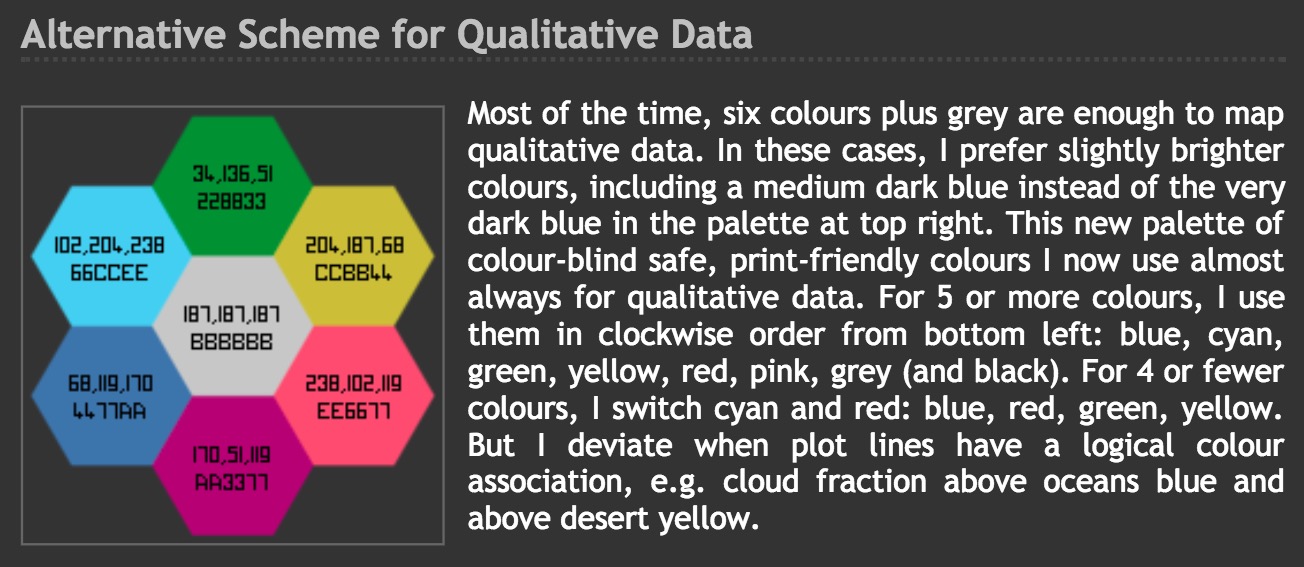
## For colors, I use this:

https://personal.sron.nl/~pault/



## Data prep

**Run: createFrequencies-Bacheler.R**

Needs: baseRscript.R

Needs: RResistanceMutations.r

Needs fasta files

writes: freqPatTs\_Bacheler\_Threshold1.csv

**Run: createOverviewDF-BachelerFilter.R**

needs: freqPatTs\_Bacheler\_Threshold1.csv

The main dataframe with data is called "OverviewDFilter"

written as OverviewSelCoeff\_BachelerFilter.csv

## % Rscript CheckWhetherDataCompWModelBacheler.R

## "MakeTrees\_To\_Check\_Hypermutation.R"



"MildHypermutationCheck.R"

