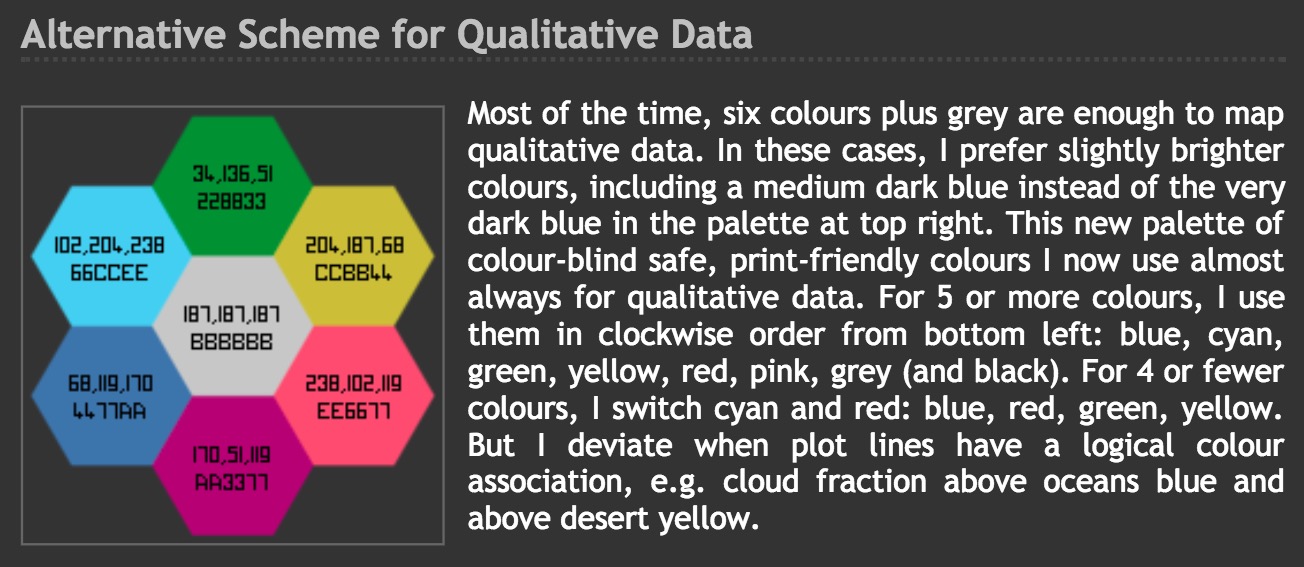
## For colors, we use this:

(set in baseRscript.R)

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

writes: needs: BachelerCountData\_Threshold1.csv

**Run: createOverviewDF-BachelerFilter.R**

needs: freqPatTs\_Bacheler\_Threshold1.csv

The main dataframe with data is called "OverviewDFilter"

written as OverviewSelCoeff\_BachelerFilter.csv

## CheckWhetherDataCompWModelBacheler.R

All code in Simulations folder

uses: ./Code\_and\_shellscript/HIVevolution\_HIV1site5000

based on ./Code\_and\_shellscript/HIVevolution\_HIV1.cpp

uses: ./Code\_and\_shellscript/tempscript.sh

uses: ./Code\_and\_shellscript/make\_HIV1site

creates: SimFreqs172.csv / SimFreqs173.csv / SimFreqs174.csv

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



"MildHypermutationCheck.R"

