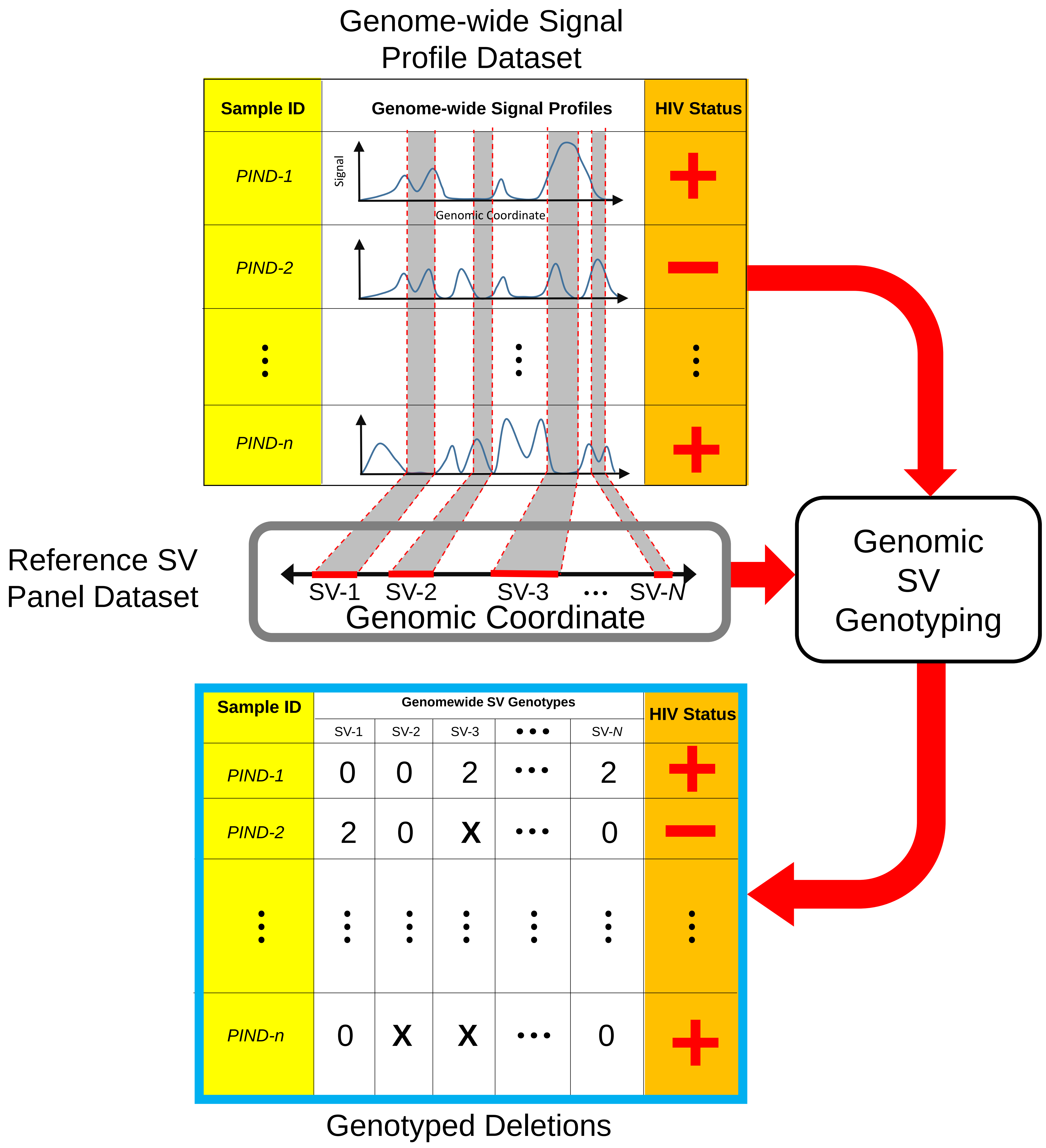
PrivaSeq2 Figures

--Attacker Strikes Back--

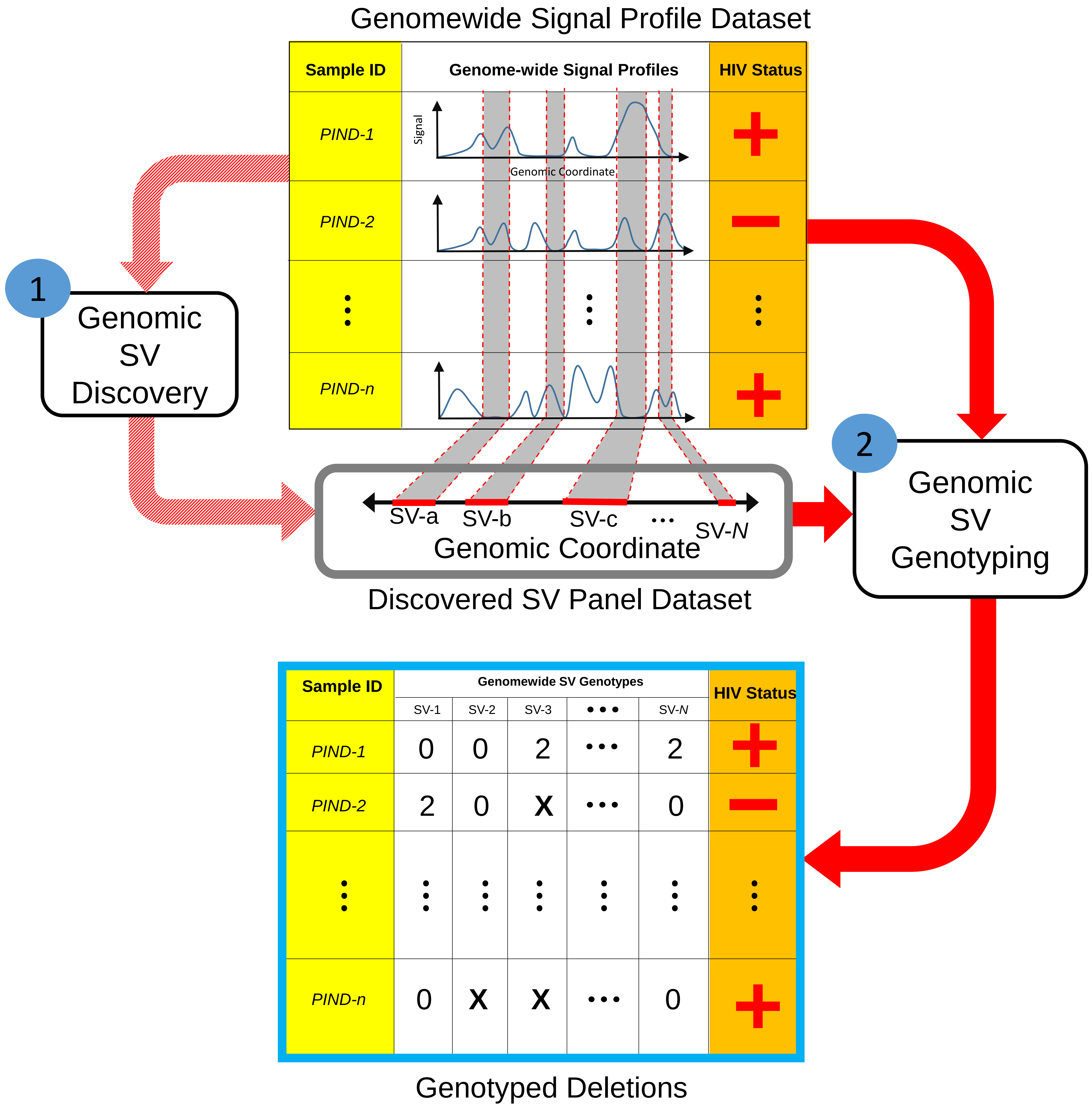
Fig 1a



Sample ID	Genome-wide Signal Profiles	HIV Status
PIND-1		+
PIND-2		−
⋮	⋮	⋮
PIND-n		+

Sample ID	Genomewide SV Genotypes					HIV Status
	SV-1	SV-2	SV-3	• • •	SV-*N*	
PIND-1	0	0	2	• • •	2	+
PIND-2	2	0	X	• • •	0	−
⋮	⋮	⋮	⋮	⋮	⋮	⋮
PIND-n	0	X	X	• • •	0	+

