**The Roles of Prostate and Testes in Male Reproductive System**

**Abstract:**

This study uses differential expression analysis and enrichment analysis to prove whether or not the prostate and testes are in charge of making the same amounts of primary components of semen, such as sperm, proteolytic enzymes and fibrinolysin. Prostate and testes are both parts of the male reproductive system and seminal plasma, or semen, consists of a complex mixture of fluids that is produced from the testes, epididymis and male accessory glands. The results will show that the prostate and testes are two drastically different tissues in regard to their functions in the male reproductive system, and sperm and fibrinolysin are exclusively generated in testes and prostate, respectively.

**Introduction:**

Prostate and testes are both parts of the male reproductive system. The testes, located in the scrotum, generate sex hormones and produce sperm for reproduction. The prostate is an exocrine gland which produces fluid using the hormones generated from testes for semen. The semen, or seminal plasma, consists of a complex mixture of fluids that is produced from the testes, epididymis and male accessory glands. The prostate plays a significant role in ejaculation (Verze, P., et al). This study is intended to discover if prostate and testes are in charge of making the same amounts of major elements of seminal plasma, such as sperm, proteolytic enzymes and fibrinolysin. To test the hypotheses, I will be using the data of three replicates of human prostate and three replicates of human testis to run a differential expression analysis and an enrichment analysis. Before the analysis, the data sets will go through the data process of quality control, trimming, mapping and feature counting.

**Method:**

The 6 single-end mRNA-Seq raw datasets (three of the prostate and three of the testis) were from *Georgetown-Ross-Human-Mapping-Data* library.

FastQC, a read quality report tool, version 0.11.3 from Globus Genomics (pilot1.globusgenomics.org) was used for quality control. In total, six replicates of prostate and testis were fed into the FastQC separately. And default settings were used for all six replicates.

I used TrimGalore (version: 0.4.3, Cutadapt version: 1.2.1) from Globus Genomics (pilot1.globusgenomics.org) to trim the replicates separately. Default settings were used. Using the default settings, bases that had a Phred quality score less than 20 and adaptor sequence “AGATCGGAAGAGC” were trimmed. Reads that were less than 20 bases after trimming and those that did not have a paired-end mate were filtered out.

Read mapping was performed using HiSAT(version: 2.0.5, samtools version: 1.2) from Globus Genomics (pilot1.globusgenomics.org). Replicates were processed as single end reads and were aligned against GRCh38 as the reference genome. The parameter of primary alignments was set to 5. All the other parameters were set as default.

I assembled and counted the transcript using featureCounts (subread-1.4.6-p1), a tool that measures gene expression in RNA-Seq experiments from BAM files. All six mapped replicate BAM files were put into featureCounts. Build-in index was used to fit the reference. GFF gene identifier under featureCounts parameters settings was set to “gene\_name” so the name index column was “gene\_name” instead of “gene\_id”. This helped minimize steps for the following differential analysis. All the other parameters were set as default. A tabular file, which had the counts of all the features for six replicates were created in this step. The same tabular file was used in the differential analysis.

Plot graphs were created and differential analysis were performed in RStudio (Version 1.2.5033) using R language. Pheatmap was used to make heatmap, RColorBrewer was in charge of color palettes for heatmaps, BiocManager was used for managing Bioconductor packages, org.Hs.eg.db was used for converting gene names and ENSEMBL ids in mapping, DESeq2 was utilized for visualization and analysis of RNAseq data, and genefilter was relied upon for filtering genes in high-throughput datasets. EdgeR, statmod for statistical modeling and org.Hs.eg.db were used in differential expressions analysis. The tabular file obtained from the previous step was imported into RStudio for plot making and differential analysis. A text file which consisted of gene symbols and their logFC, logCPM, PValue and FDR was made after the analysis.

Enrichment analysis was performed using Enrichr, a web-based enrichment analysis tool (amp.pharm.mssm.edu/Enrichr/). Significant genes were selected based on the threshold in which logFC was either greater than 5 or less than -5 (excluding less differentially expressed genes) and FDR was less than 0.05. GO molecular function and GO Biological Processes were both under the category of Ontologies, and KEGG Human was under pathway category. The bar chart was ordered by combined scores in descending order.