## Create SingleSiteFrequencySpectra

**Run: analyseAndFigures-Bacheler.R**

needs: OverviewSelCoeff\_BachelerFilter.csv

needs: freqPatTs\_Bacheler\_Threshold1.csv

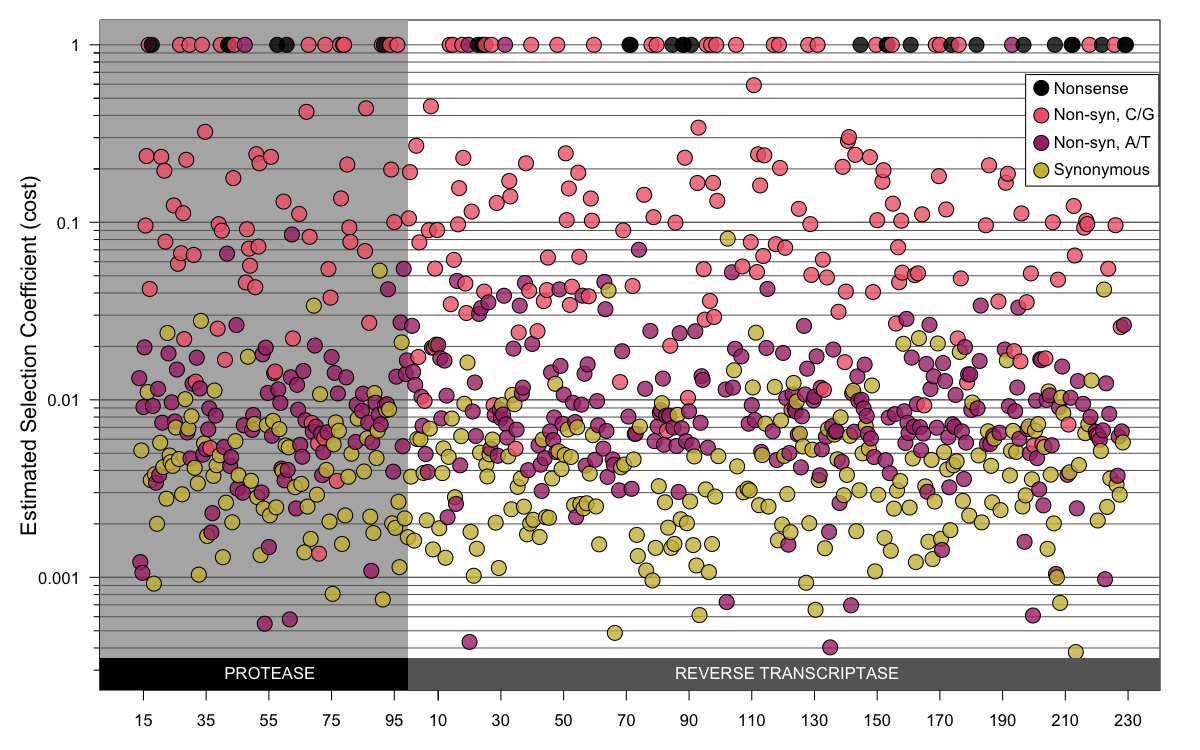
needs: "Output/SimFreqs172.csv", "Output/SimFreqs173.csv",

"Output/SimFreqs174.csv"

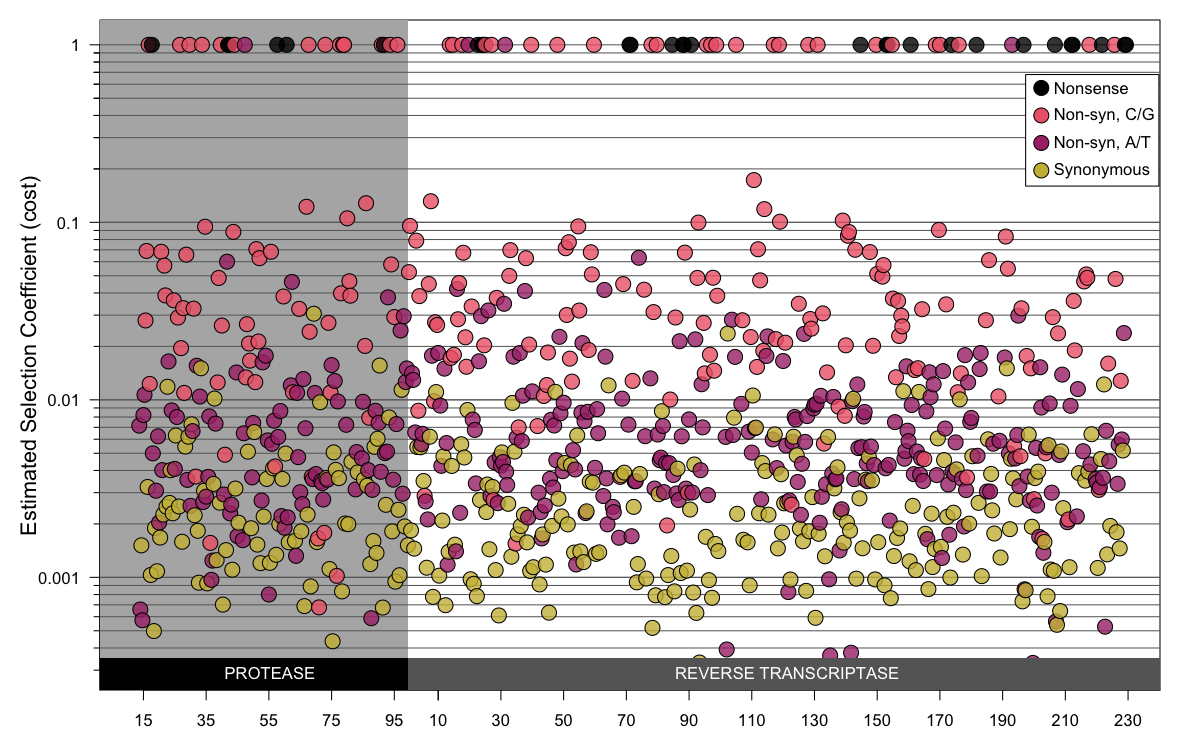
creates: SingleSiteFrequencySpectraPRO\_58\_Nov2017.pdf



