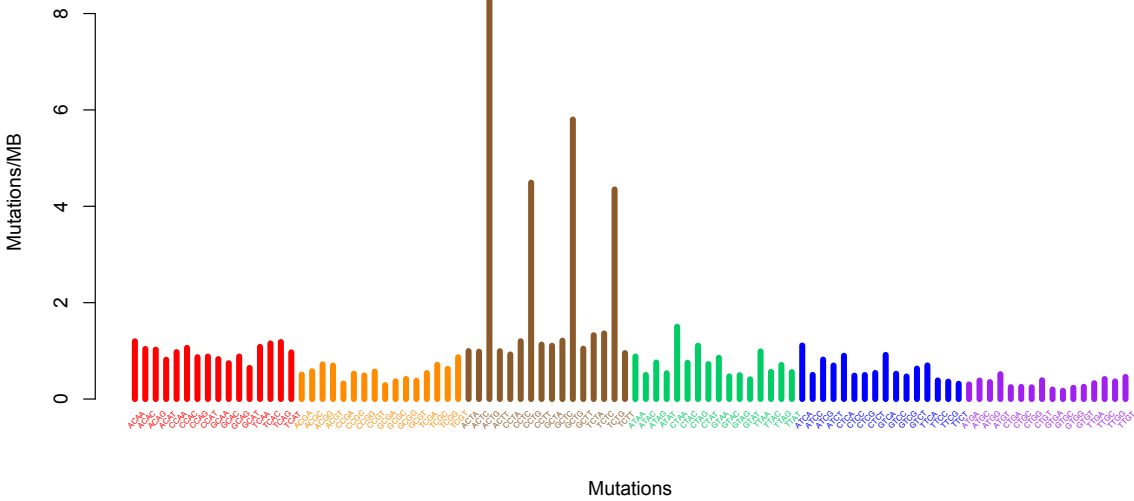


# Signature

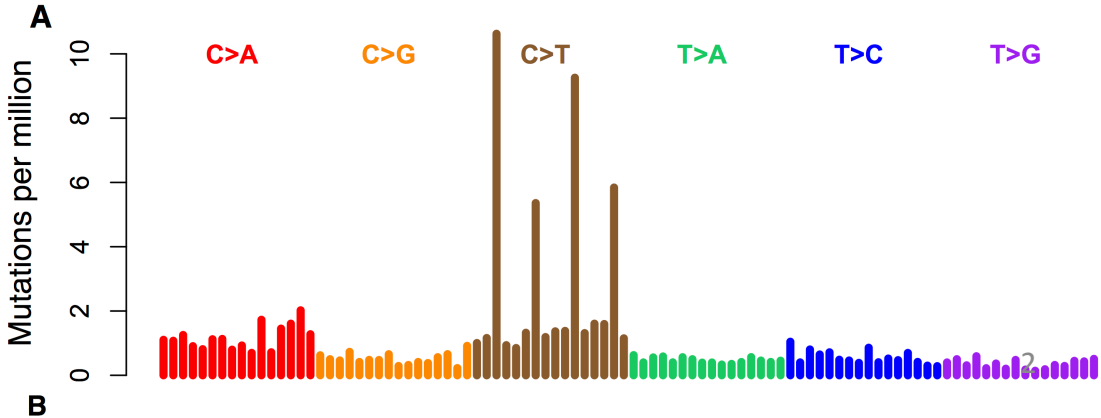
Shantao

# Similar Spectra

PCAWWG: Kidney-RCC

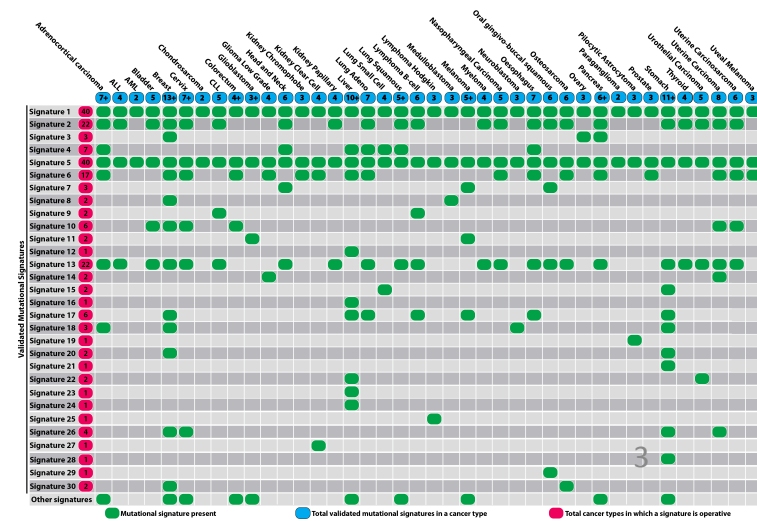


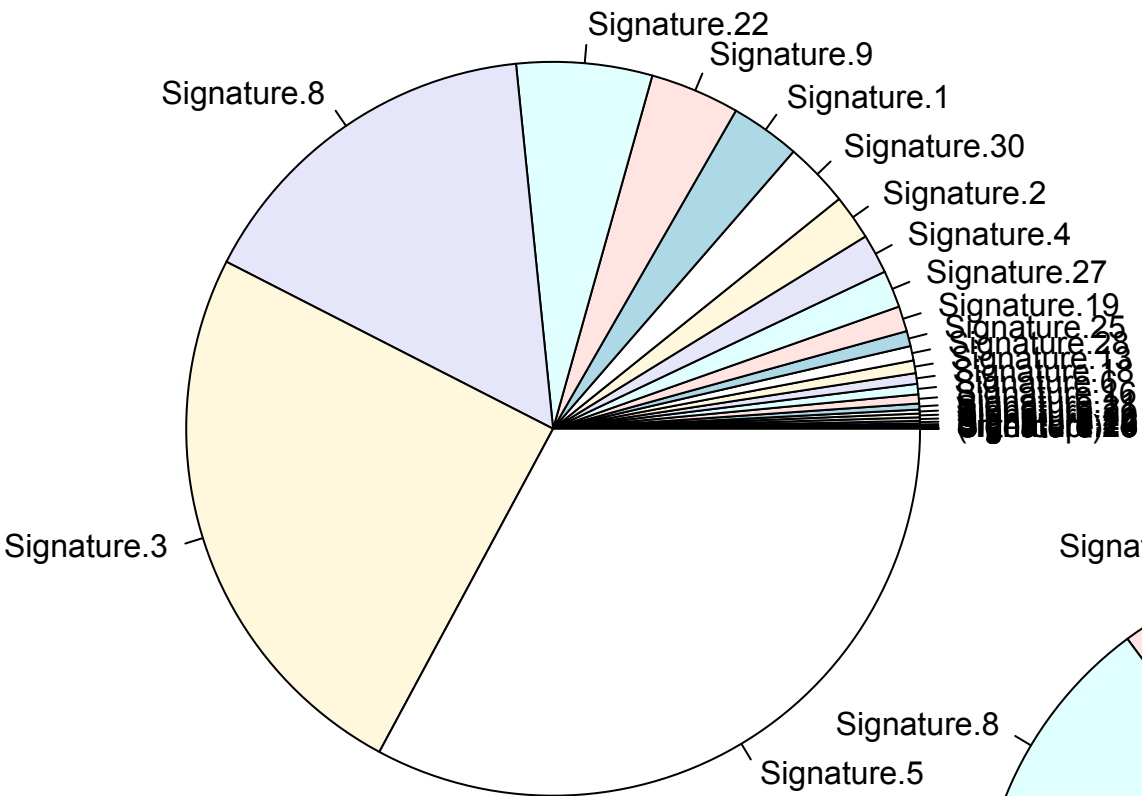
TCGA: pRCC



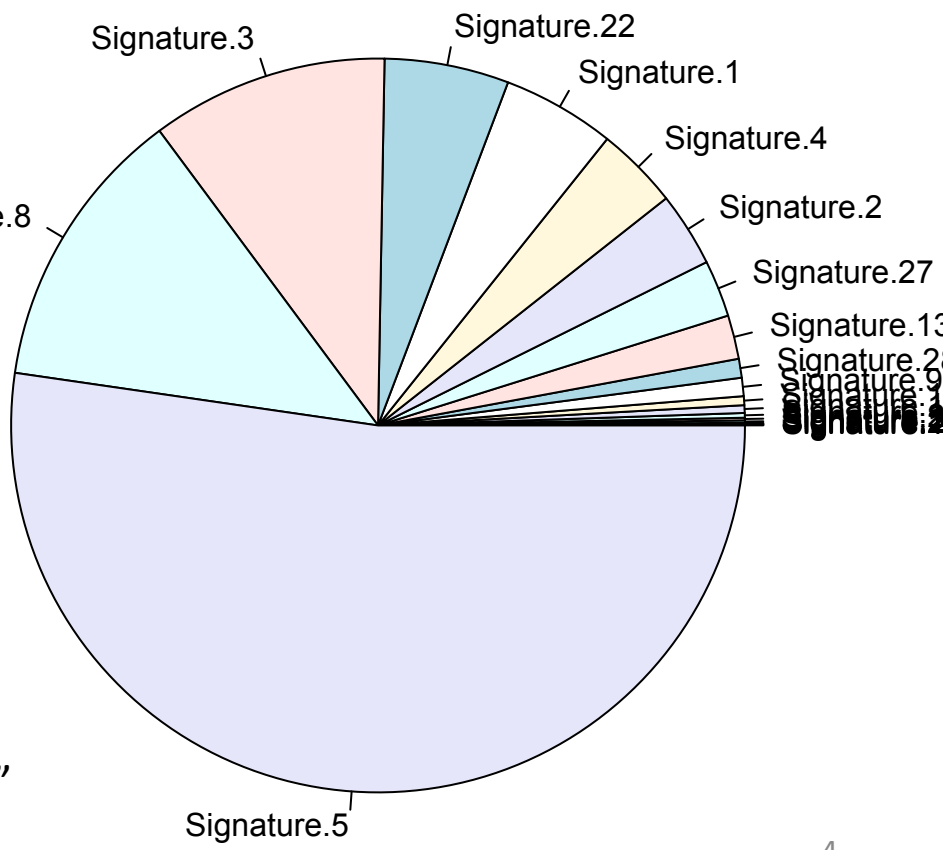
# LASSO

- Completely objective, mathematically sound
- Three scenarios
  - No prior, CV/regularization take full control
  - With prior, no penalization on Alexandrov's sig.
  - Only feed in Alexandrov's sig. (“hard prior”)





