

PrivaSeq2 Figures

--Attacker Strikes Back--

Change the ICI-\Pi figure for reporting one dot for each deletion?

Fig 1

a) GW Linking Attack Scenario

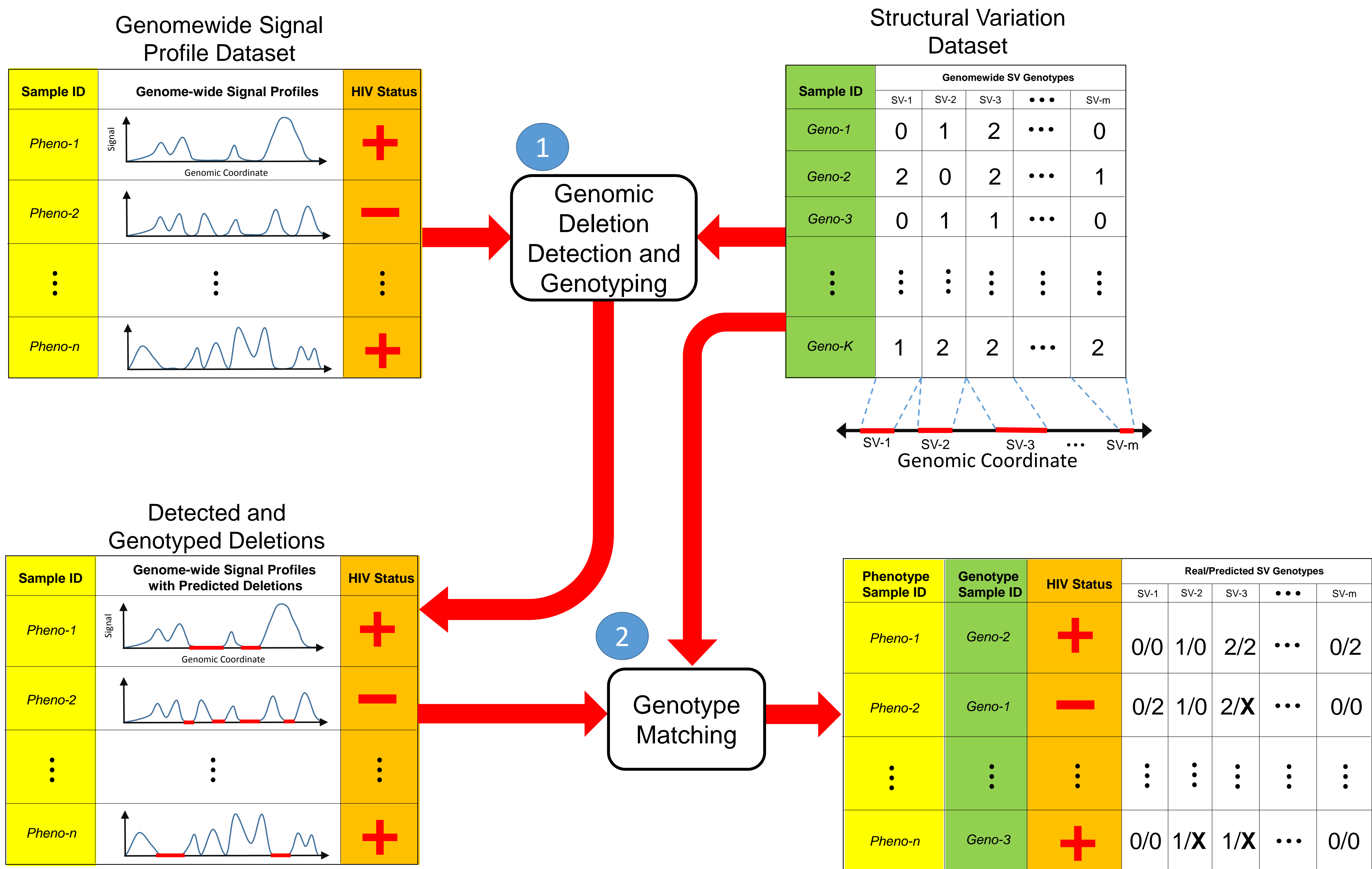


Fig 1

b) NA12878 large deletion

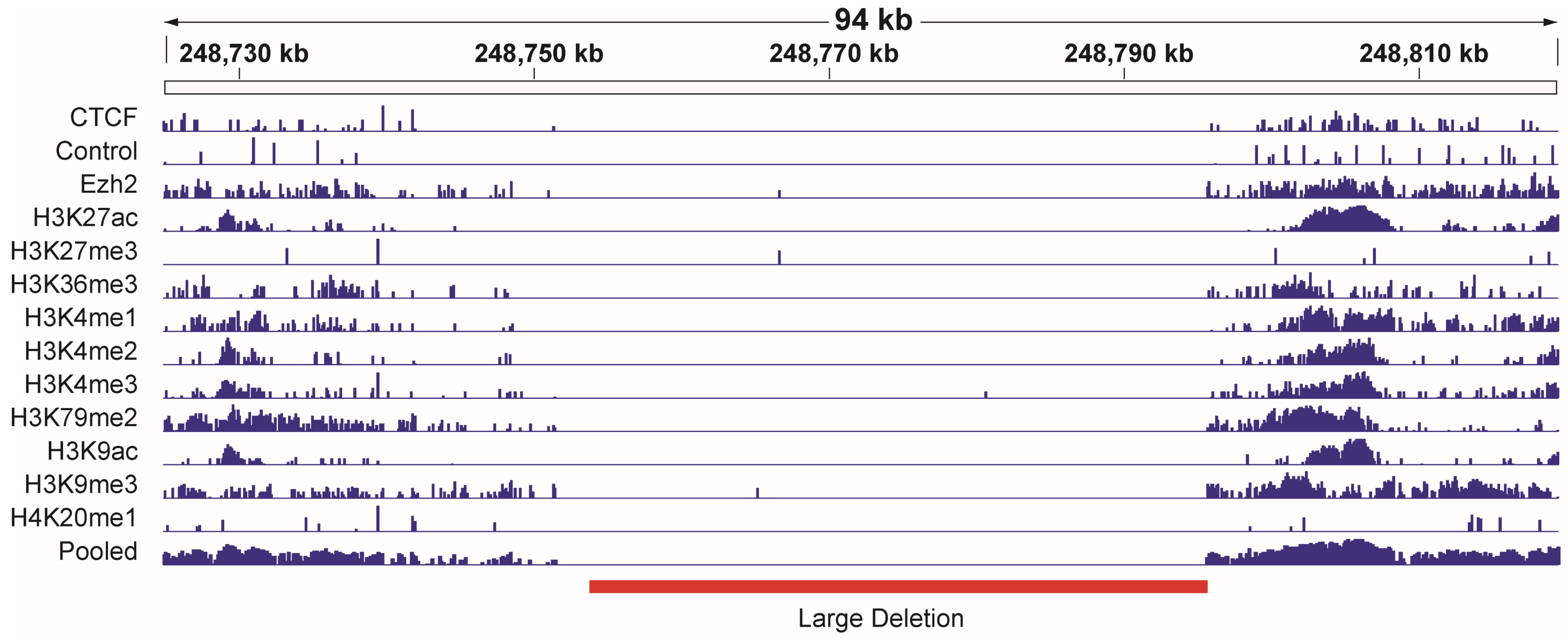


Fig 1

