5'-3' ends, splicing patterns and other post-transcriptional modifications and to quantify the changing expression levels of each transcript during development and under different conditions [10].

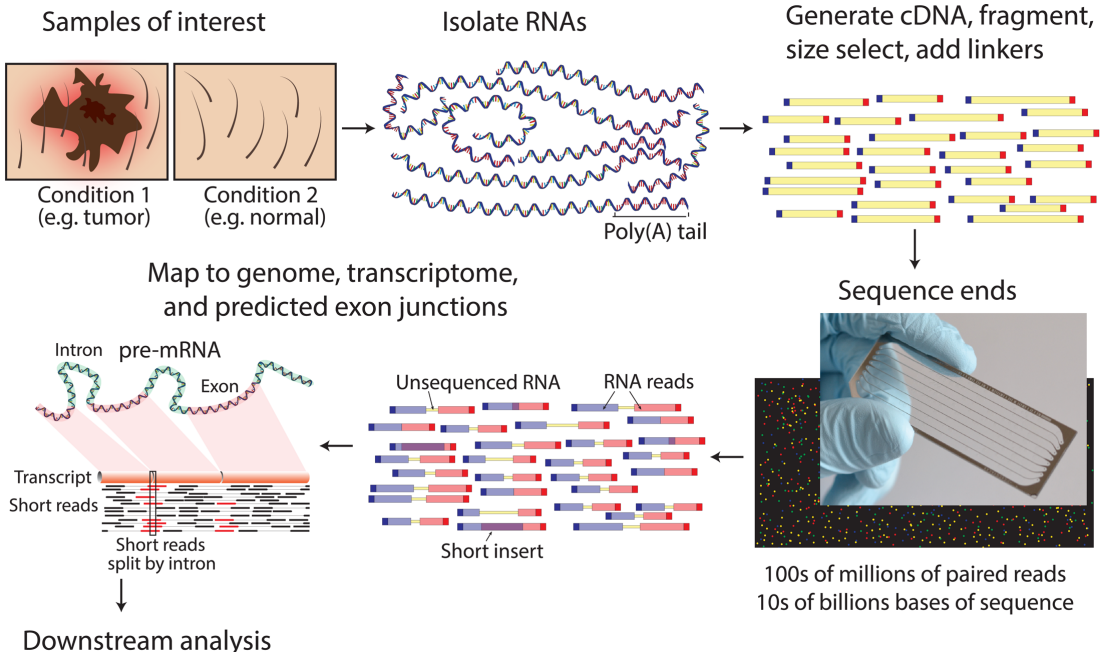


Figure 5. Example of sample preparation and RNA-seq workflow.

RNA-seq is the first sequencing based method that allows the entire transcriptome to be surveyed in a very high-throughout and quantitative manner. This method offers both single-base resolution for annotation and digital gene expression levels at the genome scale. High-throughput sequencing of RNA allows one to count and exquisitely resolve tens of thousands of RNA transcripts and isoforms within a single assay (high quality reads), with potentially far greater impact on clinical diagnostic tests. Moreover, it enables direct measurement of a single transcript changed in one of a multitude ways through alternative promoter usage, rare splicing events and gene fusion, epitranscriptome modifications, allele-specific expression or nonsense-mediated decay. This multivariate information opens up clinical applications that are beyond the scope and capability of microarrays [11].

RNA-seq in complex diseases

Genome-wide transcriptome analysis in cancer and other multifactorial diseases provides a global view of the expressed elements and networks that reshapes the biology of normal cells in their transformation and progression to aggressive diseased cells.

Ferreira et al. have characterized the CLL transcriptional landscape at high resolution by performing RNA-seq on a large cohort of CLL samples. The transcriptomic architecture of CLL refines the molecular characterization of the disease and opens new avenues for the clinical management of patients. Through RNA-seq it is possible to define global transcriptome signatures that are not restricted to gene expression patterns, and that lead therefore to a molecular characterization of the disease at higher resolution. Indeed, in addition to identifying hundreds of protein-coding genes differentially expressed between CLL and normal B cells, authors have characterized the status of other transcriptional elements, such as TEs, lncRNAs, and pseudogenes—all of which are typically invisible to DNA microarrays. Due to the increased resolution of RNA-seq, Ferreira et al. also identified genes with significantly different splicing patterns. Analysis at the level of individual transcript isoform reveals that many of the pathways that are altered in the disease are affected not only by changes in the expression of key genes in the pathway, but also changes in the relative proportion of the alternative splice isoforms produced by each gene. Moreover, RNA-seq provides the ability to quantify the alternative usage of transcription initiation sites, which information is important to understand the therapeutic role of any gene. Analysis of RNA-seq reads mapping to known TEs revealed a general trend of derepression of TE transcription in CLL compared with normal cells [12].

Many RNA-seq studies in cancer have suggested that detrimental fusion transcripts and alternative splicing may be involved in carcinogenesis of different tissues and organs. They discovered a considerable fraction of fusion transcript that is chimeric mRNAs that may alter cell functionality. A study in melanoma revealed that the occurrence of chimeric transcripts is a frequent event as well as highlighted novel genes and pathways associated to its pathogenesis [13]. Another project demonstrated that unbiased mRNA sequencing could identify a number of known clinically relevant fusion genes as well as novel fusion genes in cell lines and clinical leukemia samples. These findings indicate that RNA-seq may have a direct clinical impact for treatment selection[14].

Autoimmune diseases come from interactions between environmental, epigenetic and genetic factors that result in downstream perturbations of complex and interactive biological networks. The many existing murine lupus models have not yet led to the development of specific treatments for human lupus. Many autoimmune diseases continue to be treated with non-specific medications such as corticosteroids and chemotherapeutic drugs [15]. Challenges studying autoimmune diseases include genetic heterogeneity referring to how a set of genetic variants might define a trait onset as well as whether a single gene leads to multiple phenotypic expressions or disorders. Diverse human populations present different allelic and genotype structures depending on their evolutionary and epidemiological history. Differences in allele and genotype frequencies among populations reflect the contribution of evolutionary forces such as selection, genetic drift, mutation and migration, which might explain why some risk alleles to autoimmunity may be protective factors to infectious diseases and vice versa [15]. Additional challenges in the field of autoimmunity include the lack of specific biomarkers that can be used for diagnosis, assessment of disease activity and prediction of flares.

Materials and Methods

Mice

NZB, NZW and NZBxNZW/ F1 mice were purchased from Harlan Laboratories. F1 young mice are three months old without proteinuria; in contrast to F1 old which considered being 7-9 months old and having proteinuria greater than 30 ug/ml. Also, C57/BL6 mice are used from BRFAA animal facility. All mice are maintained in the BRFAA Animal Facility.

Isolation PBMCs from peripheral blood

200-300ul of peripheral blood are collected in eppendorf tubes. Afterwards, ficoll is added using a syringe (5ml or 10ml). Centrifugation is performed at 1800rpm for 30min at room temperature (without break). The white layer is collected and washed with 1ml PBS. PBMCs are stained in 200ul 5%FBS/PBS for LSK compartment, like bone marrow cells. (antibodies: Lin^- cocktail, $Sca\text{-}1^+$, $c\text{-}Kit^+$)

Bone Marrow Isolation Protocol

BM cells are flushed from intact tibia, femur and humerus with 5% FBS/PBS using a 27 G syringe. A 23 G syringe is used to generate single cell suspension. Hemolysis buffer (2ml for 2min) is used in order to lyse erythrocytes. Staining of the cells is performed in 500ul 5%FBS/PBS for 20min at 4^oC. The cells are collected in 2-3ml 5% FBS/PBS using a 70um mesh (all the quantities above are enough for one animal).

MACS lineage cell depletion kit

Miltenyi biotech (No 130-090-858)

Starting material: 6 x 10^8 total bone marrow cells; 50ul out of that for staining Lin- cells. Cell pellet is resuspended in 40ul MACS buffer. Then, biotin antibody cocktail is added. The mix is incubated for 10min at 4-8^oC. Next step is the addition of 40ul MACS buffer and anti-biotin microbeads. The mix is incubated for 15min at 4-8^oC. Cells are washed using MACS buffer. Cells are resuspended in MACS buffer.

Magnetic separation (LS column): the column is equilibrated with 3ml of MACS buffer. Cell suspension (1ml) is applied and allowed to pass through the column. Subsequently, 3x washes with 3ml MACS buffer are followed. Lin^- cells are collected. The cells are stained for $Sca\text{-}1^+$ and $c\text{-}Kit^+$, optionally for Lin^- (antibody dilution 1/100).

Flow Cytometry

The following markers are used for flow cytometric analysis and cell sorting (BD FACSAria II): PE/Cy7 anti-mouse/human CD11b [BioLegend], PE/Cy7 anti-mouse Ly-6G/Ly-6C (Gr-1) [BioLegend], FITC anti-mouse/human CD45R/B220 [BioLegend], FITC anti-mouse TER-119/Erythroid Cells [BioLegend], FITC anti-mouse CD16/CD32 (FcγII/III Receptor) [BD Bioscience], APC anti-mouse Ly-6A/E (Sca-1) [BioLegend], PE anti-mouse CD117 (c-Kit) [BioLegend], PercP 7-

AAD Viability Staining Solution (BioLegend), DAPI CellTrace (Invitrogen). The lineage antibody cocktail includes anti-CD11b, anti-Gr1, anti-TER-119, anti-B220, anti-CD16/CD32. Staining of BrdU is done using BrdU staining kit for flow cytometry FITC (eBioscience, 8811-6600)

Cell Cultures

Isolated LSK cells are cultured into 96-well bottom-flat dishes IMDM supplemented with 2% FBS, 10.000U/ml penicillin (Gibco, 15140), 10.000ug/ml streptomycin (Gibco, 15140), 50ng/ml rm-IL6 (Immunotools, 12340063), 20ng/ml rm-IL3 (Immunotools, 12340033), 20ng/ml rm-SCF (Immunotools, 12343323), 20ng/ml rm-TPO (Immunotools, 12343613), 50mM β-mercaptoethanol (Gibco, 1480119), 2mM L-glutamine (Gibco, 25030-081). Cells are maintained at 37^{o}C in a humidified atmosphere containing 5% CO_2. After 24hrs, cells are collected, splited and stimulated by serum from F1 young, F1 old or C57/BL6 mice. After 24hrs, BrdU assay, Cell Trace staining or Colony-forming cell assay is followed.

BrdU proliferation assay

BrdU staining kit for flow cytometry FITC(eBioscience, Cat. 8811-6600) is used. Incorporation into the newly synthesized DNA of dividing cells (during S phase) indicating what proportion of stem cells have entered or completed cell cycle over the entire labeling period (overnight).
Cultured cells are collected after 24hrs of stimulation by the serum. They are incubated with 1x BrdU Staining Buffer for 30min at room temperature (dark). Then, DNAseI is added for 1hr at 37^{o}C (dark). The cells are incubated with the BrdU staining antibody for 30min at room temperature. Collection of the cells is performed in 5% FBS/PBS for flow cytometry analysis.