No prior

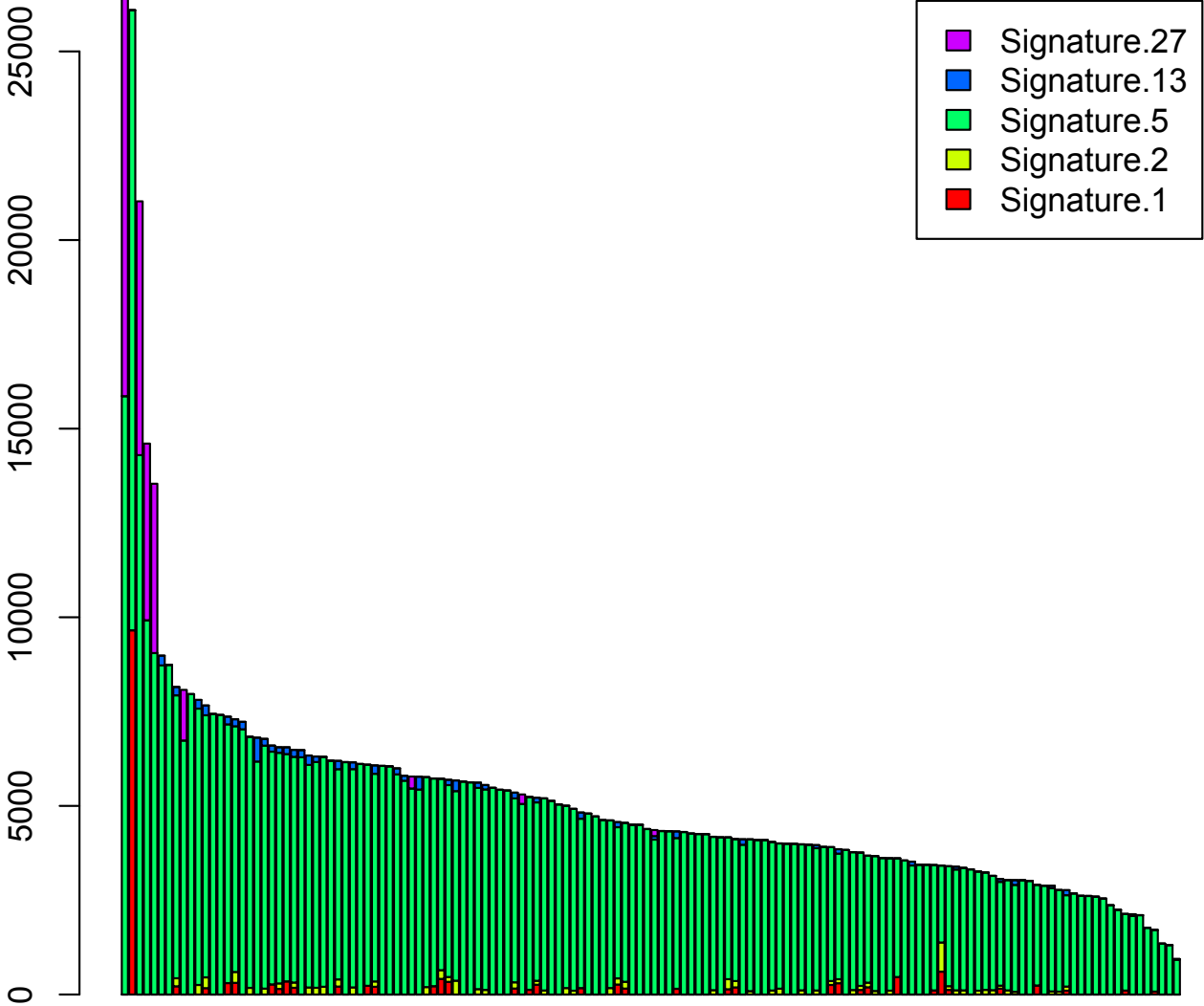


With "soft prior"

Signature.5

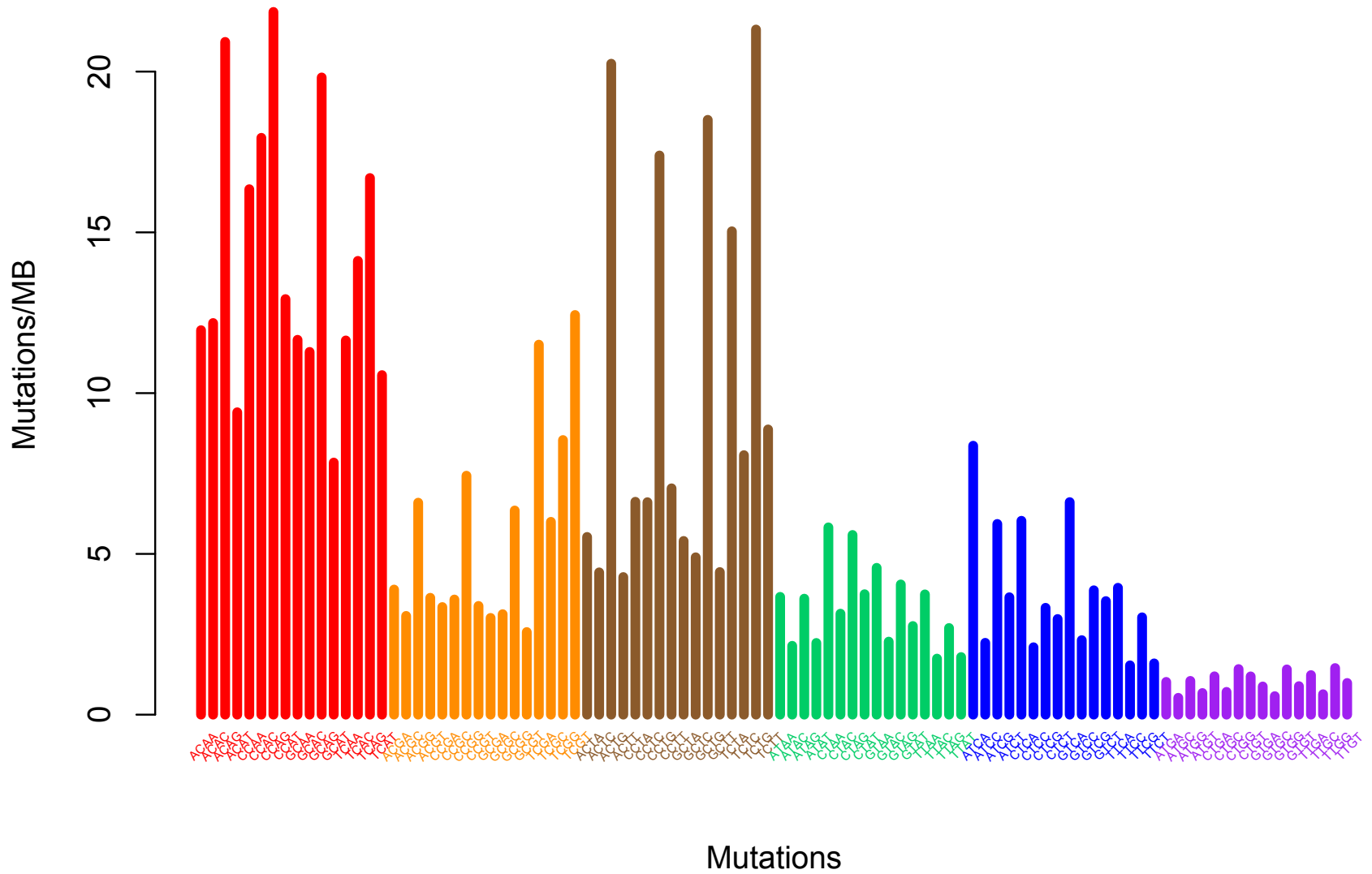
But we want 1/2/5/13/27 in RCCs

But we want 1/2/5/13/27 in RCCs....This is hard prior

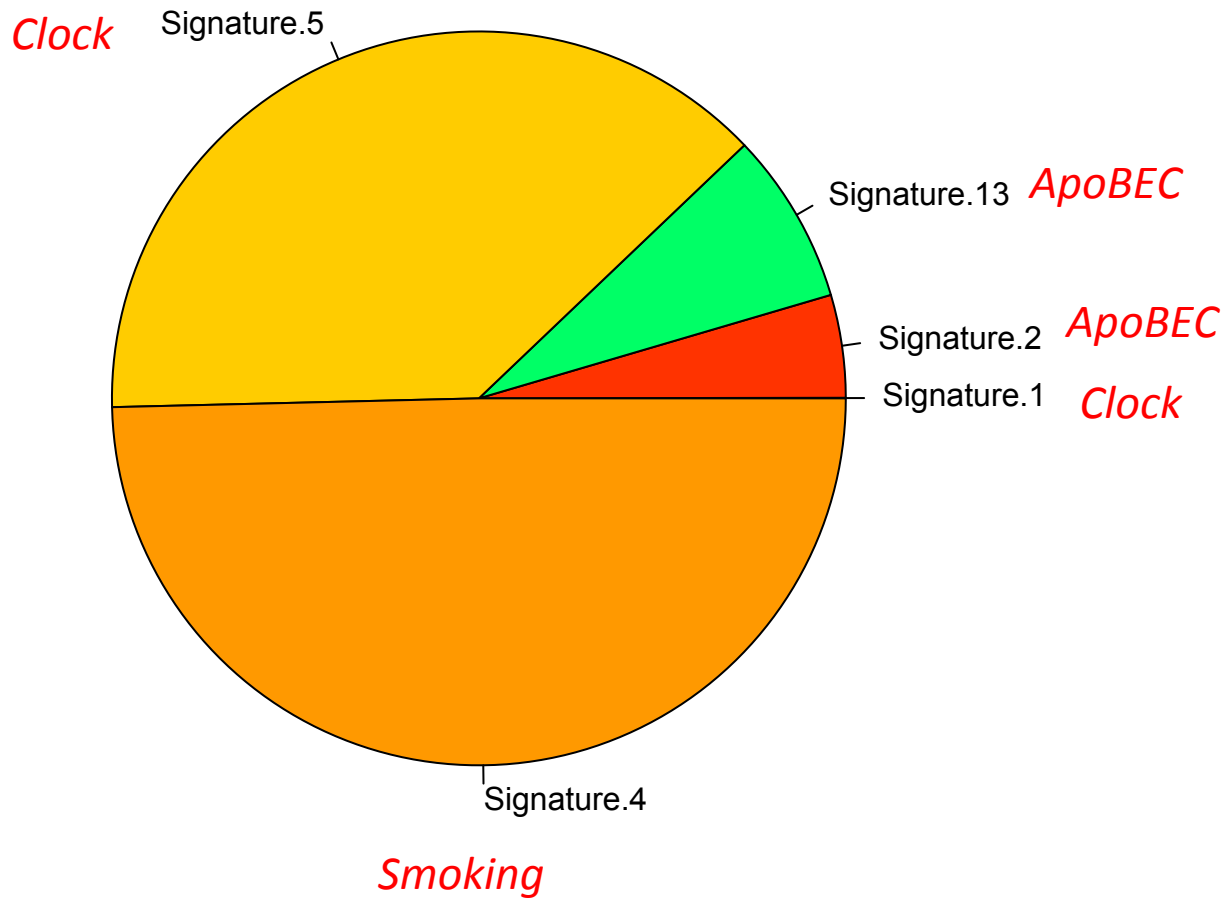


Sorry for the confusing **coloring**, I start using consistent coloring for Sig. from next slide!