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

**Run: createFrequencies-Lehman.R**

needs Data/LehmanData/PleuniAlignments (alignments thanks to Dr Scott Roy)

writes: Output/OverviewSelCoeffLehman.csv

writes: Output/OverviewSelCoeffZanini.csv

## Create ranking figures

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

needs: OverviewSelCoeff\_BachelerFilter.csv

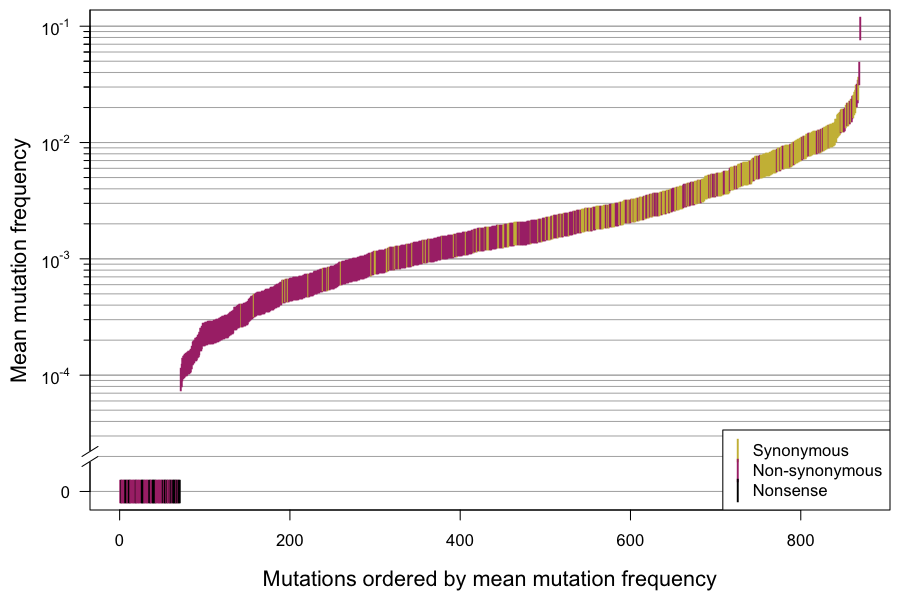
needs: Output/OverviewSelCoeffLehman.csv

