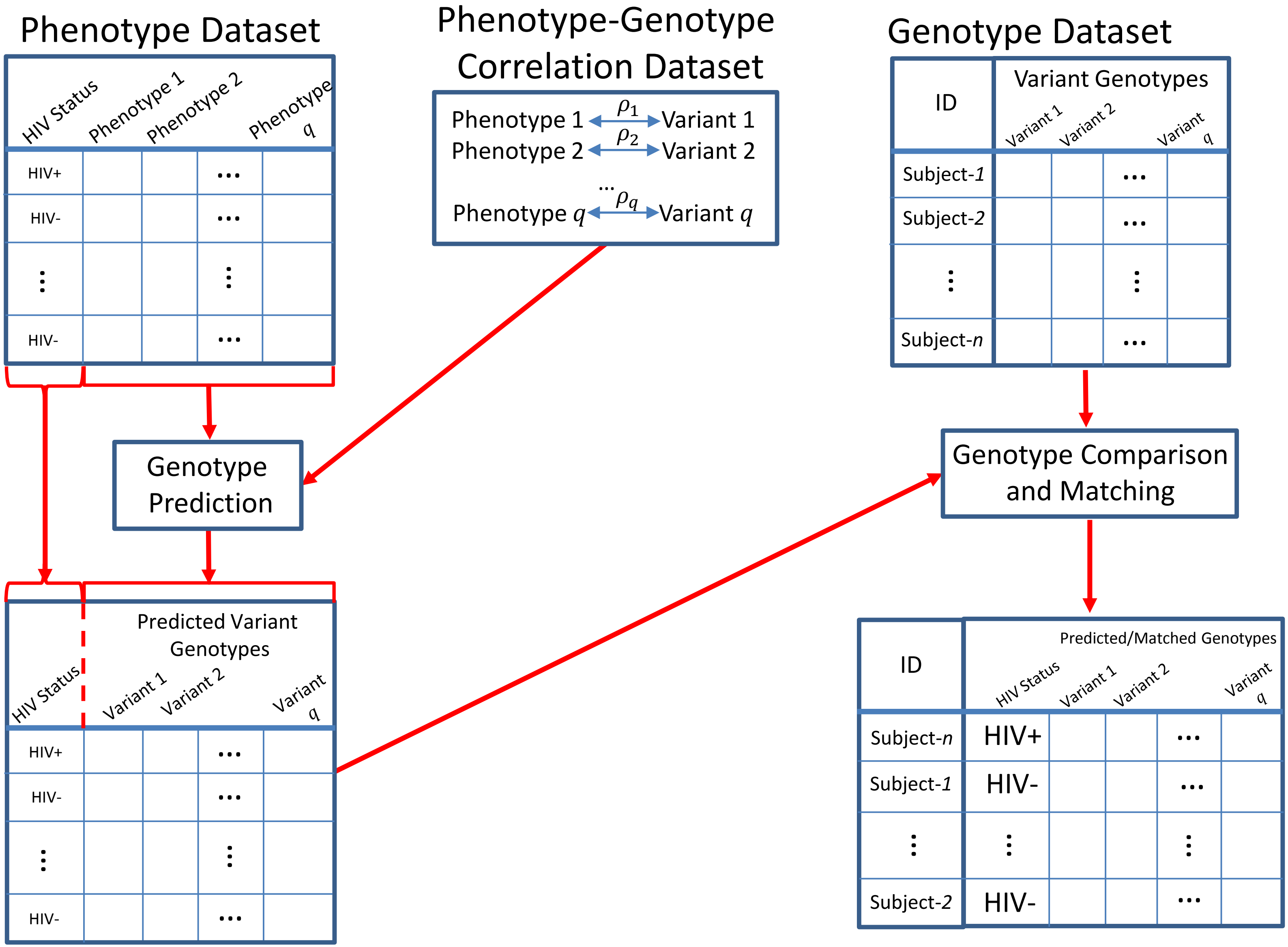
PrivaSeq Figure Pack

May 12, 2015

Fig 1a: Linking Attack Scenario



Phenotype Dataset

HIV Status	Phenotype 1	Phenotype 2		Phenotype *q*
HIV+			…	
HIV-			…	
⋮			⋮	
HIV-			…	

Phenotype-Genotype Correlation Dataset

Phenotype 1 $\xleftrightarrow{\rho_1}$ Variant 1
Phenotype 2 $\xleftrightarrow{\rho_2}$ Variant 2
…
Phenotype q $\xleftrightarrow{\rho_q}$ Variant q

Genotype Dataset

ID	Variant Genotypes			
	Variant 1	Variant 2		Variant *q*
Subject-*1*			…	
Subject-*2*			…	
⋮			⋮	
Subject-*n*			…	

Genotype Prediction

Genotype Comparison and Matching

HIV Status	Predicted Variant Genotypes			
	Variant 1	Variant 2		Variant *q*
HIV+			…	
HIV-			…	
⋮			⋮	
HIV-			…	

ID	Predicted/Matched Genotypes				
	HIV Status	Variant 1	Variant 2		Variant *q*
Subject-*n*	HIV+			…	
Subject-*1*	HIV-			…	
⋮	⋮			⋮	
Subject-*2*	HIV-			…	