Fig 1b



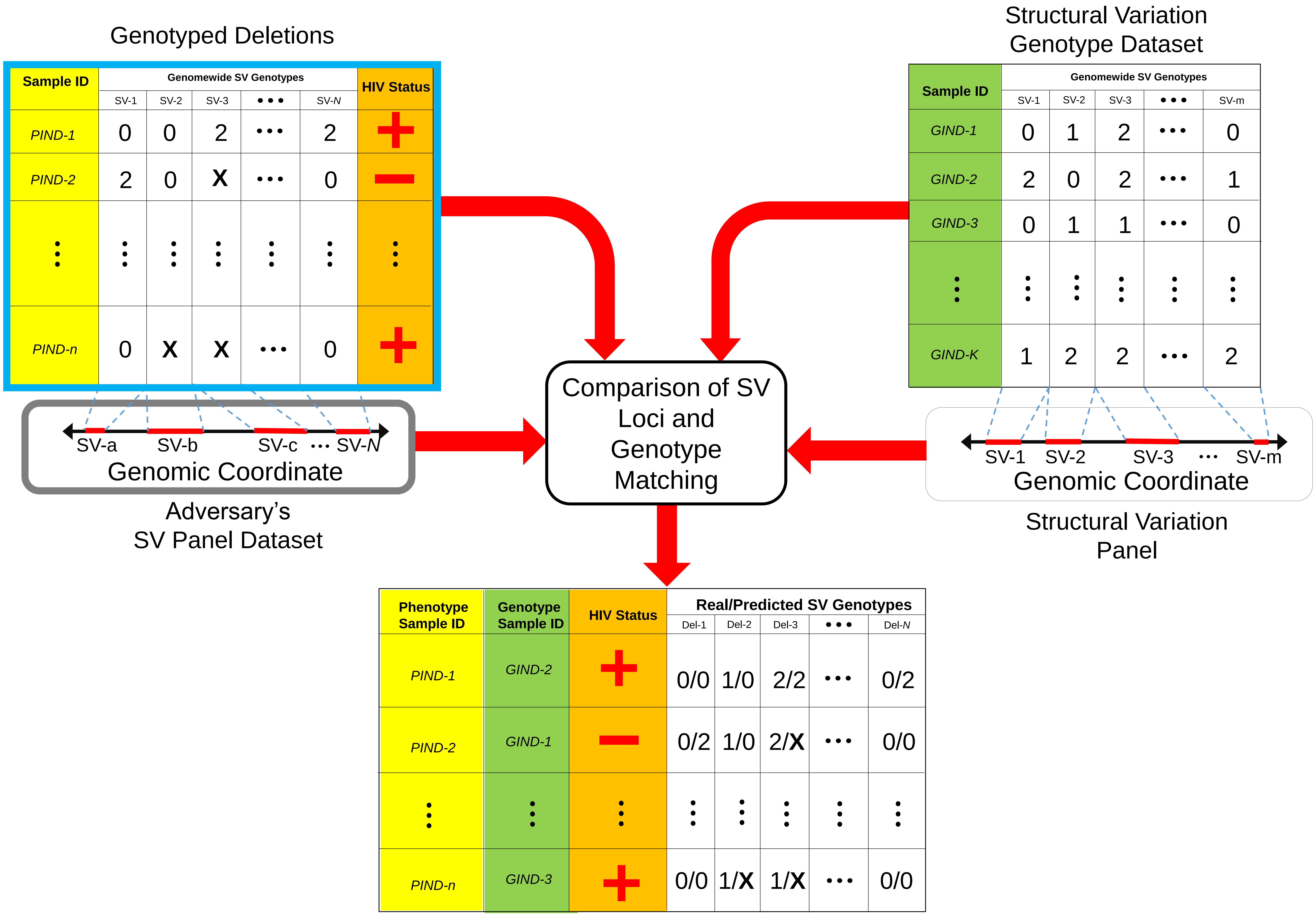
Genomewide Signal Profile Dataset

Sample ID	Genome-wide Signal Profiles	HIV Status
PIND-1		+
PIND-2		−
⋮	⋮	⋮
PIND-n		+

Sample ID	Genomewide SV Genotypes					HIV Status
	SV-1	SV-2	SV-3	• • •	SV-*N*	
PIND-1	0	0	2	• • •	2	+
PIND-2	2	0	X	• • •	0	−
⋮	⋮	⋮	⋮	⋮	⋮	⋮
PIND-n	0	X	X	• • •	0	+

Genotyped Deletions

Fig 1c



Genotyped Deletions

Sample ID	Genomewide SV Genotypes					HIV Status
	SV-1	SV-2	SV-3	• • •	SV-*N*	
PIND-1	0	0	2	• • •	2	+
PIND-2	2	0	**X**	• • •	0	−
⋮	⋮	⋮	⋮	⋮	⋮	⋮
PIND-n	0	**X**	**X**	• • •	0	+

Adversary's SV Panel Dataset

Structural Variation Genotype Dataset

Sample ID	Genomewide SV Genotypes				
	SV-1	SV-2	SV-3	• • •	SV-m
GIND-1	0	1	2	• • •	0
GIND-2	2	0	2	• • •	1
GIND-3	0	1	1	• • •	0
⋮	⋮	⋮	⋮	⋮	⋮
GIND-K	1	2	2	• • •	2

Structural Variation Panel

Phenotype Sample ID	Genotype Sample ID	HIV Status	Real/Predicted SV Genotypes				
			Del-1	Del-2	Del-3	• • •	Del-*N*
PIND-1	*GIND-2*	+	0/0	1/0	2/2	• • •	0/2
PIND-2	*GIND-1*	−	0/2	1/0	2/**X**	• • •	0/0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
PIND-n	*GIND-3*	+	0/0	1/**X**	1/**X**	• • •	0/0

Fig 1 [[ADD RNA-Seq TRACK]]

