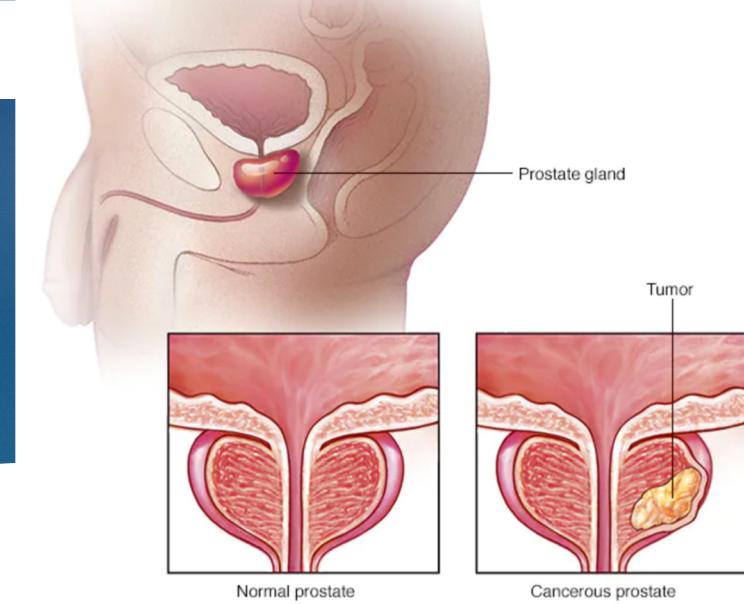


# Differential gene expression between prostate cancer of African American and European American men

TRGN514 Cheng-Hsiang Lu

# Introduction

- ▶ Prostate cancer
  - ▶ Prostate cancer is more likely to develop in older men.
  - ▶ Prostate cancer is the second most common cancer in American men.
  - ▶ Prostate cancer is the second leading cause of cancer death in American men.
- ▶ Men of African origin have 2-3 times greater chance of developing prostate cancer than Men of European origin.
- ▶ African American men are 2 times more likely to die compared to European American men with prostate cancer.



© MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH. ALL RIGHTS RESERVED.

# Article

Article

## A Systems Approach to Interrogate Gene Expression Patterns in African American Men Presenting with Clinically Localized Prostate Cancer

Gary Hardiman <sup>1,2,\*</sup> , Stephen J. Savage <sup>3,4,5</sup>, E. Starr Hazard <sup>6</sup> , Willian A. da Silveira <sup>2</sup> , Rebecca Morgan <sup>2</sup>, Adam Harris <sup>2</sup>, Melanie S. Jefferson <sup>7</sup>, Robert C. Wilson <sup>8</sup>, Susan Caulder <sup>4</sup>, Linda Ambrose <sup>4</sup>, Lewis Frey <sup>9,10</sup>, Bethany Wolf <sup>10</sup> , Sebastiano Gattoni-Celli <sup>4,5</sup> and Chanita Hughes Halbert <sup>7,11,12,13,\*</sup>

- ▶ Patients have undergone a prostate biopsy.
- ▶ 33 African American Men and 27 European American men, 60 subjects in total.
  
- ▶ My project:
  - ▶ 8 samples from African American men.
  - ▶ 8 samples from European American men.
- ▶ These samples included subjects with both high and low vitamin D levels. However, for this project, I only evaluated ancestry and included equal H/L vitamin D subjects in both groups.

# Hypothesis

- ▶ Certain genes are associated with the severity of prostate cancer which cause the racial disparities between African American men and European American men.