## Create SingleSiteFrequencySpectra

**Run: analyseAndFigures-Bacheler.R**

needs: OverviewSelCoeff\_BachelerFilter.csv

needs: freqPatTs\_Bacheler\_Threshold1.csv

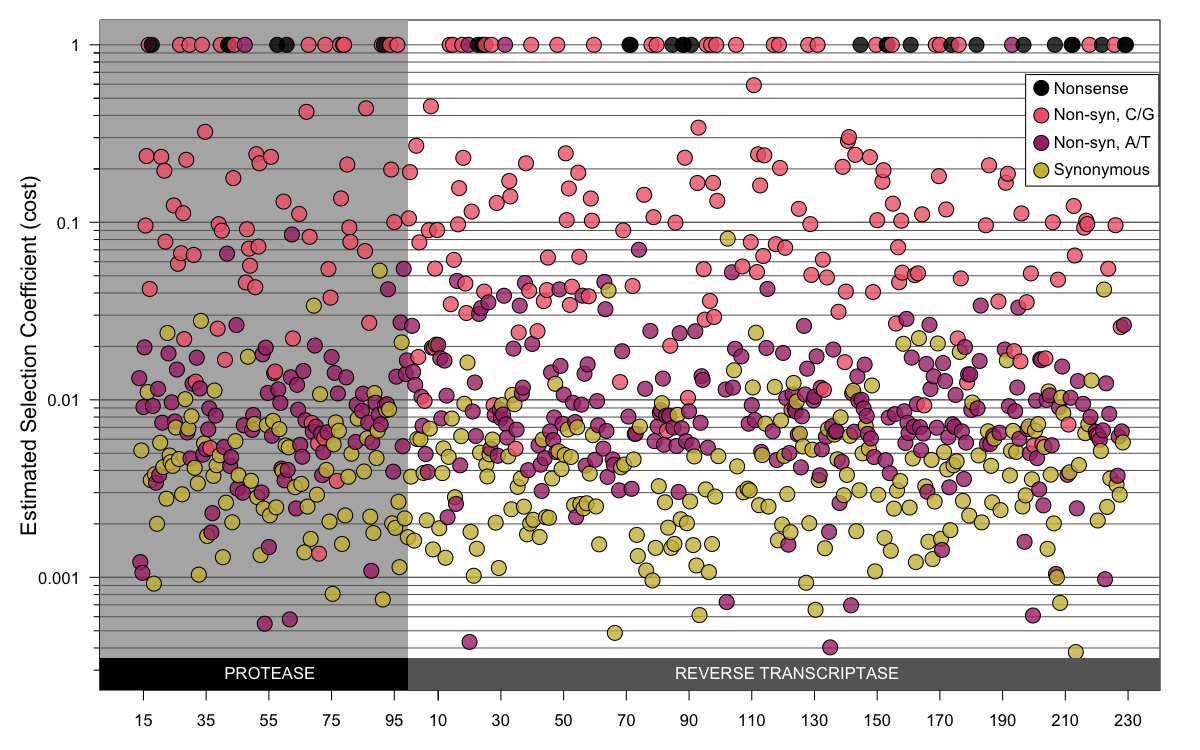
needs: "Output/SimFreqs172.csv", "Output/SimFreqs173.csv",

"Output/SimFreqs174.csv" (simulations are done with code in folder SimulationsEstimatingSelCoeffSims and "RScriptMutSel\_ImitateData.r" )

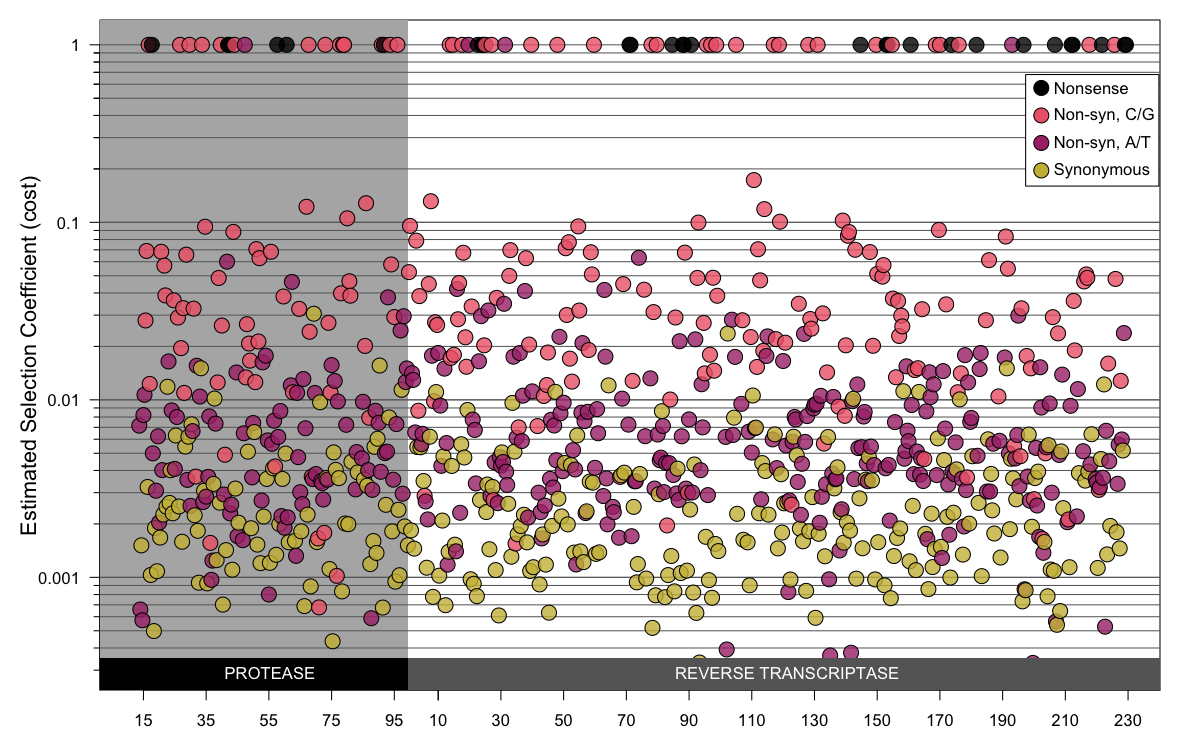
creates: SingleSiteFrequencySpectraPRO\_58\_Nov2017.pdf



