Implementation: Functions

1. **Data receiving functions:**
   - Upload data, impute data, summarize data

2. **Data handling functions**
   - Retrieval of genotype from bulk-data, caching of genotypes for quick access, retrieval of genotypes from quick-access

3. **Genetic risk scoring functions** (*lots of options*)
   - One person, one SNP:
     Trivial
   - One person, many SNPs
     Count effect-allele
     Count effect-alleles, *beta-weight
     Count effect-alleles, *beta-weight, normalized, unit-variance
     LDPred-hybrid (only for select GWAS)

Implementation: PRS-functions

Counting effect-alleles

	Effect allele	Effect Size	Person 1 genotype	Effect allele count
SNP1	G	0.97	A/G	1
SNP2	C	0.76	C/C	2
SNP3	A	0.51	G/G	0
SNP4	C	0.59	C/C	2
SNP5	G	1.1	C/C	0
SNP6	G	-0.31	G/T	1
SNP7	C	-0.56	C/C	2

Implementation: PRS-functions

Counting effect-alleles, *beta-weight

(identical to PLINK --SCORE)

	Effect allele	Effect Size	Person 1 genotype	Effect allele count	Weighted score
SNP1	G	0.97	A/G	1	0.97
SNP2	C	0.76	C/C	2	1.5
SNP3	A	0.51	G/G	0	0
SNP4	C	0.59	C/C	2	1.2
SNP5	G	1.1	C/C	0	0
SNP6	G	-0.31	G/T	1	-0.31
SNP7	C	-0.56	C/C	2	-1.1

Implementation: PRS-functions

Counting effect-alleles, *beta-weight, normalized

(Purpose: when per-SNP $mean_{population}$ = 0
-> more robust against missing SNPs)

To get this: Multiply with the frequency of the effect allele in the population:

$$\mathbf{Score_{population} = freq * 2 * beta}$$

Where *freq* is the effect allele frequency

To get this, subtract *average population score* from *weighted score*

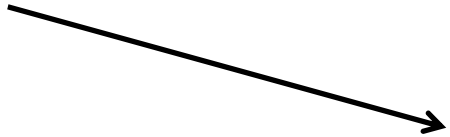
	Effect allele	Effect Size	Person 1 genotype	Effect allele count	Weighted score	Average population score	Normalized score
SNP1	G	0.97	A/G	1	0.97	0.45	0.53
SNP2	C	0.76	C/C	2	1.5	0.12	1.4
SNP3	A	0.51	G/G	0	0	0.12	-0.12
SNP4	C	0.59	C/C	2	1.2	0.21	0.97
SNP5	G	1.1	C/C	0	0	0.066	-0.066
SNP6	G	-0.31	G/T	1	-0.31	-0.27	-0.037
SNP7	C	-0.56	C/C	2	-1.1	-0.089	-1
							1.67

Implementation: PRS-functions

Counting effect-alleles, *beta-weight, normalized, unit variance

(Purpose: Z-scores -> 68% is within score of “1”)

Calculate score variance based on MAF and Hardy-Weinberg, then scale all



	Effect allele	Effect Size	Person 1 genotype	Effect allele count	Weighted score	Average population score	Normalized score	Normalized score (unit variance)
SNP1	G	0.97	A/G	1	0.97	0.45	0.53	
SNP2	C	0.76	C/C	2	1.5	0.12	1.4	
SNP3	A	0.51	G/G	0	0	0.12	-0.12	
SNP4	C	0.59	C/C	2	1.2	0.21	0.97	
SNP5	G	1.1	C/C	0	0	0.066	-0.066	
SNP6	G	-0.31	G/T	1	-0.31	-0.27	-0.037	
SNP7	C	-0.56	C/C	2	-1.1	-0.089	-1	
							1.67	**2.17**

Implementation: PRS-functions

Counting effect-alleles, *beta-weight, normalized, unit variance

(Purpose: Z-scores -> 68% is within score of “1”)

```
#per-SNP hardy-weinberg
frac_0 <- (1-effect_allele_freq)^2
frac_1 <- (1-effect_allele_freq)*(effect_allele_freq)*2
frac_2 <- (effect_allele_freq)^2
```

$$p^2 + 2pq + q^2 = 1$$

```
#per-SNP mean and standard-deviation of score
mean <- (frac_1 * 1 * effect_size + frac_2 * 2 * effect_size)
sd_per_snp_squared <- (0*effect_size - mean)^2 * frac_0 +
                      (1*effect_size - mean)^2 * frac_1 +
                      (2*effect_size - mean)^2 * frac_2

sd_per_snp <- (sd_per_snp_squared)^0.5
```

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N}} \text{"}ish\text{"}$$

”N” is fractions instead

```
#all-SNP standard-deviation of score
overall_sd <- sqrt(sum(sd_per_snp^2))
```

Implementation: PRS-functions

Counting effect-alleles, *beta-weight, normalized, unit variance

(Purpose: Z-scores -> 68% is within score of “1”)

```
#per-SNP hardy-weinberg
frac_0 <- (1-effect_allele_freq)^2
frac_1 <- (1-effect_allele_freq)*(effect_allele_freq)*2
frac_2 <- (effect_allele_freq)^2
```

$$p^2 + 2pq + q^2 = 1$$

```
#per-SNP slow simulation example with same outcome
dosages<-vector()
for(i in 1:1000){
  genotype<-sample(
    x=c(1, 0),
    size=2,
    replace=T,
    prob=c(effect_allele_freq,1-effect_allele_freq))
    dosages<-c(dosages,sum(genotype))
}
sd_per_snp <- sd(dosages * effect_size)

#all-SNP standard-deviation of score
overall_sd <- sqrt(sum(sd_per_snp^2))
```

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N}} \text{"}ish\text{"}$$

”N” is fractions instead

Implementation: PRS-functions

Counting effect-alleles, *beta-weight, normalized, unit variance

(Purpose: Z-scores -> 68% is within score of “1”)

SD = 0.777

	Effect allele	Effect Size	Person 1 genotype	Effect allele count	Weighted score	Average population score	Normalized score	Normalized score (unit variance)
SNP1	G	0.97	A/G	1	0.97	0.45	0.53	
SNP2	C	0.76	C/C	2	1.5	0.12	1.4	
SNP3	A	0.51	G/G	0	0	0.12	-0.12	
SNP4	C	0.59	C/C	2	1.2	0.21	0.97	
SNP5	G	1.1	C/C	0	0	0.066	-0.066	
SNP6	G	-0.31	G/T	1	-0.31	-0.27	-0.037	
SNP7	C	-0.56	C/C	2	-1.1	-0.089	-1	
							1.67	**2.17**