**Results**  
Quality Control (QC)

Per base sequence content, adapter content and Kmer content showed failures, and sequence duplication levels and overrepresented sequences showed warnings for prostate replicate 4a and 4c. For prostate replicate 4b, per base sequence content, adapter content, Kmer content and overrepresented sequences showed failures, and sequence duplication levels showed warning. Per base sequence content and Kmer content showed failures, and per tile sequence quality and sequence duplication levels showed warnings in all three of the testis replicates. Testis\_7a’s adapter content, and testis\_7c’s adapter content and overrepresented sequences showed failures. testis\_7b’s overrepresented sequences and adapter content, and testis\_7a’s overrepresented sequences showed warnings. The GC content, sequence length and poor-quality sequences were almost the same across both prostate and testis. However, the total sequences of prostate\_4c was a little shorter than others (Table 1).

Non-random priming during library preparation was the most likely reason for the failure in per base sequence content and Kmer content. Adapter content failure was likely due to adapter contamination. Sequence duplication levels and overrepresented sequences were expected to show warnings or failures because RNA-Seq data was used in this analysis.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Replicates** | **Total Sequences** | **Poor-quality sequences** | **Sequence length** | **%GC** |
| prostate\_4a | 20000000 | 0 | 101 | 48 |
| prostate\_4b | 20000000 | 0 | 101 | 48 |
| prostate\_4c | 17735105 | 0 | 101 | 48 |
| testis\_7a | 20000000 | 0 | 101 | 47 |
| testis\_7b | 20000000 | 0 | 101 | 48 |
| testis\_7c | 20000000 | 0 | 101 | 48 |

Table 1: Result of QC

Trimming

To remove low quality sequence information and adaptor sequences from the ends, the researcher performed individual trimming on the replicates. The average ratio of trimmed reads for prostate replicates (49.1%) was greater than it was for testis replicates (46.1%). Similarly, the average ratio of trimmed bases and the average ratio of quality-trimmed of prostate replicates were greater than they are for testis replicates (shown in Table 2). Therefore, the quality of the prostate replicates were more consistent than the testis replicates.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Replicates** | **Processed reads** | **Processed bases (bp)** | **Trimmed reads** | **Quality-trimmed (bp)** | **Trimmed bases (bp)** | **Adapter ('AGATCGGAAGAGC') trimmed (times)** |
| prostate\_4a | 20000000 | 2020000000 | 9798600 (49.0%) | 51398825 (2.54% of total) | 118079924 (5.85% of total) | 8978182 |
| prostate\_4b | 20000000 | 2020000000 | 9798600 (49.0%) | 51398825 (2.54% of total) | 118079924 (5.85% of total) | 9798600 |
| prostate\_4c | 17735105 | 1791245605 | 8730140 (49.2%) | 41835210 (2.34% of total) | 99508840 (5.56% of total) | 8730140 |
| testis\_7a | 20000000 | 2020000000 | 8850426 (44.3%) | 35688414 (1.77% of total) | 70322418 (3.48% of total) | 8850426 |
| testis\_7b | 20000000 | 2020000000 | 8551086 (42.8%) | 31188946 (1.54% of total) | 64250807 (3.18% of total) | 8551086 |
| testis\_7c | 20000000 | 2020000000 | 10252117 (51.3%) | 41702714 (2.06% of total) | 130416259 (6.46% of total) | 10252117 |