Cell trace proliferation assay

The appropriate amount of CellTrace™ Violet (Invitrogen, Catalog No. C34557) is added at the beginning of the culture. The cells are incubated for 20min at 37^{o}C, protected from light. Five times the original staining volume of complete IMDM is added to the cells in order to quench any unbound dye (incubation for 5 min). The cells are centrifuged, resuspended in fresh complete IMDM and incubated for 4 days according to the above cell culture protocol.

Colony-Forming Cell Assay

LSK cells (1 x 10^3) are plated in duplicate in 35-mm tissue culture dishes (CLS430165 SIGMA) containing 1.1ml of MethylCellulose (MethoCultTM GF M3434). After 12-15 days of incubation at 37^{o}C in 5% CO_2, CFU-M, CFU-G, CFU-GM, CFU-GEMM, BFU-E, CFU-Mix are scored under an inverted light microscope and stored in cell lysis buffer at -80^{o}C.

RNA extraction

Qiagen RNeasy Micro Kit (74004), NucleoSpin® RNA XS MN (740902.50)
Samples are mixed with lysis buffer and mercaptoethanol. 27G syringe is used to resuspend the mix. One volume of 70% ethanol is added, and the samples are transferred to the column. Centrifugations are performed at speed >8000g for 15s at room temperature. Next, 1 volume of buffer RW1 is added to the samples. DNAse enzyme with RDD buffer are mixed and added to the column. Samples are incubated for 20-25min at room temperature. One volume of buffer RW1 is added again. Columns are transferred to new 2ml collection tubes. 500ul RPE buffer is added. The columns are transferred to new 2ml collection tube and centrifuged for 5min with open lid. Elution is performed by adding 14ul RNAse free water, incubate for 2min and centrifuge for 1min (full speed).

Phenol-chloroform
Trizol (250ul) is added to the samples (mix gently). The proportion of chloroform/trizol is 0.2ml chloroform per 1ml trizol. After centrifugation at full speed, the upper phase is transferred in new tube and 0.5ml isopropanol per 1ml trizol is added. The cell pellet is washed with cold 70% ethanol (100ul). In each tube is added 10ul of DNAse/RNAse free buffer is added with 5ul DNAse and they are incubated for 30min at 37^{o}C.
Precipitation is achieved with 1/10 volume of CH_3COONa 3M pH5.2 and addition of 2.5 volumes of cold 100% ethanol. Next, incubation follows in ice for 15min. Samples are centrifuged for 15min at full speed, 4^{o}C. The cell pellet is washed with 70% cold ethanol. Elution is perfomed with 50ul RNAse free water.

Reverse Transcription
Total cell RNA is used for cDNA synthesis with SuperScript® II Reverse Transcriptase Kit (18064-014, Invitrogen™). The ingredients for first-strand cDNA synthesis are 1st mix [oligoDT (500ng/ul); dNTP Mix (10mM); sterile, distilled water] and 2nd mix [5x First-Strand Buffer; 0.1M DTT; RNAseOUT™ (40 units/ul); RT enzyme and 1ng-5ul RNA].
The program includes the following steps: Heating 1st mixture to 62^{o}C for 5min. Incubation at 42^{o}C for 2min. There is an intermediate step of 5 min at 4^{o}C. Lastly, the 2nd mixture is added and incubated at 42^{o}C for 50 min. The reaction is inactivated by heating at 70^{o}C for 15min.

Real-Time PCR
Real-time PCR is performed using the iTaq™ Universal SYBR® Green Supermix (Biorad,172-5121). mRNA levels are normalized to HPRT which gene is used as an internal control. [Conditions: Tm 60^{o}C, 50 cycles]

Gene	Forward primer	Reverse primer
Cdkn2c (p18)	CCT TGG GGG AAC GAG TTG G	AAA TTG GGA TTA GCA CCT CTG AG
Cdkn1a (p21)	CCT GGT GAT GTC CGA CCT G	CCA TGA GCG CAT CGC AAT C
Cdkn1b (p27)	TCA AAC GTG AGA GTG TCT AAC G	CCG GGC CGA AGA GAT TTC TG
Cyclin D	GCG TAC CCT GAC ACC AAT CTC	GAG CAG AAG TGC GAA GAG GAG
Nfe2l2 (Nrf2)	CTT TAG TCA GCG ACA GAA GGAC	AGG CAT CTT GTT TGG GAA TGT G
HPRT	GTG AAA CTG GAA AAG CCA AA	GGA CGC AGC AAC TGA CAT

Table 1. Primer sequences

Statistical Analysis

Data are processed in GraphPad Prism 5.0 software. Statistical analysis for comparisons between two groups is performed with Student's t-test.

RNA-seq pipeline

RNA samples -duplicates from C57/BL6 young, F1 young and F1 old mice- were sequenced at University of Geneva. The sequencing was performed with Illumina Hiseq 2000; paired end sequencing of 49 bp per pair was done. Raw data were filtered by fastqc quality control as well as by cutadapt for adapter contamination.

There were three different approaches in order to analyze the expression profile of hematopoietic stem cells.

1. The mapping was done with GEM mapper using GemTools pipeline. Next step followed with HTseq and DEseq algorithms.
2. Tophat 2.0, with mm9 genome version of genome browser, was used at the first step. Further analysis was done with Cufflinks pipeline[16] MetaSeq programs

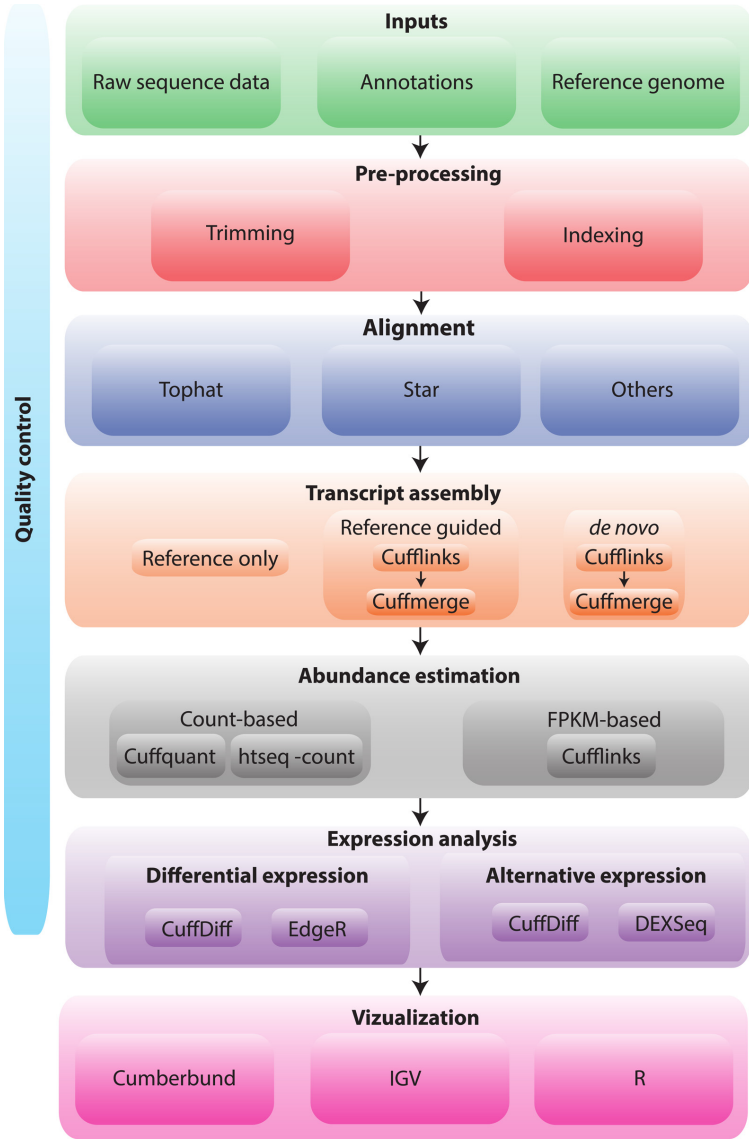


Figure 6. RNA-seq analysis workflow example[17]

3. As before, for alignment was used Tophat 2.0 but afterwards metaSeq was used in order to detect differentially expressed genes. [see Apeendix: Bioinformatics analysis]

After the above steps, the analyses that followed are listed below:

- CummeRbund: which is designed to describe and visualize the results from Cufflinks like expression plots and heatmaps.
- Ingenuity pathway analysis in order to identify biological functions or pathways enriched among the differentially expressed genes.

Results- Discussion

LSK Compartment Gating Strategy

- We used FACS cell sorting in order to isolate (from total bone marrow cells or peripheral blood mononuclear cells-PBMCs) purified LSK cells (Figure 7). On average, a naïve C57/BL6 mouse generated 80.00-100.000 LSK cells. Figure 8 shows that NZB had 6,78%, NZW 2,48%, C57/BL6 young 4,94% and C57/BL6 old 2,74%. Generally, F1 young presented with 1,86% and F1 old with 5,67%.

