Uncovering Differential Gene Expression Patterns between Human Placenta and Bone marrow

Introduction

Placenta is a highly vascularized organ essential for oxygen and nutrient transport for healthy development of the fetus. It also serves as a site for the formation of Hematopoietic Stem Cells in fetus before bone marrow becomes responsible for this function. In an adult, bone marrow is the primary organ for the production of hematopoietic stem cells. These cells act as precursor to all the blood cells including red blood cells, white blood cells and platelets.

Though placenta and bone marrow have shared function, they are active at different stages of human development. Bone marrow is a lifelong supplier of blood cells whereas placenta is responsible for the overall embryonic development including nutrient and waste exchange between mother and fetus. [1]

Each tissue expresses different genes. Genes related to hematopoiesis are highly expressed in bone marrow and gene related to growth hormone, cell cycle regulation, cellular transport and blood vessel development are highly expressed in placenta. Studying the differential gene expression helps us in the identification and quantification of the gene expression in different tissues. These studies can reveal the function of these genes in various biological processes and their connection with disease and drug treatment. This report focuses on the gene expression analysis between placenta and bone marrow. The hypothesis driving this project is that low expression of genes involved in hematopoiesis and immune regulation in placenta compared to bone marrow may not support long-term production of blood cells leading to blood disorder and immunodeficiency in infants.

Methods

1. **Data Collection**
   The RNA-Seq data of placenta and bone marrow used in this study are sourced from Uppsala Biobank. 3 biological replicates for each sample were chosen.

2. **Preprocessing**
   - **Quality Control:**
     Quality check was performed using **FastQC** tool to evaluate the samples based on per base sequence quality and content, adapter content, sequence length distribution, overrepresented sequences, etc.

   - **Trimming:**
     Quality control was followed by trimming of low-quality reads (Q<30), short reads (<50 bp) were eliminated using **Trim Galore** for proper mapping.

   - **Mapping:**
     Filtered reads with high quality were then mapped to **GRCh38** reference genome by **HISAT2**. It allows counting of reads aligned to each gene in reference. The resulting file is stored in BAM format.

   - **Counting:**
     featureCounts is used to count the reads mapped to each gene. This results in a matrix with each row as gene and values in column shows the number of reads mapped to the gene.

3. **Differential Expression Analysis**
   edgeR package in **RStudio** was used to perform differential gene expression analysis.

   Normalization
   The data was normalized to correct the variations in library size. This adjusts the compositional differences in library size to make expression level more comparable between samples.

   Statistical Testing
   The **exactTest** function was used to compare bone marrow and placenta samples. It estimates the dispersion to test for differences in expression levels between bone marrow and placenta.

   Threshold for significance
   Threshold of FDR <0.05 was set to identify statistically significant genes. Additionally, genes showing log2 fold change less than -1 or greater than 1 were considered significant.

Results

1. **Preprocessing**

   Raw RNA Seq data was subjected to multiple preprocessing steps to ensure good quality reads for differential gene expression analysis. FastQC showed high sequence quality for most data but few indicated uneven nucleotide distribution across bases, and adapter contamination was noticed in some samples.

   Trim Galore trimmed these low-quality reads and samples contaminated with adapters were cleaned for accurate results.

   HISAT2 aligned these reads to human reference genome.
   Bone marrow samples showed 99% read alignment.
   Placenta samples showed 99% read alignment.

2. **Differential Expression Analysis**

   DGE list containing gene expression count matrix was created for each tissue sample and normalization was performed to extract top differentially expressed genes. 6186 genes were highly expressed in placenta and 4496 genes were highly expressed in bone marrow. The expression of **DLK1, EGFL6, COL4A2, CAPN6, COL4A1** was more seen in placenta and **MPO, AZU1, CTSG** were bone marrow specific genes.

3. **Visualization**

- **Multi-Dimensional Scaling Plot (Figure 1):**
  MDS shows replicates of same sample clustered together indicating similarity in their expression pattern.