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

**Run: createFrequencies-Stanford.R"**

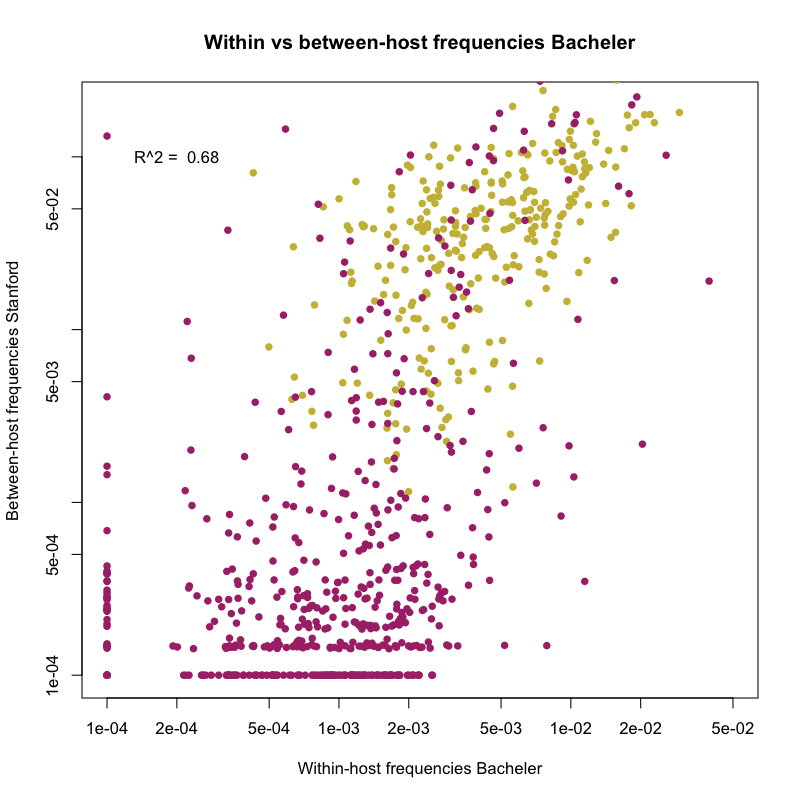
needs: Data/StanfordData/subtypeB-pr\_naive\_aligned.fasta