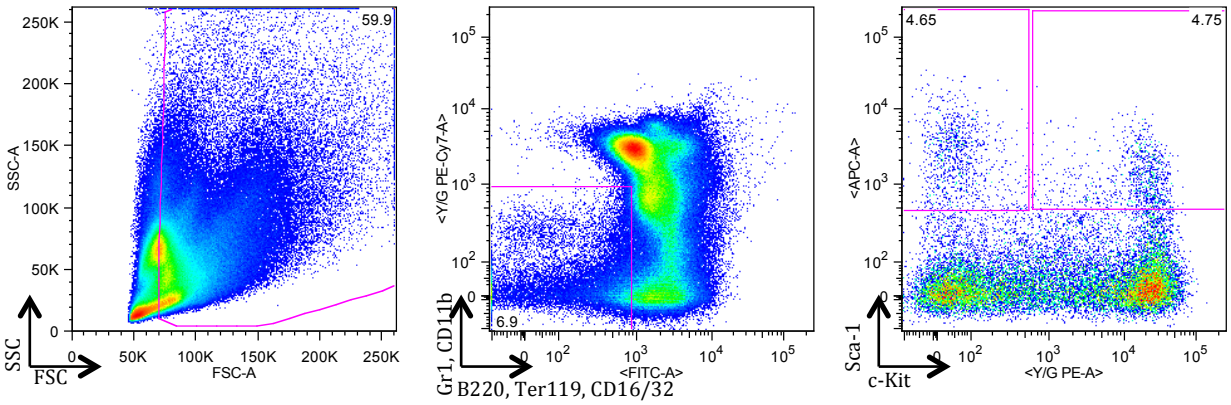
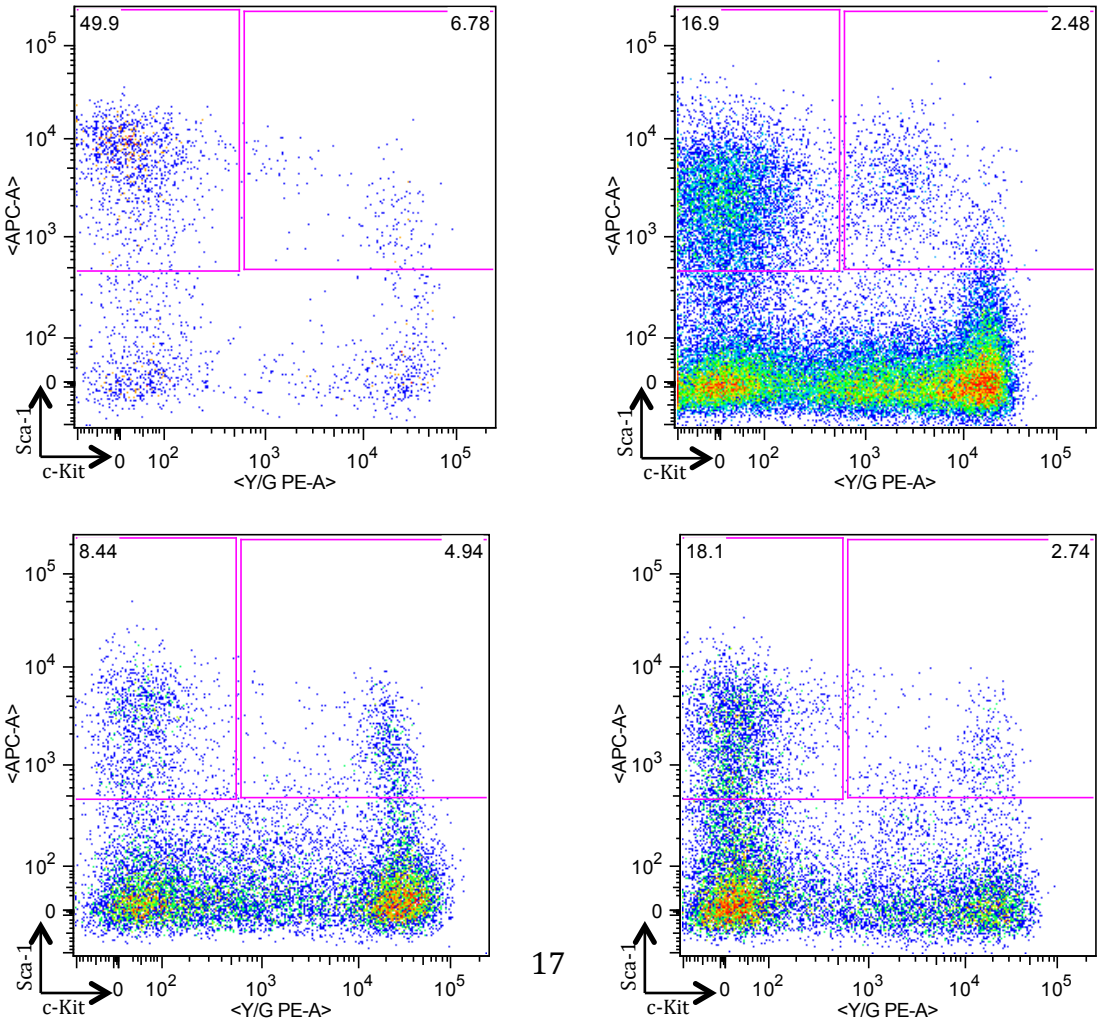


Figure 7. Total bone marrow or blood cells (left), choosing the Lin$^-$ cells (center) and lastly, the double c-Kit$^+$ Sca-1$^+$



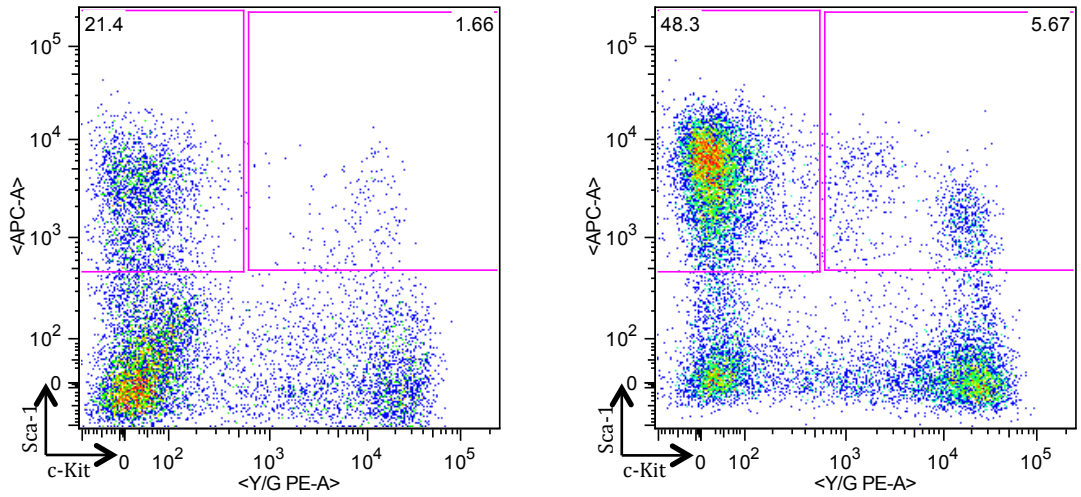


Figure 8. Representative examples for LSK percentages in bone marrow. NZB (up-left), NZW (up-right), C57/BL6 young (middle-left), C57/BL6 old (middle-right), F1 young (down-left) and F1 old (down-right)

- Whole bone marrow cells of F1 young, F1 old and C57/BL6 mice present the same trend as the LSK compartment. They are significantly enhanced in F1 old mice compared to F1 young (Figure 9). Also, we observed similar phenomenon in the $Lin^{-}Sca\text{-}1^{+}c\text{-}Kit^{-}$ compartment, where cells contain early myeloid-committed precursors. They were significantly increased compared to F1 young and their counterpart C57/BL6 old (Figure 10).

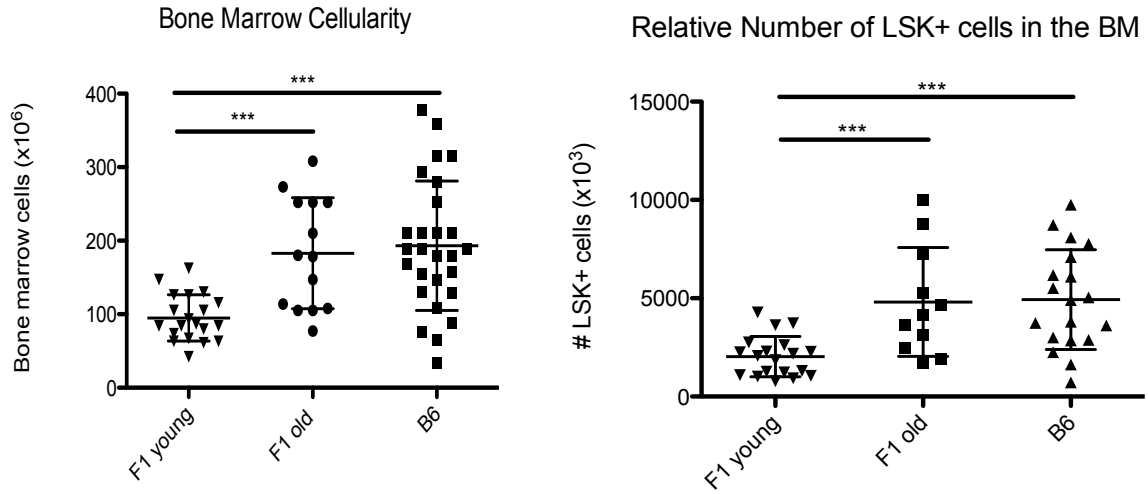


Figure 9. Relative number of total BM cells (left) and LSK compartment (right)

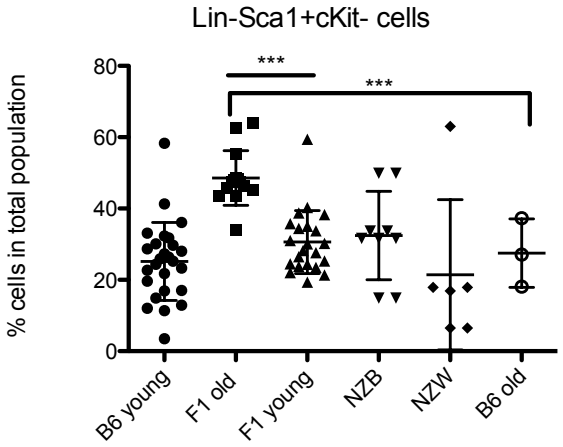


Figure 10. Lin$^-$Sca-1$^+$c-Kit$^-$ compartment in total bone marrow cells

- MACS column was used in order to decrease the number of cells and the time for the FACS sort. This MACS column is used to deplete the Lin$^+$ cells, therefore decreasing the total number of cells and enriching the Lin$^-$ cells. As Figure 11 depicts, after column, the Lin$^-$ cells had almost two-fold increase after column; and the FACS sort was 30 min shorter.

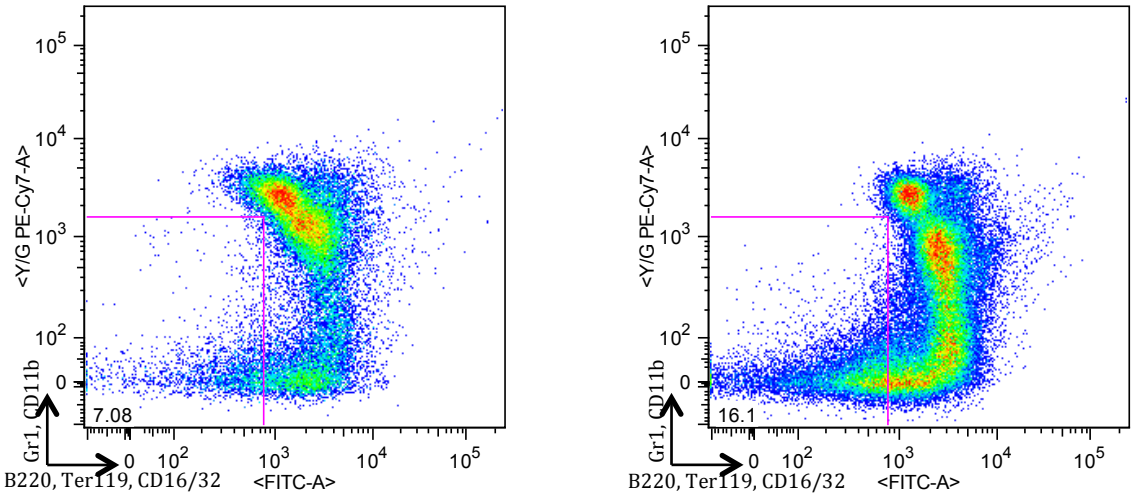


Figure 11. Before (left) and after (right) MACS Lin$^-$ cell depletion

Cell Cultures

We queried what is the impact of the inflammatory F1 serum in B6 derived LSK cells and if there is any difference stimulated with F1 either young or old serum. In order to answer these questions, we cultured sorted healthy LSK cells (naïve C57/BL6 mice) and after 24 hours, we stimulate them with F1 young or F1 old. C57/BL6 serum was also used as internal control followed by proliferation assays and colony-formation assay (Figure 12).