# Sequencing source

Series GSE138503					
Sample	Sample name	SRA	SRR ID	Single end or paired end	Sample new name
GSM4109968	AA: 2973_7	SRX6957518	SRR10238300	paired end	AHS1
GSM4109975	AA: 2980_14	SRX6957525	SRR10238307	paired end	AHS2
GSM4109994	AA: 3001_38	SRX6957544	SRR10238326	paired end	AHS3
GSM4109997	AA: 3027_41_S2	SRX6957547	SRR10238329	paired end	AHS4
GSM4109996	AA: 3026_40_S1	SRX6957546	SRR10238328	paired end	ALS1
GSM4109999	AA: 3029_43_S4	SRX6957549	SRR10238331	paired end	ALS2
GSM4110002	AA: 3033_49_S8	SRX6957552	SRR10238334	paired end	ALS3
GSM4110003	AA: 3034_52_S9	SRX6957553	SRR10238335	paired end	ALS4
GSM4109969	EA: 2974_8	SRX6957519	SRR10238301	paired end	EHS1
GSM4109973	EA: 2978_12	SRX6957523	SRR10238305	paired end	EHS2
GSM4109976	EA: 2981_15	SRX6957526	SRR10238308	paired end	EHS3
GSM4109990	EA: 2997_34	SRX6957540	SRR10238322	paired end	EHS4
GSM4109967	EA: 2972_6	SRX6957517	SRR10238299	paired end	ELS1
GSM4110001	EA: 3031_45_S6	SRX6957551	SRR10238333	paired end	ELS2
GSM4110006	EA: 3037_56_S12	SRX6957556	SRR10238338	paired end	ELS3
GSM4110008	AA: 3039_58_S14	SRX6957558	SRR10238340	paired end	ELS4