c) NA12878 large deletion

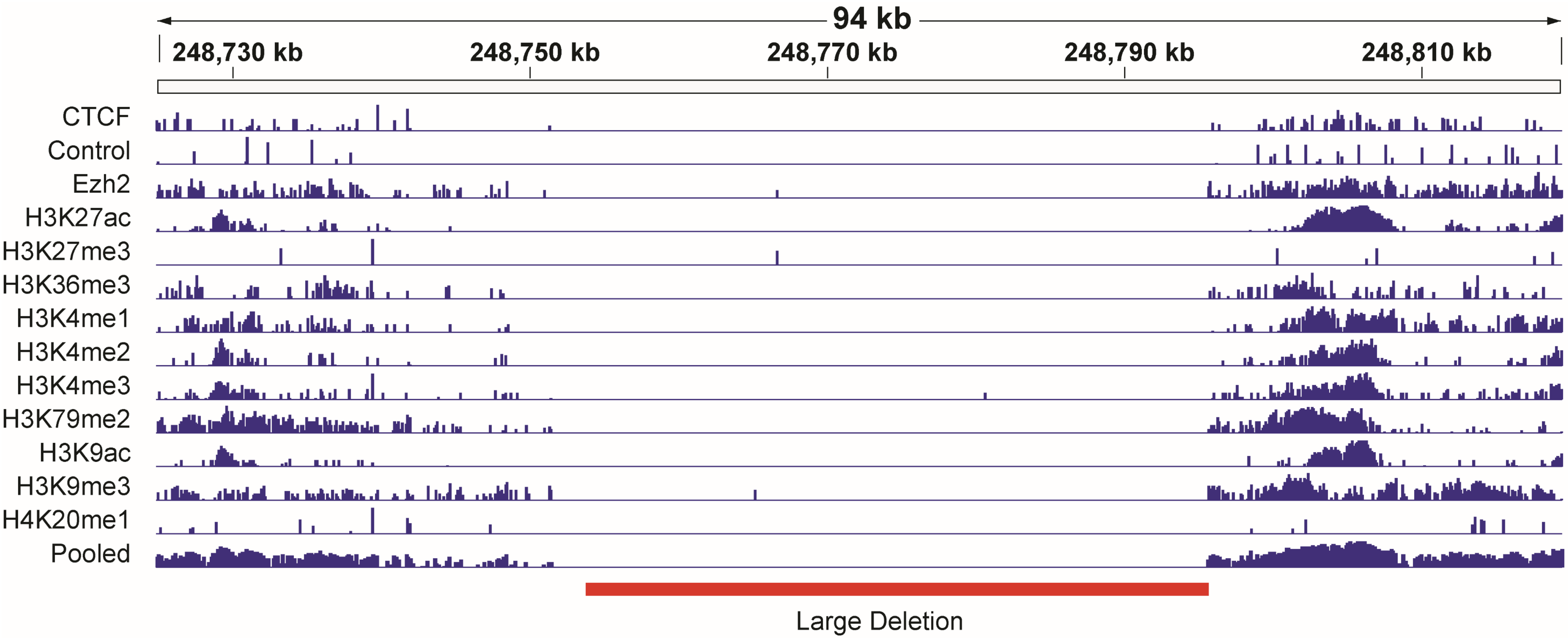


Fig 1

b) Detection and Genotyping of small and large deletions from signal profiles

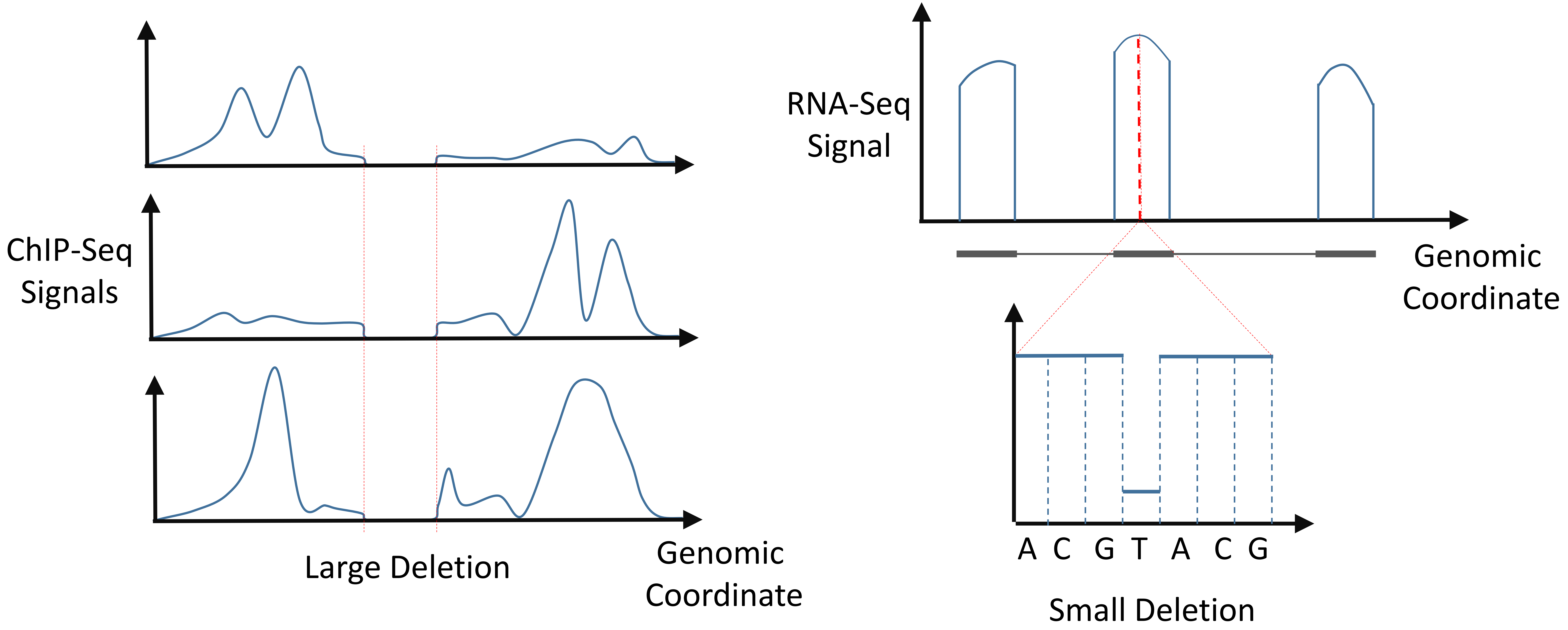
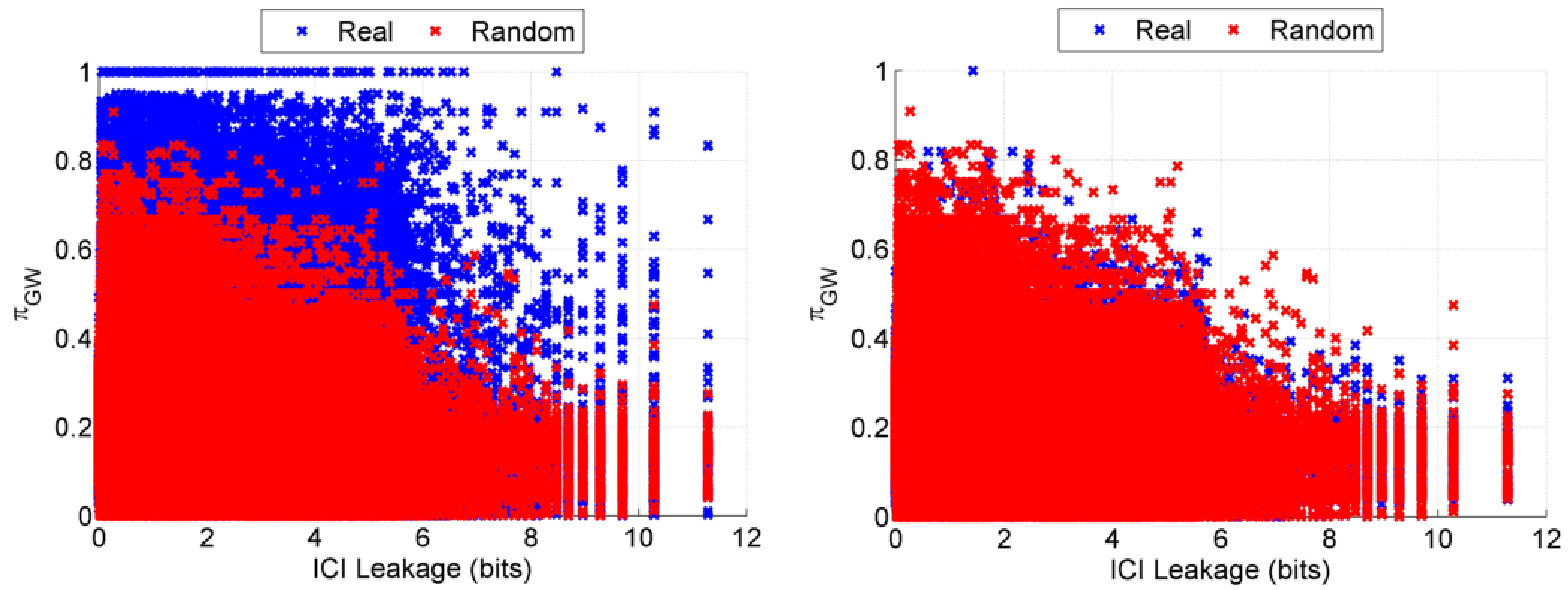


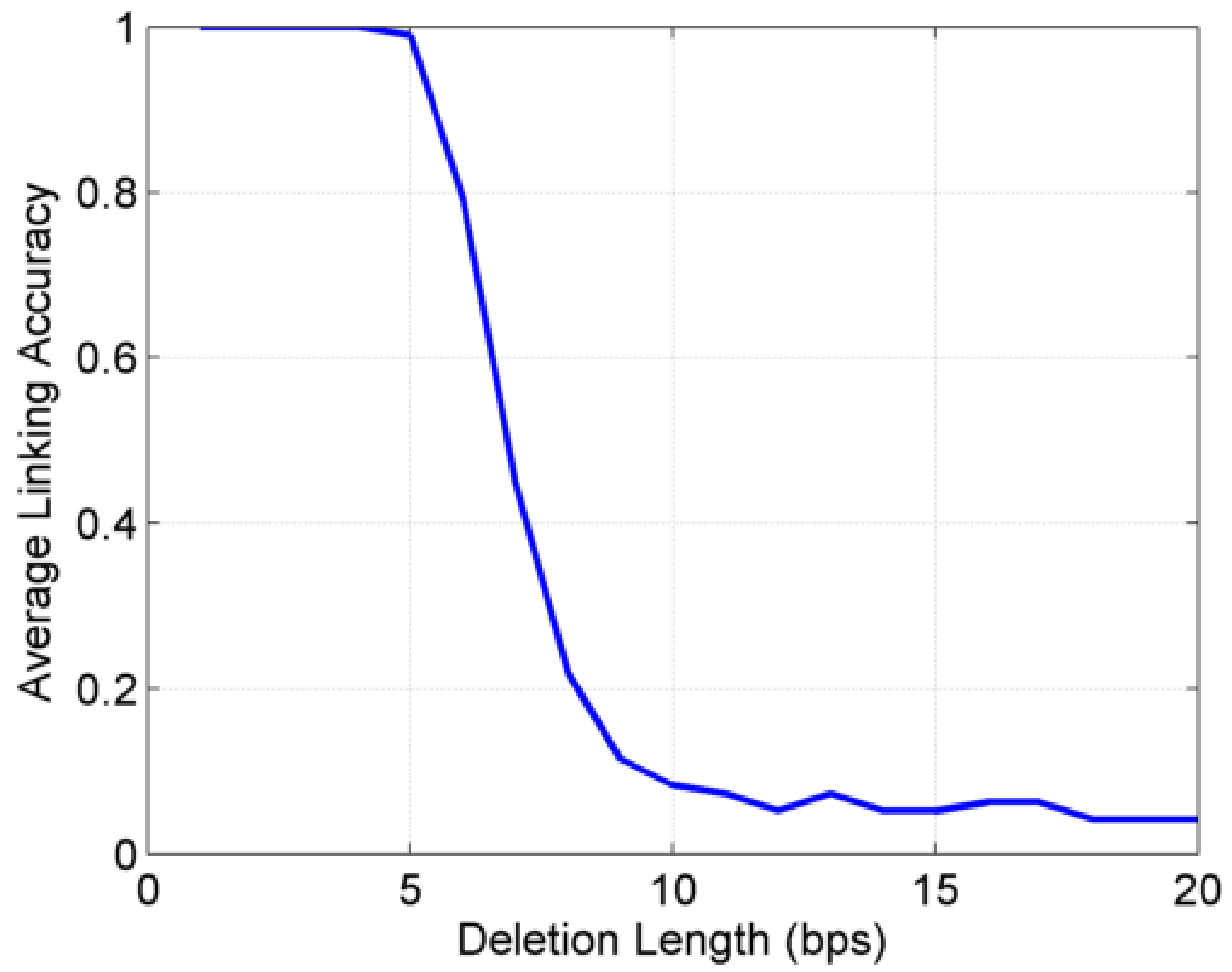
Fig 2

a) GEUVADIS ICI-\Pi_GW estimates:


a) Genotying Only

b) After Anonymization

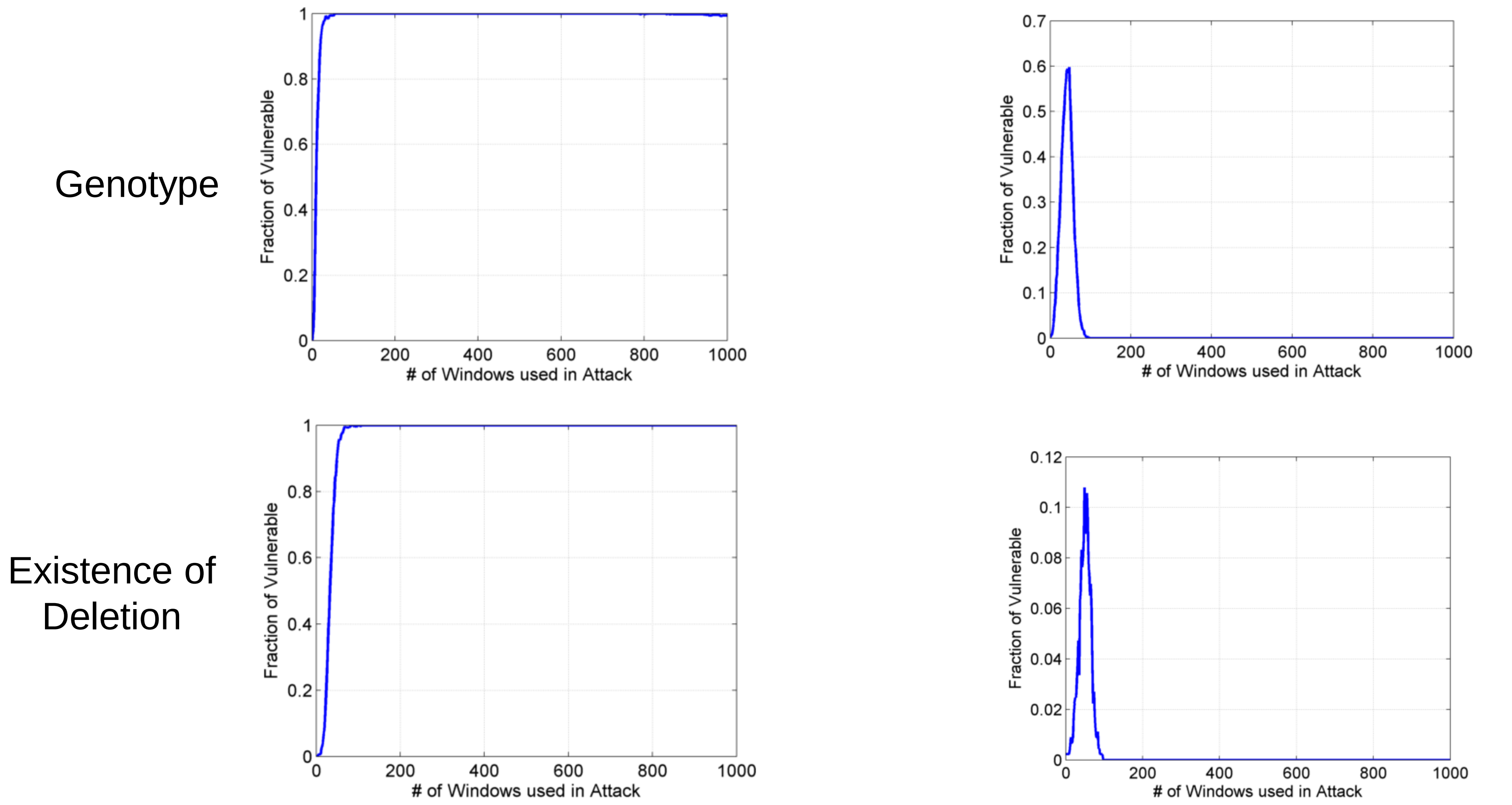
e) Indel length vs vulnerability statistics