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

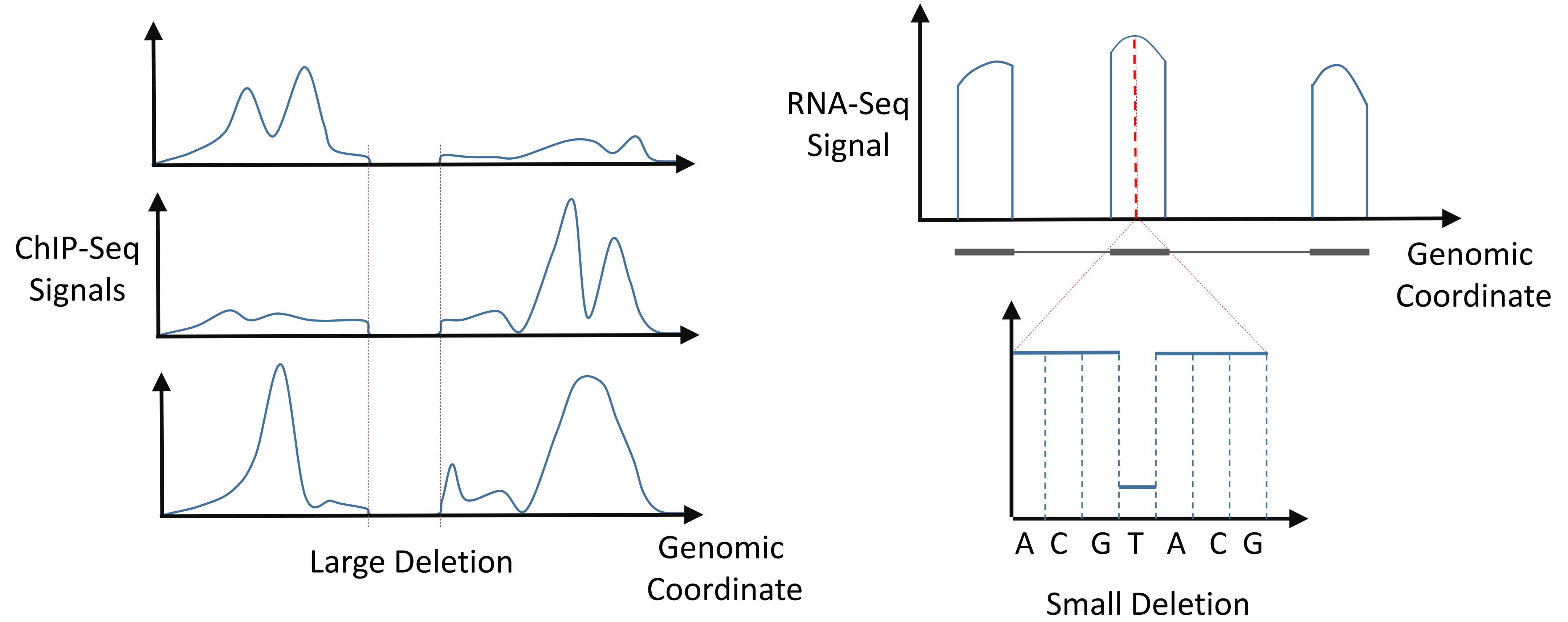
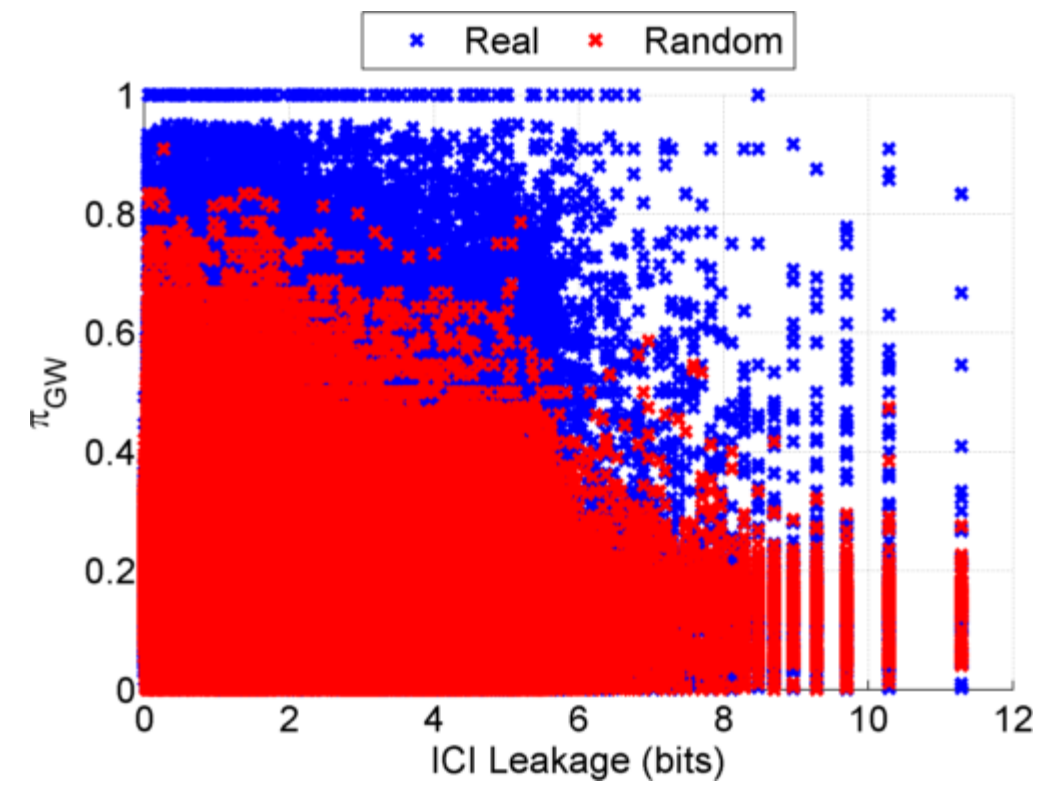