- **Principal Component Analysis Plot (Figure 2):**
  Bone marrow and placenta samples show highly distinct gene expression profile.

- **Heatmap (Figure 3):**
  Top 100 differentially expressed genes were studied in the heatmap. Bone marrow specific genes and placenta specific genes clustered separately.

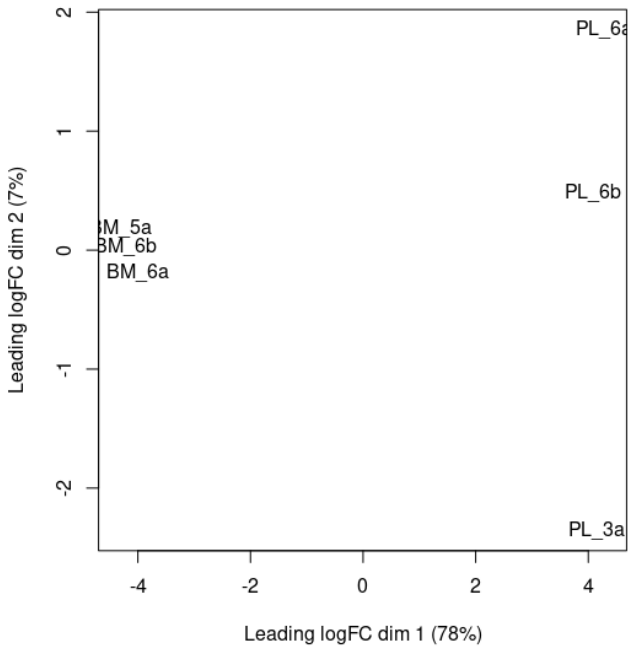

Figure 1. MDS Plot

Bone marrow and placenta forms two separated clusters indicating strong difference in overall gene expression. Tight clustering of sample indicates low variability within each group.

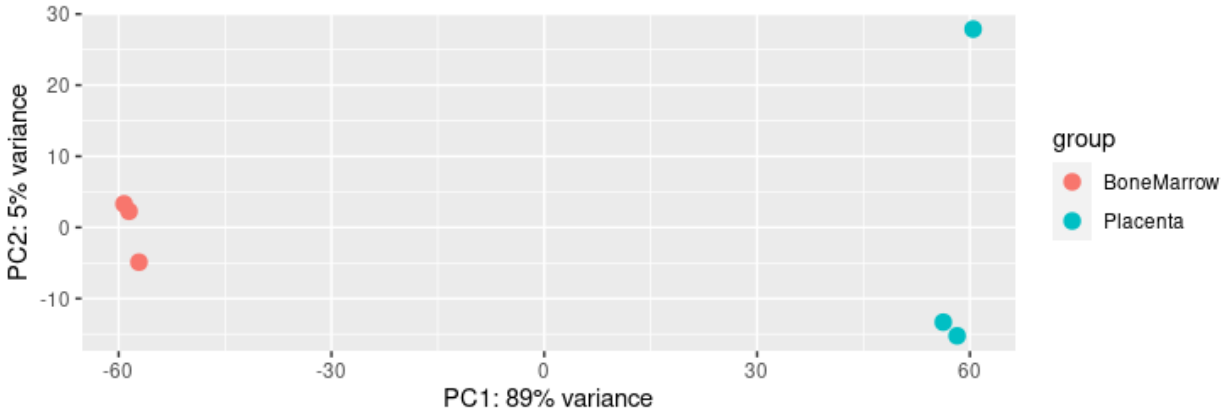

Figure 2. PCA Plot

Bone marrow samples are grouped together on the left whereas placenta samples are clustered on the right demonstrating high consistency within groups. Tissue type (PC1 with 89% variance) separates the two samples.