# Lung-SCC

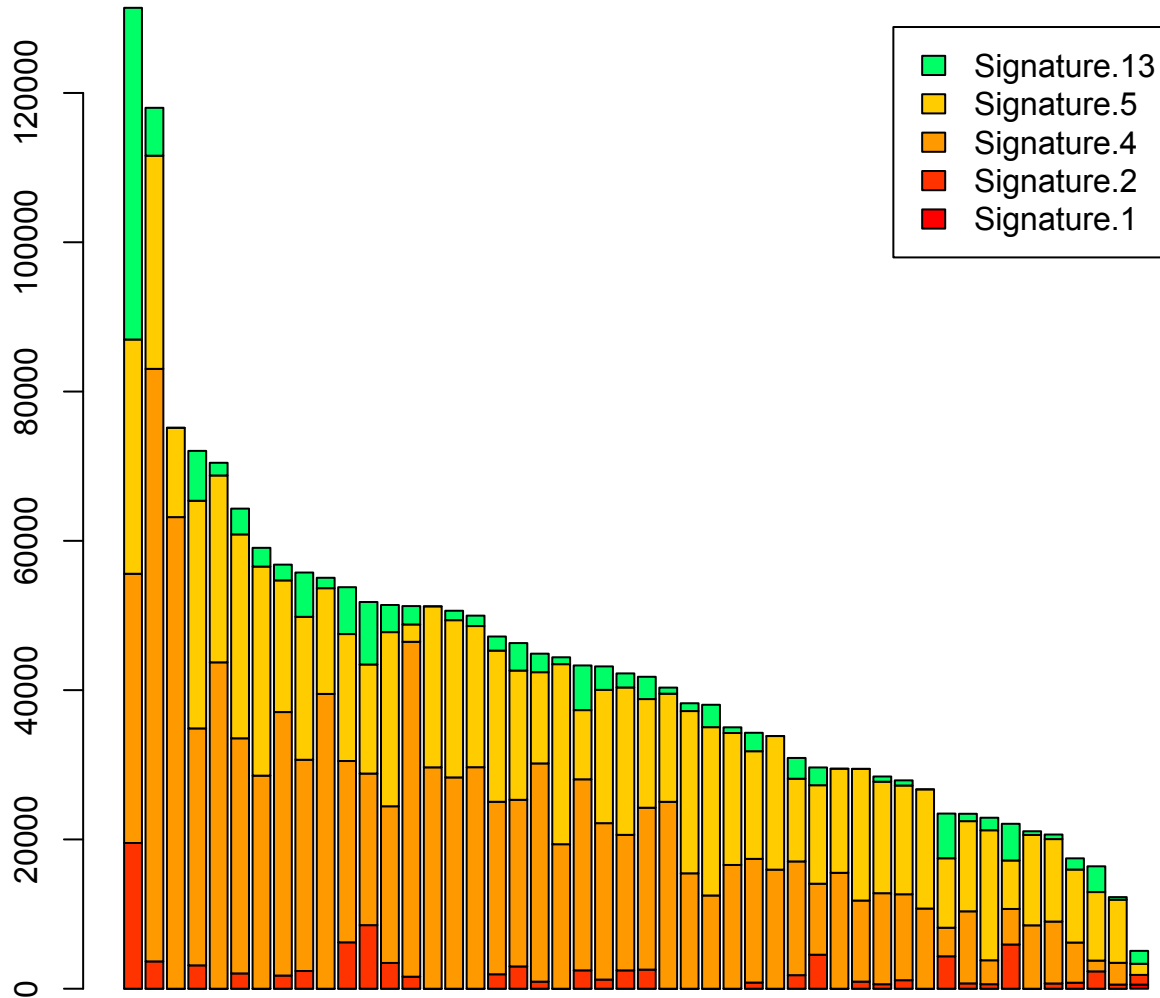


# With Hard Prior



# Everyone Smokes?

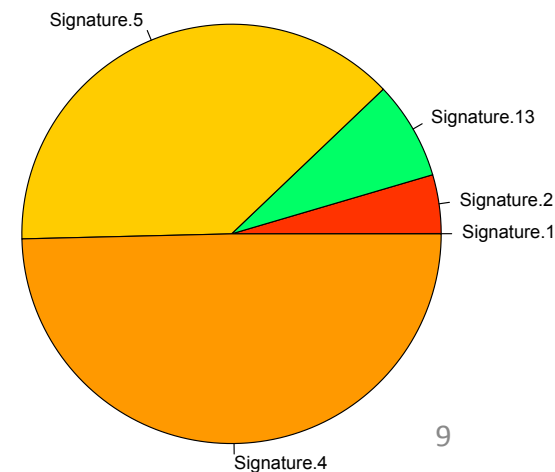
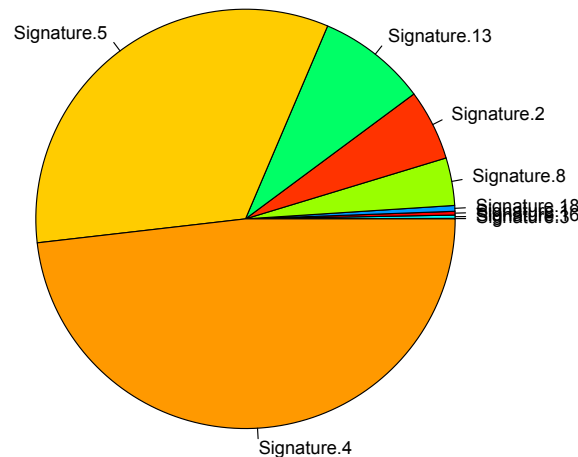
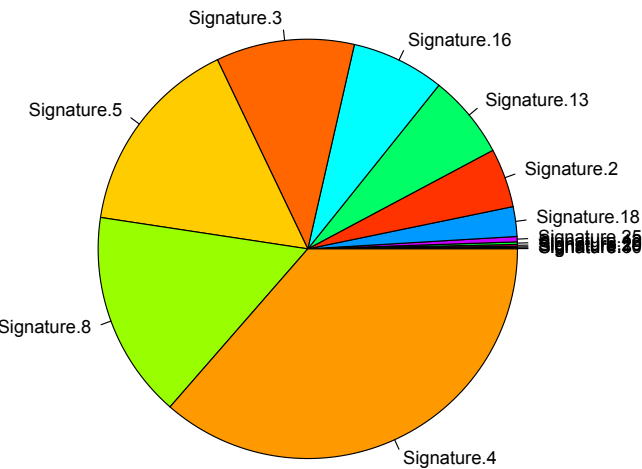
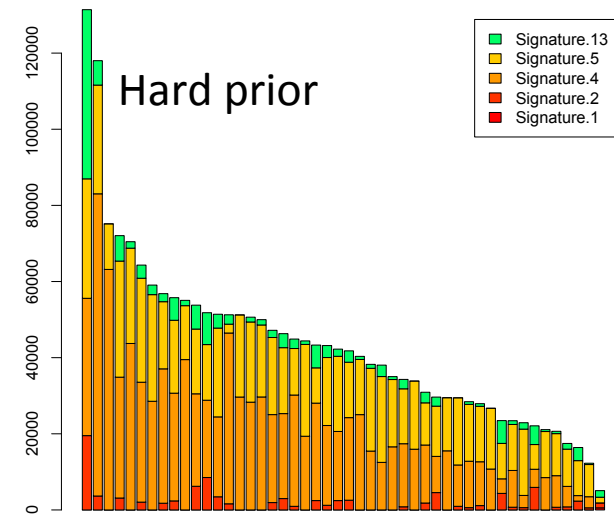
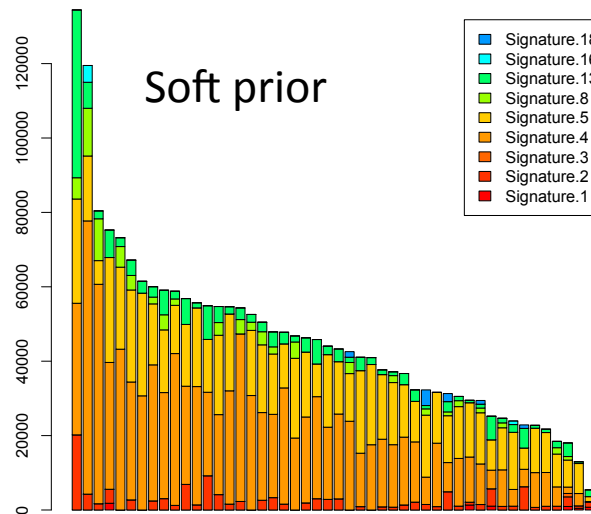
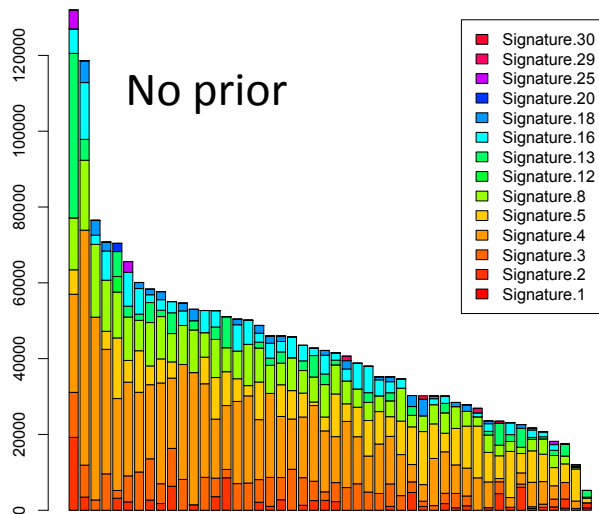
Bummer: everyone has every signature here!





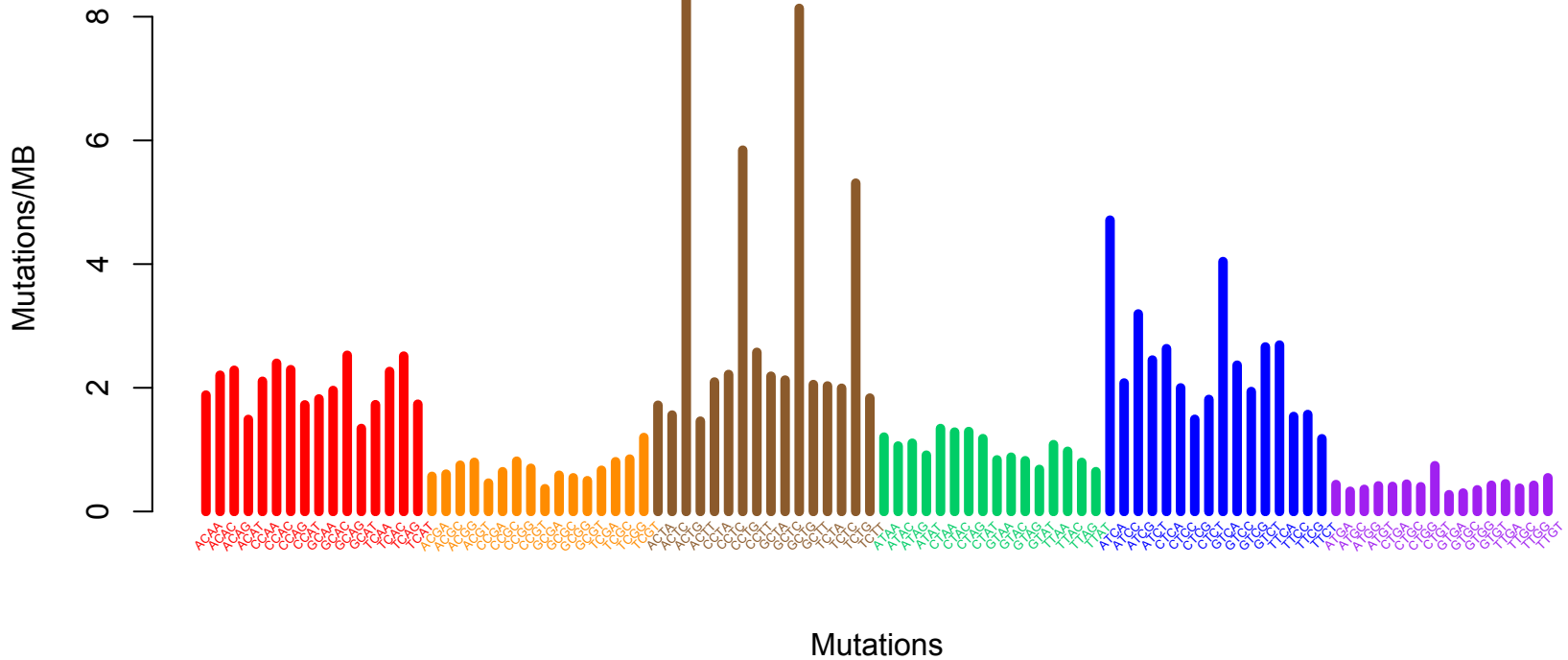
# Three Scenarios

Suppressed Sig.8 (unknown), 3 (DSB) and 16(unknown)...soft prior seems to work well here



