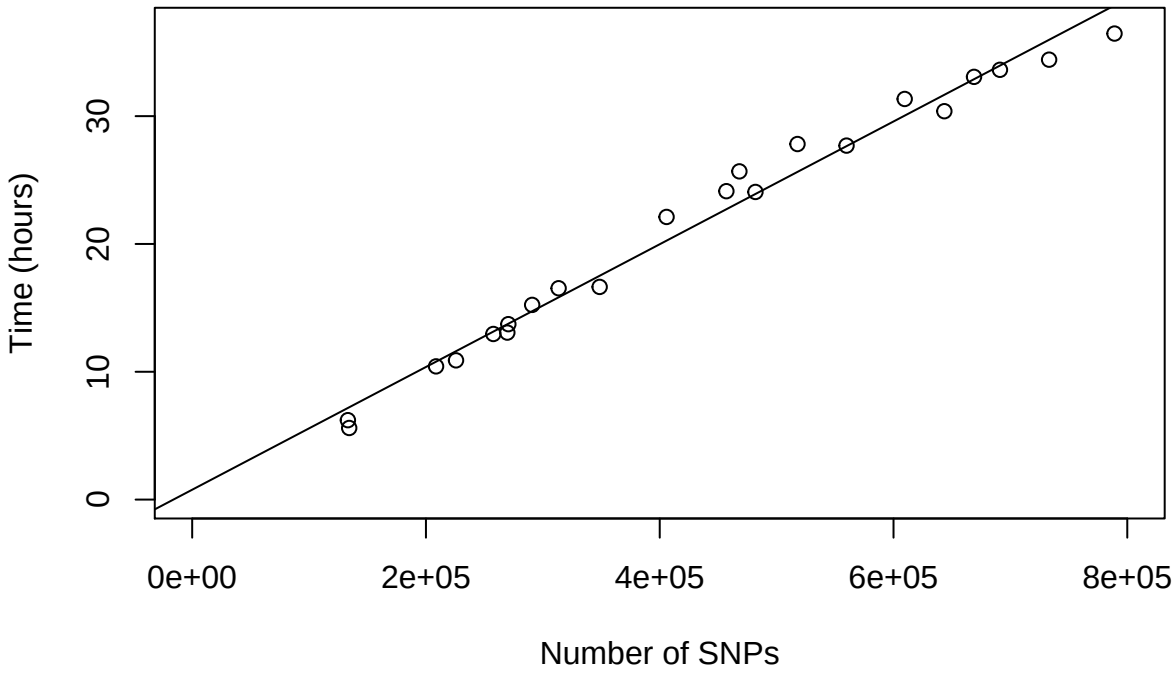
Time required for plink to run on Biobank

2016-08-04

Timings for:

- 120000 samples
- 15 covariates
- linear model
- 9478363 SNPs
- Single core per chromosome

chr	hrs	nsnp
1	34.418889	733132
2	36.462222	789072
3	33.074444	668873
4	33.626667	691061
5	31.344167	609559
6	30.385556	643488
7	27.695556	559809
8	27.818889	517857
9	22.114722	405848
10	24.067778	481873
11	25.678889	468149
12	24.130000	457017
13	16.639444	348590
14	16.533889	313495
15	13.723333	270486
16	15.231944	290855
17	12.955278	257642
18	13.065000	269623
19	10.892778	225733
20	10.421389	208646
21	6.212222	133218
22	5.601667	134337



Total time for a single core: 472.09 hours.

Predicted time for 28 cores: 16.86 hours.

Note: Running chr 22 again on its took 3.75 hours, substantially less than when things are run together. This is probably because 16 cores reading in data concurrently led to a much longer read-in time. Reducing the biobank data size in terms of number of SNPs would probably improve things, these analyses were run on the unfiltered data.