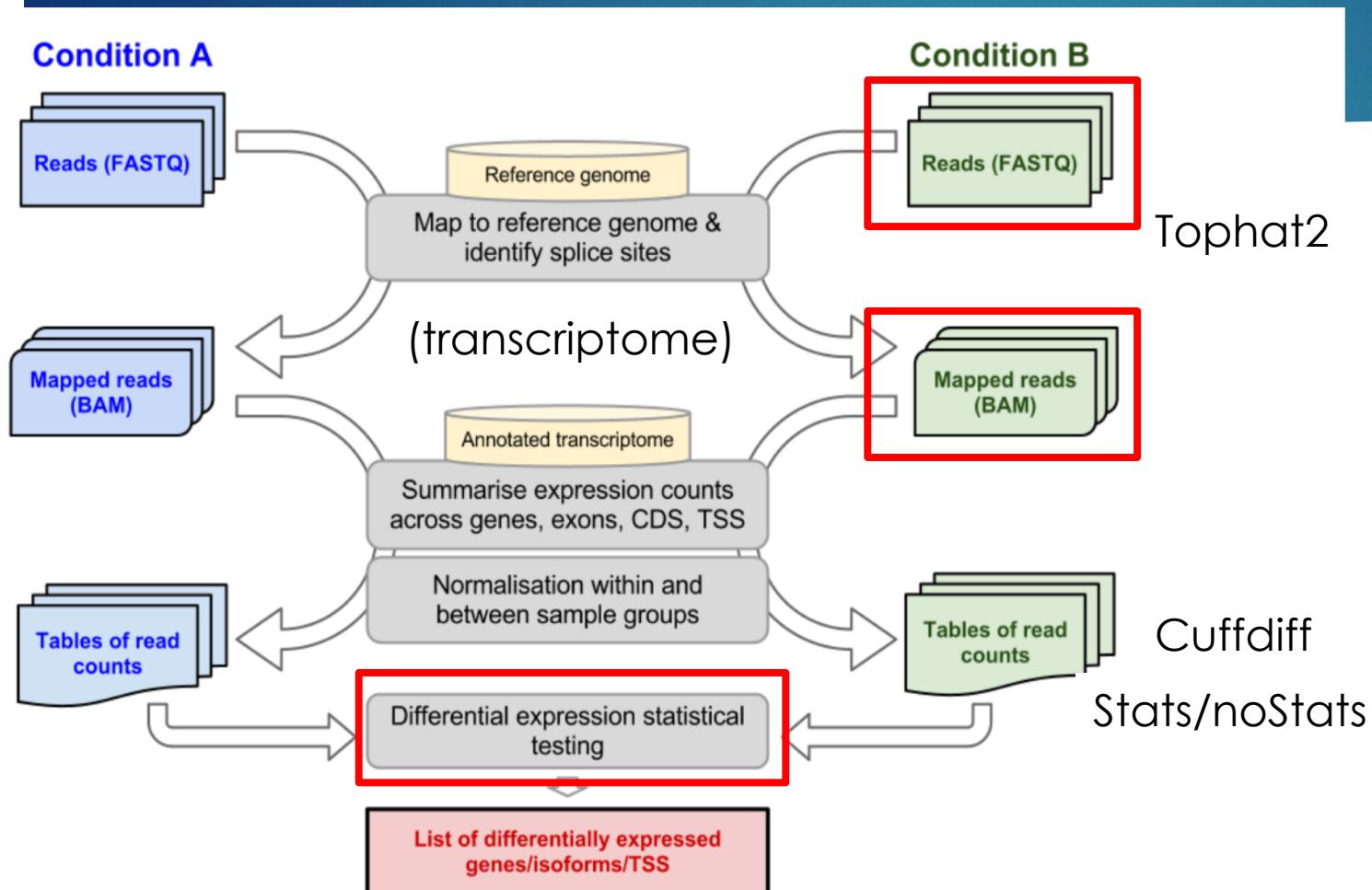
# Pre-alignment sequencing metrics

Read per samples		
<b>Maximum</b>	67,887,728	AHS2
<b>Minimum</b>	56,652,656	ALS3
<b>Mean</b>	63,855,032	

African American  
 European American



# Pipeline

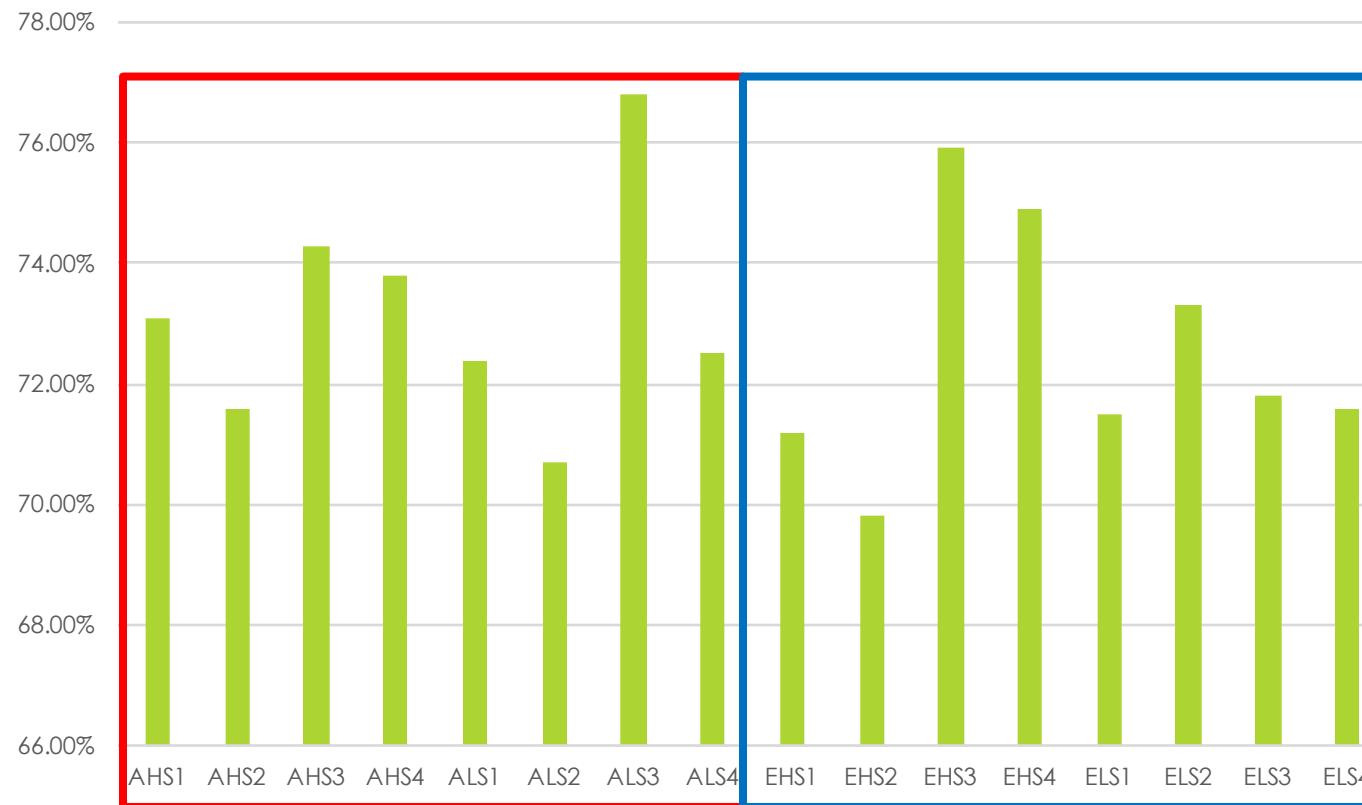


- ▶ Ref: Homo\_sapiens.GRCh38.104.chr.gtf
- ▶ ensembl.GRCh38.104.fa
- ▶ Cuffdiff
  - ▶ Stats: gene\_exp.diff
  - ▶ noStats: genes.fpkm\_tracking