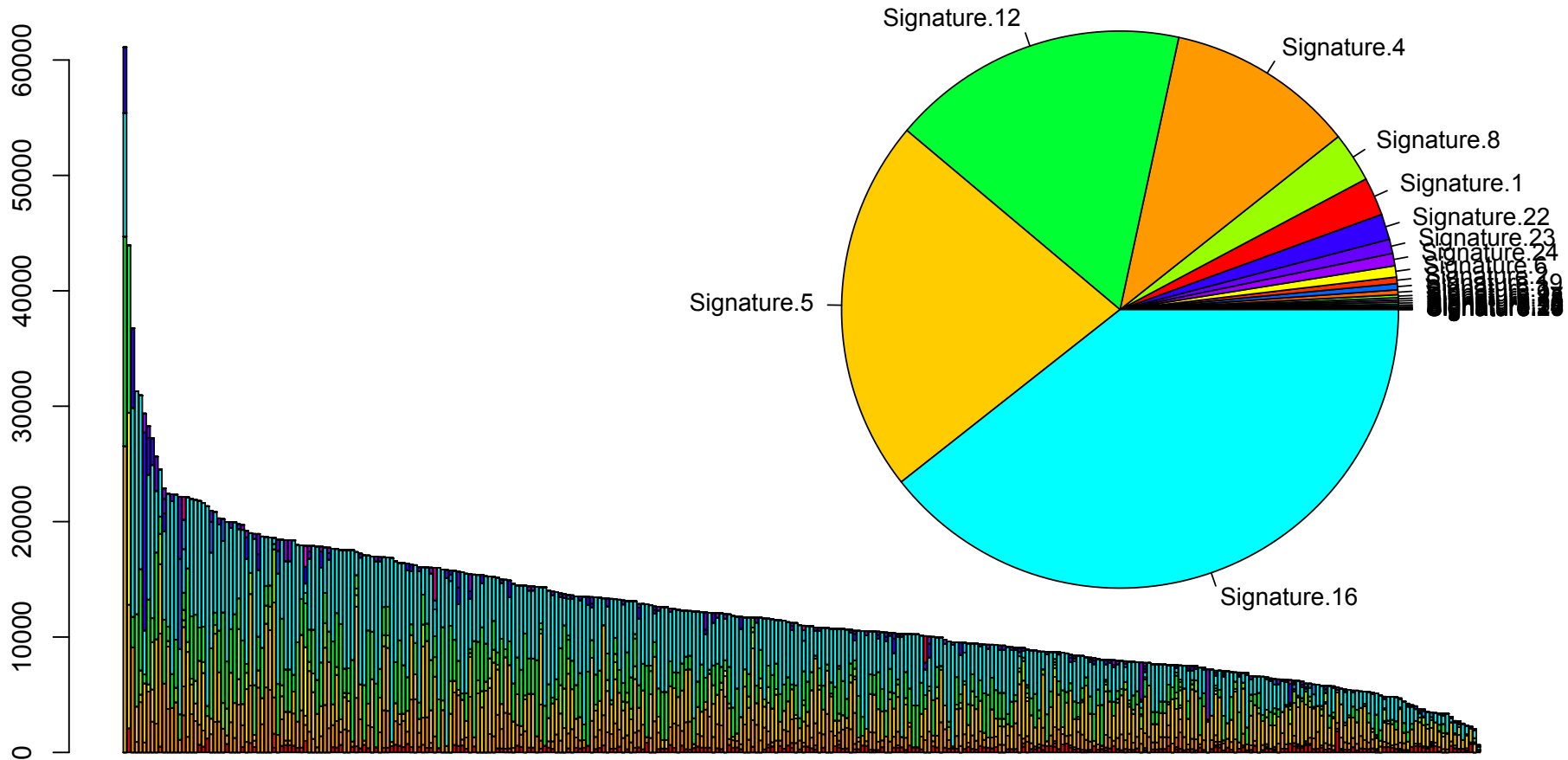
# Liver CA

Has smoking history documented and shows Sig4

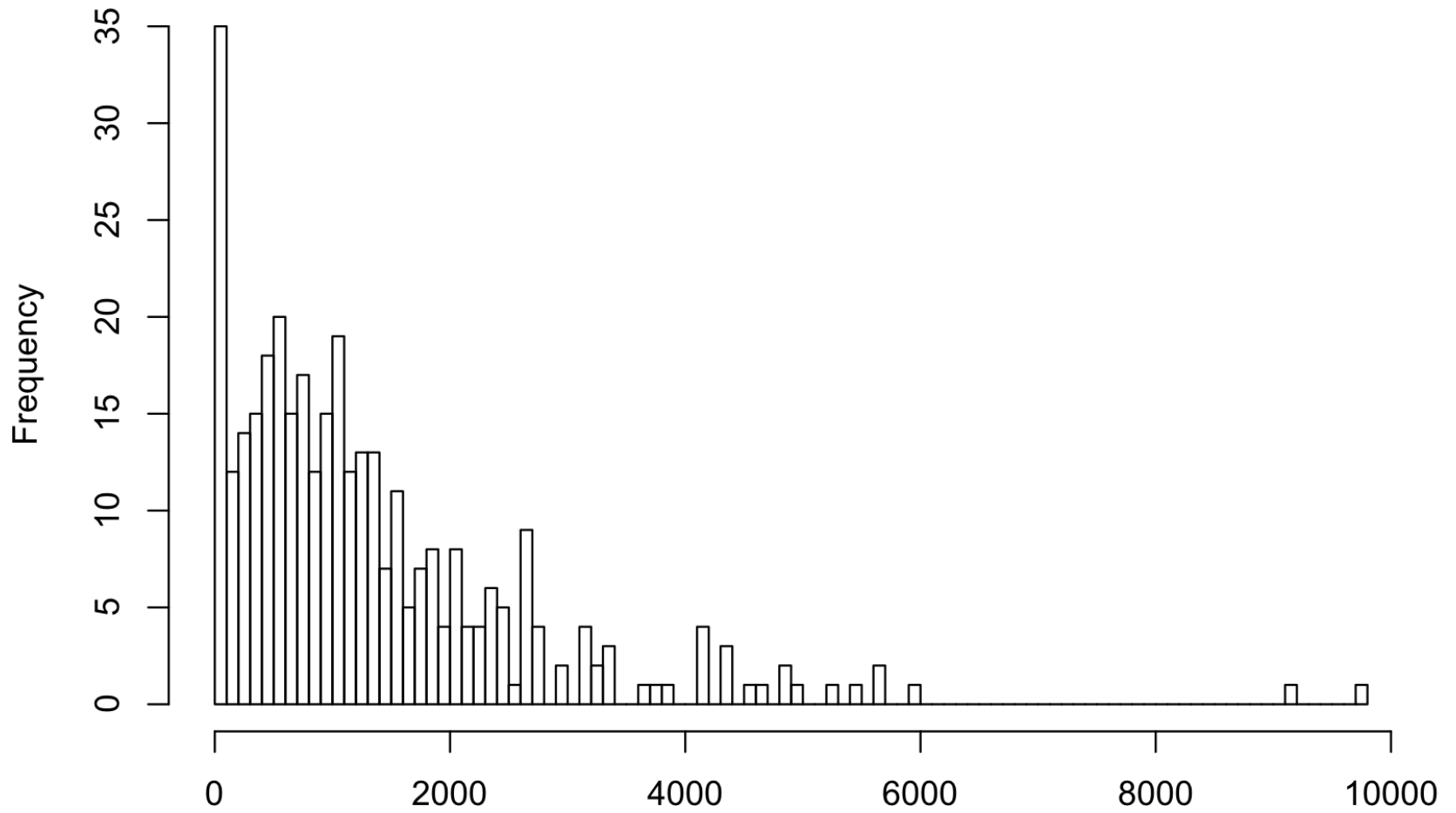


# Soft Prior

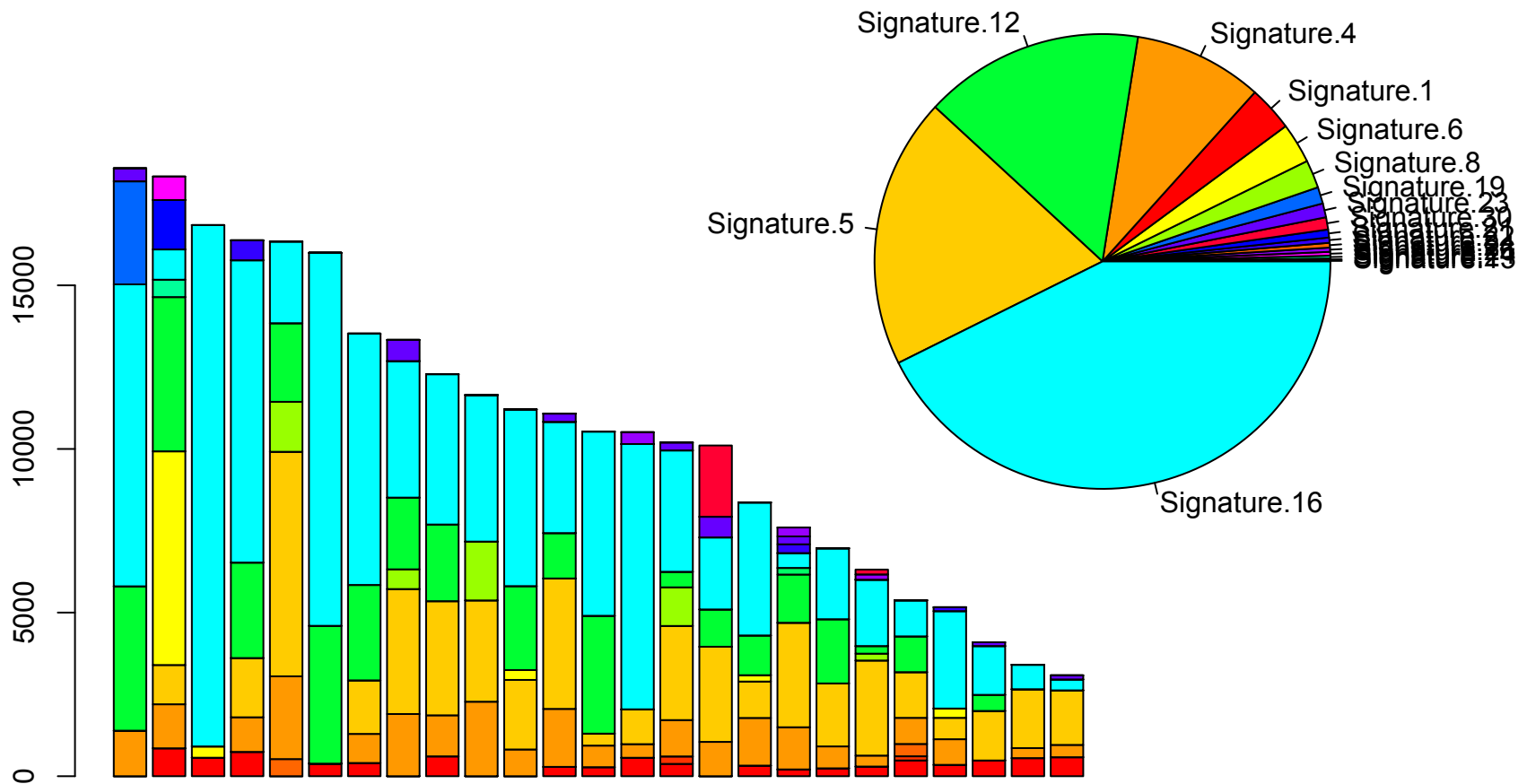
Sig 16/12 become the backbone

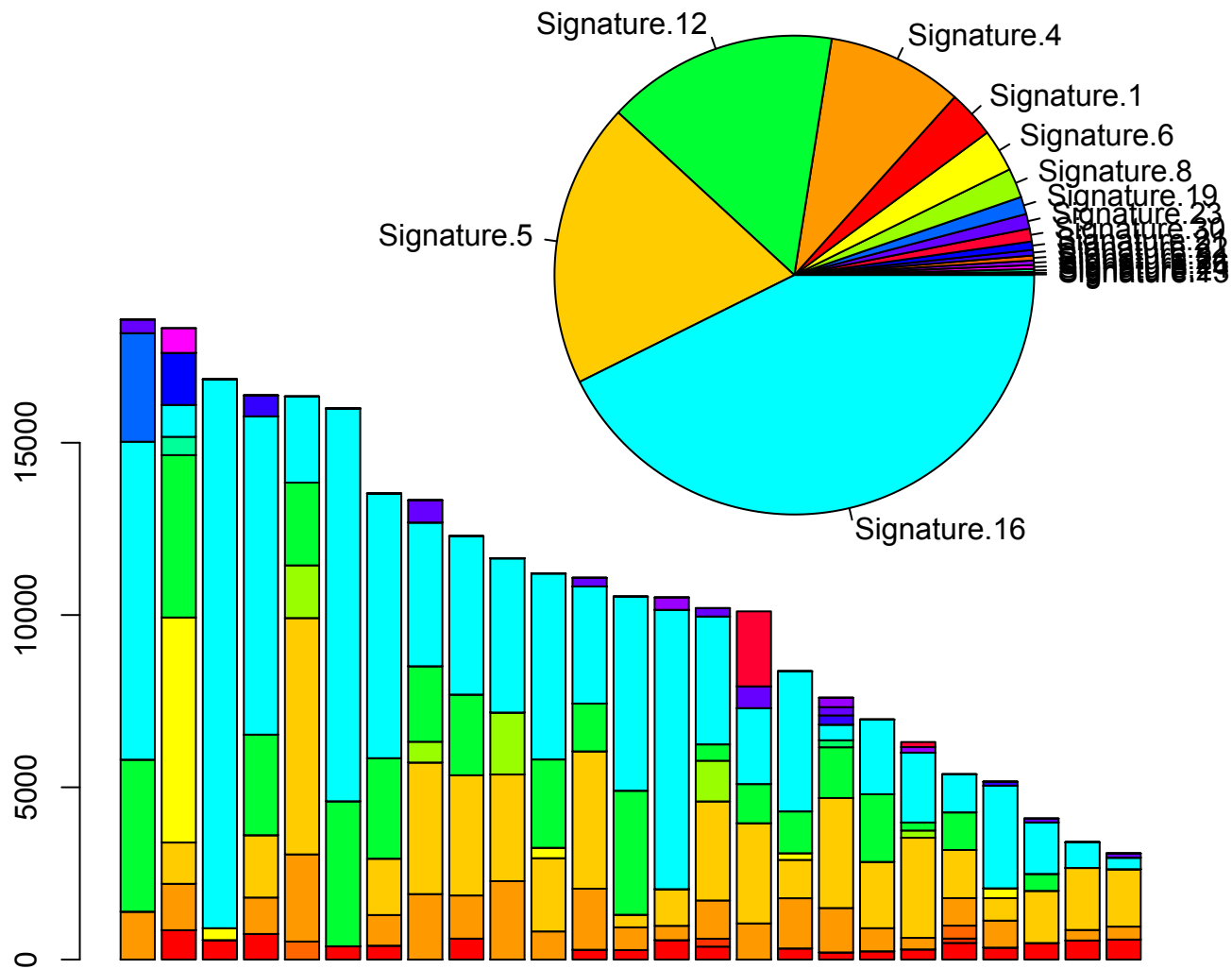
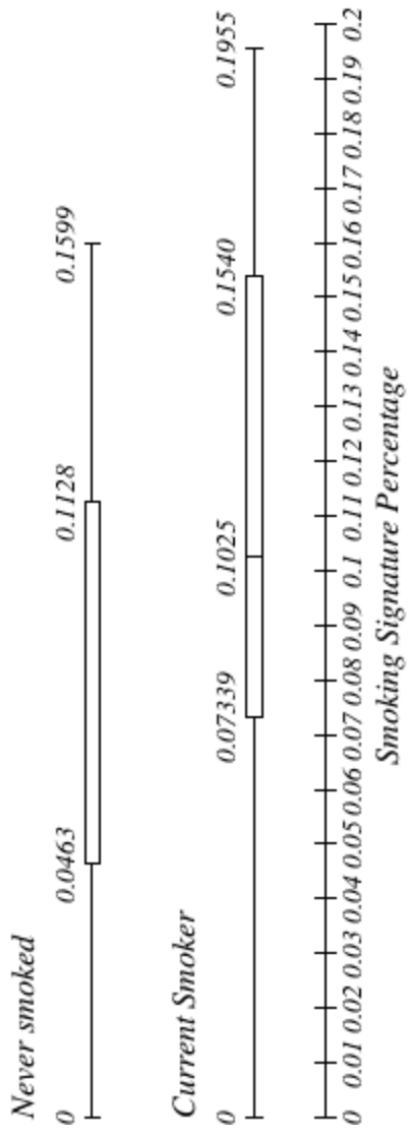


# Sig 4

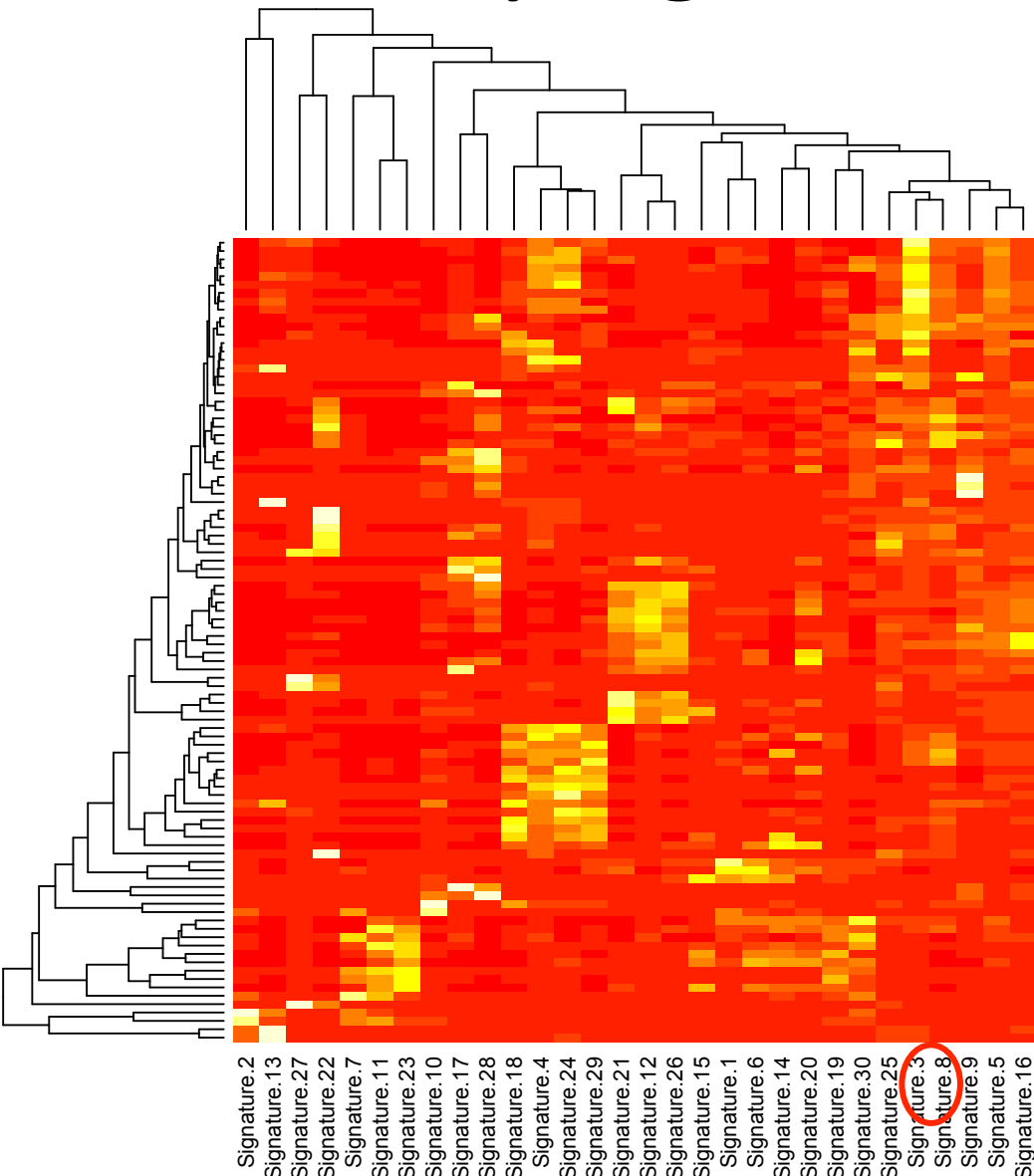


# 25 Samples w/ Smoking Records





# Why Sig. 3 and 8 Show up



C>G  
T>G  
[C]G>[G]G  
[C]G>[T]G (methylation)

• Sig3 is claimed to be associated with DSB repair failure (BRAC cancers)

• Contaminates pRCC too!

• Alexandrov in a 16' paper explicitly says “remove Sig3 based on bio. Knowledge”

• Sig3 is kind unique, similar to Sig5, close to Sig5 (clock!)