c, d) GEUVADIS extremity attack: c) Genotyping only
d) Joint Discovery + Genotyping

d) Genotyping Only (1kG MAF>0.01)

e) Discovery+Genotyping



f,g) Min # of identifying variants

f) Genotyping only (1kG MAF>0.01)

g) Discovery + Genotyping

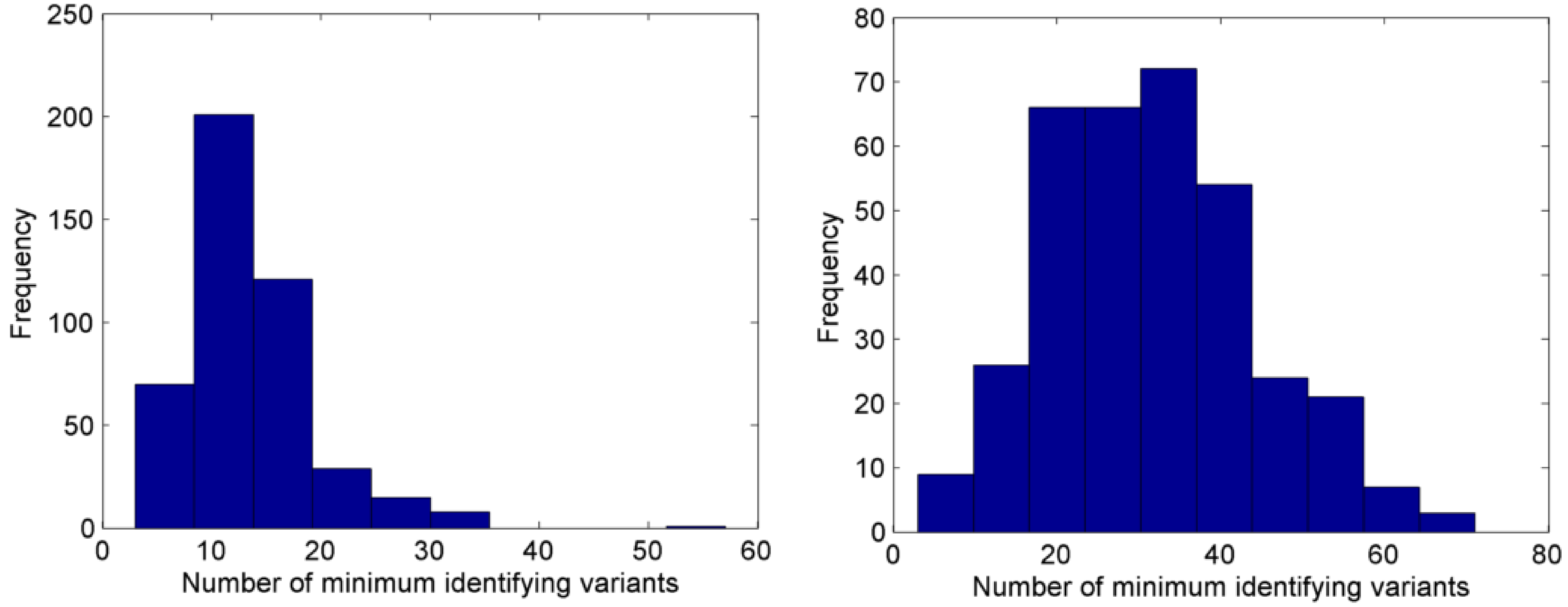
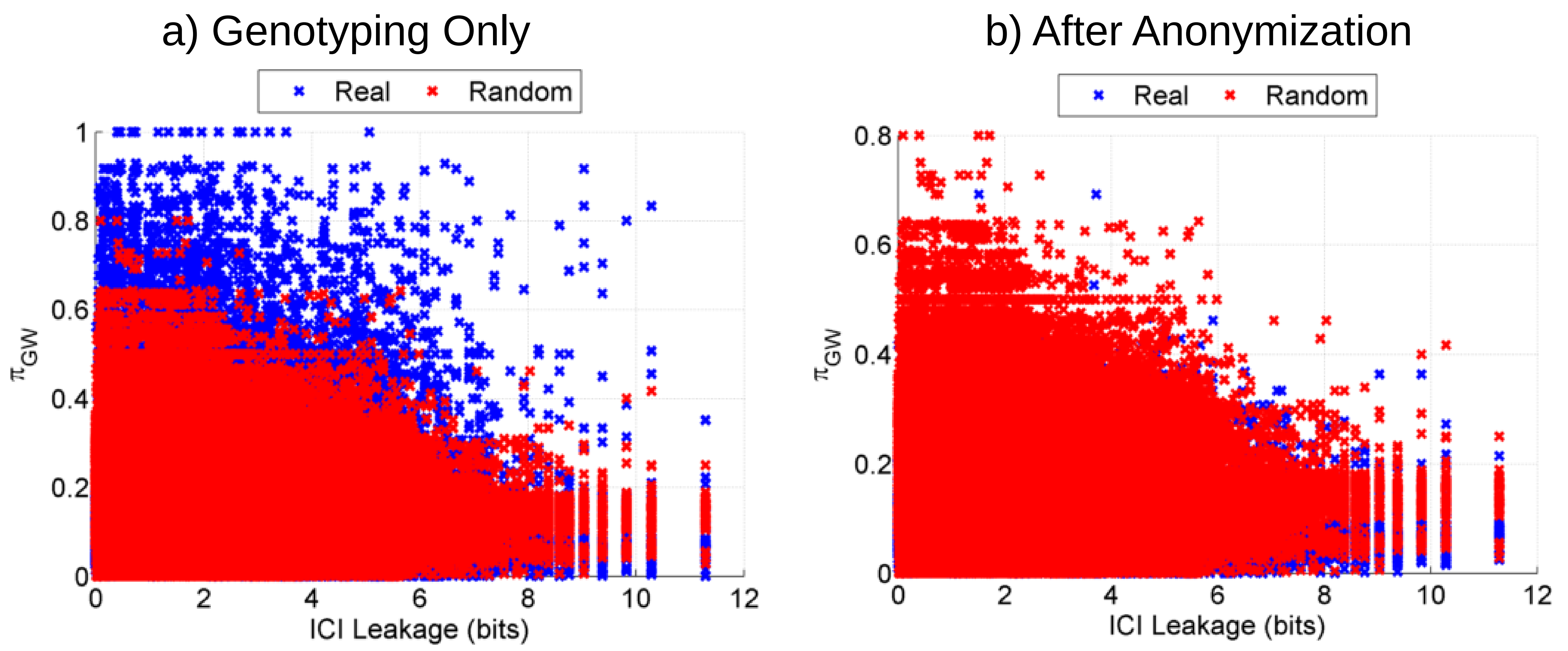


Fig 3

a, b) Gtex ICI-\Pi_GW estimates: a) known indels, b) De-novo indels
(1kG indels, GTex mapped reads)


a) Genotyping Only

b) After Anonymization

Gtex extremity attack: known indels (1kG MAF>0.01)
c) Genotyping d) d_1,2 vs Vulnerability

