

Differential gene expression in males vs females for head and neck squamous cancers (HNSC)

Czuee Morey

May 27, 2016

The gene expression data for HNSC was taken from the Cancer genome atlas study [1]. The data contained 566 samples for RNA-seq from males and females. In some cases, normal samples from the same individual were also present. The data was treated suitably for analysis. In part 1, the normal tissues were removed and only tumor samples were considered. In part 2, the samples for which both tumor and normal tissue are present were considered. The log of gene expression(+1) was considered for all cases, and the expression was centered around the mean in some cases.

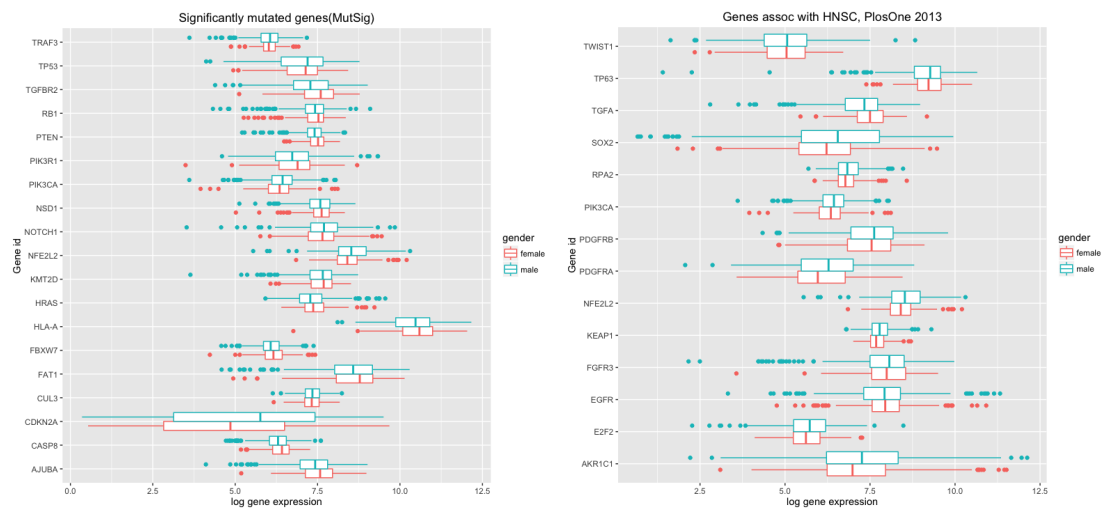


Figure 1: Significantly mutated genes [1] and genes associated with HNSC [2] do not show differential expression between males and females.

1 Difference in gene expression between tumor samples from males vs females.

1.1 Selecting genes with highest variability

The genes were filtered to select only genes that would have the highest variance in their gene expression and are more likely to have differential expression between males and females. The genes were sorted by the standard deviation (SD) of gene expression from all samples. The figure 2 shows the plot with ordered

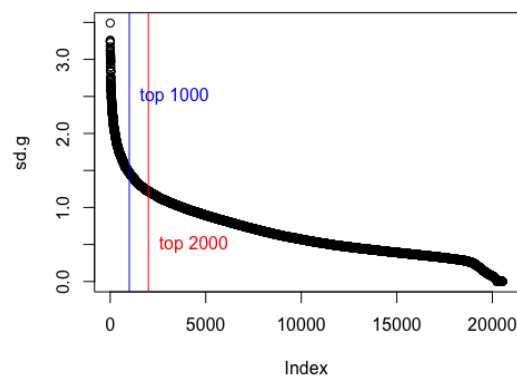


Figure 2: Plot for Standard deviation of gene expression ordered from highest to lowest. The vertical lines indicate the position for first 1000 and 2000 genes that were selected for further study.

1.2 Genes with significant difference in means between males and females

References

- [1] Network, T. C. G. A. (2015) *Comprehensive genomic characterization of head and neck squamous cell carcinomas*. Nature 517, 576–582.
- [2] Vonn Walter et al (2013) *Molecular Subtypes in Head and Neck Cancer Exhibit Distinct Patterns of Chromosomal Gain and Loss of Canonical Cancer Genes* Plos One 8(2), e56823.

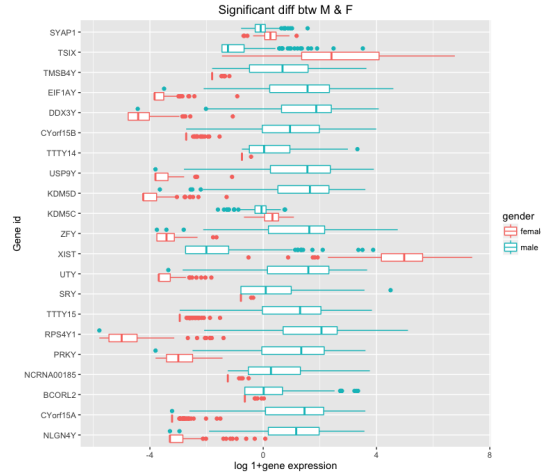


Figure 3: The genes that have a difference between their means in males and females greater than the SD of the entire distribution were selected. 21 genes were found.

Table 1: The genes that have a difference between their means in males and females greater than the SD of the entire distribution were selected. 19 descriptions - NCRNA00185 and TTTY14 are the same gene. Also CYorf15A and B have the same description. All the genes are X and Y linked only.

Approved Symbol	Approved Name	Chromosome
DDX3Y	DEAD-box helicase 3, Y-linked	Yq11
EIF1AY	eukaryotic translation initiation factor 1A, Y-linked	Yq11.223
KDM5C	lysine demethylase 5C	Xp11.22-p11.21
KDM5D	lysine demethylase 5D	Yq11
NLGN4Y	neuroligin 4, Y-linked	Yq11.221
PRKY	protein kinase, Y-linked, pseudogene	Yp11.2
RPS4Y1	ribosomal protein S4, Y-linked 1	Yp11.3
SRY	sex determining region Y	Yp11.3
SYAP1	synapse associated protein 1	Xp22.31
TMSB4Y	thymosin beta 4, Y-linked	Yq11.221
TSIX	TSIX transcript, XIST antisense RNA	Xq13.2
TTY14	testis-specific transcript, Y-linked 14 (non-protein coding)	Yq11.222
TTY15	testis-specific transcript, Y-linked 15 (non-protein coding)	Yq11.1
USP9Y	ubiquitin specific peptidase 9, Y-linked	Yq11.2
UTY	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	Yq11.221
XIST	X inactive specific transcript (non-protein coding)	Xq13.2
ZFY	zinc finger protein, Y-linked	Yp11.3