Fig 1b: Datasets

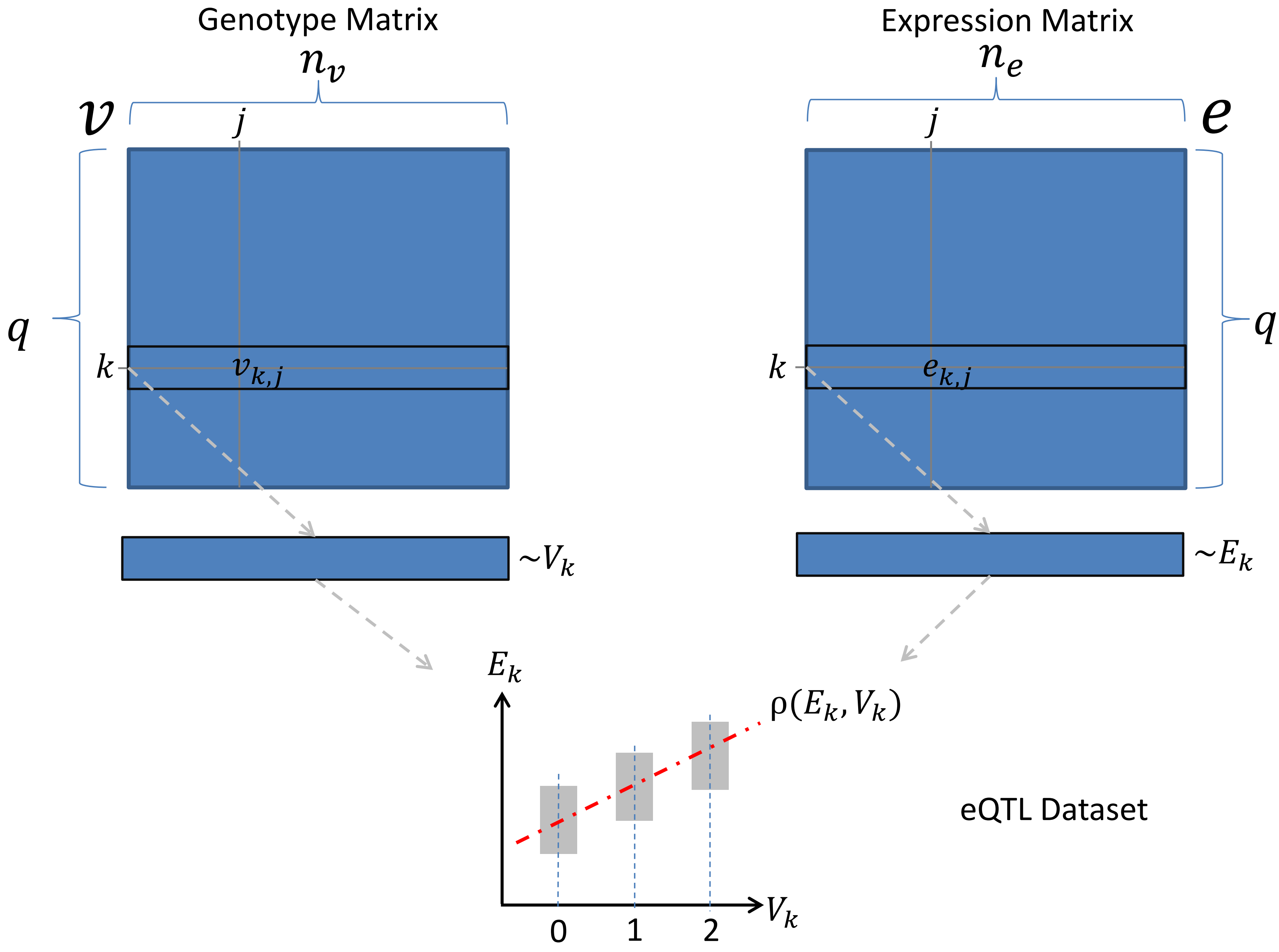
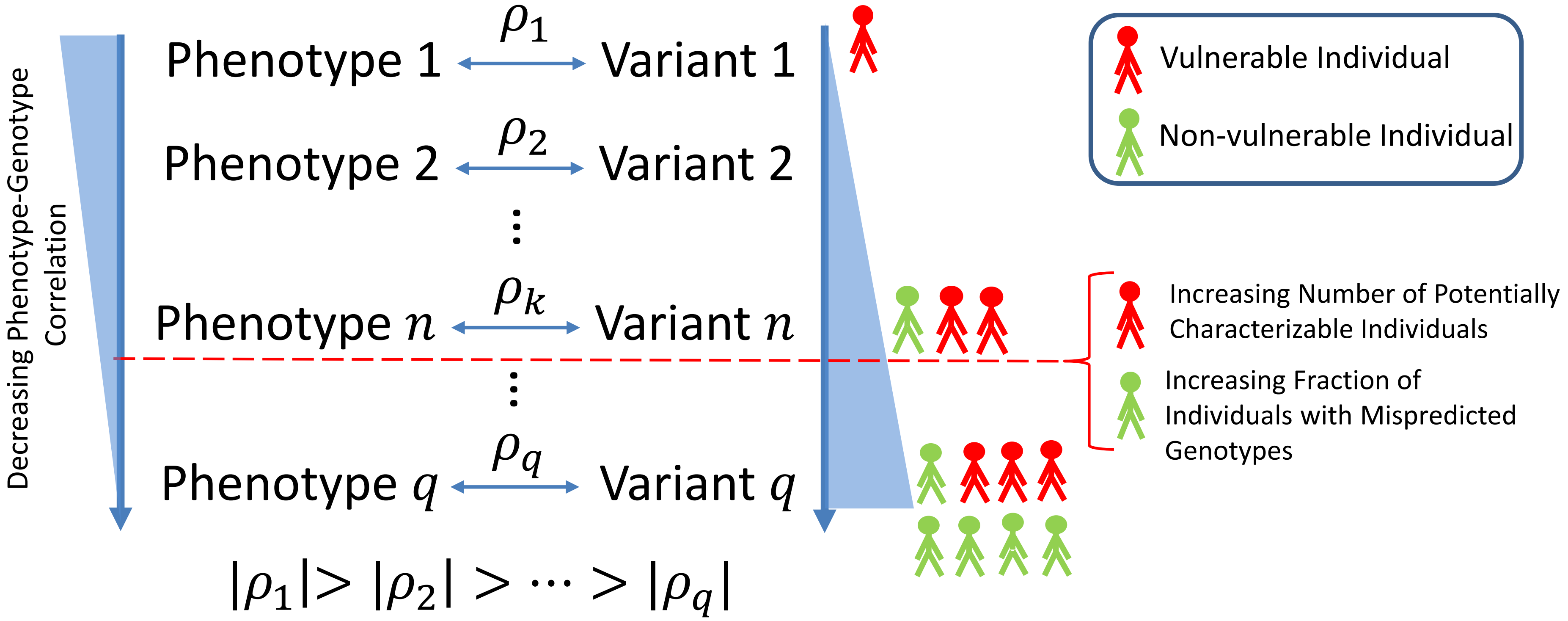