Table 2 Result of Trimming

Read Mapping

The overall alignment rates passed 91.11% for all the replicates. Around 80% of the reads were aligned exactly one time and around 14% of the reads were aligned more than one time. Only less than 9% of the reads were not aligned whatsoever (shown in Table 3).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Replicates** | **Aligned 0 time** | **Aligned 1 time** | **Aligned >1 time** | **Overall alignment rate** |
| prostate\_4a | 1197273 (6.02%) | 15887871 (79.88%) | 2804880 (14.10%) | 93.98% |
| prostate\_4b | 1744772 (8.89%) | 15213808 (77.56%) | 2657286 (13.55%) | 91.11% |
| prostate\_4c | 1033279 (5.90%) | 14117440 (80.62%) | 2360173 (13.48%) | 94.10% |
| testis\_7a | 643313 (3.23%) | 16445837 (82.66%) | 2806068 (14.10%) | 96.77% |
| testis\_7b | 639814 (3.21%) | 16400579 (82.39%) | 2866703 (14.40%) | 96.79% |
| testis\_7c | 821917 (4.18%) | 16058630 (81.68%) | 2780194 (14.14%) | 95.82% |

Table 3 Result of Mapping

Transcript Assembly & Quantitation

Most reads- over 60%- were assigned to genes/proteins. Features not aligning and multi-mapping were the major reasons for failure of assigning reads (shown in Table 4).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Replicates** | **Assigned** | **Unassigned** | | | |
| **Ambiguity** | **Multi-Mapping** | **No Features** | **Unmapped** |
| prostate\_4a | 14540291 | 973065 | 2250553 | 2341049 | 1197273 |
| prostate\_4b | 13949280 | 902616 | 2206675 | 2197076 | 1744772 |
| prostate\_4c | 13314568 | 824463 | 1914469 | 1633238 | 1033279 |
| testis\_7a | 14381754 | 909980 | 2271235 | 3103136 | 643313 |
| testis\_7b | 14590359 | 941366 | 2360339 | 2846627 | 639814 |
| testis\_7c | 15097497 | 937926 | 2171849 | 1998061 | 821917 |

Table 4 Report of Transcript Assembly & Quantitation (using featureCounts)

Differential Analysis

The differential analysis showed that prostate and testis were expressed significantly differently from each other. The replicates of the same organ were expressed much more closely to each other compared to the replicates of the other organ (shown in Figure 1 and Figure 2), which was expected. The biological coefficient of variation was valid (shown in Figure 3). Most of the genes expressed either in the prostate or in the testis, while only a small number of genes appeared in both body parts (show in Figure 4). In the top 50 differentially expressed genes, there were more highly expressed genes in the testis than in the prostate (shown in Figure 5).

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

Figure 2 PCA Plot of Prostate vs Testis

Both principal components separated testis group and prostate group well, meaning the genes expressed in prostate and testis are highly different.

Figure 3 BCV Plot of Prostate vs Testis

The common is normal (around 0.4).

Figure 4 CPM plot of Prostate vs Testis

Most of the genes are differentially expressed.

Figure 1 Gene Expression Heatmap of Prostate vs Testis

The distance is closer within the same organ, and the distance is farther within the different organs.

Enrichment Analysis

In order to find out what different biological functions and biological process the two tissues have, an enrichment analysis was performed on the three replicates of the prostate and the three of the testis; 870 significant genes were chosen in total. As shown in Figure 6, flagellated sperm motility (GO:0030317), cell wall macromolecule catabolic process (GO:0016998), spermatogenesis (GO:0007283) and male gamete generation (GO:0048232) were the four most significant GO biological processes the genes pointed to.

**A picture containing building, door

Description automatically generated**

Figure 5 Heatmap of Most Differentially Expressed Genes in Prostate and Testis

The top 50 most differentially expressed genes in prostate and testis are presented.