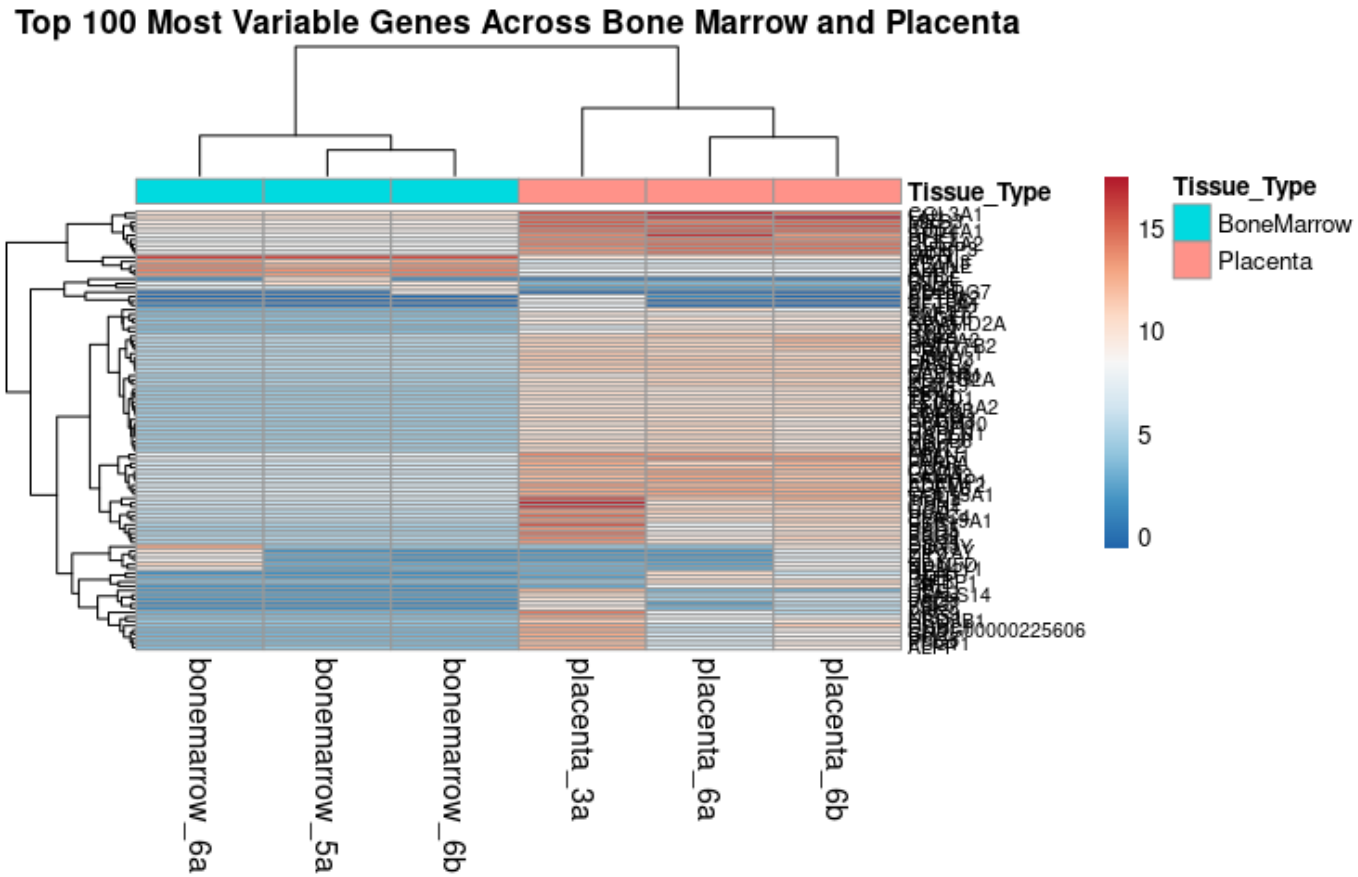

Figure 3. Heatmap

Genes such as DLK1, EGFL6, COL4A2, CAPN6 showed strong expression pattern in placenta reflecting their role in vascularization, tissue remodeling, cell differentiation in placenta for fetal development. MPO, AZU1, CTSG genes were upregulated in bone marrow as these genes are related to hematopoiesis and immune defense.