Fig 2a



Decreasing Phenotype-Genotype Correlation

Phenotype 1 $\xleftrightarrow{\rho_1}$ Variant 1

Phenotype 2 $\xleftrightarrow{\rho_2}$ Variant 2

$\vdots$

Phenotype n $\xleftrightarrow{\rho_k}$ Variant n

$\vdots$

Phenotype q $\xleftrightarrow{\rho_q}$ Variant q

$$|\rho_1| > |\rho_2| > \cdots > |\rho_q|$$

Vulnerable Individual

Non-vulnerable Individual

Increasing Number of Potentially Characterizable Individuals

Increasing Fraction of Individuals with Mispredicted Genotypes

Fig 2b

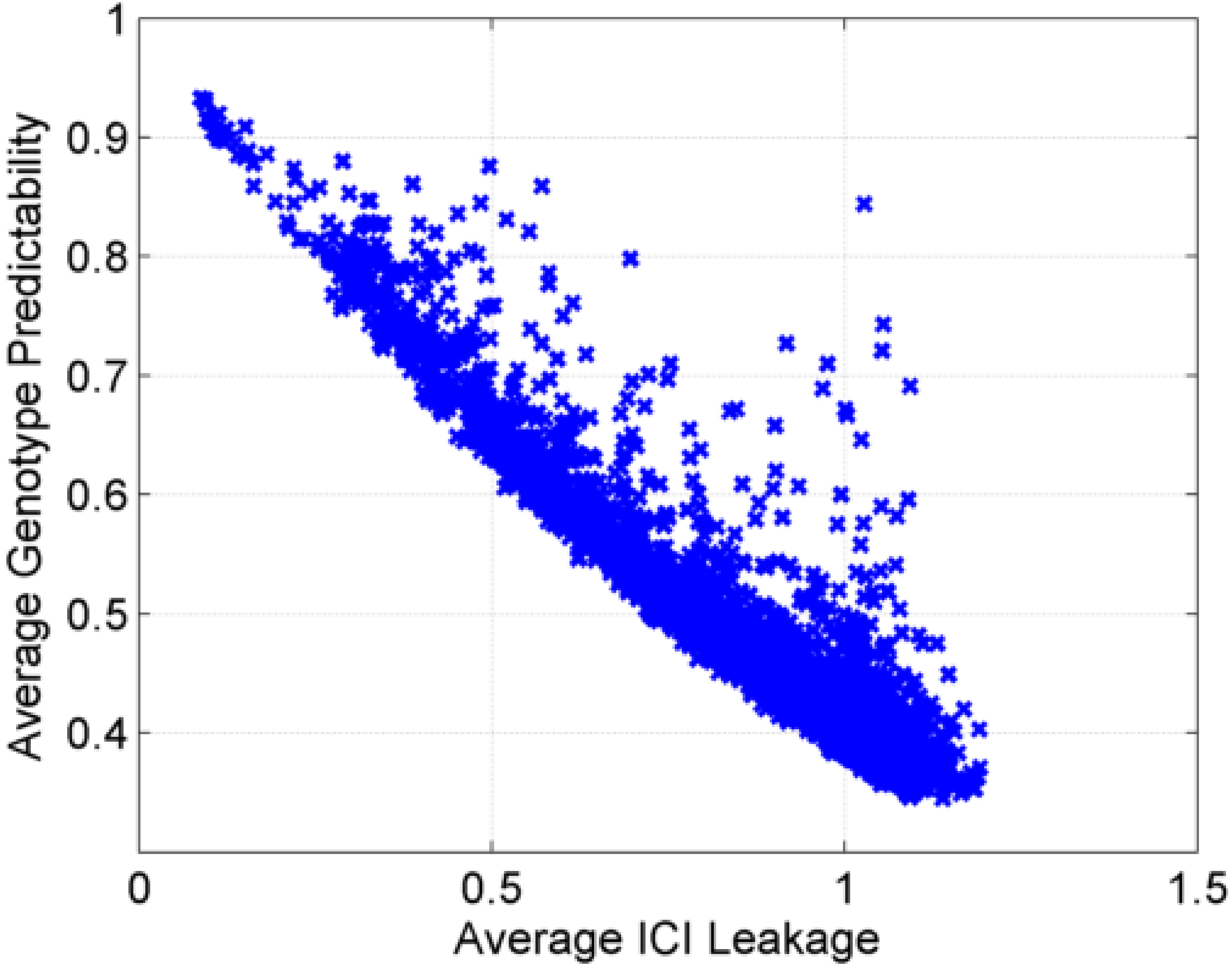

Average Genotype Predictability
1
0.9
0.8
0.7
0.6
0.5
0.4
0
0.5
1
1.5
Average ICI Leakage

Fig 2cde: Predictability vs Information leakage: Multiple eQTLs

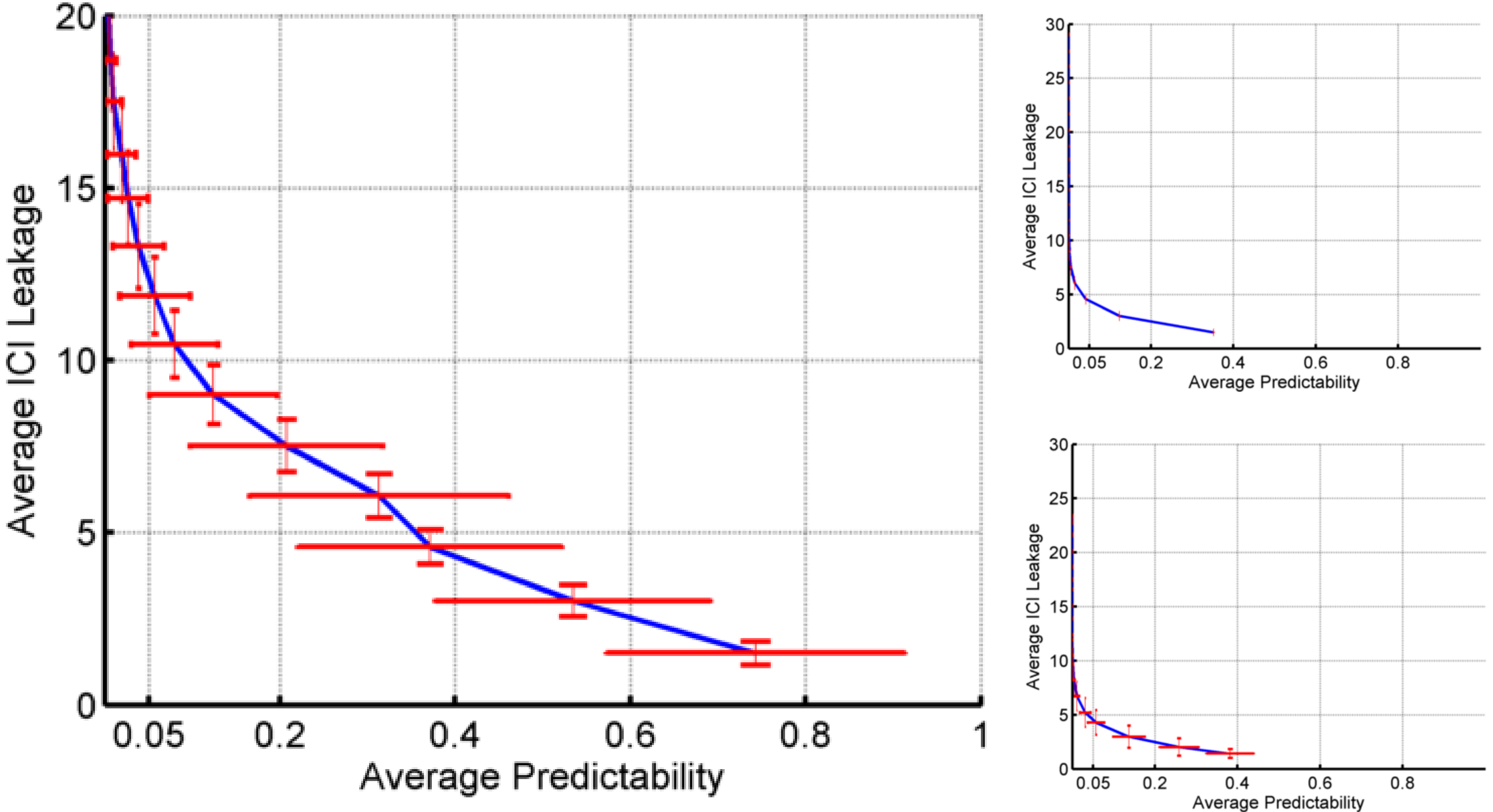
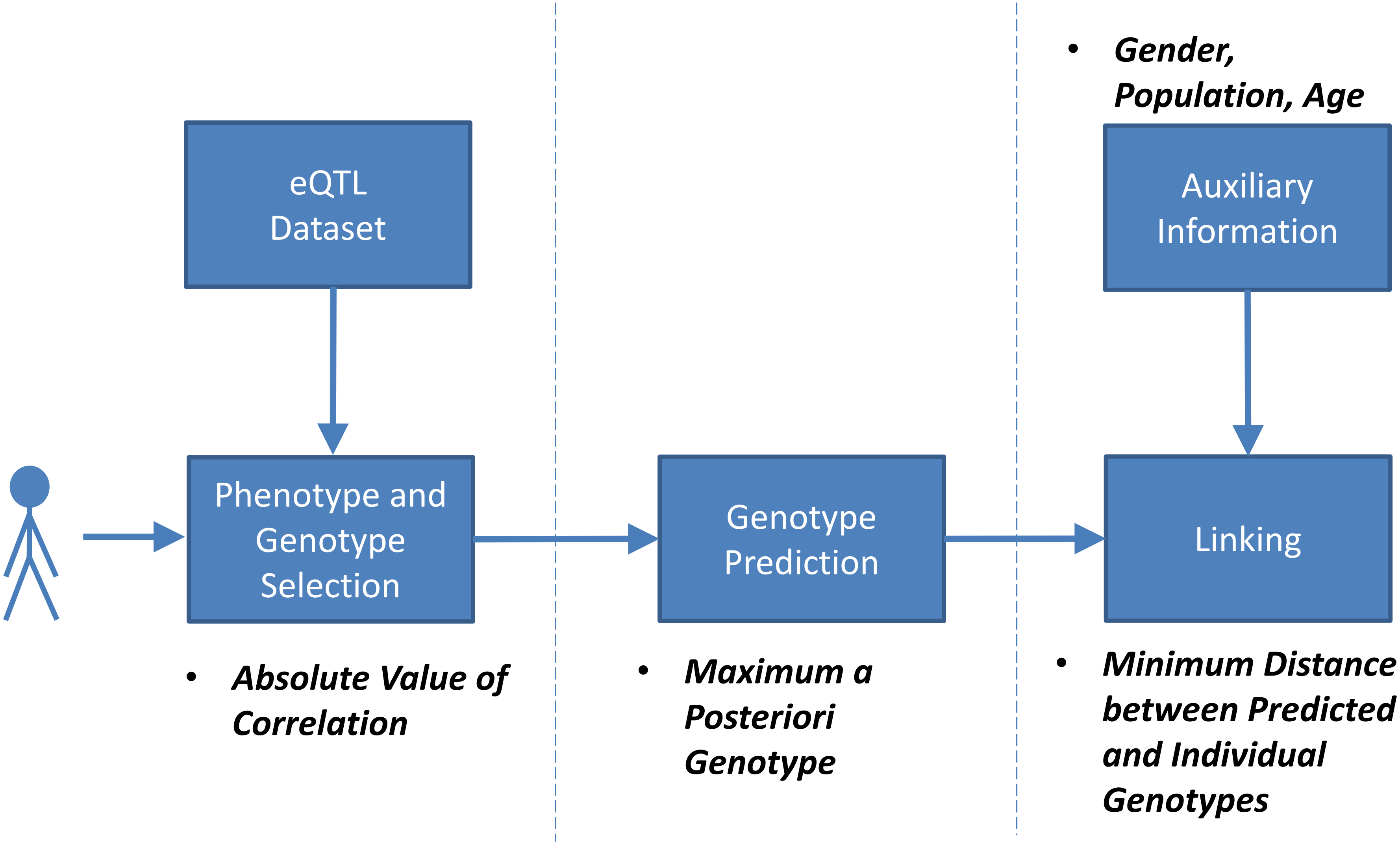