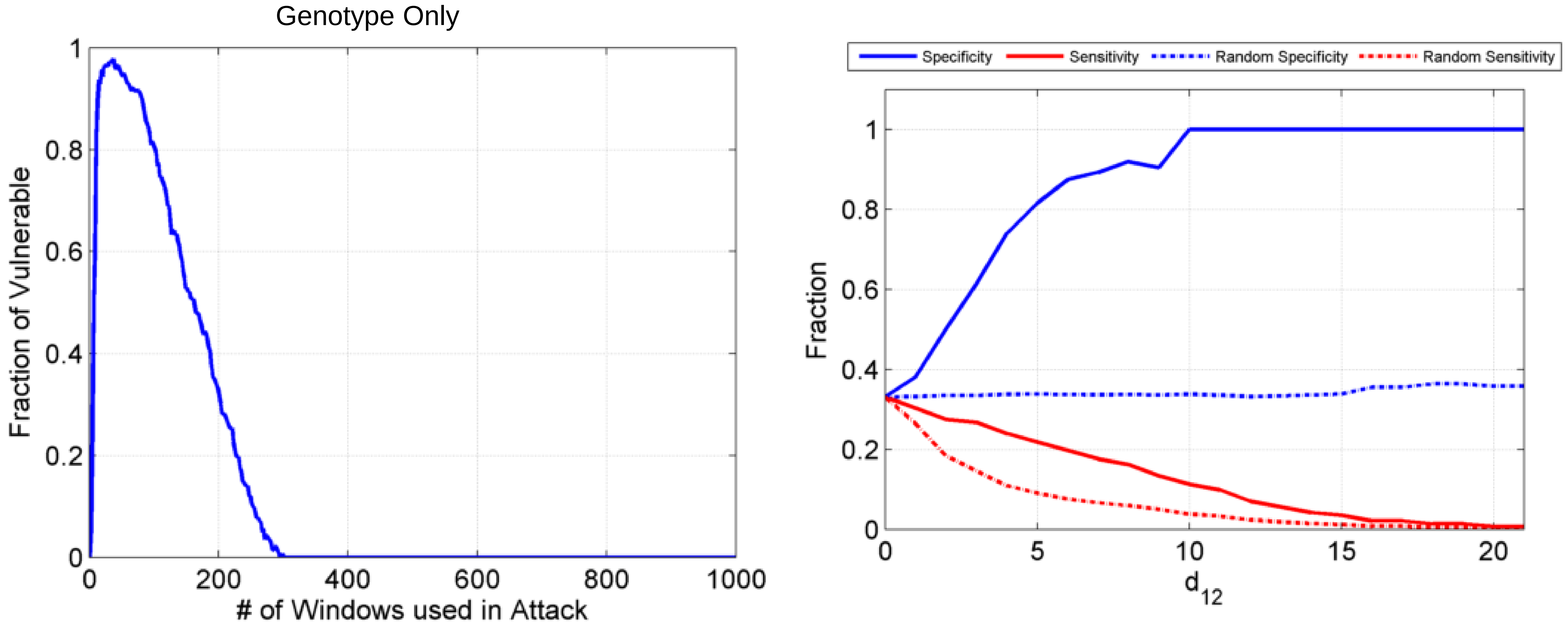
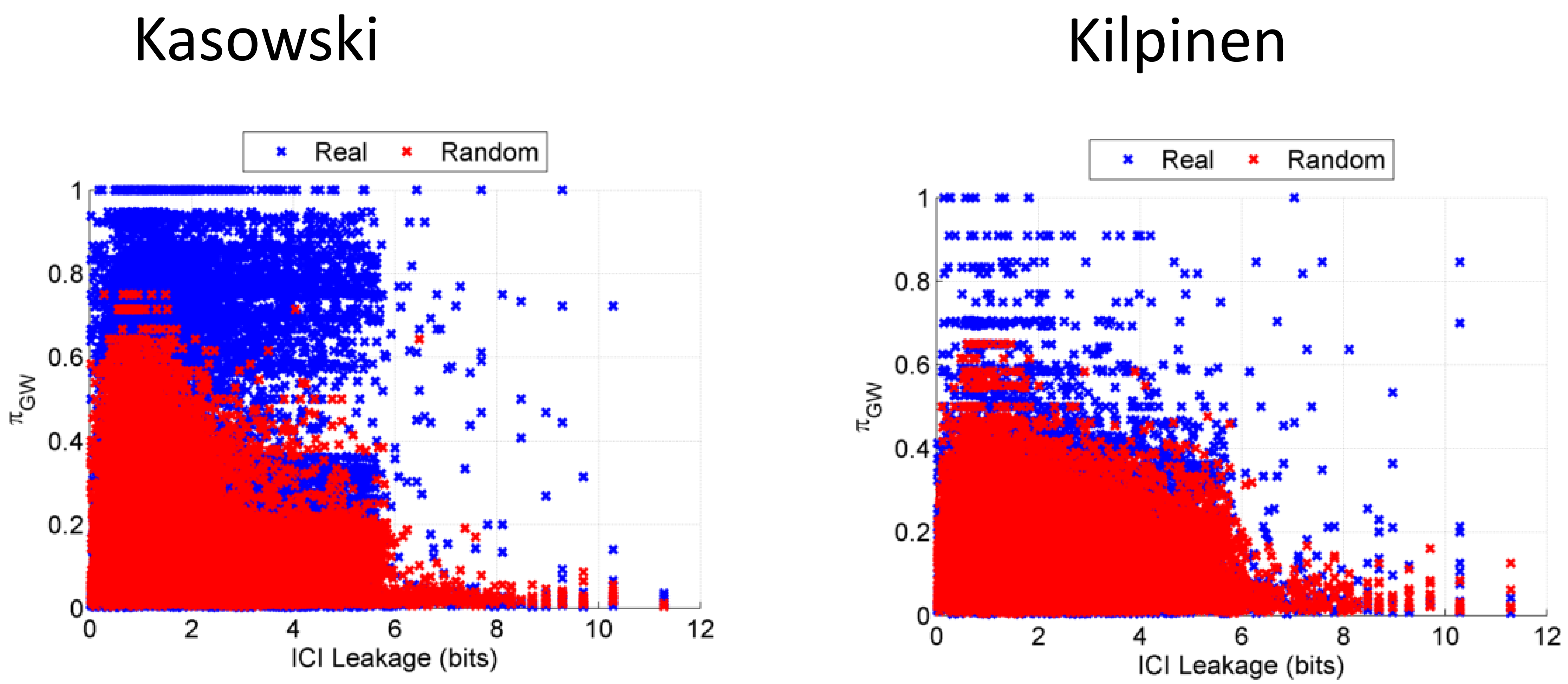
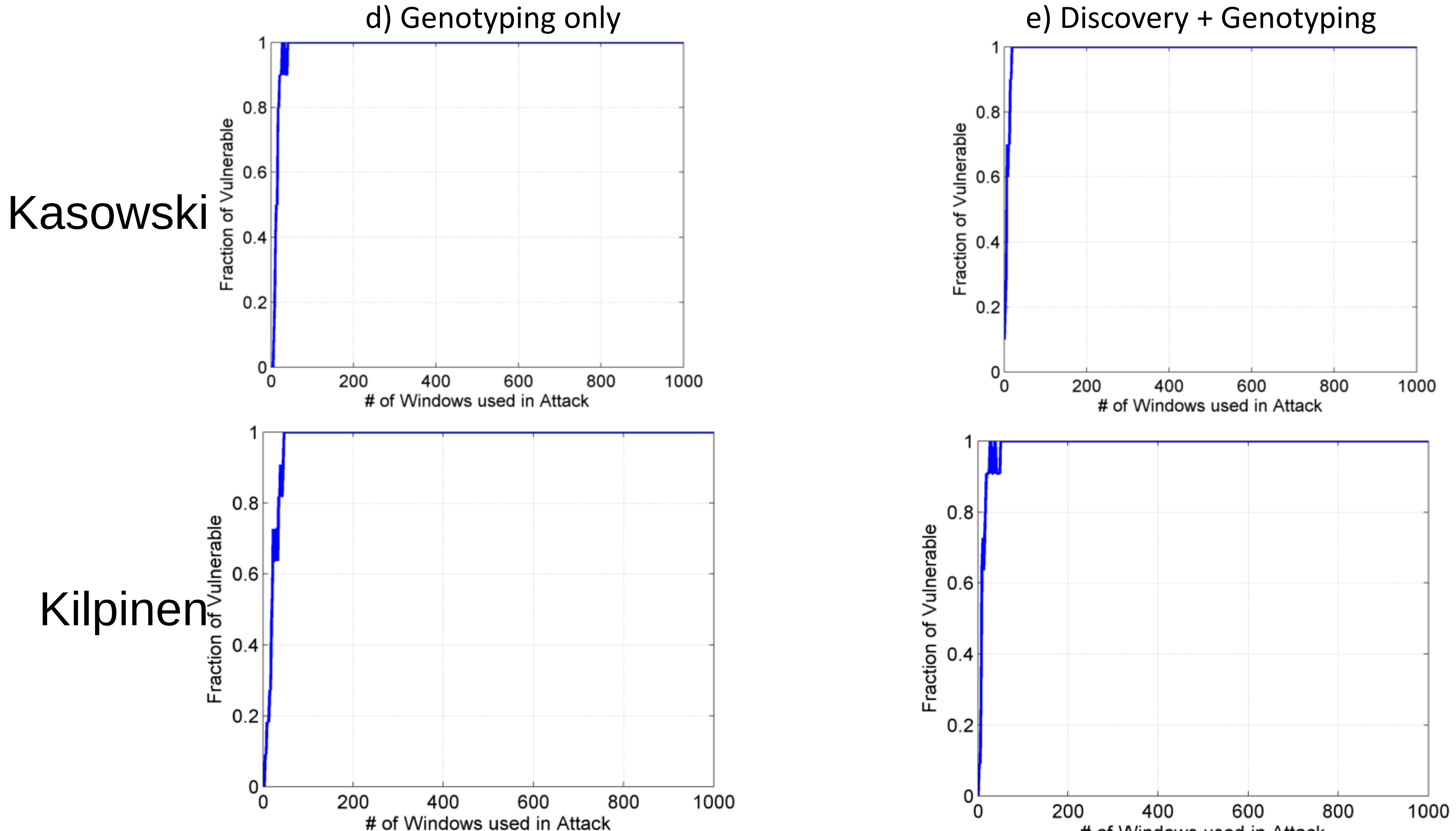