# Alignment metrics and statistics

□ African American  
□ European American

% Alignment to Transcriptome



<b>Maximum(%)</b>	76.80%	ALS3
<b>Minimum(%)</b>	69.80%	EHS2
<b>Mean(%)</b>	72.83%	

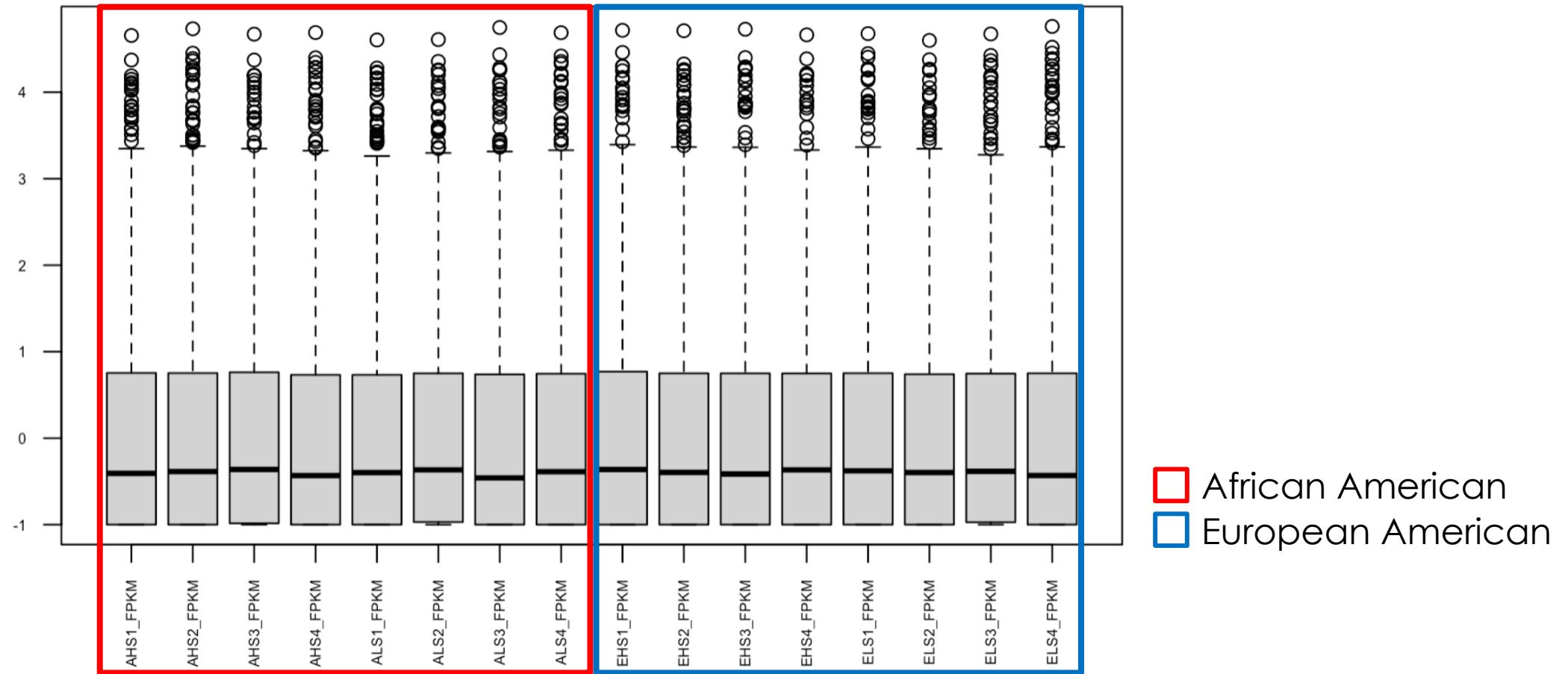
Samples	% Alignment	% Discordance
AHS1	73.10%	1.60%
AHS2	71.60%	0.30%
AHS3	74.30%	0.70%
AHS4	73.80%	0.30%
ALS1	72.40%	0.30%
ALS2	70.70%	0.30%
ALS3	76.80%	0.30%
ALS4	72.50%	0.30%
EHS1	71.20%	0.70%
EHS2	69.80%	0.30%
EHS3	75.90%	0.50%
EHS4	74.90%	0.80%
ELS1	71.50%	0.50%
ELS2	73.30%	0.30%
ELS3	71.80%	0.30%
ELS4	71.60%	0.40%