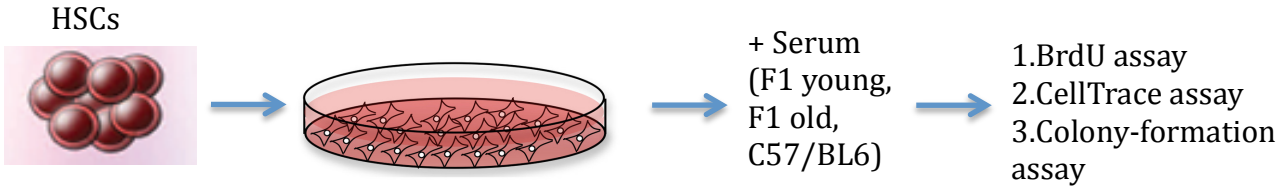


Figure 12. LSK culture *in vitro* setup

BrdU proliferation assay

- BrdU is used in order to detect the proliferating cells after the serum stimulation. BrdU is an analog of thymidine, which can be incorporated into newly synthesized DNA of replicating cells (S phase of cell cycle), substituting thymidine during DNA replication. The gating strategy that was followed during flow cytometry is shown at Figure 13.

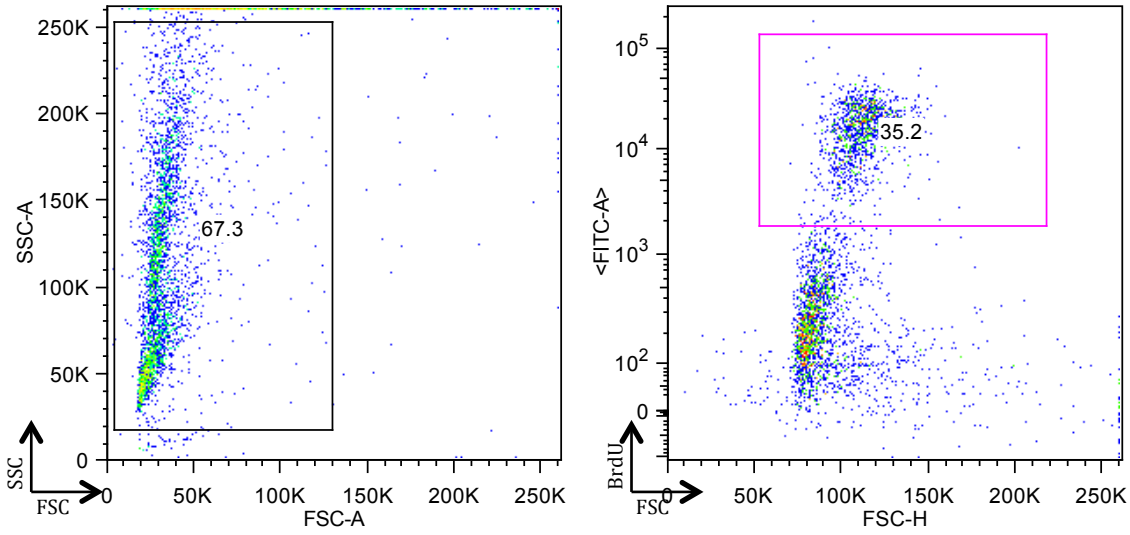


Figure 13. Gating strategy for BrdU staining; FITC⁺ cells represents the proliferative cells

- It should be noted that the number of cells that we end up to assay is very low, as the upper-right panel at Figure 14. The rest panels as well as Figure 15 confirm our hypothesis. LSK cells stimulated by F1 young serum had almost 10% higher proliferation. The inflammatory serum prompts HSCs to proliferate more compared to C57/BL6 serum (only 13,1% of them proliferate). In the negative control, without stimulus in the culture, only 7% of HSCs were proliferating. There seems to be a trend that F1 young serum has greater effect on the healthy LSK than F1 old serum; but due to the number of experiments there is no statistical significance between these two conditions (Figure 15).

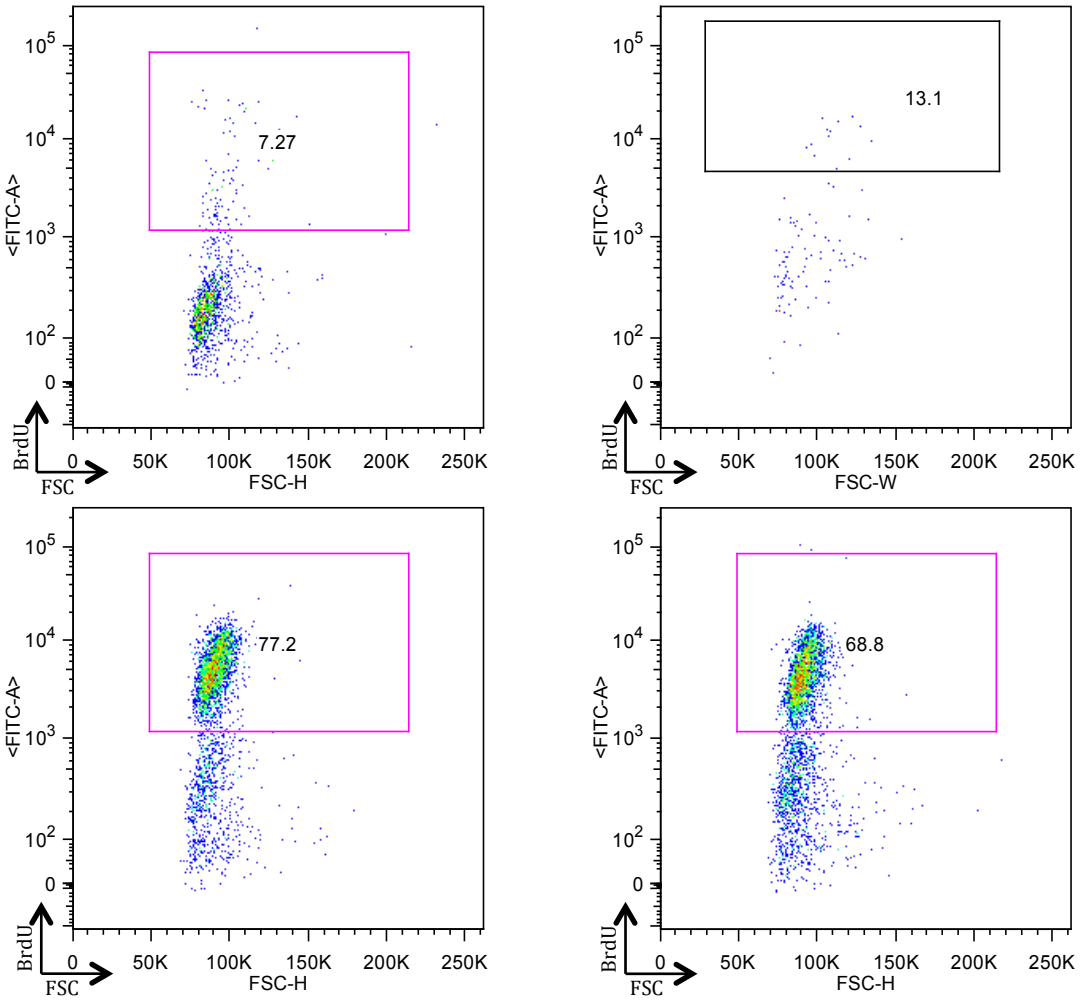


Figure 14. HSCs *in vitro* without serum stimulation (up-left), with C57/BL6 (up-right), F1 young (down-left) and F1 old (down-right) serum

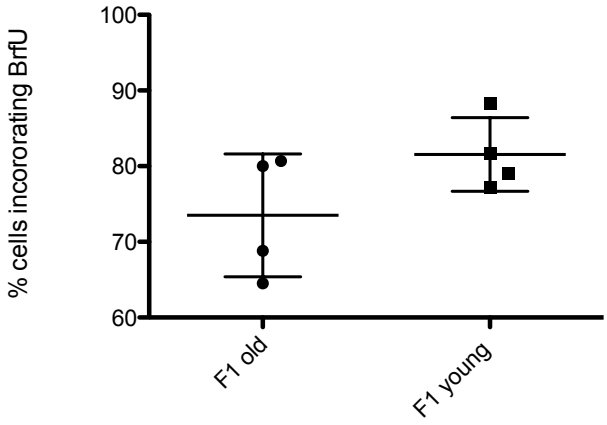


Figure 15. Cell proliferation comparison after F1 young and F1 old serum stimulation

Cell trace proliferation assay

- Based on BrdU results, we wondered what altered the cell proliferation during the culture. CellTrace is used for *in vitro* labeling of cells in order to trace multiple generations by flow cytometry. Figure 16 shows the analysis of CellTrace assay, where the proliferative cells stimulated by F1 young are 84,1% whereas only the 42% of cells stimulating with F1 old serum are proliferating during the *in vitro* culture.

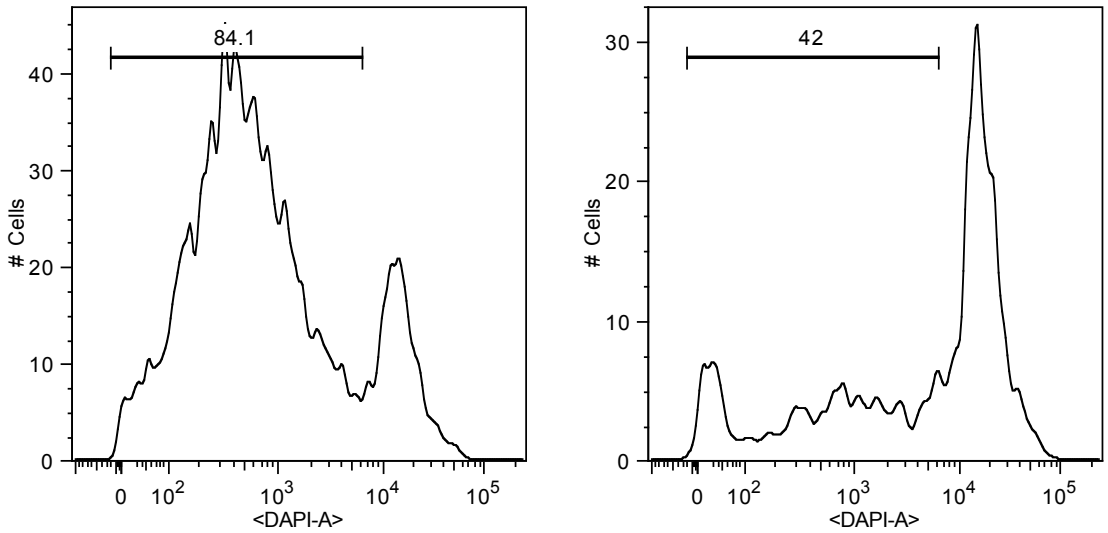


Figure 16. Proliferation of cells stimulated by F1 young (left) and F1 old (right) serum

- In parallel with CellTrace, we used 7-AAD to detect cell death between the samples during the culture (Figure 17). Only 4,55% of the cells with F1 young were detected positive for death; in contrast to 16,4% of the cells stimulated by F1 old were dead. Combination of these two results agrees with our working hypothesis.