Tables

Table1. Gene Expression Regulation in Placenta-Bone marrow

Regulation Status	Number of Genes
Upregulated in Placenta	6186
Upregulated in Bone marrow	4496
Not significant	66,629

Table 2. Top Genes with Differential Expression in Placenta and Bone marrow

Gene	LogFC	LogCPM	p-value	FDR
DLK1	14.62288	9.978944	3.174601e-75	2.454316e-70
EGFL6	15.03279	8.701953	2.706645e-71	1.046267e-66
COL4A2	11.95712	10.400691	9.248565e-69	2.383386e-64
CAPN6	15.42297	7.145907	104617e-68	7.933302e-64
COL4A1	12.21574	10.897552	5.950892e-68	9.201389e-64
PAGE4	15.46058	7.183490	3.187225e-67	4.106793e-63
MPO	-10.56083	11.579160	8.343941e-67	9.215406e-63
AZU1	-11.63981	9.744137	1.631443e-66	1.576607e-62
LIPG	12.71926	7.186288	3.880884e-66	3.333722e-62
CTSG	-10.51636	9.637341	6.096452e-66	4.713228e-62

Table 3. Differentially expressed genes and their function

Gene	**Function of Gene**	**LogFC**
DLK1	Negative Regulator of hematopoietic stem cells	14.62288
EGFL6	Angiogenesis, tissue development and remodeling	15.03279
CAPN6	Regulation of microtubule dynamics and cytoskeletal organization	15.42297
COL4A1	Helps in the formation of type IV collagen in basement membrane	12.21574
MPO	Boosts innate immunity for host defense	-10.56083
AZU1	Involved in immune defense	-11.63981
CTSG	Destruction of pathogen and inflammation regulation	-10.51636

References

[1] Placenta as a source of hematopoietic stem cells
https://pmc.ncbi.nlm.nih.gov/articles/PMC3586314/#:~:text=HSCs%20can%20be%20found%20in,human%20term%20placenta%20was%20unexpected

[2] DLK1 is a negative regulator of hematopoietic stem and progenitor cells
https://www.researchgate.net/publication/229156097_Dlk1_is_a_negative_regulator_of_emerging_hematopoietic_stem_and_progenitor_cells

[3] COL4A1 Gene

https://medlineplus.gov/genetics/gene/col4a1/#conditions

[4] MPO plays an important role in innate immunity for host defense against invading microorganisms
https://pmc.ncbi.nlm.nih.gov/articles/PMC7988577/#:~:text=As%20a%20component%20of%20the,O2)%E2%80%90mediated%20reactions.

[5] Role of AZU1 gene in bone marrow
https://www.proteinatlas.org/ENSG00000172232-AZU1

ADDITIONAL WORK

Further analysis is required to investigate the differential expression analysis of genes in bone marrow and placenta.

1. **Pathway Analysis**

Enrichment analysis will be performed using **DAVID** and **Enrichr** to identify major biological processes and pathways regulated by these differentially expressed genes. This is help us understand cellular component of the tissues, their molecular function and biological processes specific to placenta and bone marrow.

2. Analysis of protein interaction network:

Identification of key protein hubs and interaction between top DEGs will be studied using **Cytoscape.** This tool will help in visualizing and understanding protein hubs in networks of bone marrow and placenta.

3. **Gene-Disease Enrichment**

This aims to check is the differentially expressed genes between placenta and bone marrow are involved in causing blood related disorder or immunodeficiency.

4. **Literature Review**

Publications will be reviewed to validate the results by cross referencing with published articles stating any relation of DEGs in bone marrow and placenta with blood disorder or immune dysfunction.