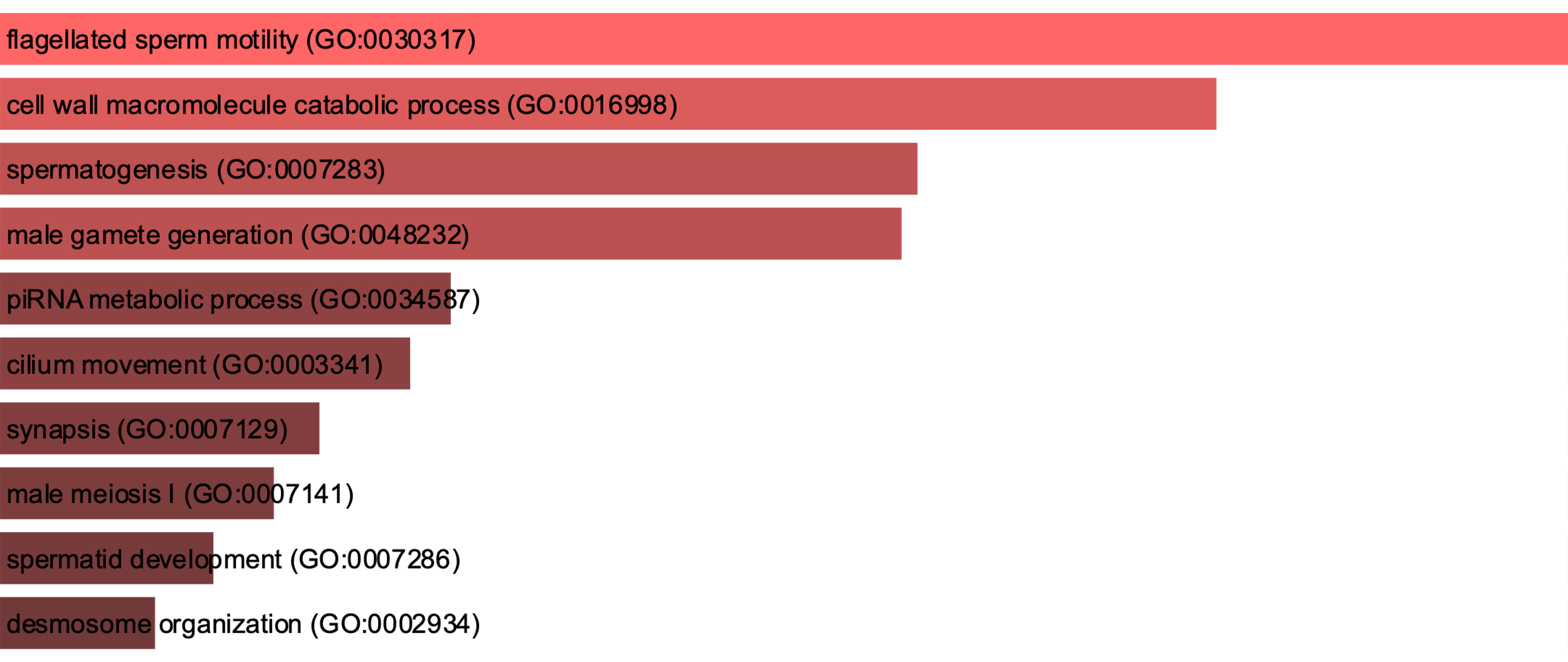