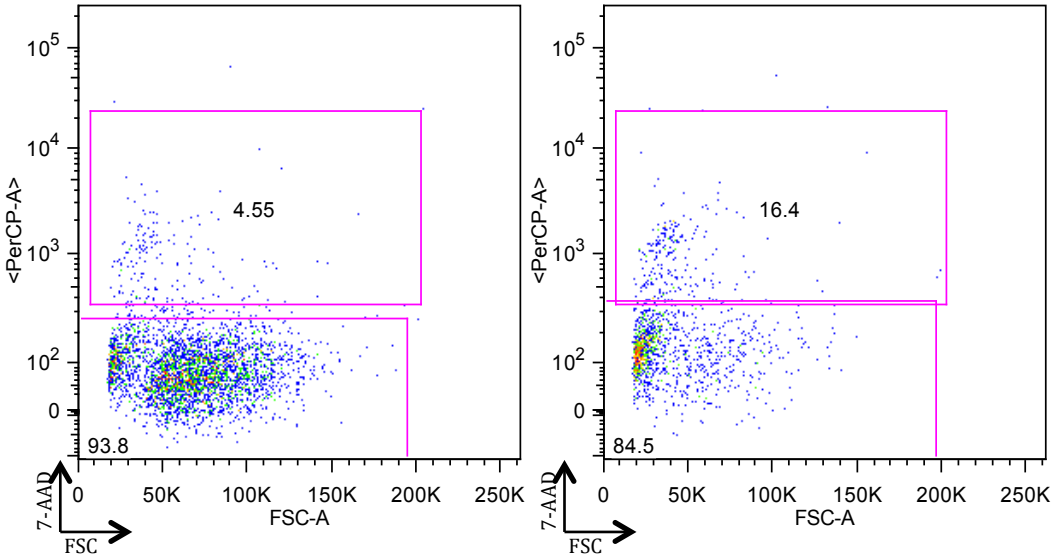


Figure 17. Cell death. Cells with F1 young (left) and F1 old (right) serum

Colony-Forming Cell Assay

To determine the lineage potential of these stimulated hematopoietic stem cells, we performed colony-formation cell assay with or without serum during the semi-solid culture. After 12-15 days, we counted the colonies of each class under inverted microscope. The figures 18 and 19 show the average number of each group between the two different states, stimulated with F1 young or F1 old serum. There seems to be a drift for cells after F1 young serum to skew to myeloid lineage. Because of the number of the experiments, statistical analysis found no significance between the two conditions.
HSCs exposed to inflammation selectively lose lymphoid potential and produce myeloid lineage cells. Findings show that HSCs in mice with active lupus have undergone similar functional alteration as are seen in chronically infected animals. These findings include a greatly expanded HSC population in bone marrow (Figure 9), abnormal accumulation of HSCs in periphery and a biased hematopoietic output skewing toward the myeloid than the lymphoid lineage [18].

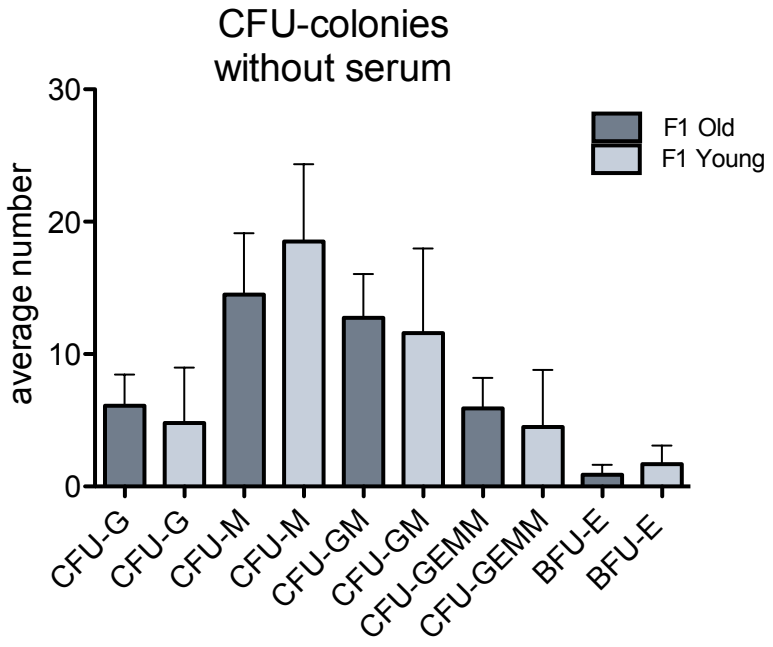


Figure 18. Average number of CFU colonies (n=5)

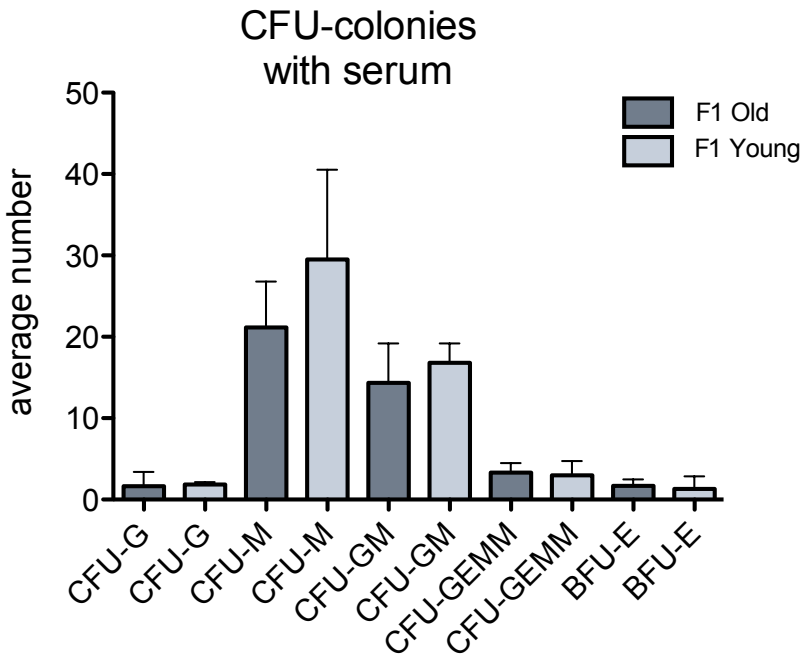


Figure 19. Average number of CFU colonies with serum (n=3)

The findings from the cell cultures are in agreement with our working hypothesis. All indicate that F1 young serum boost LSK cells to proliferate more than the F1 old or the healthy serum. We have excluded the case that cells die with 7-AAD experiment. Moreover, MethoCults exhibit the potential of LSK cells stimulated with different serum each time. F1 young stimulus favors the differentiation towards myeloid lineage. Although there is an obvious trend, statistical analysis didn't conclude to any significance. Therefore there is need for more experiments in order to determine clearly this phenomenon.

Real-Time PCR

Studies propose that the subfractionation of the HSC compartment into dormant and activated phenotypes with distinct rates of cell cycle enter, comprising ~5-10% and 90-95% of the HSC pool, respectively. Dormant HSCs are computed to divide only once every 145d or more, and appear to be enriched for long-term reconstitution potential. This small population of cells may represent a reservoir of HSC activity kept aside in the adult BM to be called upon only by severe hematopoietic injury thus ensuring the maintenance of blood homeostasis.

Studies of aged mice show an overall decrease in HSC cell cycle activity, with old HSCs undergoing fewer cell divisions than young HSCs as assessed by BrdU. These results suggest that Cdks and other activators of the cell cycle may undergo functional decline or that the activity of certain cell cycle checkpoint mechanisms such as CKIs may increase with age, hence delaying HSC entry into the cell cycle [1].

Chronic infections cause dormant HSCs to enter the cell cycle and may accelerate HSC senescence, because HSCs that have been exposed to infections and inflammatory signals manifest functional changes characteristic of natural aging HSCs such as impaired self-renewal capacity and biased hematopoietic output favoring the myeloid over the lymphoid lineage.
p21 and p18 impose a significant impact on HSC quiescence and self-renewal capability. Overexpression of p18 reduces HSC engraftment and defective expression of p18 in lupus mice is at least one of the factors that enhance self-renewal capacity of HSCs (Figure 20) [18].
Mice lacking the G1 checkpoint regulator p21 display a higher rate of HSC proliferation and differentiation and a lower self-renewal capacity, suggesting that p21 is required for maintaining HSC quiescence and that, in its absence, HSCs rapidly proliferate and differentiate to more-committed lineages [19].

Nrf2 is a b-Zip transcription factor, which is ubiquitously expressed as a master regulator in the antioxidant response pathway. Modulation of the redox balance plays a pivotal role in maintaining HSC functions. Nrf2, also, regulates proliferation and differentiation of HSPCs in a cell-intrinsic manner. It acts as a negative regulator of cell-cycle entry in HSCs, actively maintaining the balance between HSC quiescence and self-renewal [20].

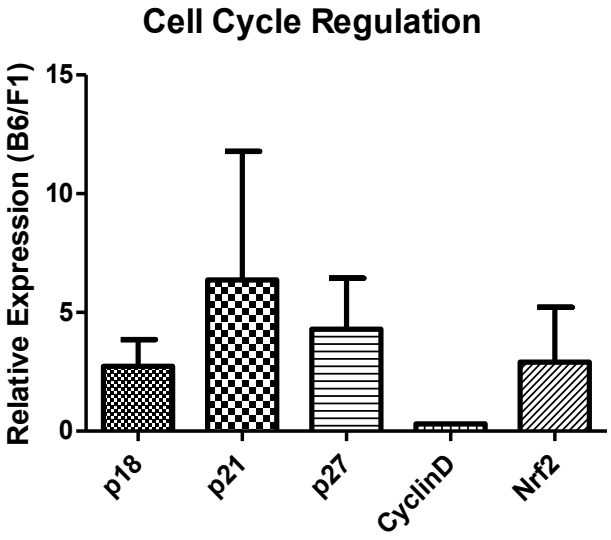


Figure 20. Relative gene expression of cell cycle regulators in HSCs of F1 old

Gene expression results indicate that cell cycle is impaired in HSCs of F1 old mice compared to B6 healthy control. Both CDKs and Nrf2 are downregulated in F1 mice, in contrast with Cyclin D, which prompts cell cycle proliferation. These results in combination with the literature confirm that HSCs of F1 old are pushed to proliferate in order to respond to chronic inflammation.