Average discordance = 0.49%

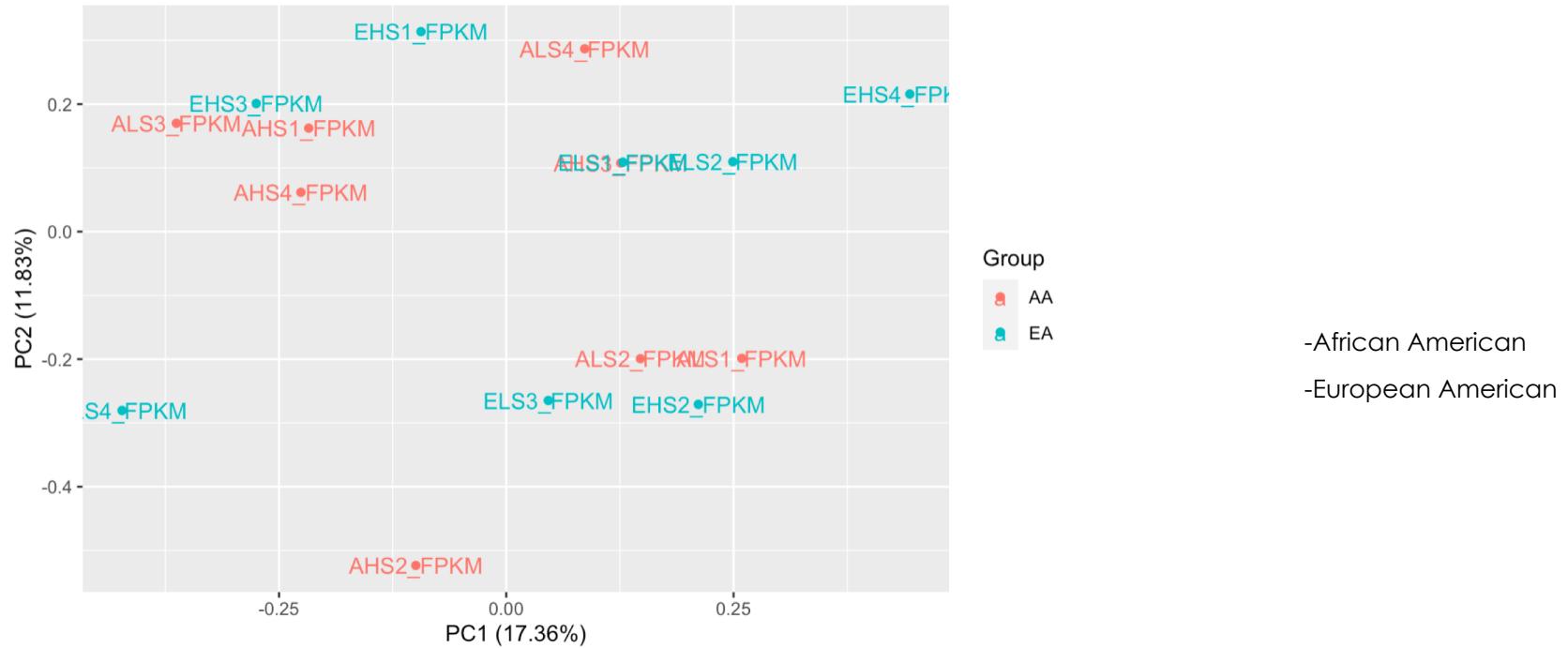
# Steps of processing data

- ▶ Get the FPKM values for all of my samples in genes\_fpkm.tracking file(60590 genes)
- ▶ Remove all zero data(42910 genes).
- ▶ Add a pseudo count 0.1 to each of my value.
- ▶ Log10 scale.