creates: EstSelCoeffAbramPRO\_Nov2017.png



creates: EstSelCoeffZanPRO\_Nov2017.png (same as previous but using different mutation rates)



## Prep Zanini and Lehman data

**Run: createFrequencies-ZaniniFiles.R**

needs: Zanini datafiles from Data/ZaniniNeherData

writes: freqPatTs\_Zanini.csv

**Run: createOverviewDF-ZaniniFilter.R**

needs: freqPatTs\_Zanini.csv

needs: HIVMutRates /HIVMutRates.csv

writes: Output/OverviewSelCoeffZanini.csv

createFrequencies-Lehman.R

**writes:** Output/OverviewSelCoeffLehman.csv

writes: Output/OverviewSelCoeffZanini.csv

## Create ranking figures

**Run: "ranking\_Ordered\_Figure1.R"**

needs: OverviewSelCoeff\_BachelerFilter.csv

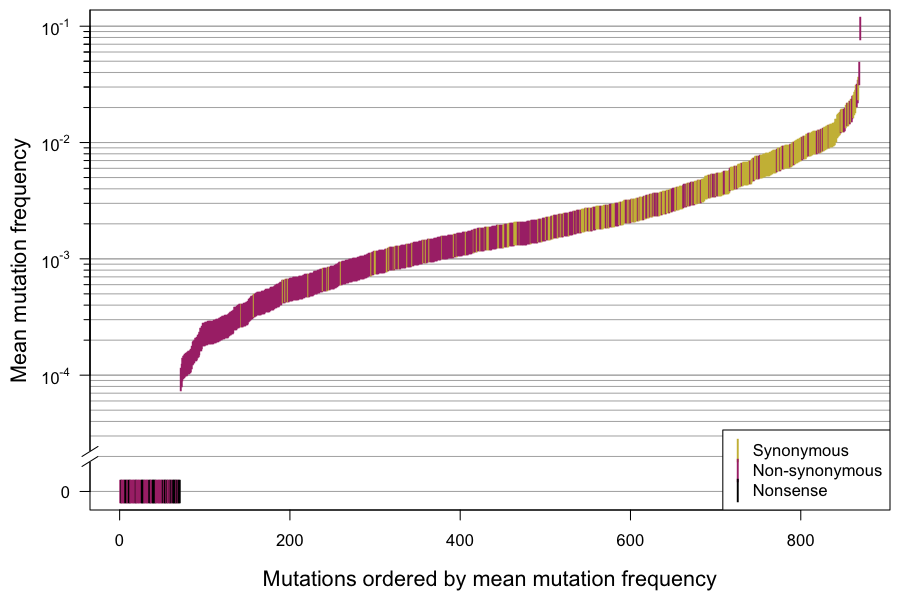
needs: Output/OverviewSelCoeffLehman.csv