• I have more thoughts on this Signature correlation thing. More to come.

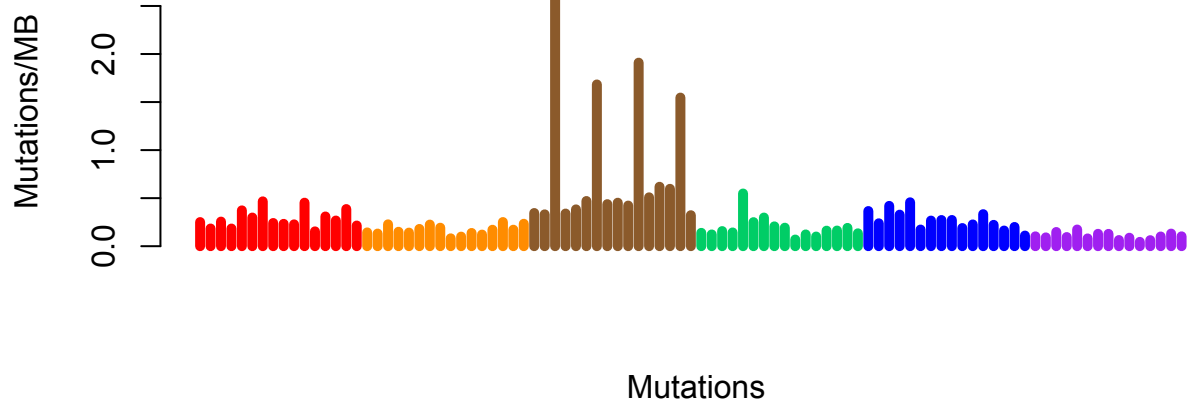
# Summary

- Prefer **soft prior**: better fit & gives leeway for “*other signatures*”
  - Seems to work in SCC, but not in RCC
  - Also the Alexandrov’s results are largely based on WXS
- Next step: run through all types, assign probabilities to each SNPs
- Also signature analysis looks interesting, also not the priority at this moment

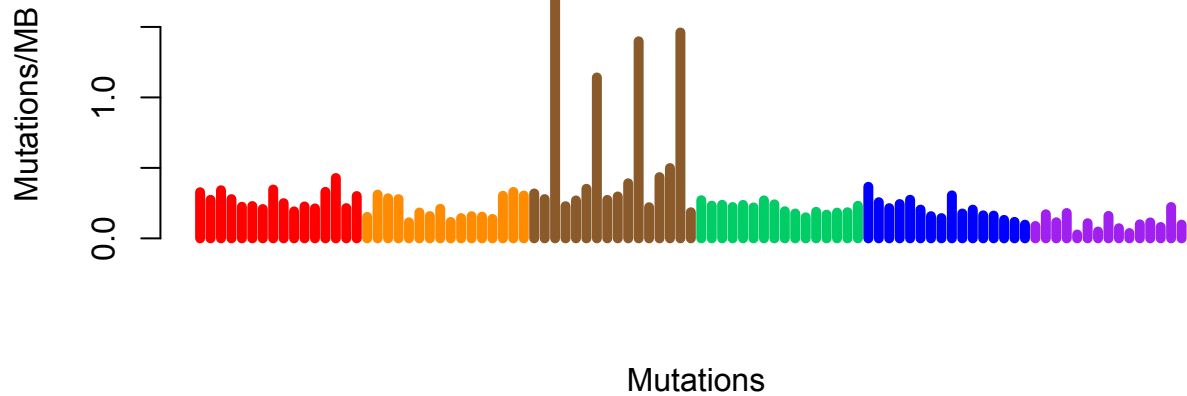


# Neutral vs. Nominal

Neutral

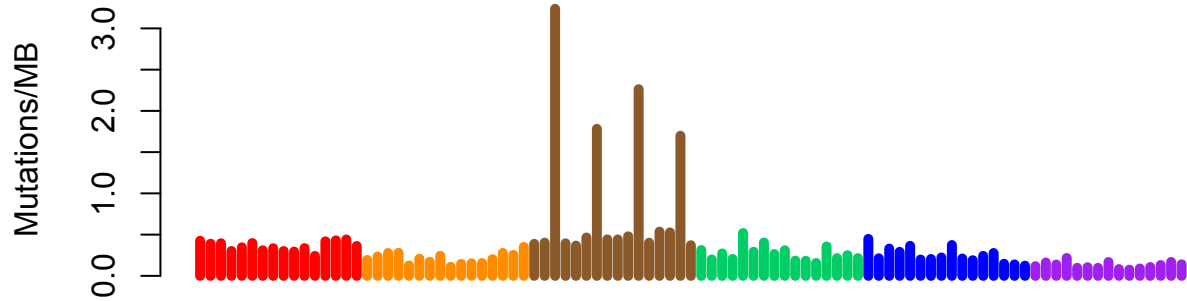


Nominal



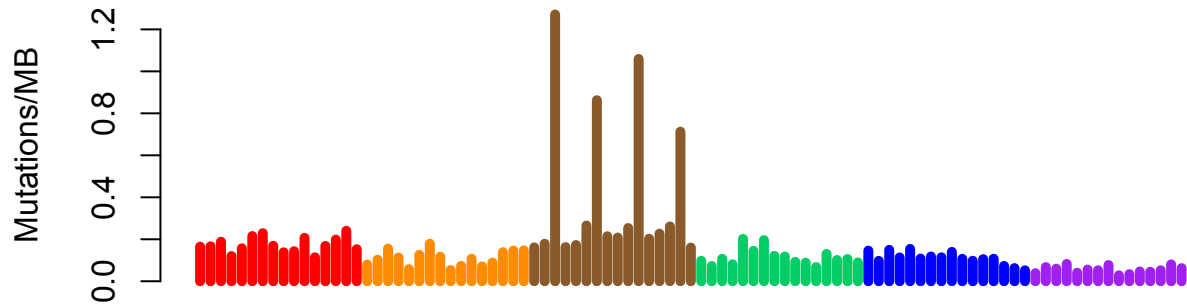
# Neutral vs. Nominal

Neutral

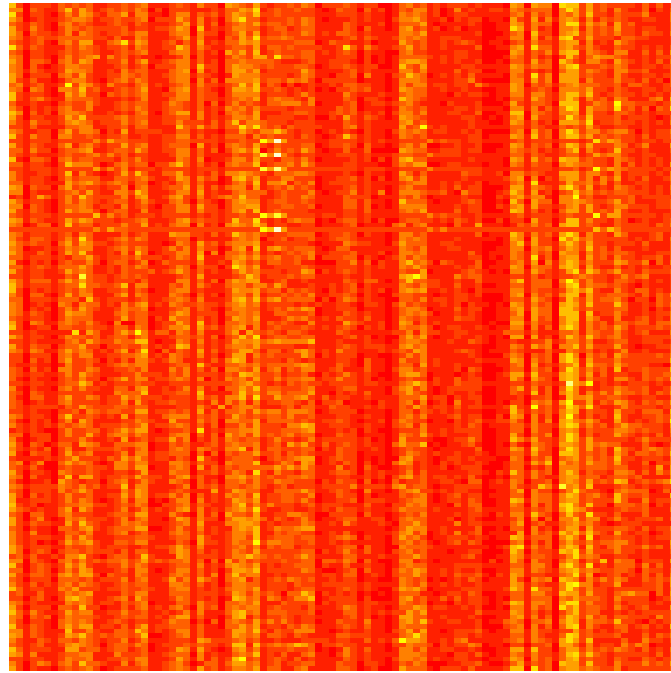
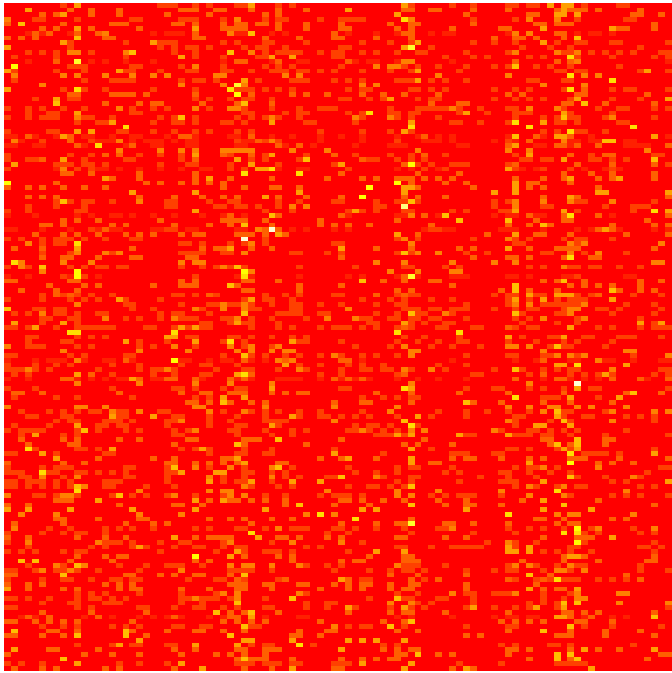
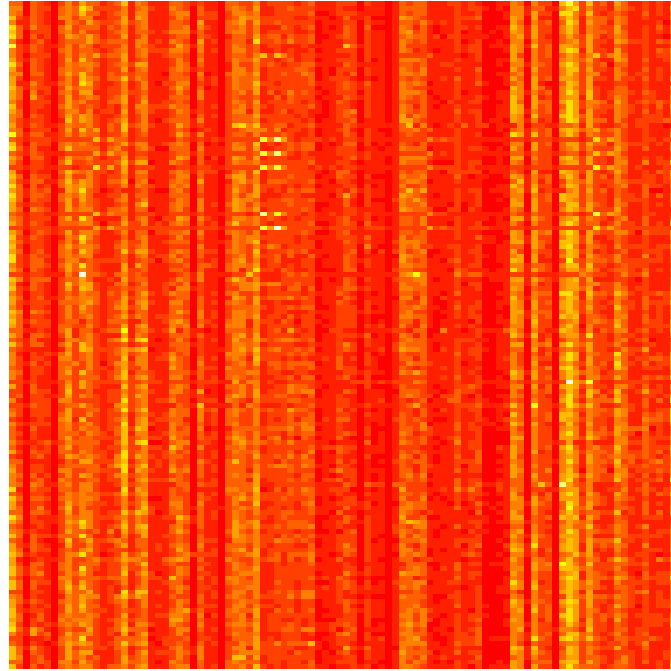
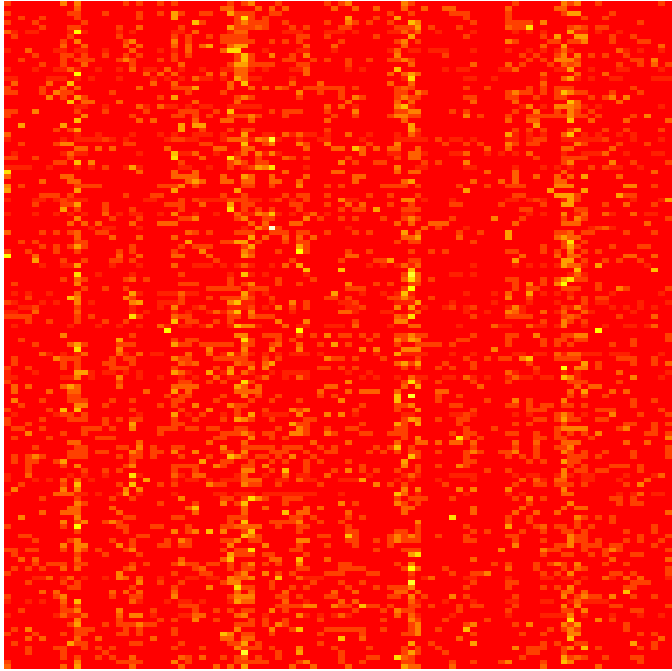