Genotype Only

Fig 4

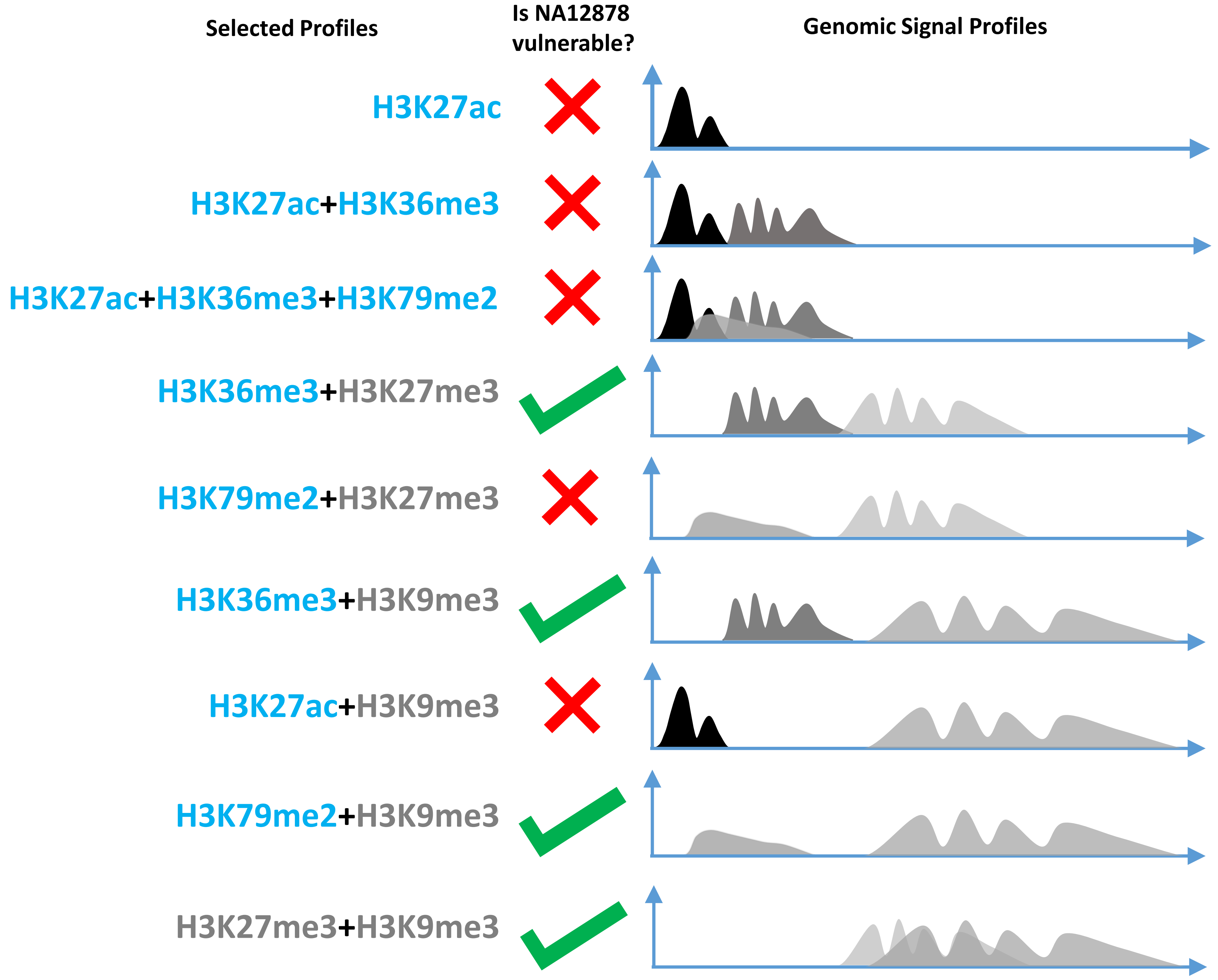
a) Genotyping large deletions



c,d) Extremity Attack Accuracy:
c) Genotyping Only
d) Joint Discovery and Genotyping



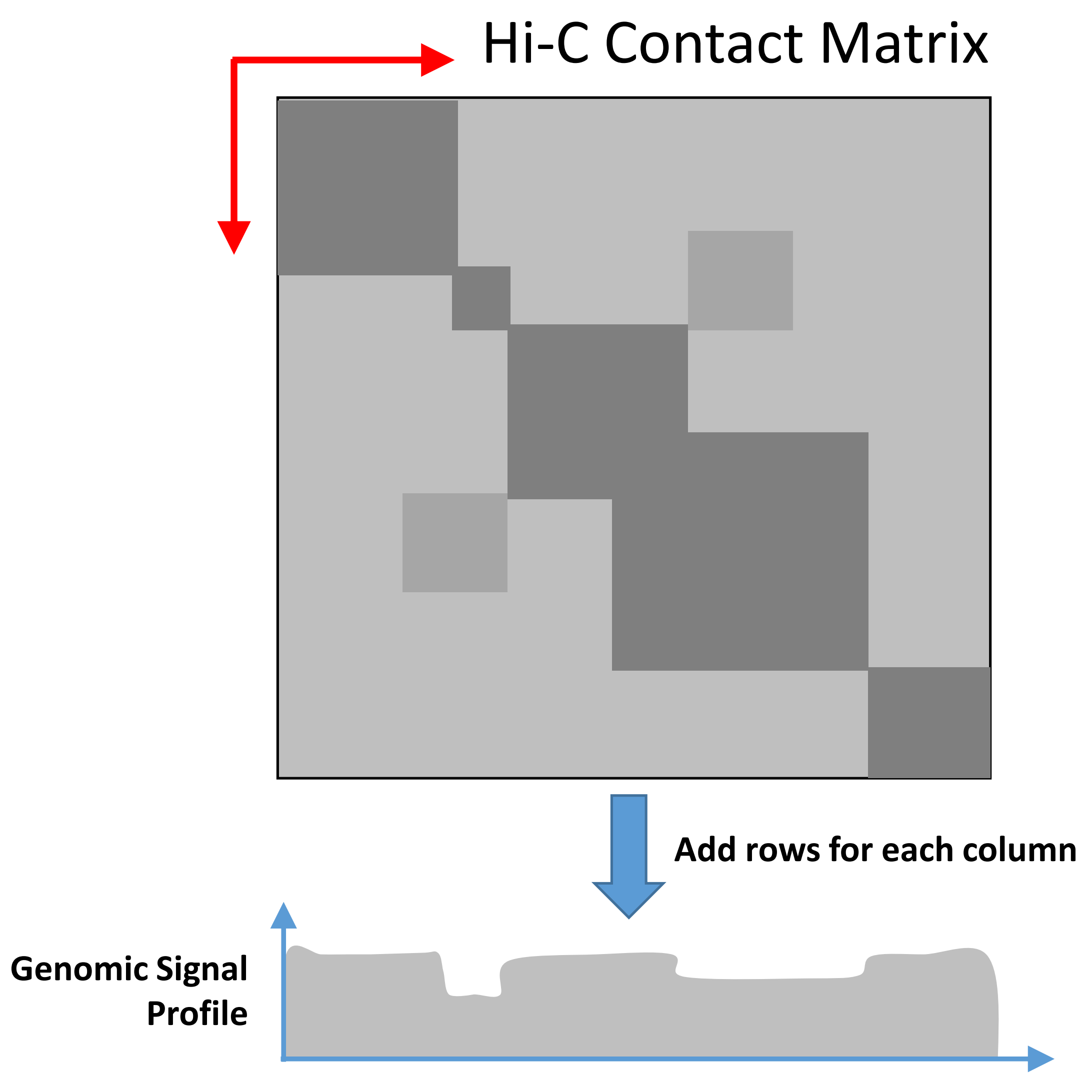
e) Combinations of datasets imply different vulnerability levels (known variants, variant existence)



Selected Profiles	Is NA12878 vulnerable?
H3K27ac	✗
H3K27ac+H3K36me3	✗
H3K27ac+H3K36me3+H3K79me2	✗
H3K36me3+H3K27me3	✓
H3K79me2+H3K27me3	✗
H3K36me3+H3K9me3	✓
H3K27ac+H3K9me3	✗
H3K79me2+H3K9me3	✓
H3K27me3+H3K9me3	✓

Fig 5

a) Hi-C analysis b) Per SV ICI-\Pi_GW estimates: NA12878 Hi-C data (different resolutions)