RNA-seq pipeline

The analysis of RNA-seq data targets to transcript discovery, genome annotation, studying the mechanisms of gene regulation, differential gene expression analysis, allele-specific expression analysis, detection of RNA editing, gene fusion detection and other types of variant detection.

- The raw data after RNA-seq were checked with FastQC program for the quality of the sequencing [Appendix]. Figure 21 is an example depicting the quality of a sample. The y-axis on the graph shows the quality scores. The higher the score the better the base call is. The background of the graph divides the y-axis into very good quality calls (green), calls of reasonable quality (orange) and calls of poor quality (red). The quality of calls on most platforms will degrade as the run progresses, so it is common to see base calls falling into the orange area towards the end of a read. The next figure depicts the per sequence quality score which allows to check if a subset of the sequences have universally low quality values. It is often the case that a subset of sequences will have universally poor quality; since they are poorly imaged (on the edge of the field of view), however these should represent only a small percentage of the total sequences.

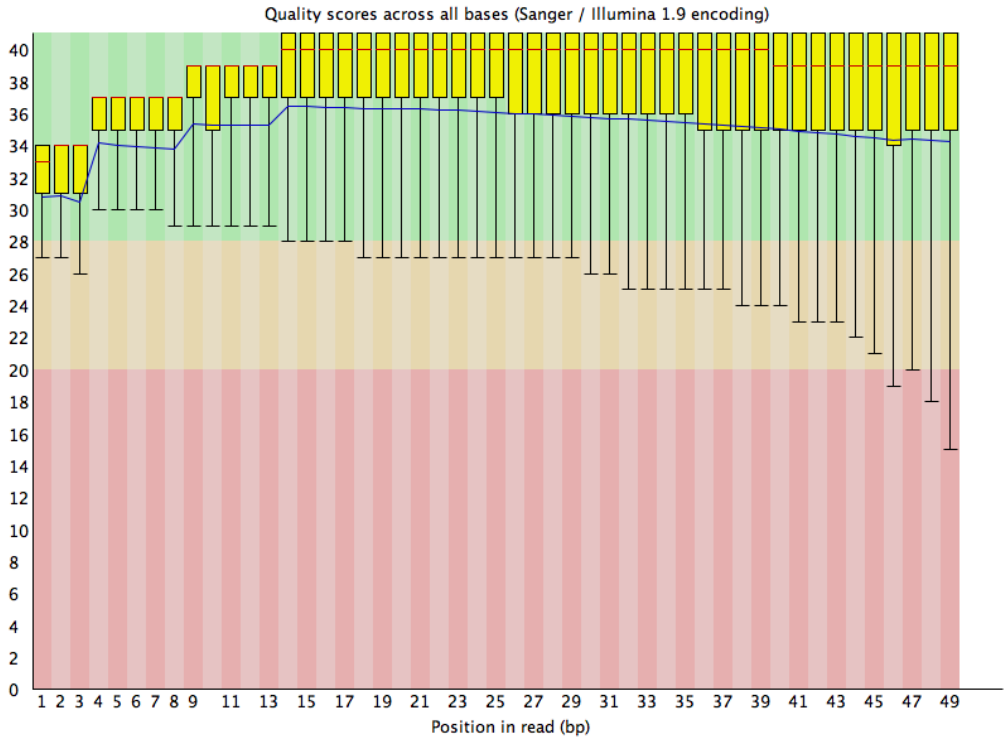


Figure 21. Overview of the range of quality values across all bases at each position in the FastQ file.

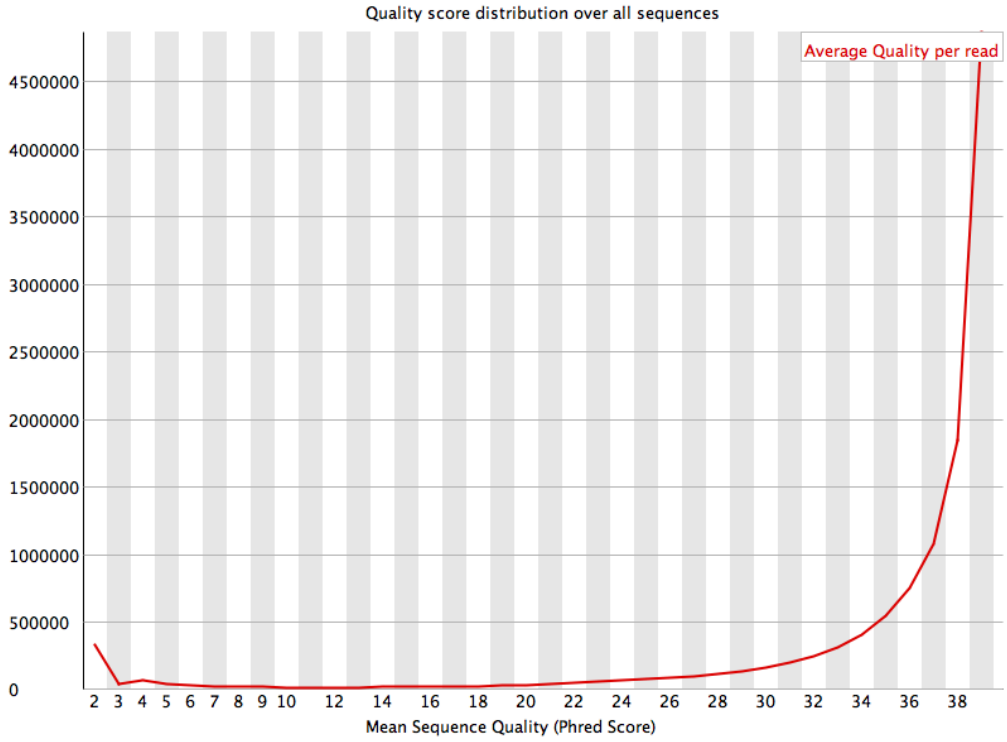


Figure 22. The per sequence quality score report

- As mentioned in "Materials and Methods", we followed three different approaches in order to analyze the data. The discussion of the expression profile below refers to F1 young versus to F1 old comparison. There are also comparisons between C57/BL6 and F1 young, as well as C57/BL6 and F1 old (not shown). The first approach resulted in 243 differentially expressed genes (DEGs) in LSK compartment between F1 young and old mice. In the second workflow, data were aligned by Tophat2 and analyzed with Cufflinks pipeline, gave us 547 DEGs between these two conditions. By the third approach, which was used also Tophat2 for alignment, and subsequently MetaSeq R package, we came up with 362 differentially expressed genes. Each pipeline list was filtered with threshold 1.5x fold change and p value 0.05; except the last one, which was filtered with 2x fold change. The different results between the approaches are expected because the algorithms use different parameters. We choose to focus on the second workflow, which results are listed below.

- Heatmap in Figure 23 is a visualization of all differentially expressed genes between F1 young and F1 old hematopoietic stem cells. Some areas are evident where there are genes expressed in HSCs in F1 young upregulated, whereas the expression in F1 old HSCs is downregulated and vice versa. These findings confirm our hypothesis that hematopoietic stem cells may have intrinsic –genetic or epigenetic- alterations, which can affect the next lineages.

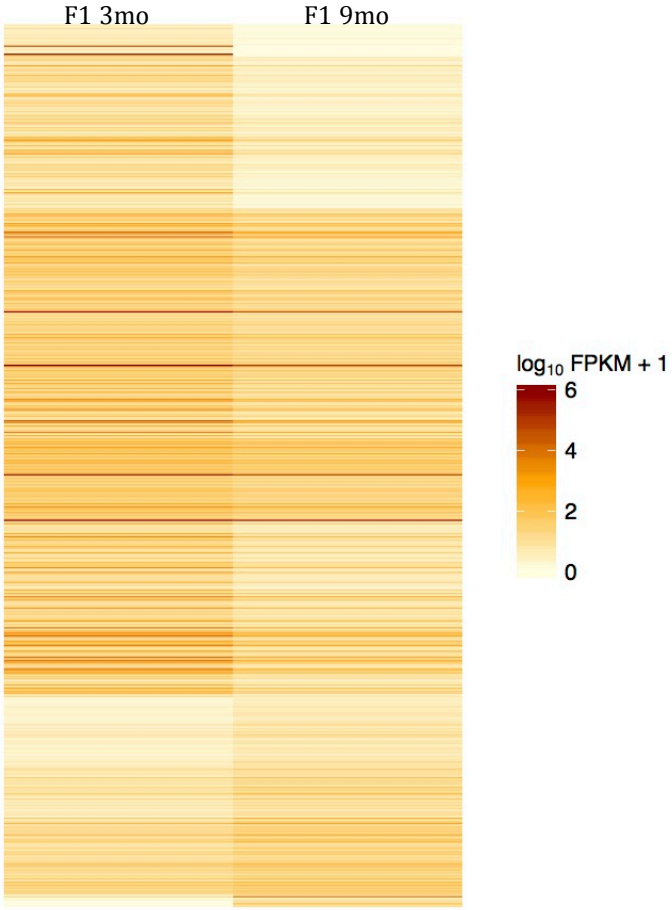


Figure 23. Differentially Expressed Genes visualized at heatmap

Top DOWN	Top UP
Sec14l5	Gzmb
Il5ra	Ppp3cc
Fut1	Slc22a3
Prlr	Rps23
2300002M23Rik	Nrn1
Tnfrsf17	1100001G20Rik
Stac2	Klrg1
Krt222	Fgf3
Igj	Ear2,Ear3
F9	Gzmk
Oosp1	Rpl37
Wnt10a	Ear6,Ear7
Pls1	Epx
Derl3	Spib
Chst1	Ssx2ip
Chac1	Fpr1
Bhlhe41	Pum1
Bmp6	Chd7
Eaf2	Rpl35
Gpm6a	Wnk1

Table 2. Top down- and up- regulated interesting genes

In table 1, there are the top twenty upregulated and downregulated of the differentially expressed genes in hematopoietic stem cells between young and old F1 mice. They contain crucial transcription factors i.e. for development as Wnt10a, Bmp6; genes involved in inflammation and immune response as Il5ra; as well as factors responsible for cell homeostasis as Slc22a3.