Mutations

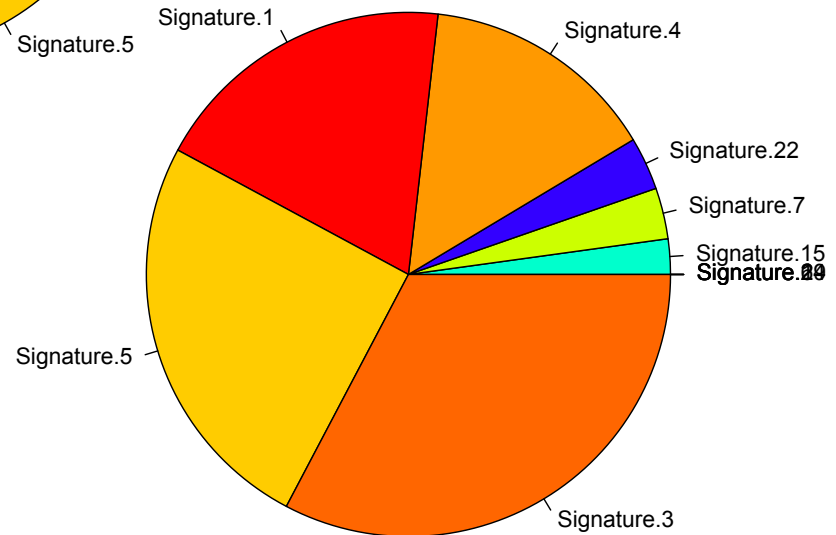
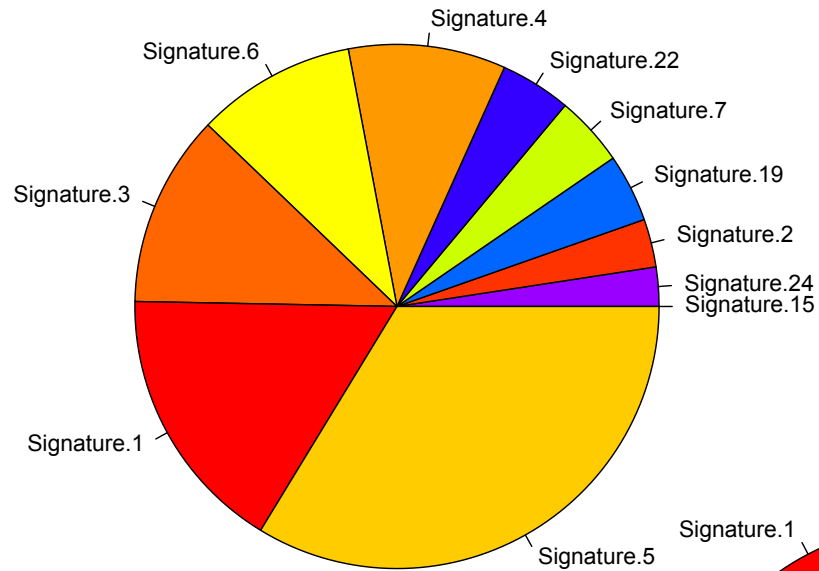
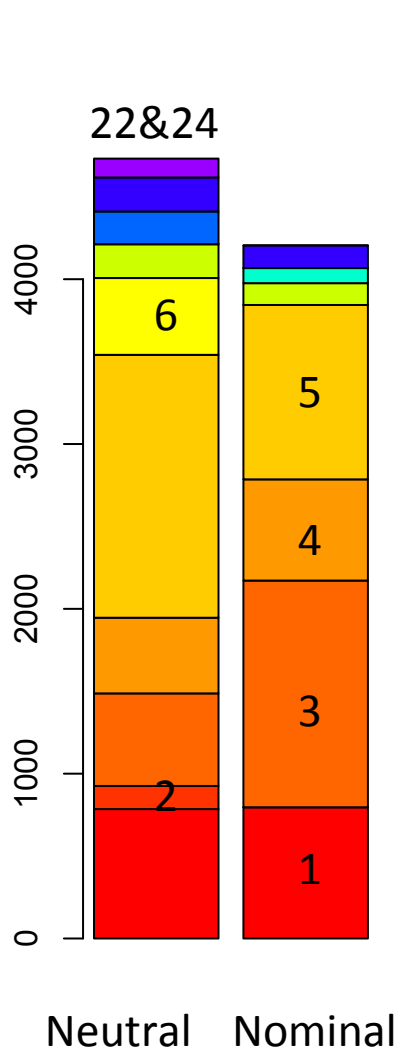
Nominal



Mutations

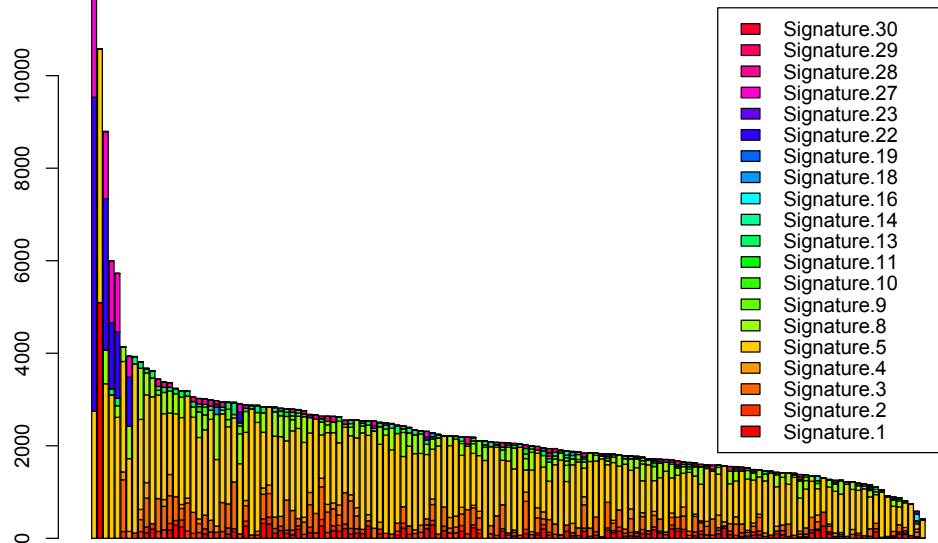


# Signature identified

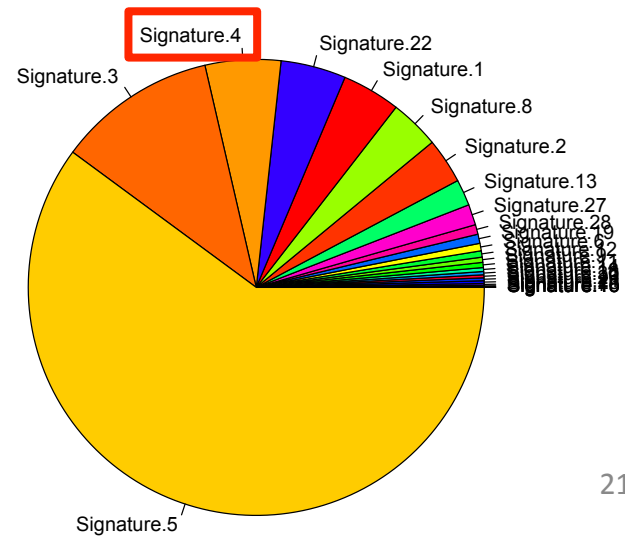
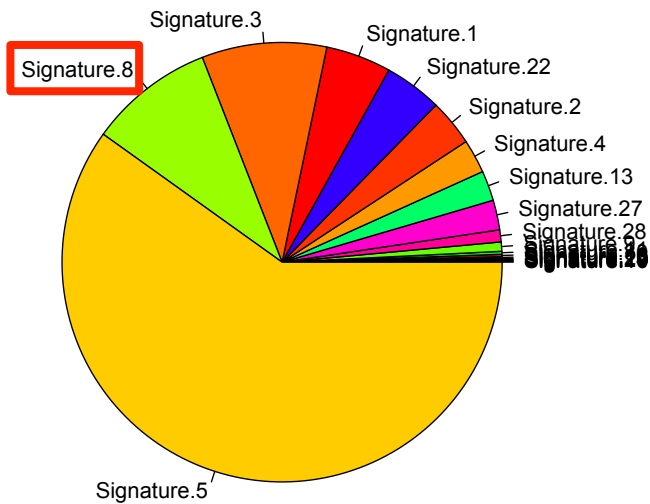
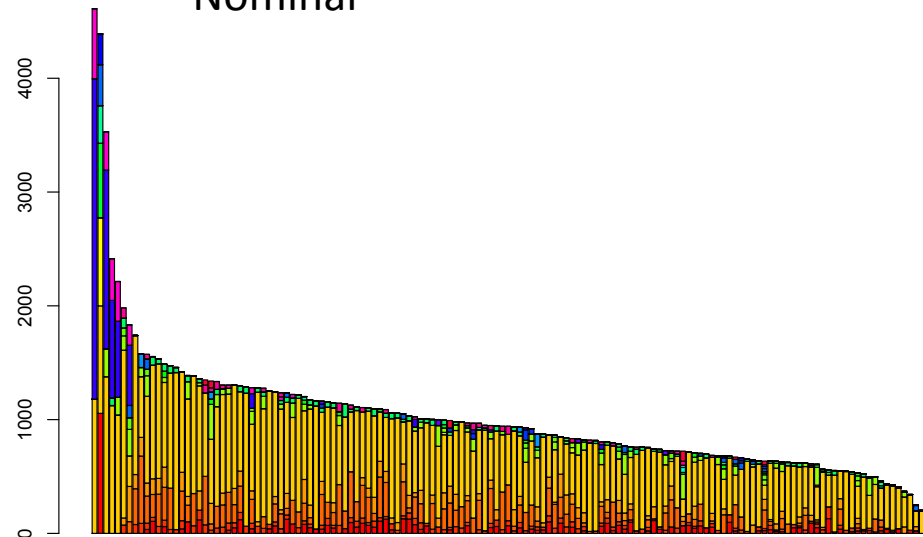