Fig 3: 3 Steps of Linking Attack



eQTL Dataset

Phenotype and Genotype Selection

- ***Absolute Value of Correlation***

Genotype Prediction

- ***Maximum a Posteriori Genotype***

- ***Gender, Population, Age***

Auxiliary Information

Linking

- ***Minimum Distance between Predicted and Individual Genotypes***

Figure 4a: MAP Genotype Prediction Accuracy

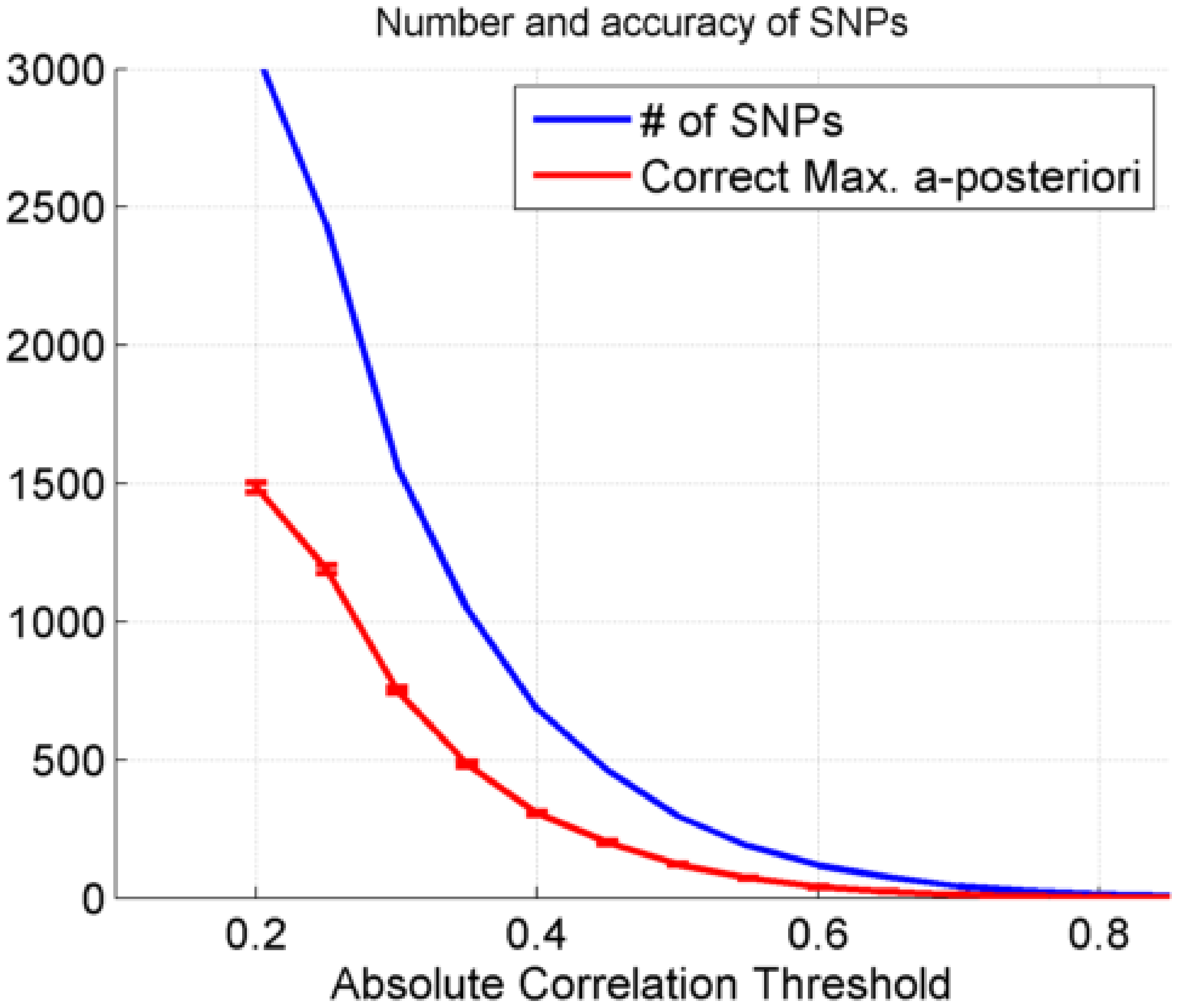


Fig 4b: Fraction of Vulnerable Individuals

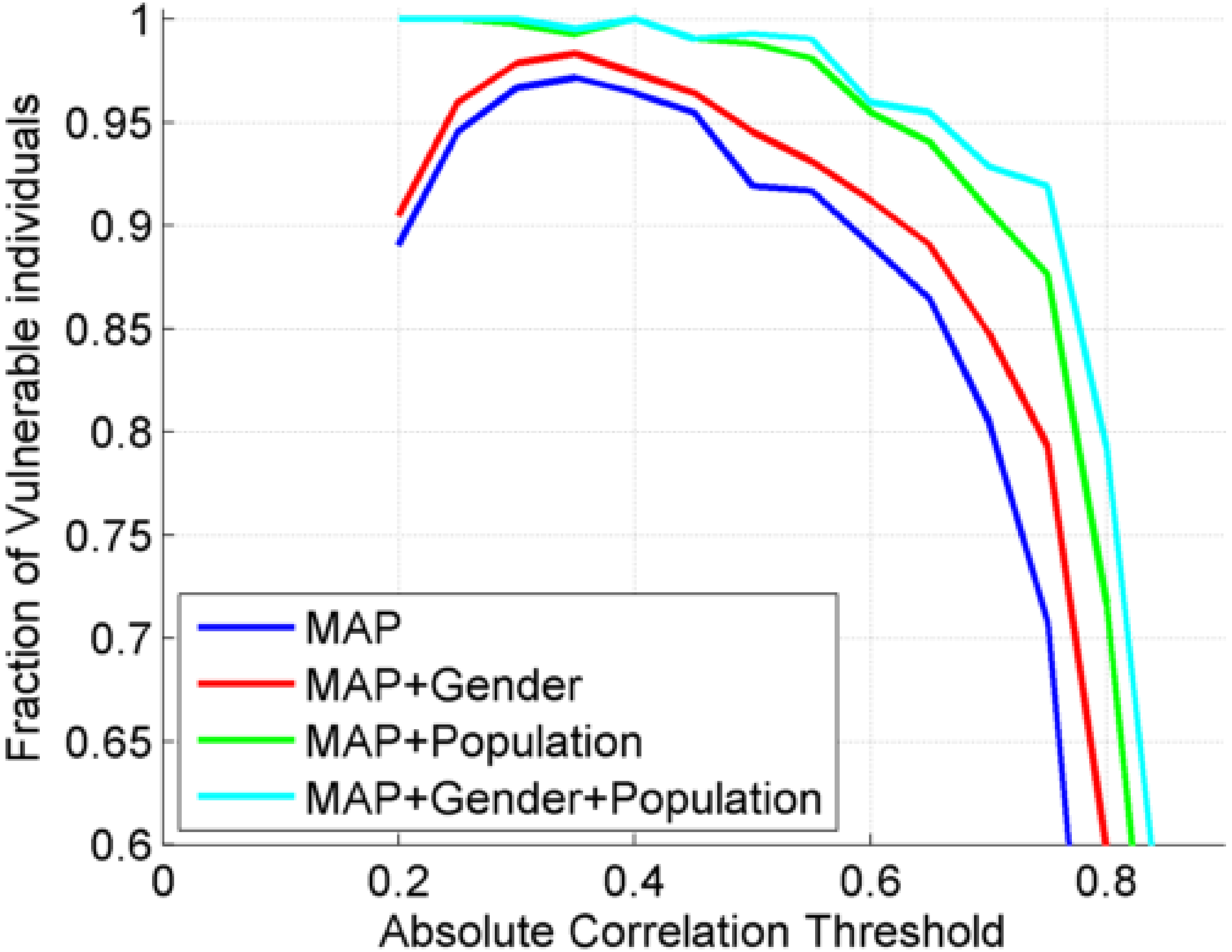
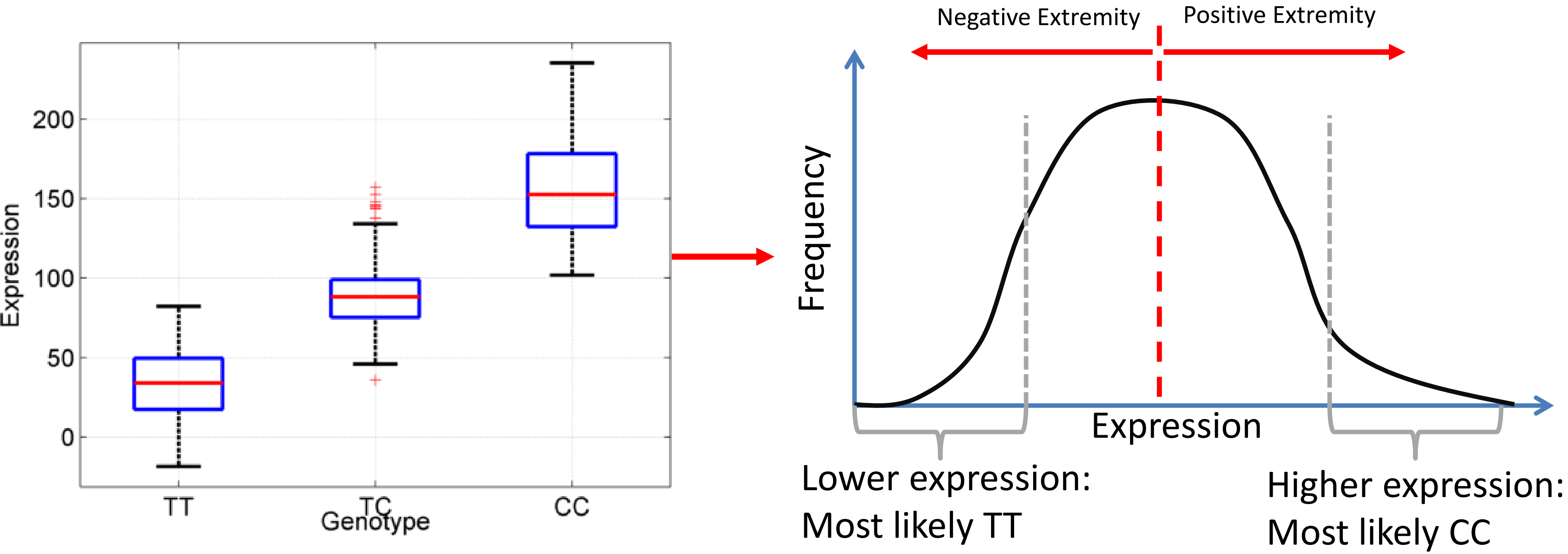


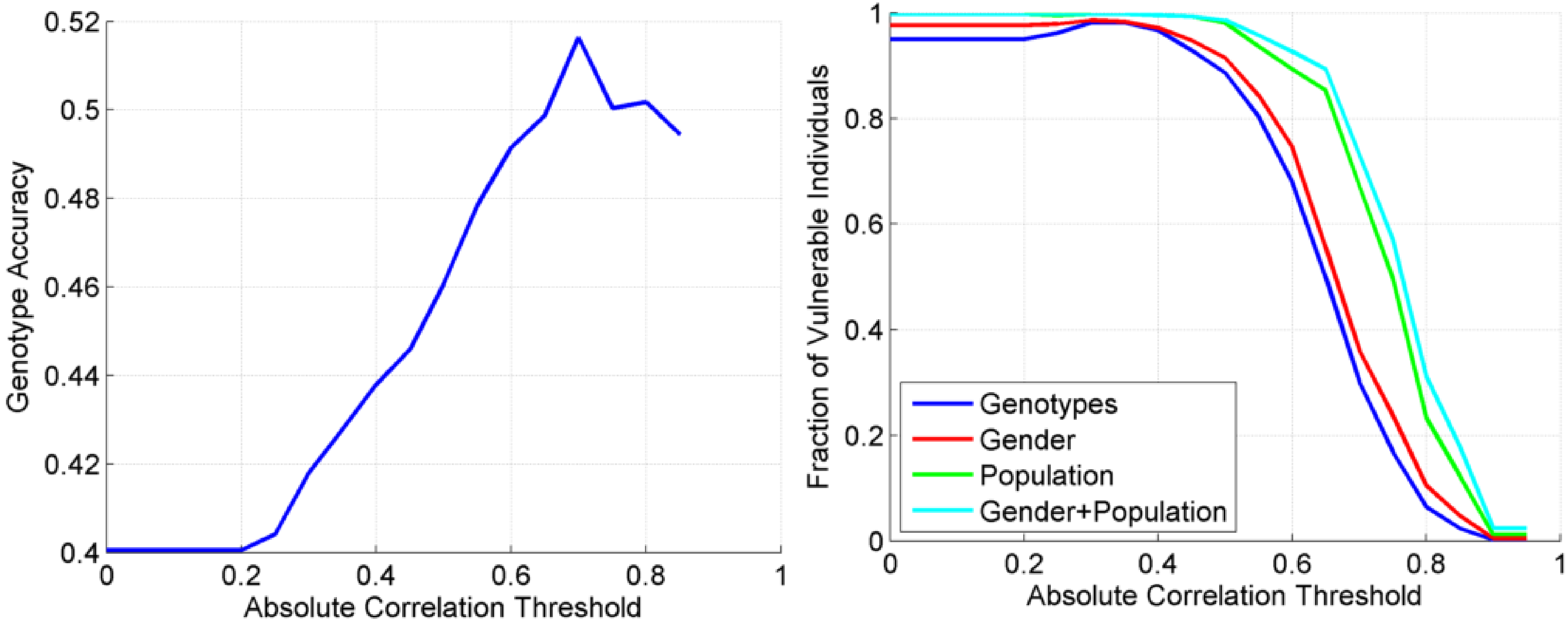
Fig 5a: Extremity based genotype prediction