- IPA is a very useful tool for identifying relationships, mechanisms, functions and pathways of relevance. It can build a complete regulatory picture and a better understanding of the biology underlying a given gene expression set. It categorizes its findings based on a in-house curated ontology, covering entities (proteins and other molecules), relationships between the entities and functional information (pathways, biological process, disease, etc). IPA functional analysis organizes biological information for a high-level overview. It provides access to detailed information on functions and the molecules involved. Functional analysis categories contain high quality gene ontology information as well as curated IPA content ie diseases and disorders

Canonical pathways (signaling and metabolic) are generated prior to data input, based on the literature. Upon input genes, networks are generated de novo based on input. The figures below (24-27) present pathway networks extracted from differentially expressed genes in HSCs between F1 young and F1 old mice

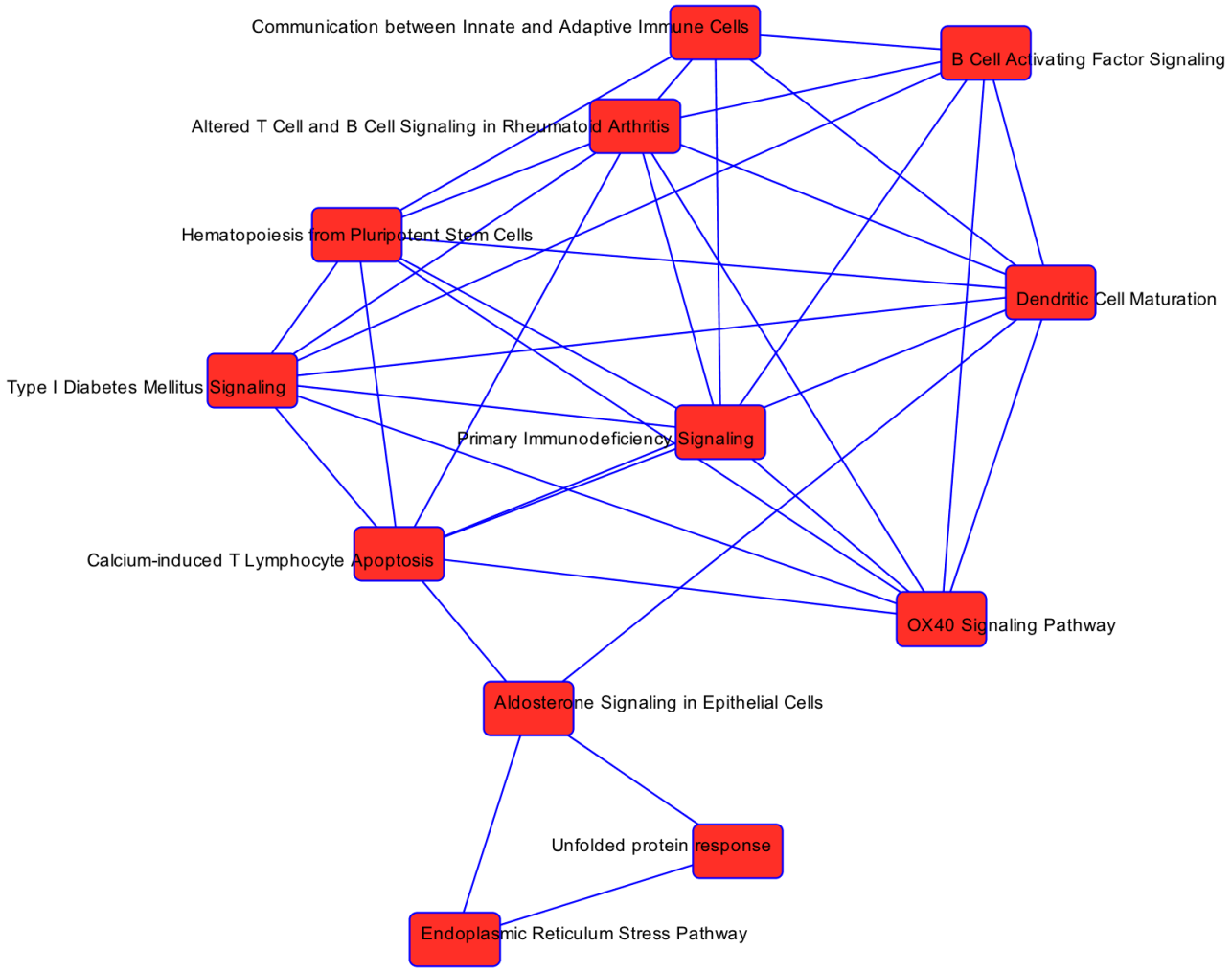


Figure 24. The first 12 canonical pathways

The molecules which are implicated are crucial for antigen-presenting response, cell homeostasis all of them linking with NFκB, crucial molecule for cell development.