# In noncoding

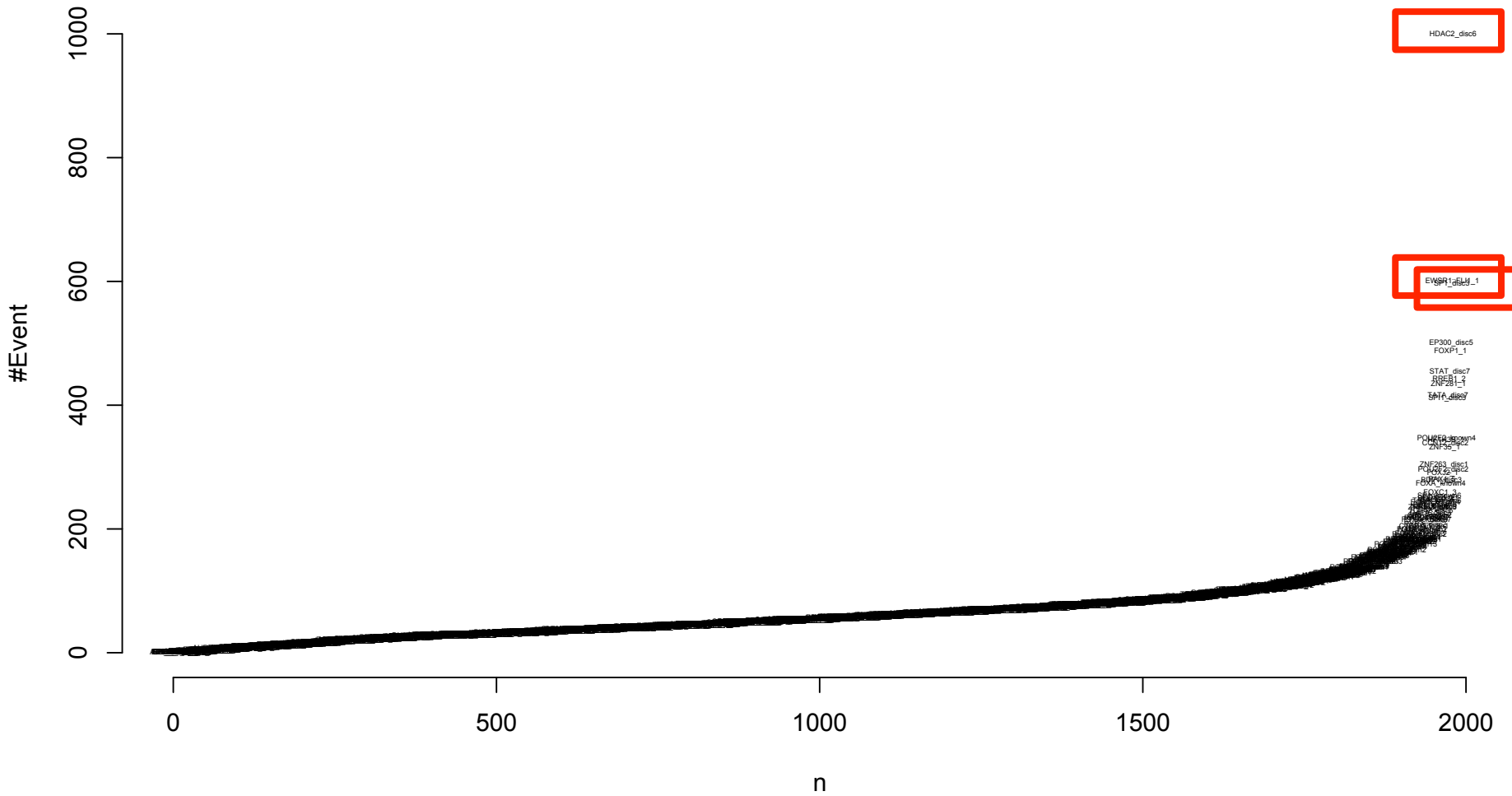
Neutral



Nominal

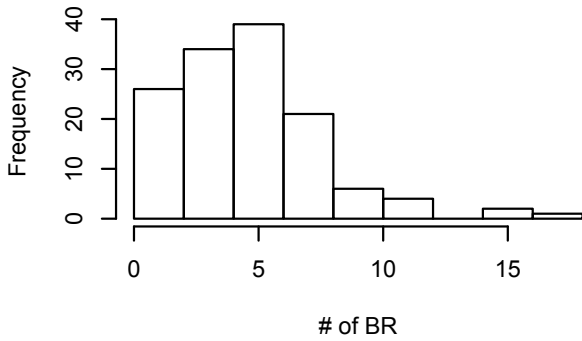


# Motif BR

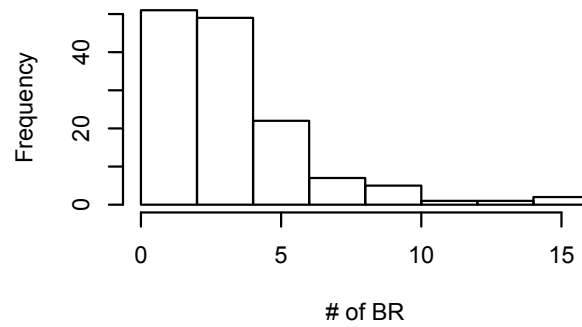


# Top 3 motifs

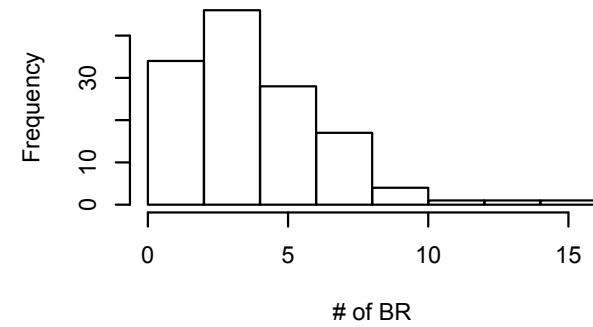
Freq. of HDAC2 BR



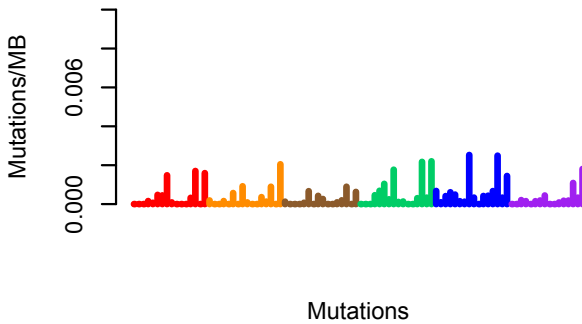
Freq. of EWSR1 BR



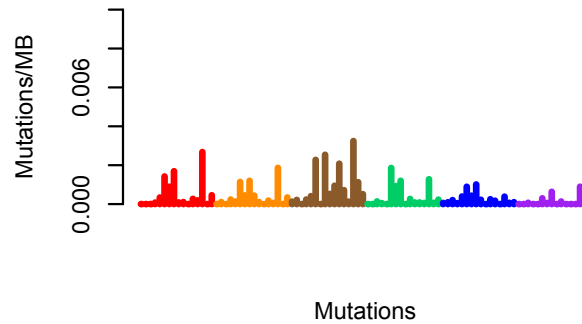
Freq. of SP1 BR



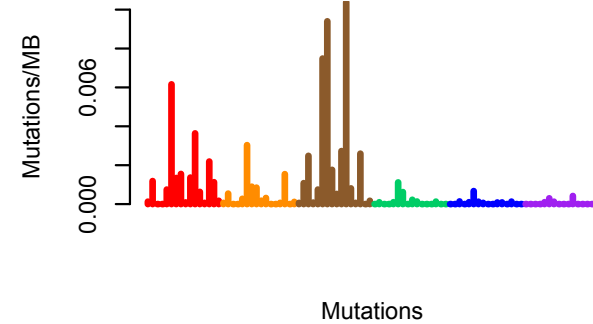
HDAC2 (N=675)

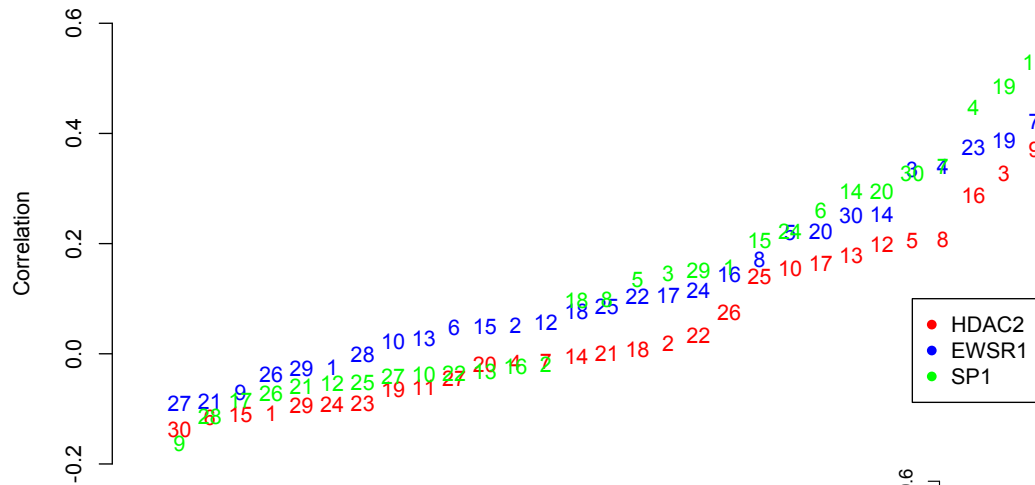


EWSR1 (N=514)

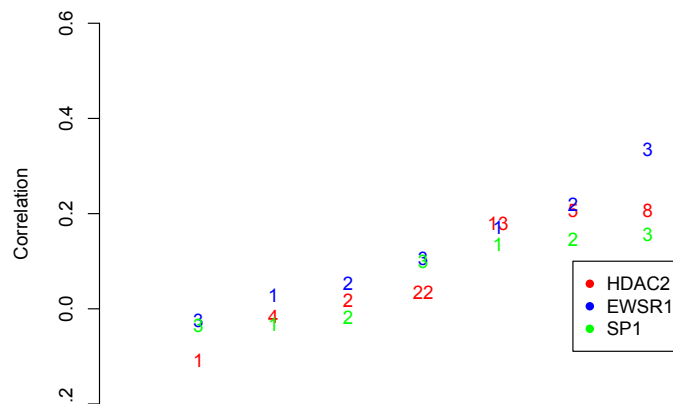


SP1 (N=571)

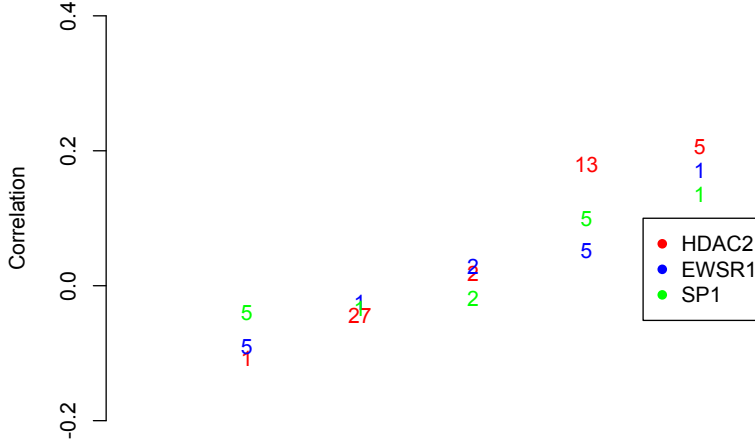




LASSO top Signatures=1,2,3,4,5,8,13,22



Alexsandrov's Signatures=1,2,5,8,13,27





# Questions

- Biological Stories?
  - Do we have a hypothesis? Why should signature differ?
- How to compare them?