needs: Output/OverviewSelCoeffZanini.csv

creates: F1-ordered-Nov2017Lehman-v3.png

creates: F1-ordered-Nov2017Zanini-v3.png

creates: F1-ordered-Nov2017Bacheler-v3.png



## Create DFE figures

**Run: MakeDFEHistsLogScale.R**

needs: OverviewSelCoeff\_BachelerFilter.csv

needs: Output/OverviewSelCoeffLehman.csv

needs: Output/OverviewSelCoeffZanini.csv

creates DFE\_log\_Nov2017\_Bacheler\_nonsyn\_.pdf

creates DFE\_log\_Nov2017\_Bacheler\_syn\_.pdf

creates DFE\_log\_Nov2017\_Lehman\_nonsyn\_.pdf

creates DFE\_log\_Nov2017\_Lehman\_syn\_.pdf

creates DFE\_log\_Nov2017\_Zanini\_nonsyn\_.pdf

creates DFE\_log\_Nov2017\_Zanini\_syn\_.pdf

## Create AA transition figure

**Run: AA.transition.figure.r**

needs: OverviewSelCoeff\_BachelerFilter.csv

creates: aachangesNS\_2017Nov.pdf

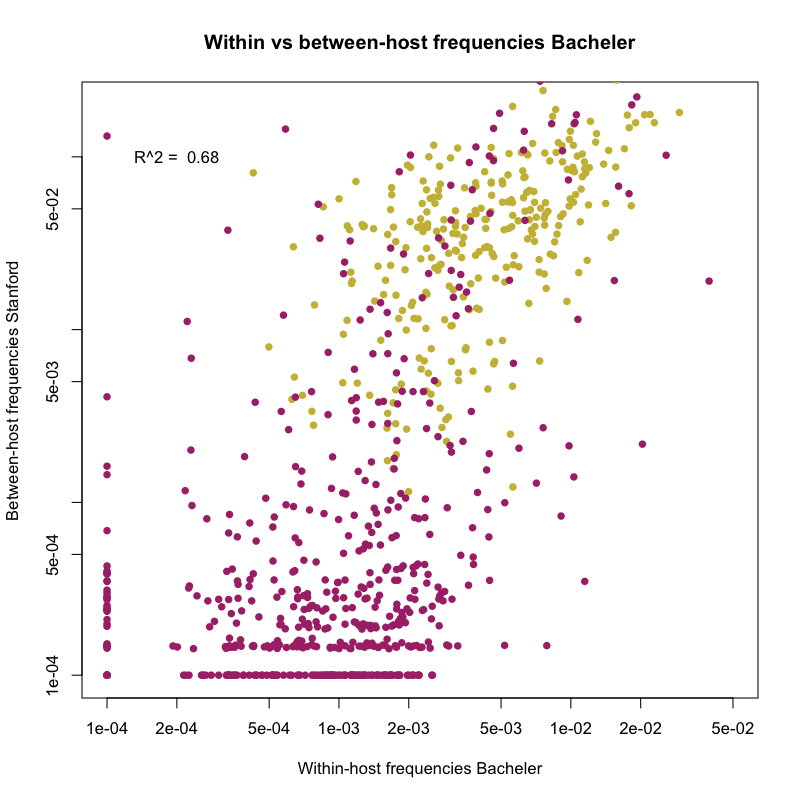


## Stanford epidemic data comparison

**What** createFrequencies-Stanford.R"

freqPatTs\_Stanford.csv

"CalculatePlotR2StanfordBacheler.R"



## Outlier analysis and location effect

"ConfIntervalsOutliersAnalysis.R"

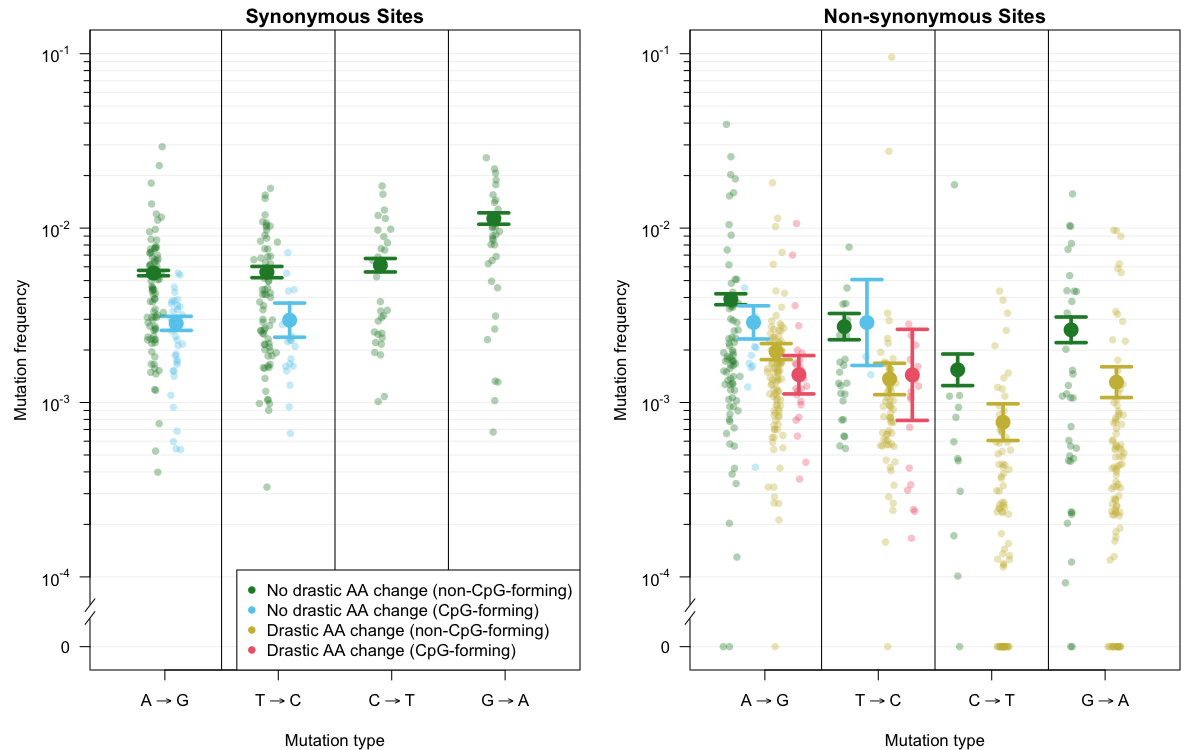
"RandomizeLocation.R"