needs: Data/StanfordData/subtypeB-rt\_naive\_aligned.fasta

creates: freqPatTs\_Stanford.csv

**Run: CalculatePlotR2StanfordBacheler.R"**

creates: StanfordVsBacheler2017Nov27.png



## Outlier analysis and location effect

ConfIntervalsOutliersAnalysis.R

RandomizeLocation.R

## GLM code

prepareDataForGLM.R

needs: BachelerCountData\_Threshold1.csv

needs: OverviewSelCoeff\_BachelerFilter.csv

needs: Pol\_SHAPE.csv

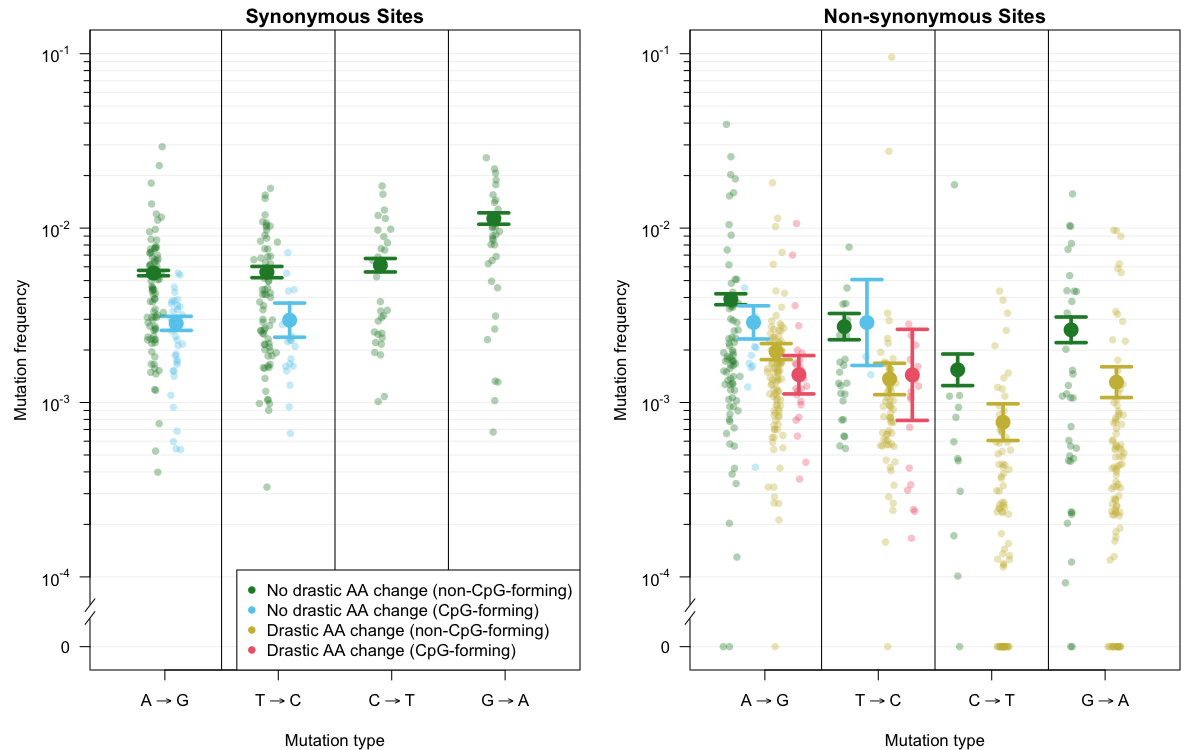
writes: datFitModel.csv

Run: GLM\_ModeledFreqsNov2017.r

writes: GLMResultsText.txt

needs: helperFunctionsForGLMPlots.R

creates: modeled\_freqs\_Sep2017\_2.png



creates: modeled\_sels\_May2017.pdf

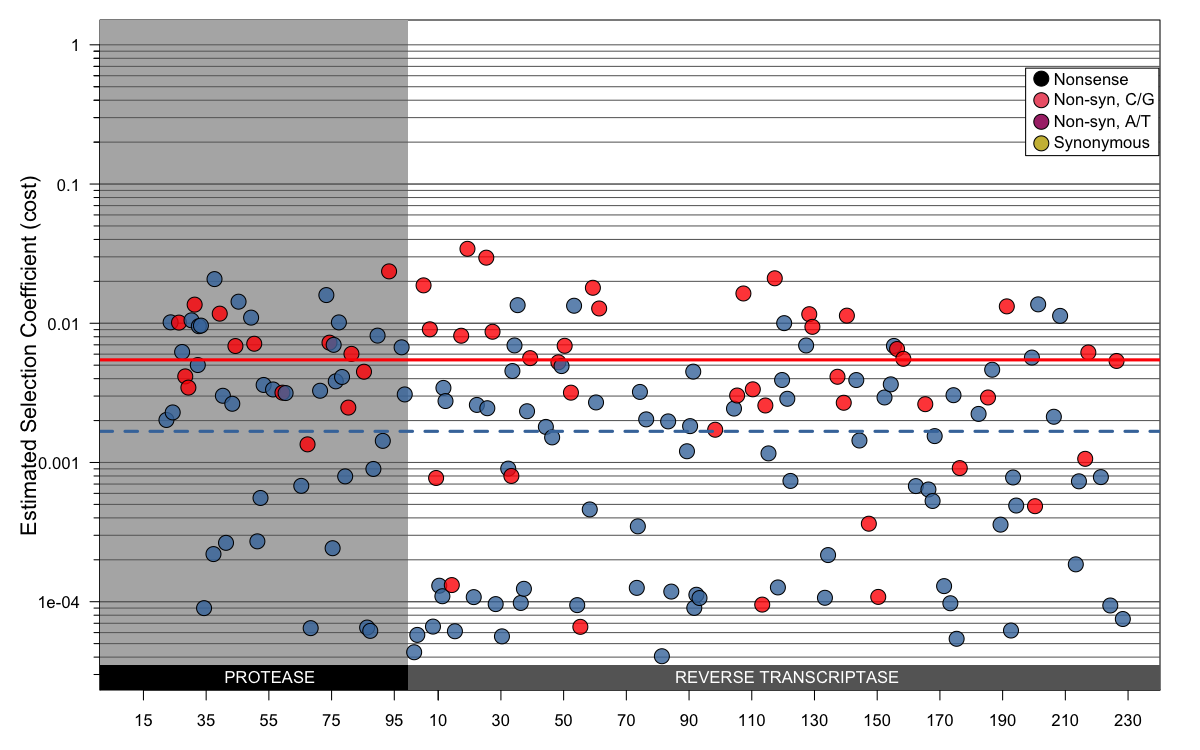
## Mutrates figure

"MutratesFig.R"



## Results not shown:

"CpGEffectZaniniData.R"



"MostCostlyAA\_ZaniniFS7.R"