# Box and whisker plot



# PCA plot



# Filter the data

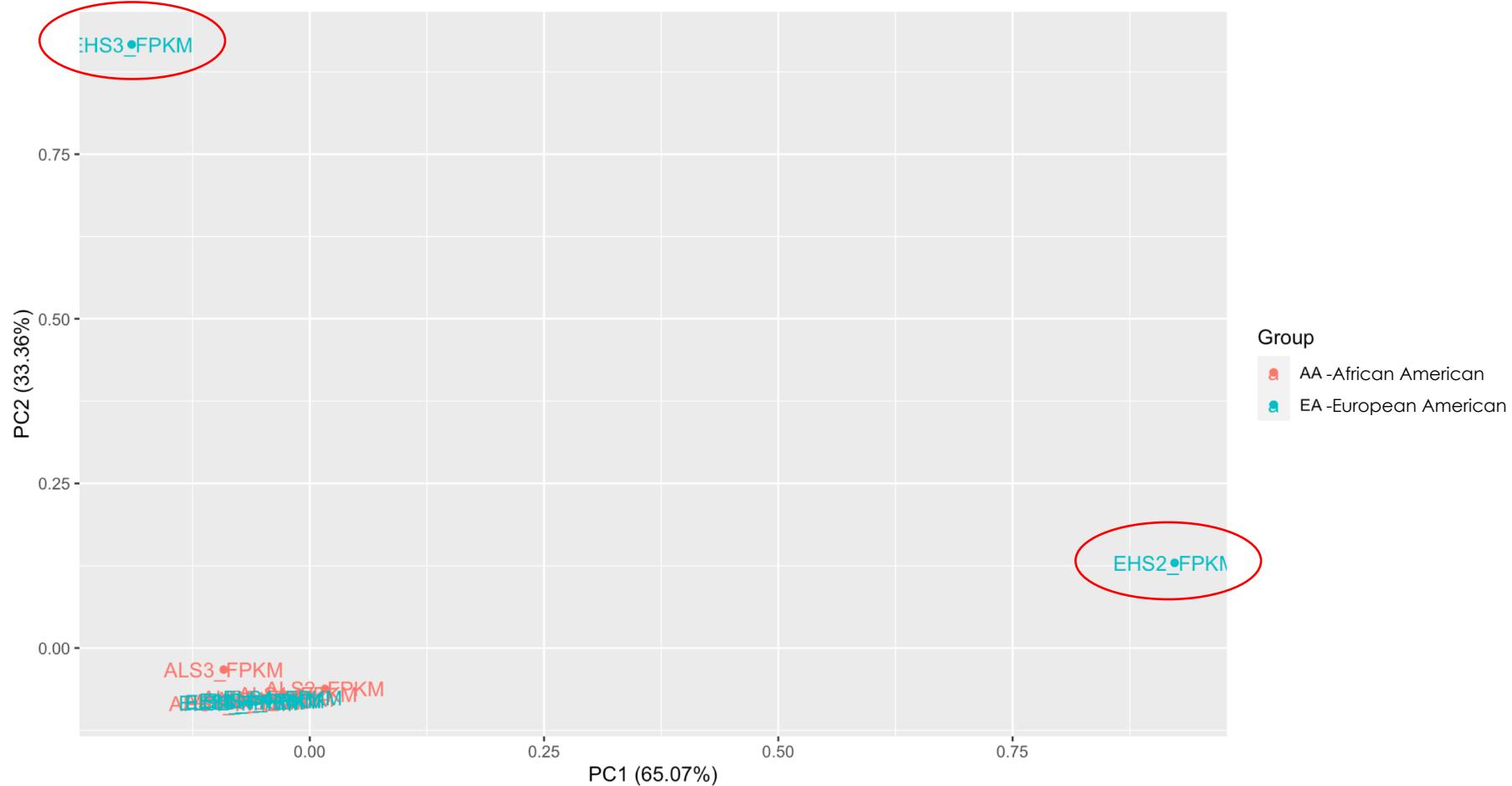
With p-value  
 $< 0.05$

Log2 fold  
difference  
 $\geq +1.5$  or  $\leq -1.5$

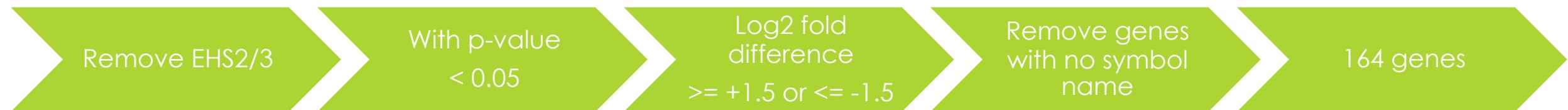
143 genes

- 92 genes are upregulated