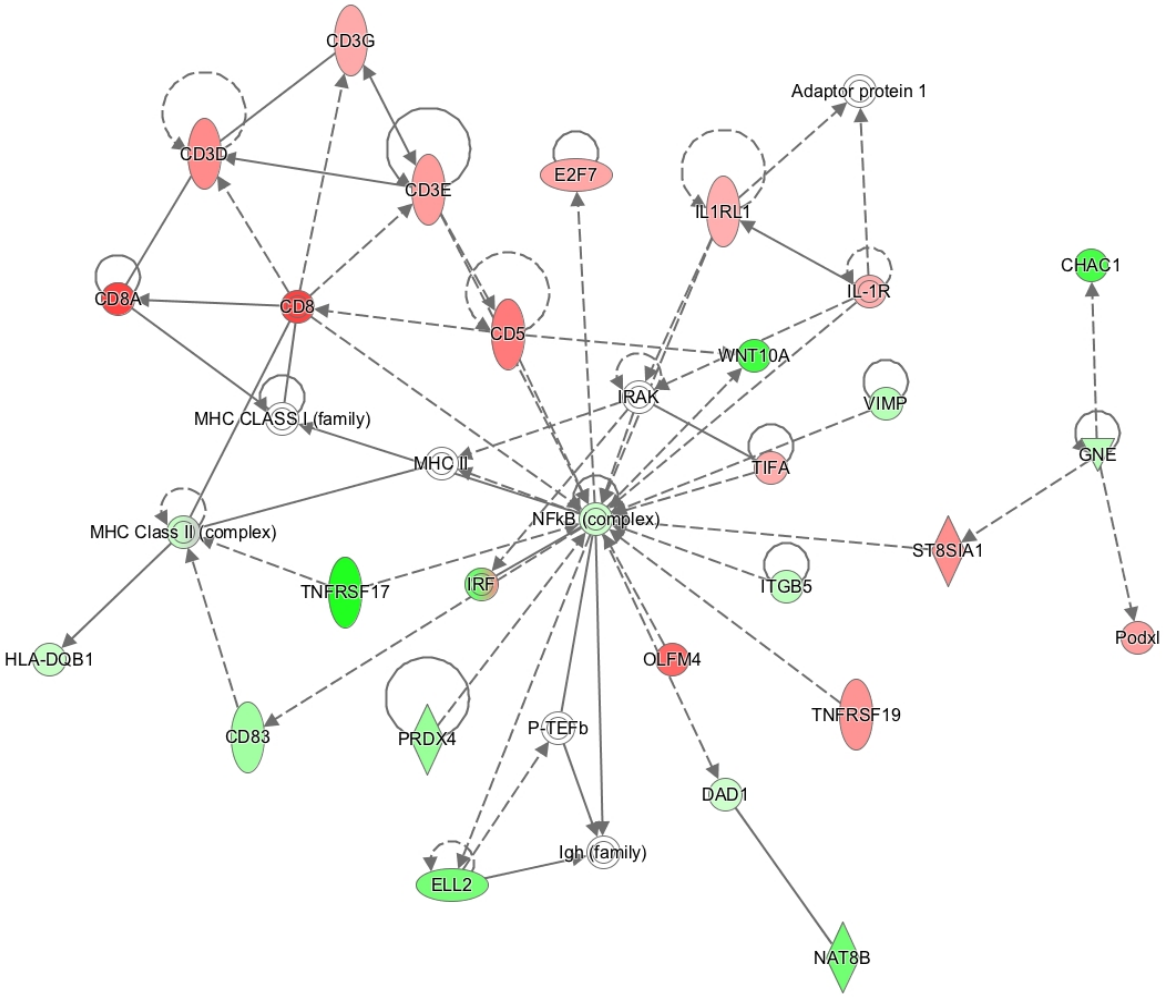


Figure 25. Hematological disease network

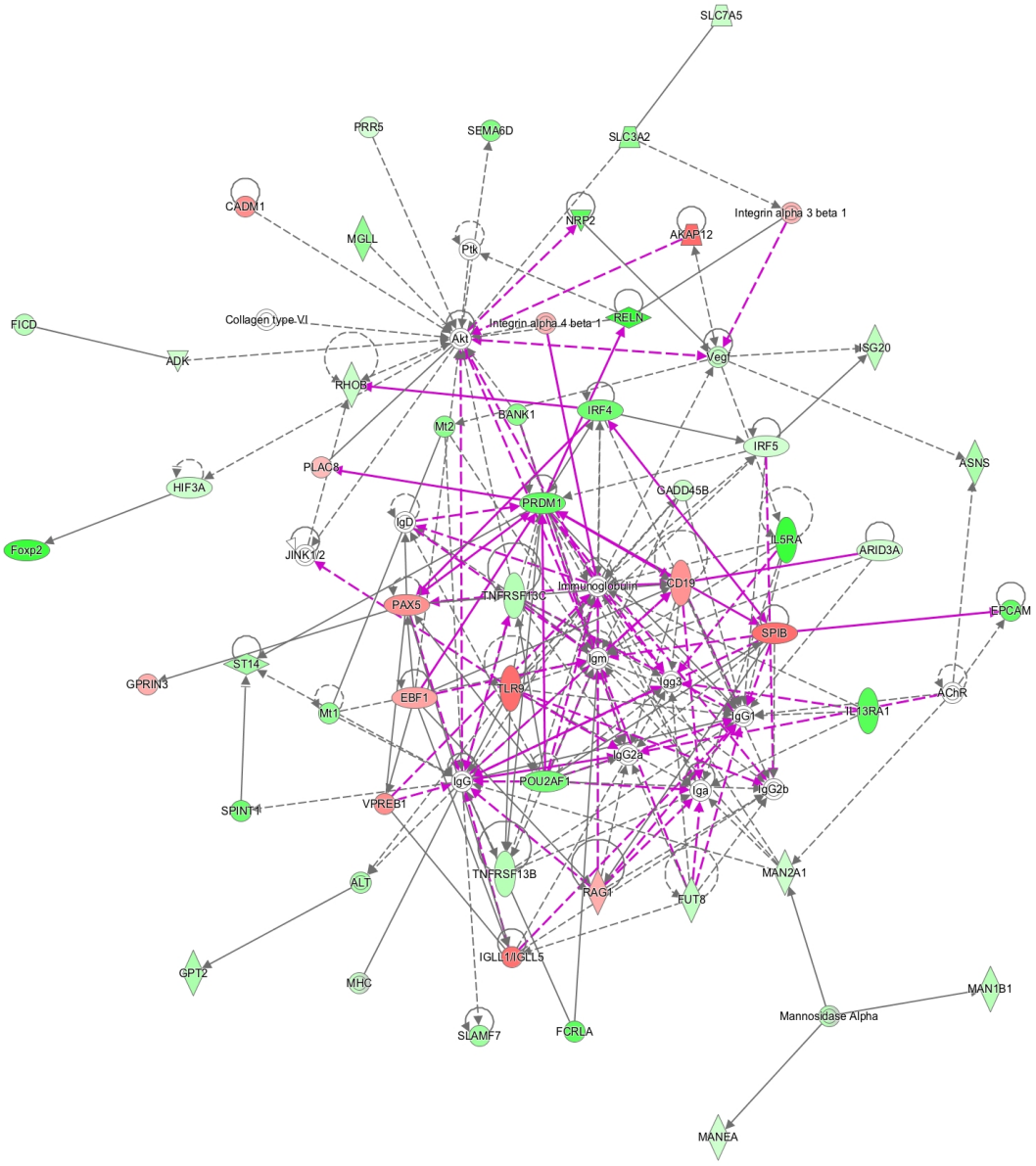


Figure 26. Immunological and inflammatory response network

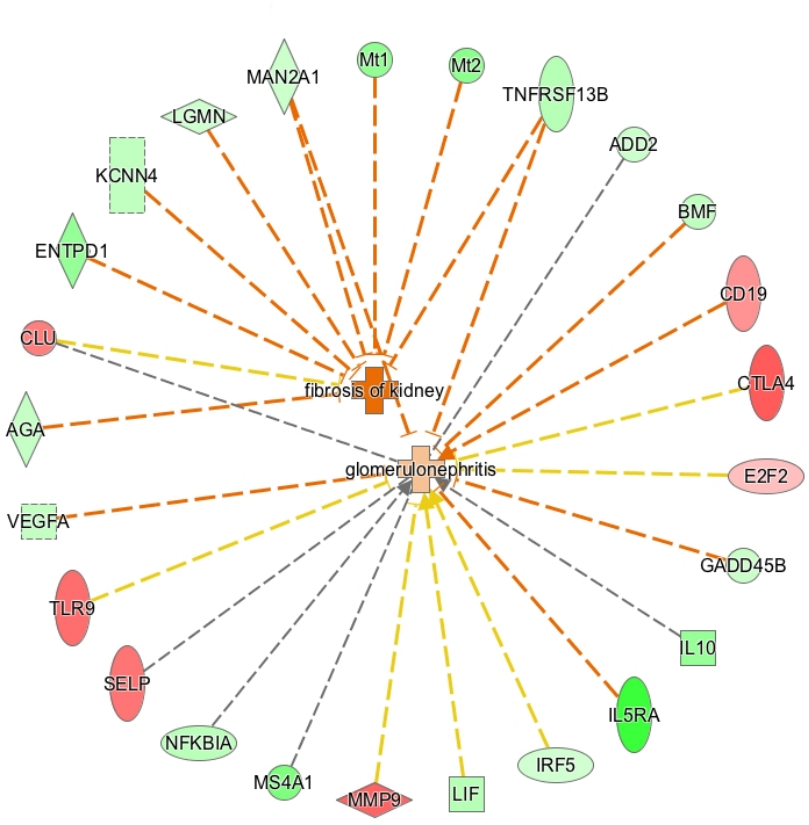


Figure 27. Networks that are involved in kidney fibrosis and glomerulonephritis

The table below presents some target genes of significant differentially expressed genes. Considering their function in conjunction with the disease relevance, we plan to find new therapeutic targets or diagnostic biomarkers. Further analysis with functional experiments *in vitro* and *in vivo* is needed.

Transcriptional Regulators	**Developmental Regulators**	**Immunological Regulators**	**Long non-coding RNAs**
Ezh1/2	Vegfa	Ctla4	Lincrna-Cox2
Prdm1	Cdkn2c	Ly6ca2	Il1bos
Kdm2a	Bmp6	Maf	
Ehd1	Epcam	IfnaR	
E2f family	Cdk6	Ifna1	
Bcl2	Casp8ap2	Irf8	
Hif1a	Cccnd1	Vimp	
Pou2af1	Cxcl10	Nfkbia	
	Lamp2	H2-Aa	
	S100A-family	H2-Ab1	

Table 3. Differentially expressed target genes in F1 old/F1 young genesets

Appendix

Bioinfomatics analysis

We are interested to find differentially expressed genes between F1 young (3mo) and F1 old (9mo) mice. Samples from C57/BL6 mice were also used as internal control. The analysis, which is presented below, refers to the comparison between diseased mice of different age/symptoms.
Please find all the extras files in this link below
https://www.dropbox.com/sh/m9xl95rw0twp3gt/AADpLFbf6jaGoc4Ca54jPzEua?dl=0

Useful tools

- ✓ <cksum> check the file's quality after delivering from sequencing (i.e. damaged)
- ✓ <fastqc> check the alignment quality (alignment tools: GemTools, Tophat2)
- ✓ <samtools sort 'input_name'.bam 'output_name'.bam> sort the bam files by name because of paired-end sequence
- ✓ <samtools view -h 'input_name'.bam > 'output_name'.sam> convert bam to sam files

Workflow (1)

1. HTseq-count

Measurement of the relative expression of genes by making initial measurements of correlations in the genomic region of the gene.

<htseq-count -a 1 sorted_input.sam ~/Mus_musculus_UCSC_mm9-1/Mus_musculus/UCSC/mm9/Annotation/Genes/genes.gtf > output_file.out>

2. DESeq

Transcripts normalization based on the size of each library of sequences.

Workflow (2)

1. Tophat2

It should be noted that reads with multiple correlations of the sequences in different parts of the genome were rejected, and kept only those which had greater reliability score.

```
tophat2 -p 20 -o tophat2_${input}
~/Mus_musculus/UCSC/mm9/Sequence/Bowtie2Index/genome
~/fastq/${input}_1.fastq ~/fastq/${input}_2.fastq 2>${input}.ERR
```

2. Cufflinks

```
cufflinks -p 18 ~/accepted_hits.bam -o ~/$output
```

3. Cuffmerge

cuffmerge -p 18 -o output_merge -s
~/Mus_musculus/UCSC/mm9/Sequence/WholeGenomeFasta/genome.fa -g
~/Mus_musculus/UCSC/mm9/Annotation/Genes/genes.gtf GTF_list.txt
GTF_list.txt includes all the directories from output transcripts.gtf from previous step (cufflinks)

4. Cuffdiff

cuffdiff -p 20 -o 'output_directory' --labels 3moF1,9moF1 -b
~/Mus_musculus/UCSC/mm9/Sequence/WholeGenomeFasta/genome.fa -u
~/cuffmerge/output_merge/merged.gtf
~/tophat/tophat2_F13MA_150226_5/accepted_hits.bam,~/tophat/tophat2_F13MB_150226_6/accepted_hits.bam
~/tophat/tophat2_F19MA_150226_7/accepted_hits.bam,~/tophat/tophat2_F19MB_150226_8/accepted_hits.bam

5. Cummerbund: visualization of the results

Workflow (3)

1. Tophat2

tophat2 -p 20 -o tophat2_${input}
~/Mus_musculus/UCSC/mm9/Sequence/Bowtie2Index/genome
~/fastq/${input}_1.fastq ~/fastq/${input}_2.fastq 2>${input}.ERR

2. metaSeq https://pythonhosted.org/metaSeq/

On our final lists include genes that exhibit differential expression between the two conditions.
log(expr(treatment)/expr(control)) > 1.5 (or 1) -> upregulation
log(expr(treatment)/expr(control)) < -1.5 (or 1) -> downregulation

References

1 Pietras, E. M., Warr, M. R. & Passegué, E. Cell cycle regulation in hematopoietic stem cells. *The Journal of cell biology* **195**, 709-720 (2011).

2 King, K. Y. & Goodell, M. A. Inflammatory modulation of HSCs: viewing the HSC as a foundation for the immune response. *Nature Reviews Immunology* **11**, 685-692 (2011).

3 Seita, J. & Weissman, I. L. Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **2**, 640-653 (2010).

4 Geiger, H., de Haan, G. & Florian, M. C. The ageing haematopoietic stem cell compartment. *Nature Reviews Immunology* **13**, 376-389 (2013).

5 Fleming, H. E. *et al.* Wnt signaling in the niche enforces hematopoietic stem cell quiescence and is necessary to preserve self-renewal in vivo. *Cell stem cell* **2**, 274-283 (2008).

6 Morgan, D. O. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annual review of cell and developmental biology* **13**, 261-291 (1997).

7 Banchereau, R., Cepika, A.-M. & Pascual, V. Systems approaches to human autoimmune diseases. *Current opinion in immunology* **25**, 598-605 (2013).

8 Mohan, C. & Putterman, C. Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nature Reviews Nephrology* **11**, 329-341 (2015).

9 Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics* **12**, 87-98 (2011).

10 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57-63 (2009).

11 Van Keuren-Jensen, K., Keats, J. J. & Craig, D. W. Bringing RNA-seq closer to the clinic. *Nature biotechnology* **32**, 884-885 (2014).

12 Ferreira, P. G. *et al.* Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome research* **24**, 212-226 (2014).

13 Costa, V., Aprile, M., Esposito, R. & Ciccodicola, A. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. *European Journal of Human Genetics* **21**, 134-142 (2013).

14 Lilljebjörn, H. *et al.* RNA-seq identifies clinically relevant fusion genes in leukemia including a novel MEF2D/CSF1R fusion responsive to imatinib. *Leukemia* **28**, 977-979 (2014).

15 Pascual, V., Chaussabel, D. & Banchereau, J. A genomic approach to human autoimmune diseases. *Annual review of immunology* **28**, 535 (2010).

16 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* **7**, 562-578 (2012).

17 Griffith, M., Walker, J. R., Spies, N. C., Ainscough, B. J. & Griffith, O. L. Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. *PLoS Comput Biol* **11**, e1004393 (2015).

18 Niu, H. *et al.* The function of hematopoietic stem cells is altered by both genetic and inflammatory factors in lupus mice. *Blood* **121**, 1986-1994 (2013).
19 Yahata, T. *et al.* Accumulation of oxidative DNA damage restricts the self-renewal capacity of human hematopoietic stem cells. *Blood* **118**, 2941-2950 (2011).
20 Tsai, J. J. *et al.* Nrf2 regulates haematopoietic stem cell function. *Nature cell biology* **15**, 309-316 (2013).