Lower expression:
Most likely TT

Higher expression:
Most likely CC

Fig 5bc


0.52
0.5
0.48
0.46
0.44
0.42
0.4
Genotype Accuracy
0
0.2
0.4
0.6
0.8
1
Absolute Correlation Threshold
1
0.8
0.6
0.4
0.2
0
Fraction of Vulnerable Individuals
Genotypes
Gender
Population
Gender+Population
0
0.2
0.4
0.6
0.8
1
Absolute Correlation Threshold

Supplementary Figures

Fig S1: eQTL properties

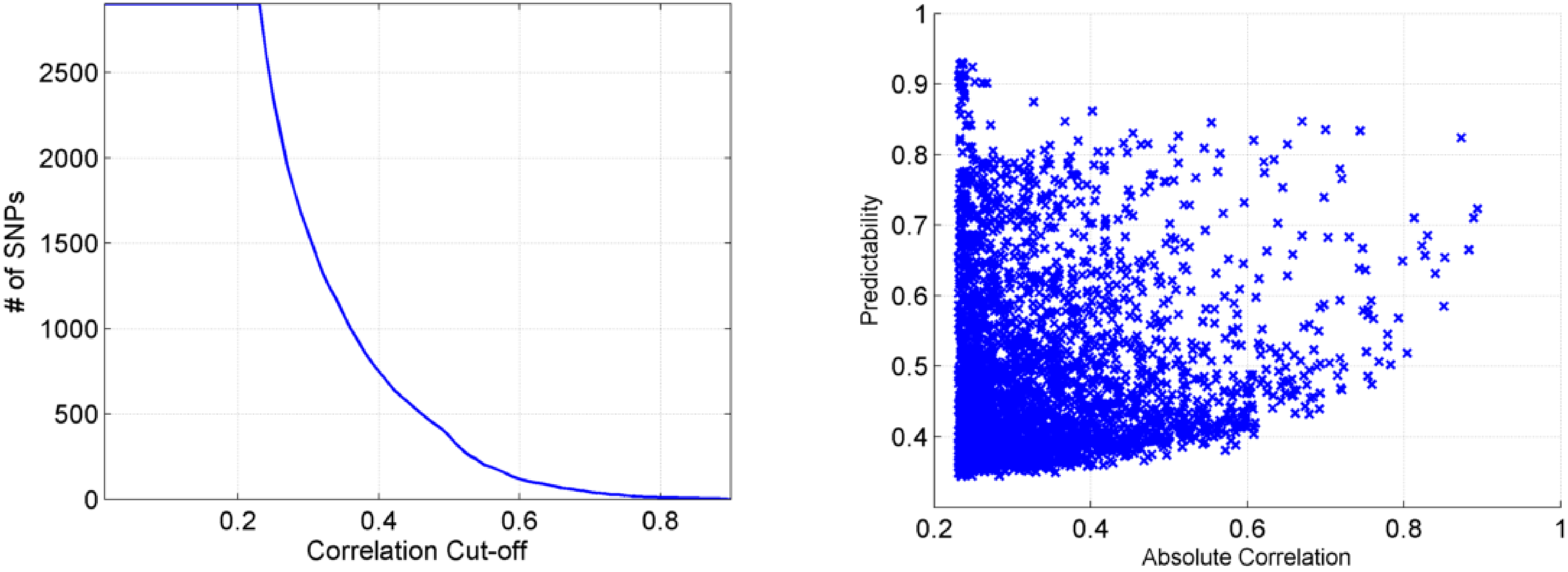
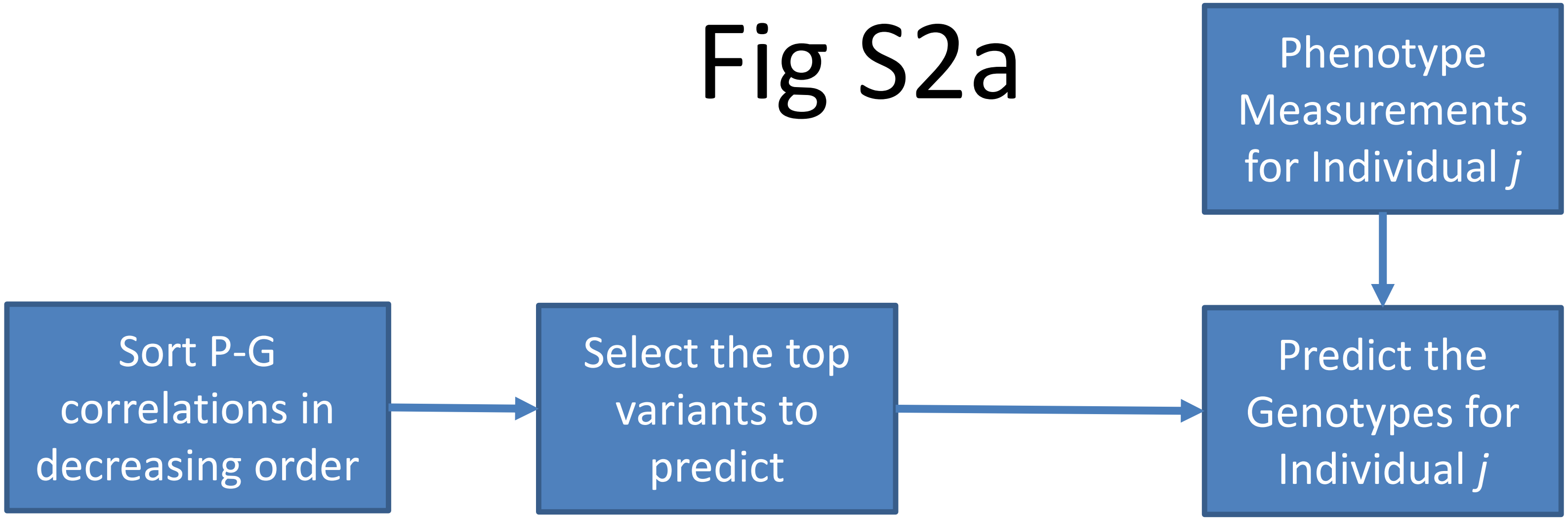