## GLM code

"prepareDataForGLM.R"

GLM\_ModeledFreqsNov2017.r"

"helperFunctionsForGLMPlots.R"



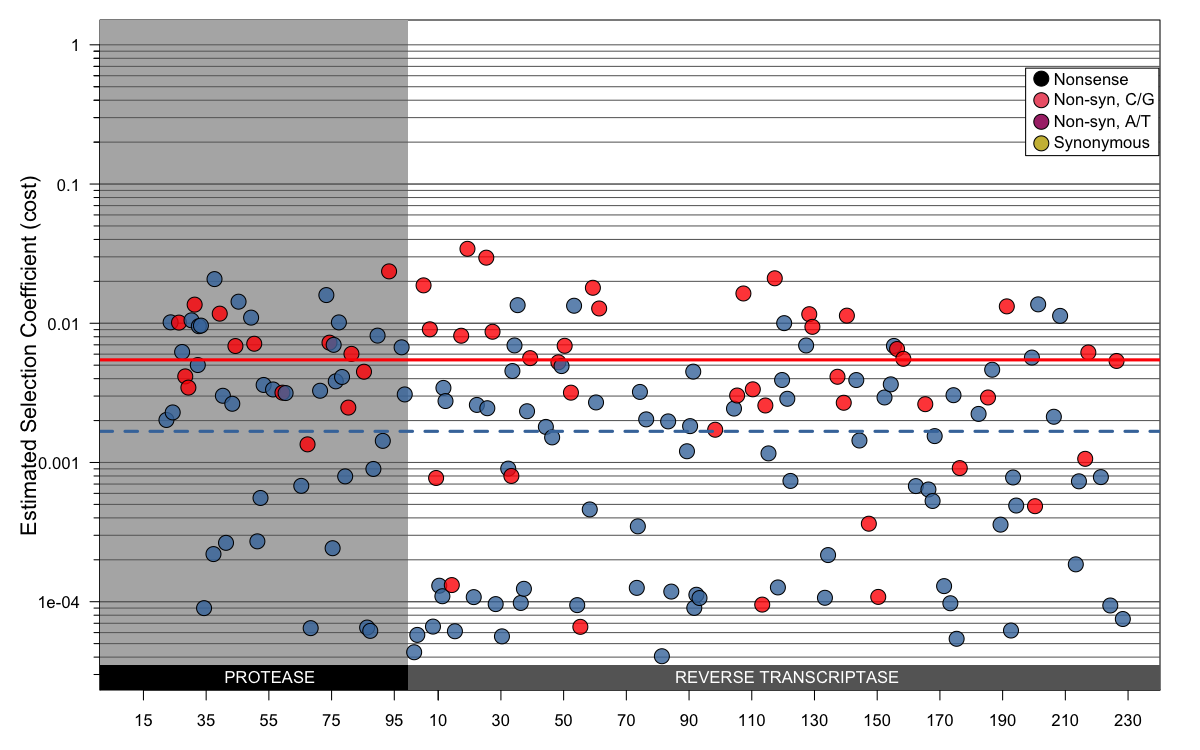
## Mutrates figure

"MutratesFig.R"



## Results not shown:

"CpGEffectZaniniData.R"



"MostCostlyAA\_ZaniniFS7.R"