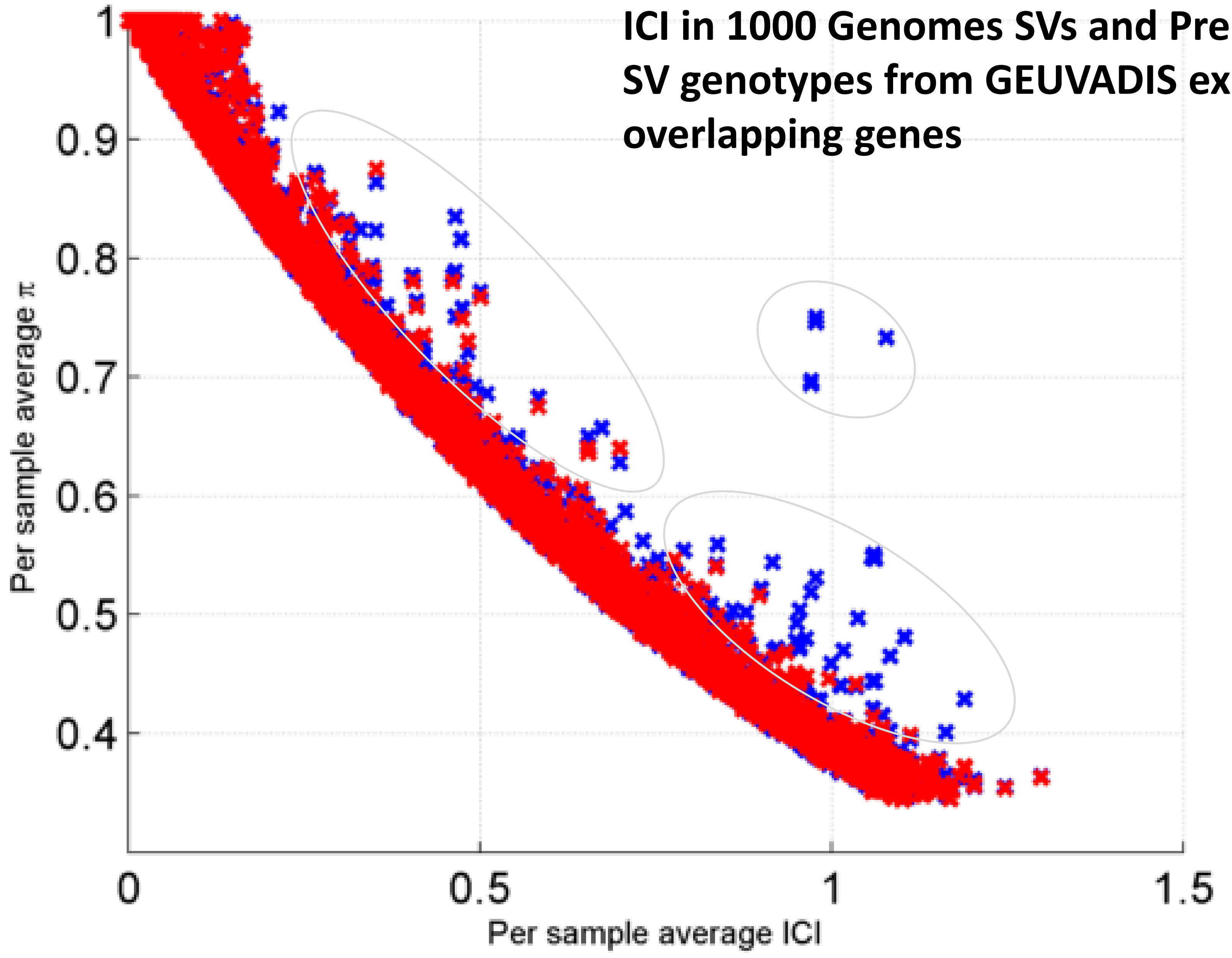


Resolution (kbs)	Is NA12878 vulnerable among 1kG samples?
1	✓
5	✓
10	✓
25	✗
50	✗

Supplementary Figures

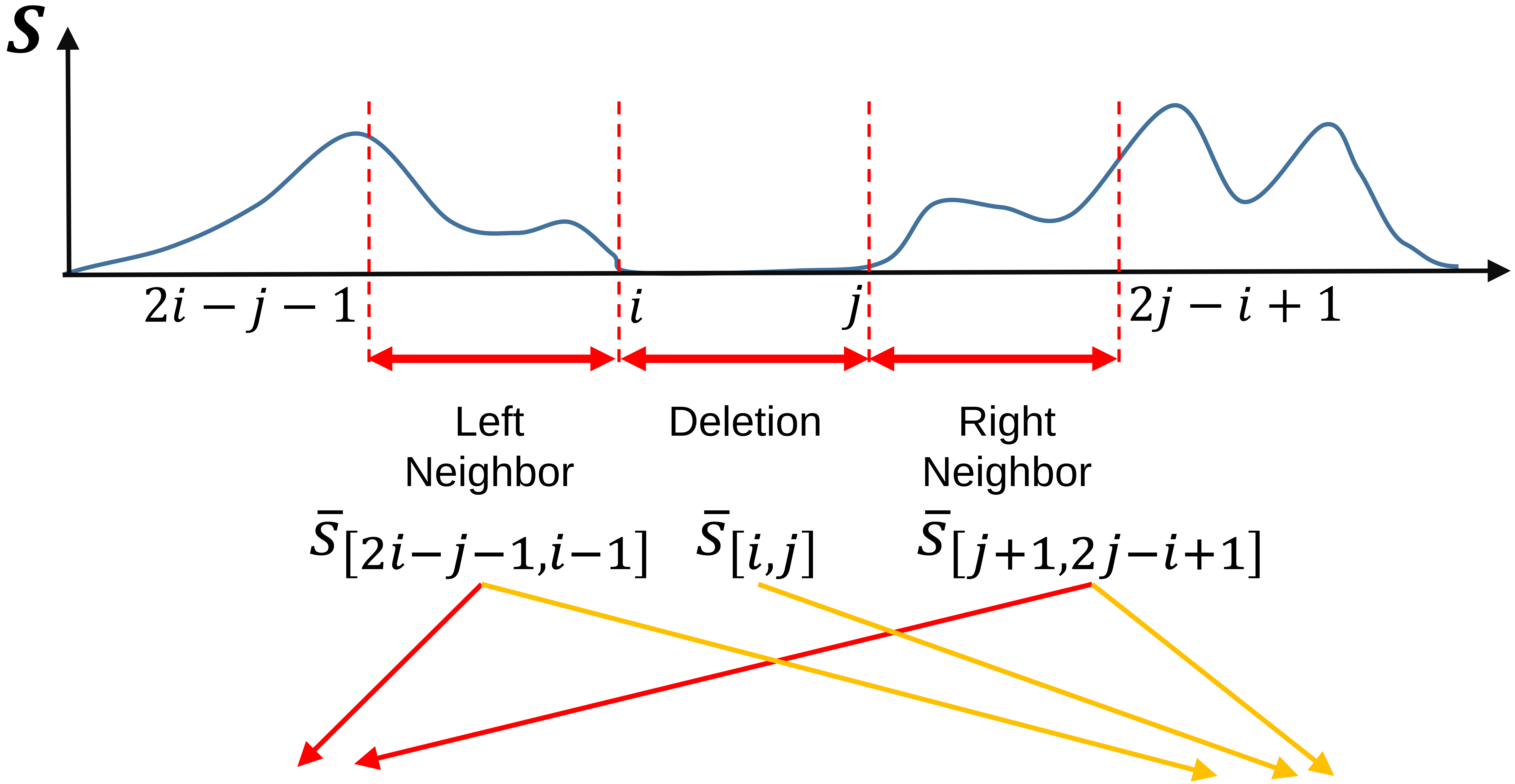
Fig S1 Sample-wide predictability of 1kG SVs

b) $\boldsymbol{\pi_{SW}}$ vs ICI for 1kG Structural Variants



ICI in 1000 Genomes SVs and Predictability (π) of SV genotypes from GEUVADIS expression levels of overlapping genes

Fig S2 Features for genotyping deletions



$\eta_{[i,j]}$: How balanced are the neighbor signal levels?

$\tau_{[i,j]}$: How high are the neighbor signal levels?

$\rho_{[i,j]}$: How deep is the dip in the signal?

Fig S3

Allele Frequency Distribution of Large and Small Deletions

Fig S4

a)

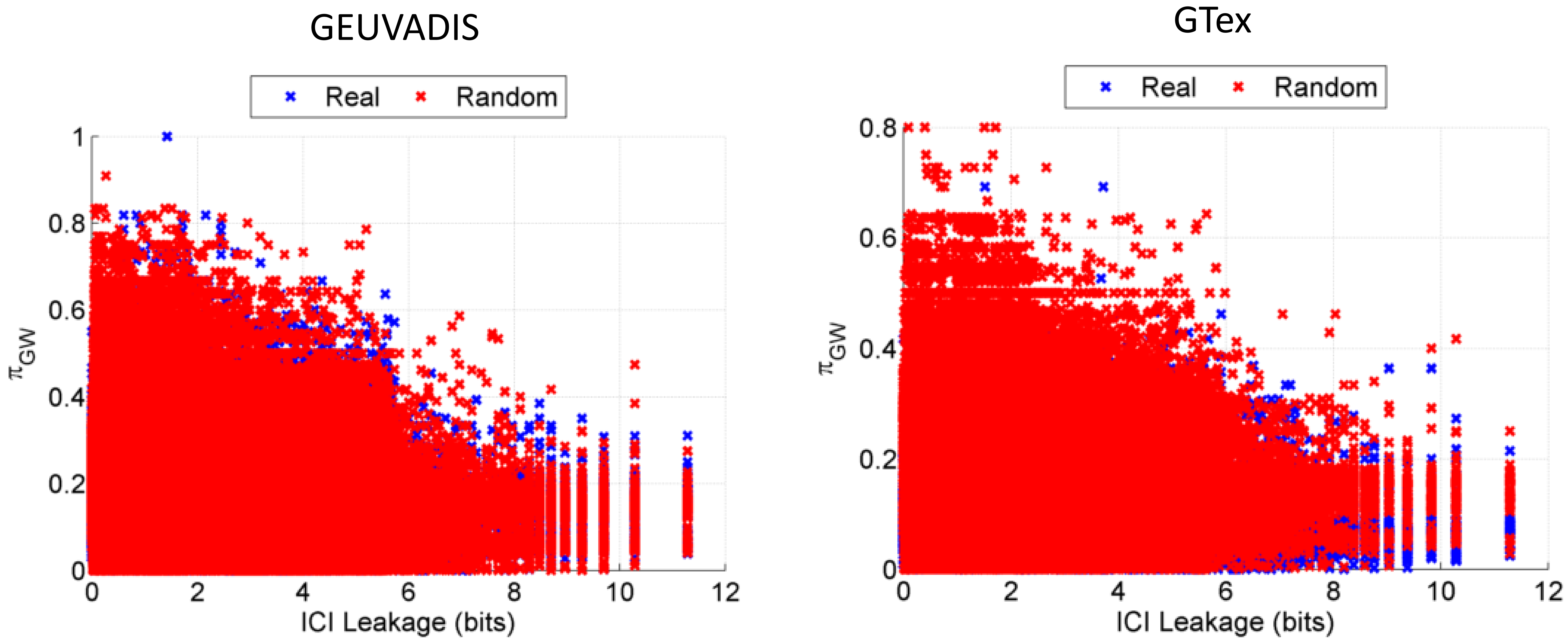
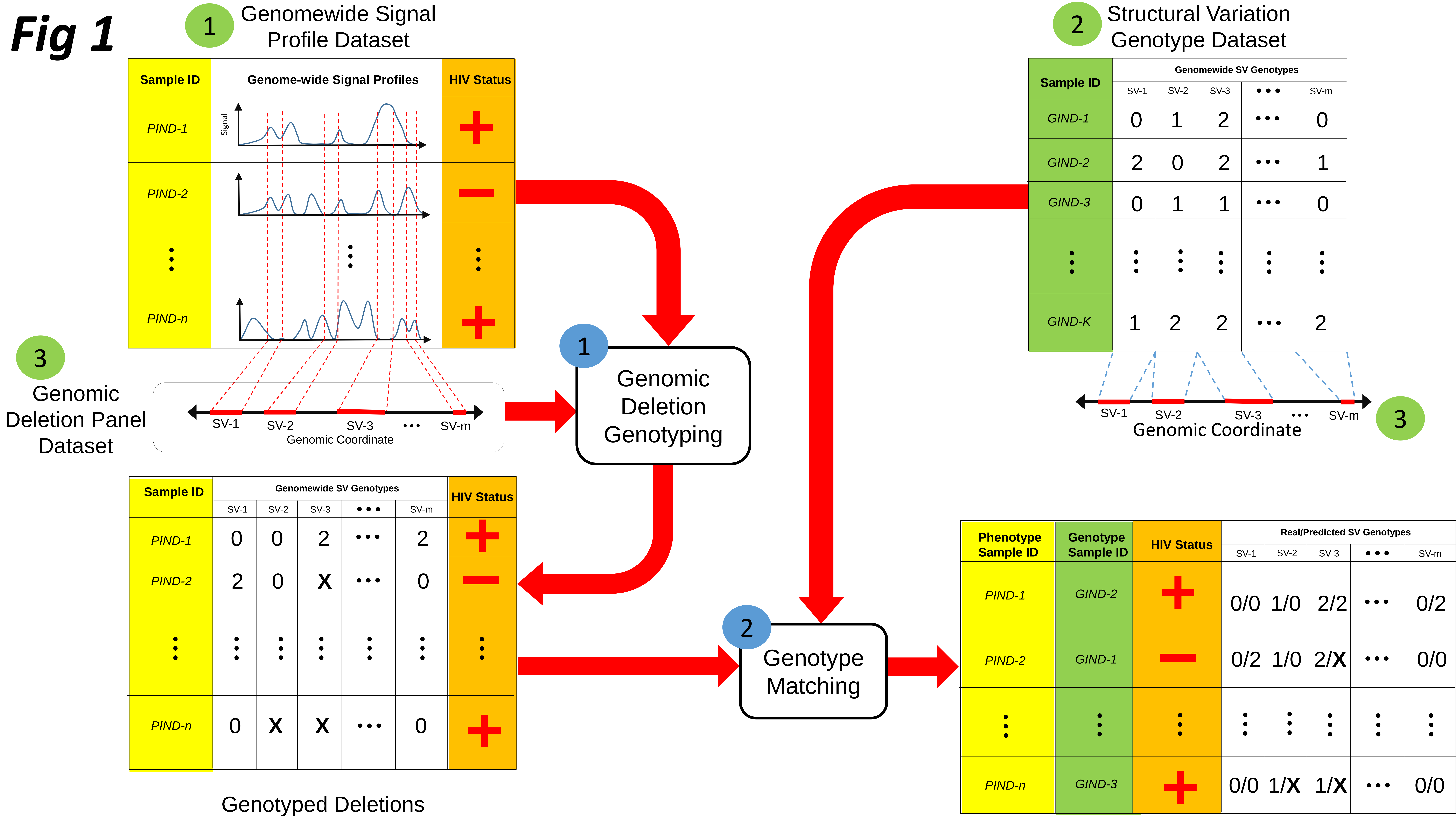