Fig S2a

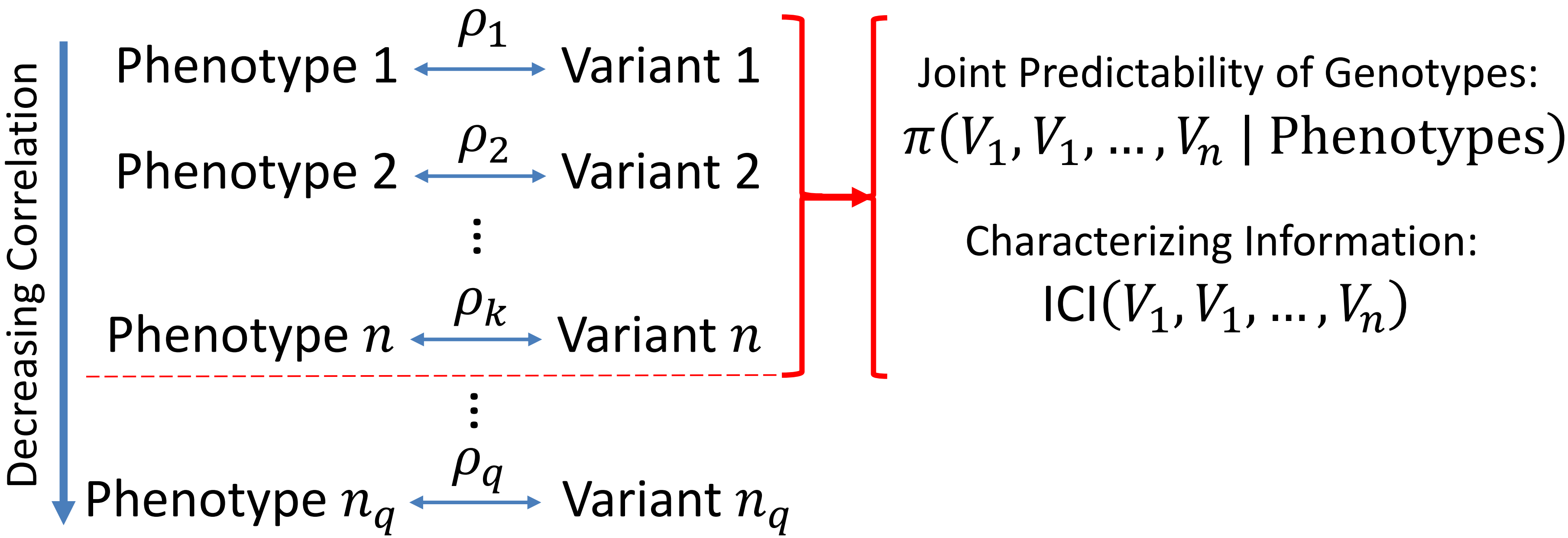


Phenotype Measurements for Individual *j*

Sort P-G correlations in decreasing order

Select the top variants to predict

Predict the Genotypes for Individual *j*



Decreasing Correlation

Phenotype 1 $\xleftrightarrow{\rho_1}$ Variant 1

Phenotype 2 $\xleftrightarrow{\rho_2}$ Variant 2

⋮

Phenotype n $\xleftrightarrow{\rho_k}$ Variant n

⋮

Phenotype n_q $\xleftrightarrow{\rho_q}$ Variant n_q

Joint Predictability of Genotypes:

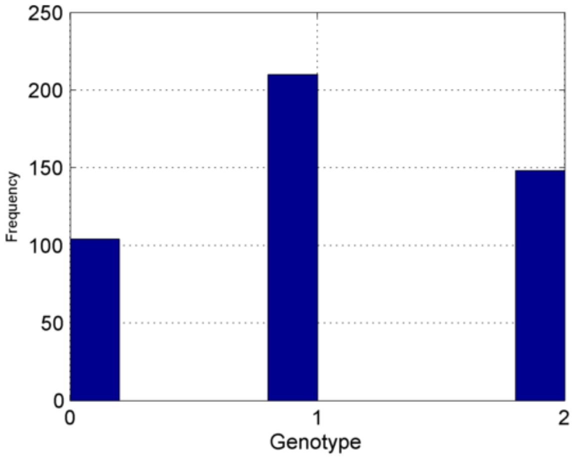
$\pi(V_1, V_1, \dots, V_n \mid \mathrm{Phenotypes})$

Characterizing Information:

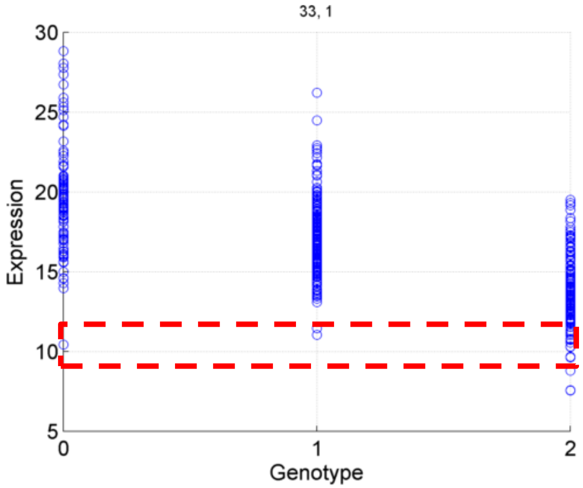
$\mathrm{ICI}(V_1, V_1, \dots, V_n)$

Fig S2b: Illustration of *a priori*, *a posteriori* distributions

Prior: $p(V_k)$

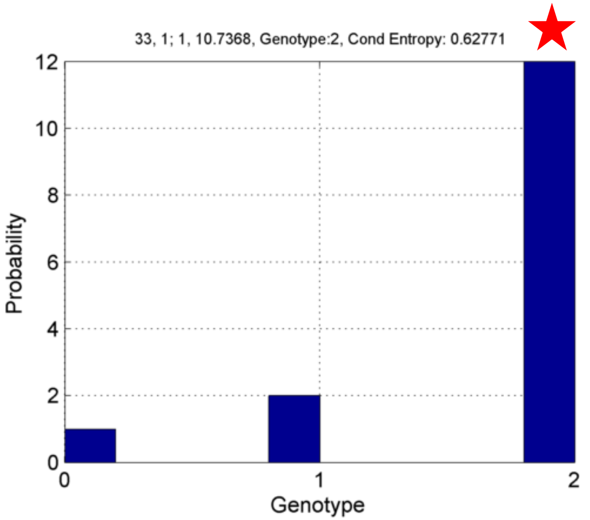


Joint: $p(V_k, E_k)$



Posterior:

$p(V_k \mid E_k = 10)$



$v'_{k,j} = 2$

Fig S3

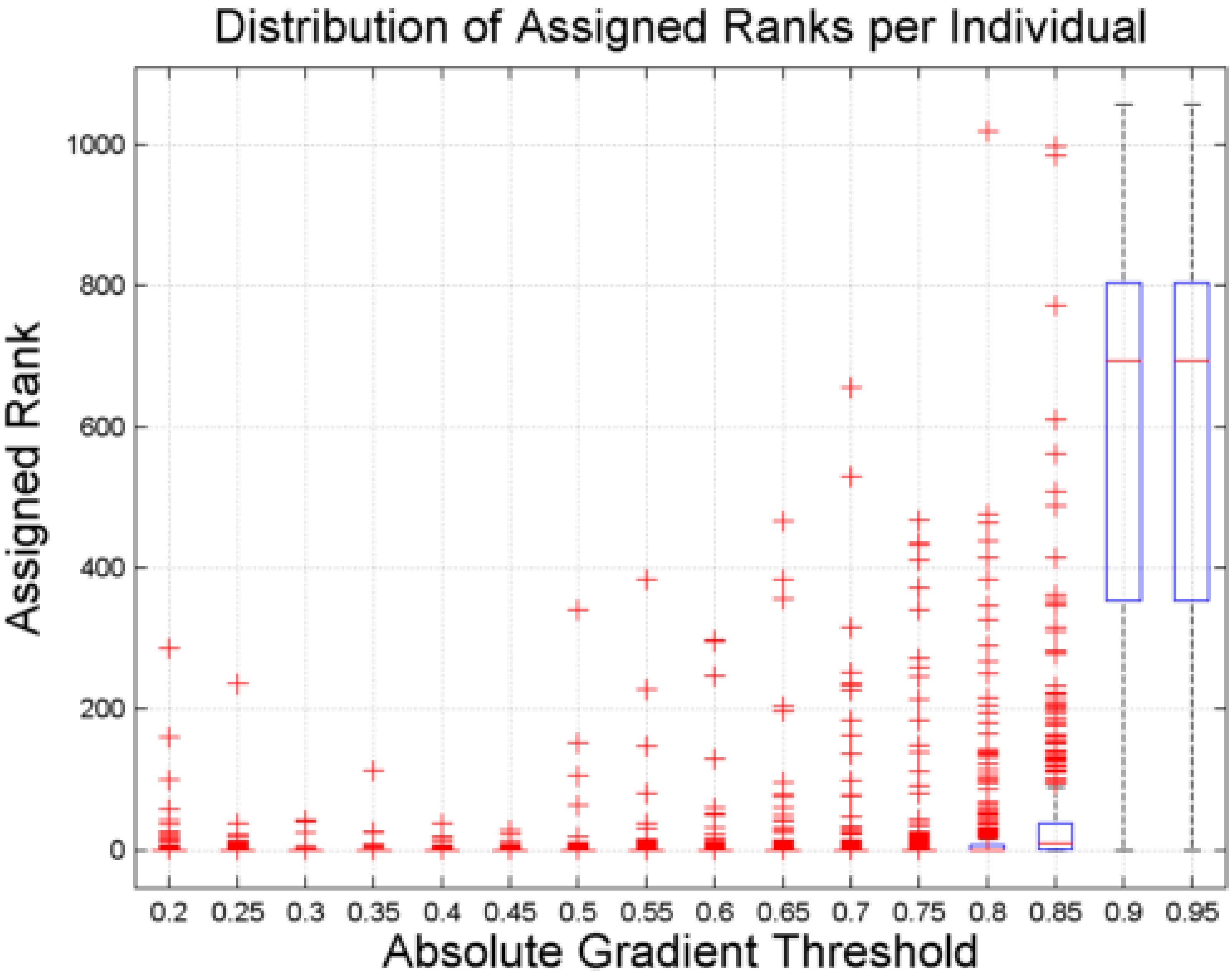
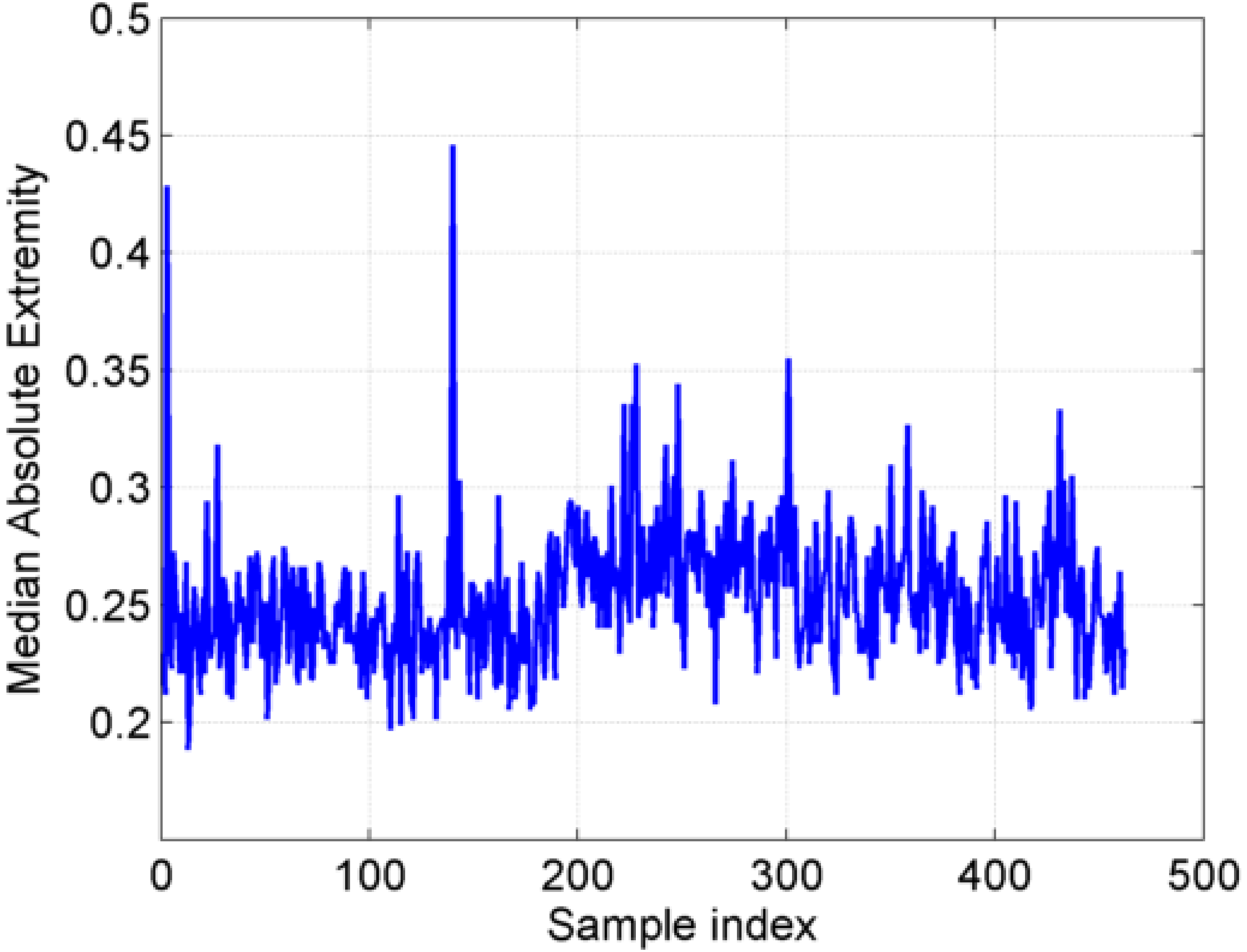

Distribution of Assigned Ranks per Individual
Assigned Rank
1000
800
600
400
200
0
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
Absolute Gradient Threshold

Fig S4


0.5
0.45
0.4
0.35
0.3
0.25
0.2
Median Absolute Extremity
0
100
200
300
400
500
Sample index