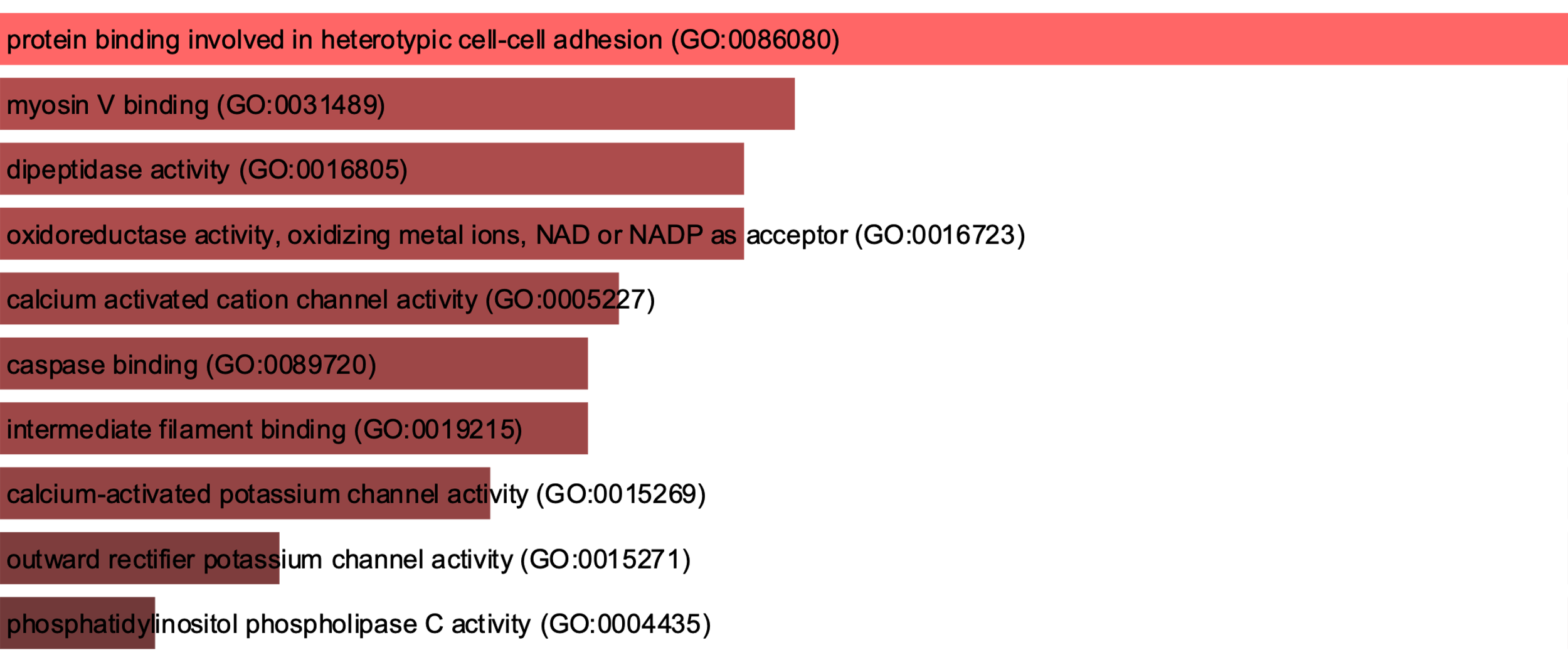