- 51 genes are downregulated

# PCA - Filtered Genes

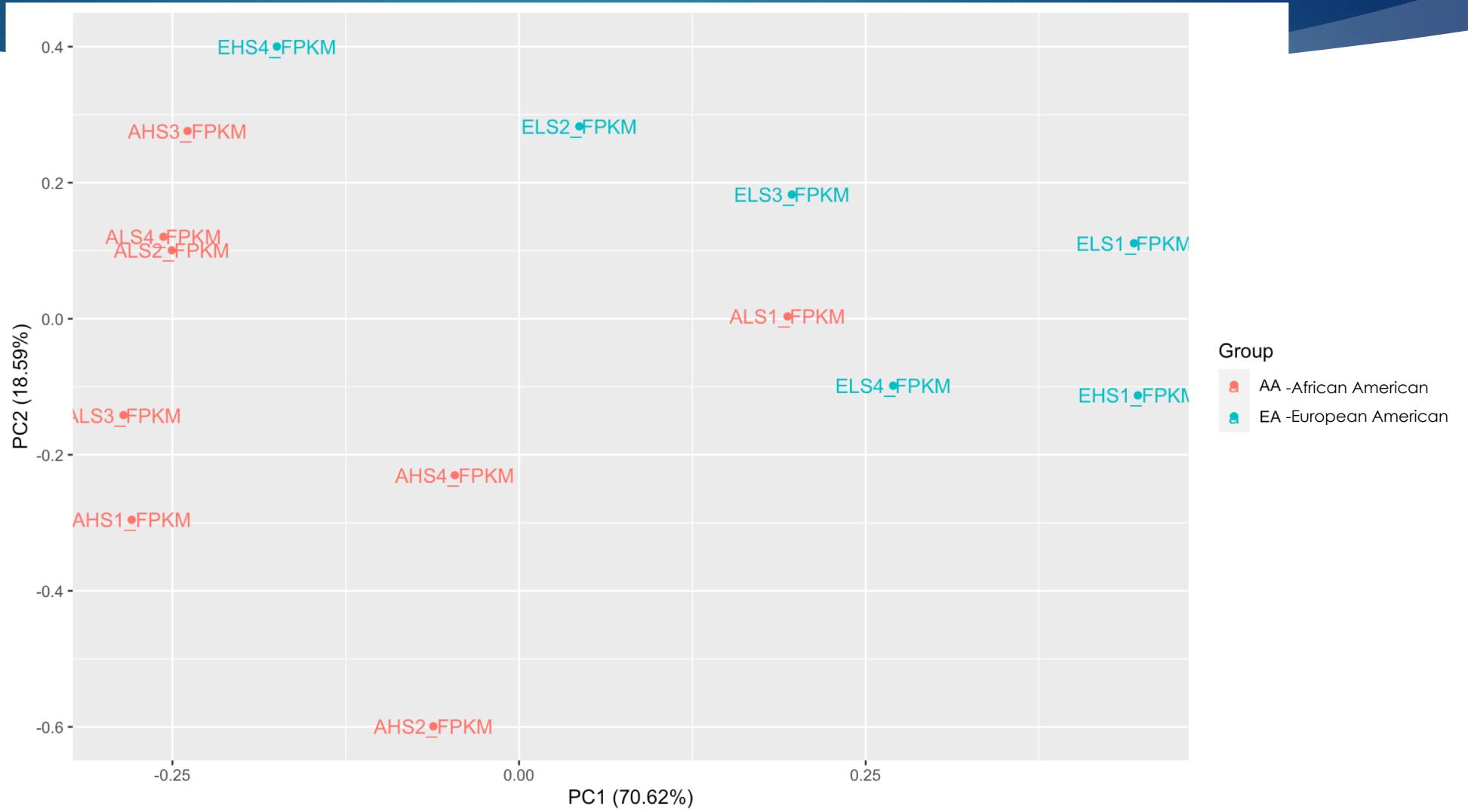


# Remove EHS2/3 and genes without symbol name



- 51 genes are upregulated.
- 113 genes are downregulated.

# PCA - Filtered Genes



# Dendrogram

Euclidian - Ward's

