Tasks:

ROC

1. Static:
   1. create a gold-standard pool of pathogenic variants from:
      1. GS parameters (het = 0.5) 🡪 allele frequency 🡪 take ones under cutoff.
      2. Find patients that are sick according to (+) variant in ANY of the genes
      3. Find mean variants that pass, and mean patients that are sick.
      4. MOST interested in vector of variants that pass
   2. Static ROC curve comes from all spaces in allele frequency [0,1] 🡪 variants classifications
   3. Don’t worry about het/prev/pen yet!
   4. Collect data on: X/1000 variants qualify across Y genes and Z% of patients have HCM, with abs(Z1) number healthy and abs(Z2) number sick.
   5. ROC curve is defined as TPR/FPR
   6. True positives are when you actually call a variant the right thing.
2. Dynamic
   1. Allow individual points to be plotted for het/prev/pen that sketch a new ROC curve.
3. Notes
   1. Resilience people at the low TPS/FPS side
   2. Incidentalome people at the high TPS/FPS side.

Kill all multi-indels?

Look at which cohorts you're downloading from.

Which population are contained.

Barplot + cutoffs:

Space them out, label with penetrance values

If any subpopulation has a high frequency, then this variant is not pathogenic;

More sensitive way of finding variants – ancestrally informative

ACMG for all thingies, not just BRCA1

Scriptize your downloading of ExAC/1000G.