GEUVADIS
Real
Random
π_{GW}
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
12
ICI Leakage (bits)
GTex
Real
Random
π_{GW}
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
12
ICI Leakage (bits)

Fig 1



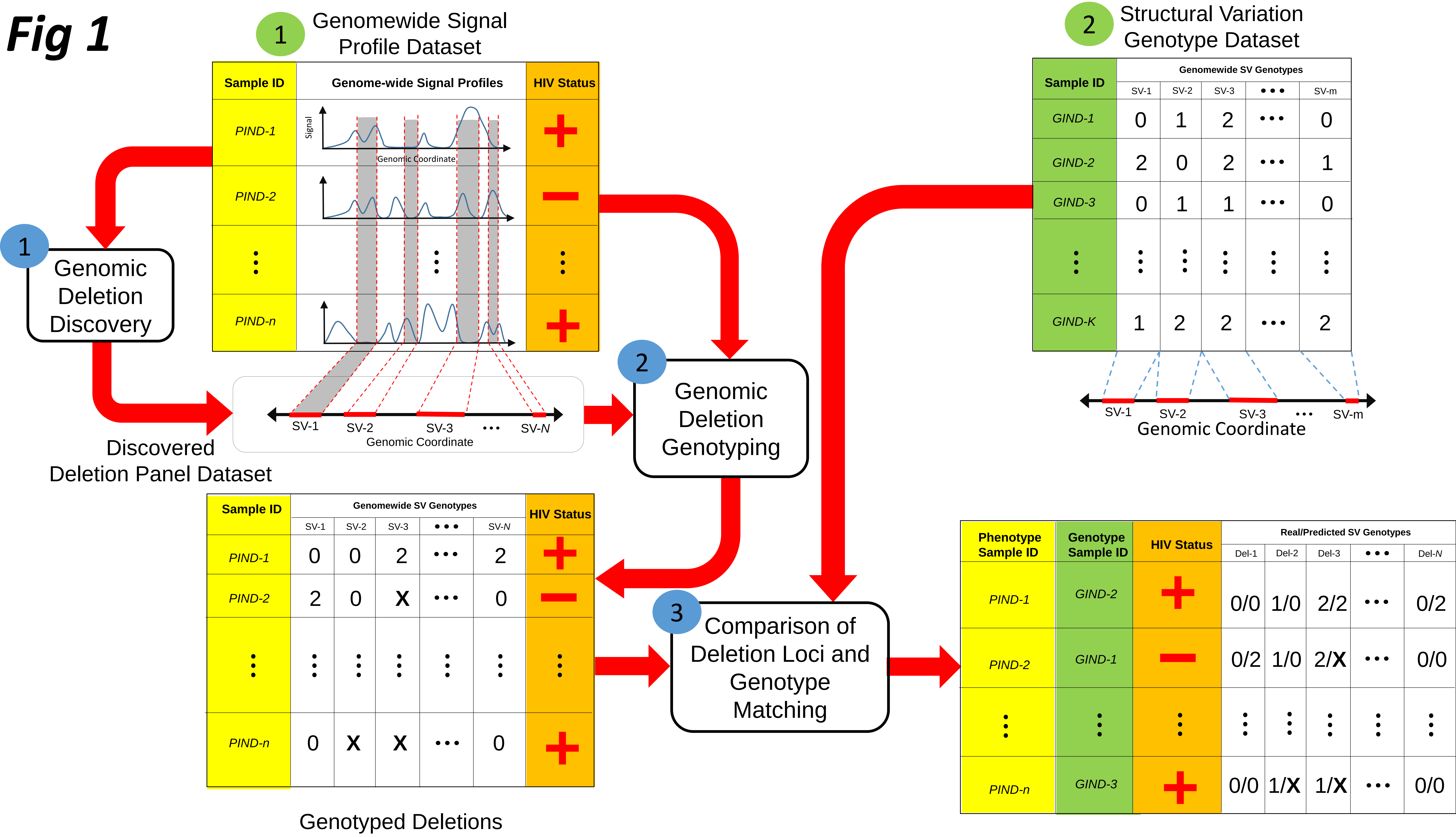
Sample ID	Genome-wide Signal Profiles	HIV Status
PIND-1		+
PIND-2		−
⋮	⋮	⋮
PIND-n		+

Sample ID	Genomewide SV Genotypes				
	SV-1	SV-2	SV-3	•••	SV-m
GIND-1	0	1	2	•••	0
GIND-2	2	0	2	•••	1
GIND-3	0	1	1	•••	0
⋮	⋮	⋮	⋮	⋮	⋮
GIND-K	1	2	2	•••	2

Sample ID	Genomewide SV Genotypes					HIV Status
	SV-1	SV-2	SV-3	•••	SV-m	
PIND-1	0	0	2	•••	2	+
PIND-2	2	0	**X**	•••	0	−
⋮	⋮	⋮	⋮	⋮	⋮	⋮
PIND-n	0	**X**	**X**	•••	0	+

Phenotype Sample ID	Genotype Sample ID	HIV Status	Real/Predicted SV Genotypes				
			SV-1	SV-2	SV-3	•••	SV-m
PIND-1	*GIND-2*	+	0/0	1/0	2/2	•••	0/2
PIND-2	*GIND-1*	−	0/2	1/0	2/**X**	•••	0/0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
PIND-n	*GIND-3*	+	0/0	1/**X**	1/**X**	•••	0/0

Fig 1



Sample ID	Genome-wide Signal Profiles	HIV Status
PIND-1		+
PIND-2		−
⋮	⋮	⋮
PIND-n		+

Sample ID	Genomewide SV Genotypes				
	SV-1	SV-2	SV-3	•••	SV-m
GIND-1	0	1	2	•••	0
GIND-2	2	0	2	•••	1
GIND-3	0	1	1	•••	0
⋮	⋮	⋮	⋮	⋮	⋮
GIND-K	1	2	2	•••	2

Sample ID	Genomewide SV Genotypes					HIV Status
	SV-1	SV-2	SV-3	•••	SV-*N*	
PIND-1	0	0	2	•••	2	+
PIND-2	2	0	**X**	•••	0	−
⋮	⋮	⋮	⋮	⋮	⋮	⋮
PIND-n	0	**X**	**X**	•••	0	+

Phenotype Sample ID	Genotype Sample ID	HIV Status	Real/Predicted SV Genotypes				
			Del-1	Del-2	Del-3	•••	Del-*N*
PIND-1	*GIND-2*	+	0/0	1/0	2/2	•••	0/2
PIND-2	*GIND-1*	−	0/2	1/0	2/**X**	•••	0/0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
PIND-n	*GIND-3*	+	0/0	1/**X**	1/**X**	•••	0/0