Figure 6 Bar Chart of Top GO Biological Process

Figure 7 Bar Chart of Top GO Molecular Function

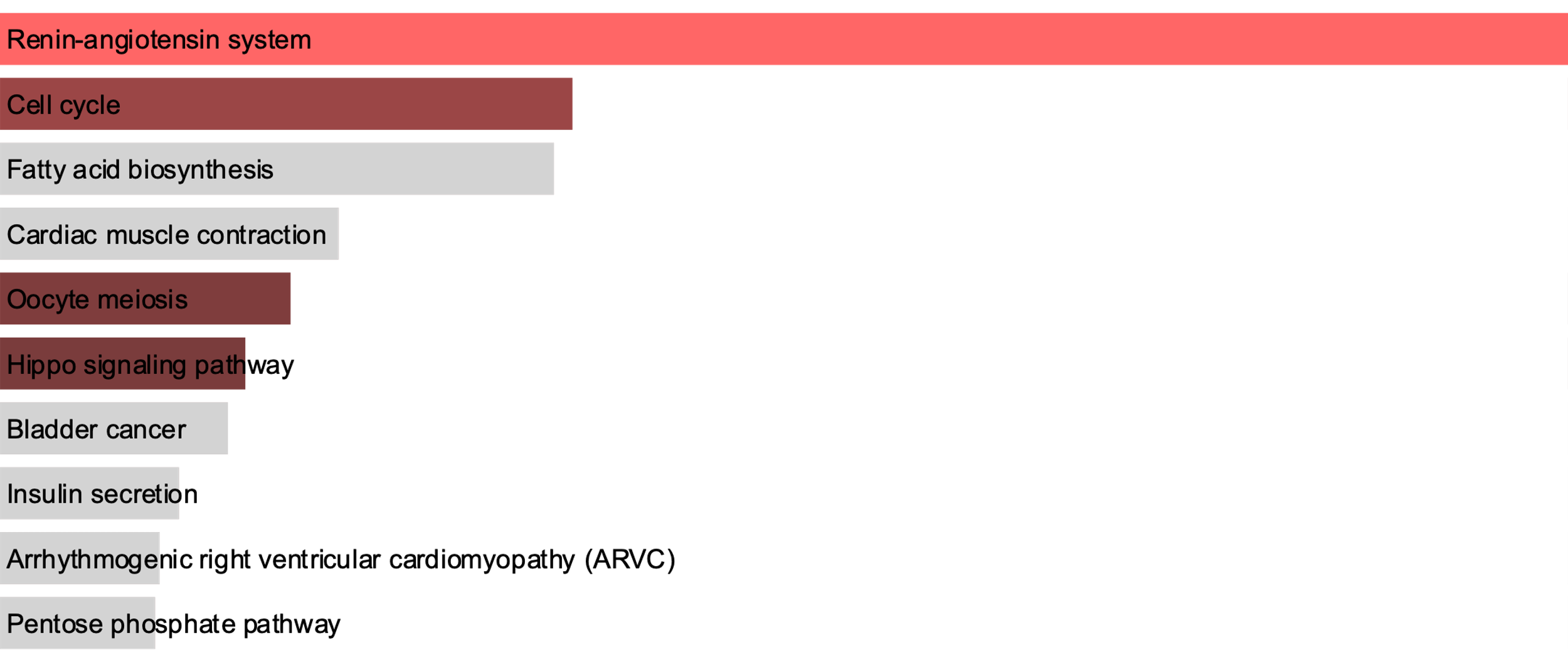


Figure 8 Bar Chart of Top KEGG Pathway

Protein binding involved in heterotypic cell-cell adhesion (GO:0086080) was the most significant GO molecular function (shown in Figure 7). Renin-angiotensin system and Cell cycle were the most different pathways between the two tissues (shown in Figure 8).

**Discussion/Conclusion:**

Flagellated sperm motility (GO:0030317) refers to any process involved in the controlled movement of a flagellated sperm cell. Cell wall macromolecule catabolic process (GO:0016998) involves the chemical reactions and pathways resulting in the breakdown of macromolecules that form part of a cell wall and it is related with meiosis. Spermatogenesis (GO:0007283) is the developmental process by which male germ line stem cells self-renew or give rise to successive cell types resulting in the development of a spermatozoa. Alone with male gamete generation (GO:0048232), they all play important roles in producing sperm.

Protein binding involved in heterotypic cell-cell adhesion (GO:0086080) is the molecular function interacting selectively and non-covalently with any protein or protein complex contributing to the adhesion of two different types of cells. Plasma, in which fibrinolysin is derived from, is highly related to the molecular function mentioned above (Hitchcock, J. K., et al, 2014). As all components of the renin-angiotensin system (RAS) have been identified in the prostate (Chow. L., et al, 2009), RAS is exclusive to the prostate.

To conclude, the prostate and testes are two drastically different tissues in regard to their functions in the male reproductive system. The results of the enrichment analysis indisputably- and predictably- show that sperm is exclusively generated in testes while fibrinolysis is exclusively generated in the prostate. However, other elements, such as proteolytic enzymes, cannot be proved to be exclusively produced in either the prostate or testes by the enrichment analysis at this point.

**References:**

Chow L, Rezmann L, Catt KJ, Louis WJ, Frauman AG, Nahmias C, Louis SN. (2009) Role of the renin-angiotensin system in prostate cancer. Mol Cell Endocrinol. 2009 Apr 29;302(2):219-29. doi: 10.1016/j.mce.2008.08.032.

Hitchcock, J. K., Katz, A. A., & Schäfer, G. (2014). Dynamic reciprocity: the role of annexin A2 in tissue integrity. Journal of cell communication and signaling, 8(2), 125–133. https://doi.org/10.1007/s12079-014-0231-0

Verze P., Cai T., Lorenzetti S. (2016) The role of the prostate in male fertility, health and disease. Nature Reviews Urology. 2016;13(7):379–386. doi: 10.1038/nrurol.2016.89.

**Supplementary Data:**

Attached in project submission