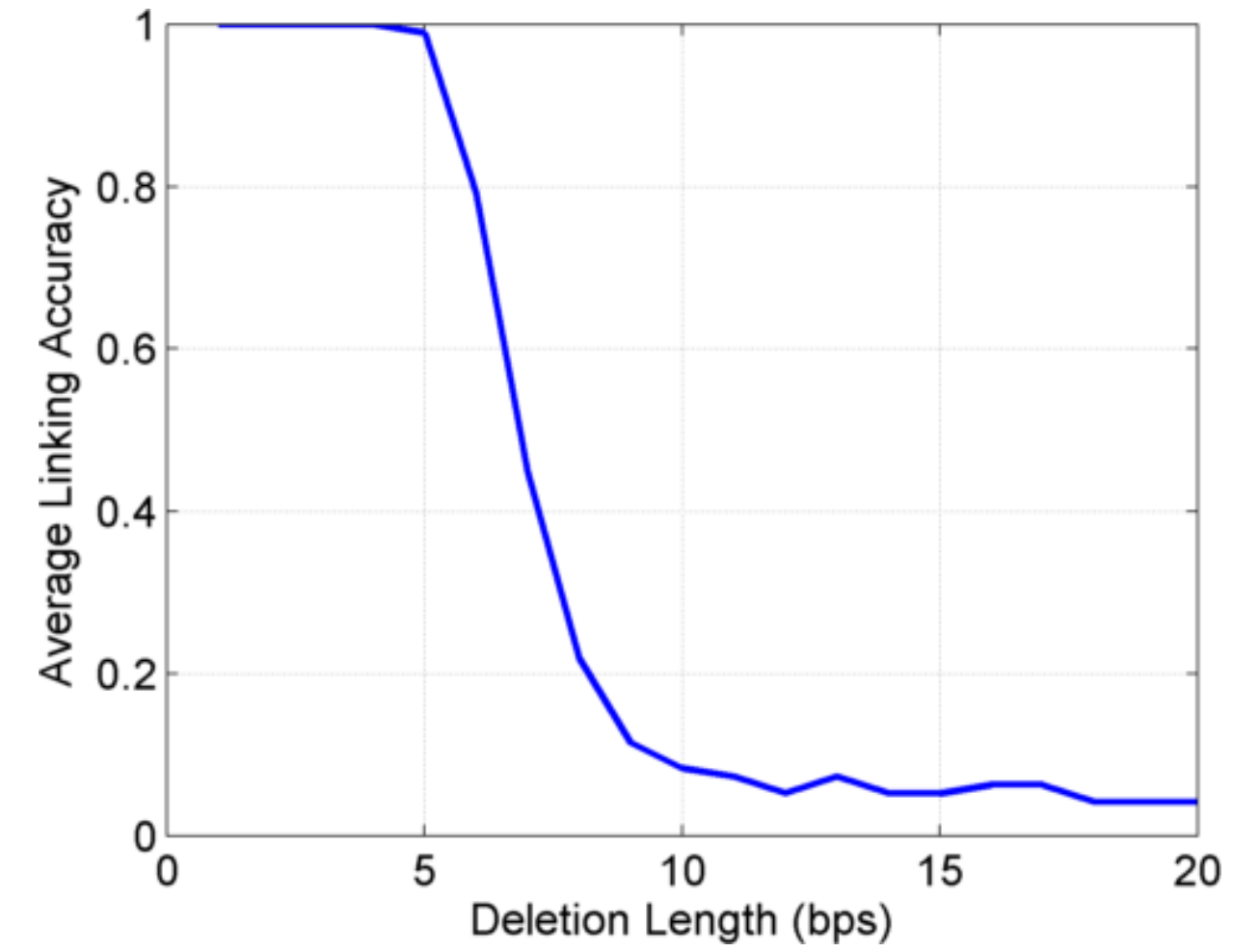
Fig 2

a) GEUVADIS ICI-\Pi_GW estimates:

a) Genotyping known deletions



e) Indel length vs vulnerability statistics



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

f,g) Min # of identifying variants

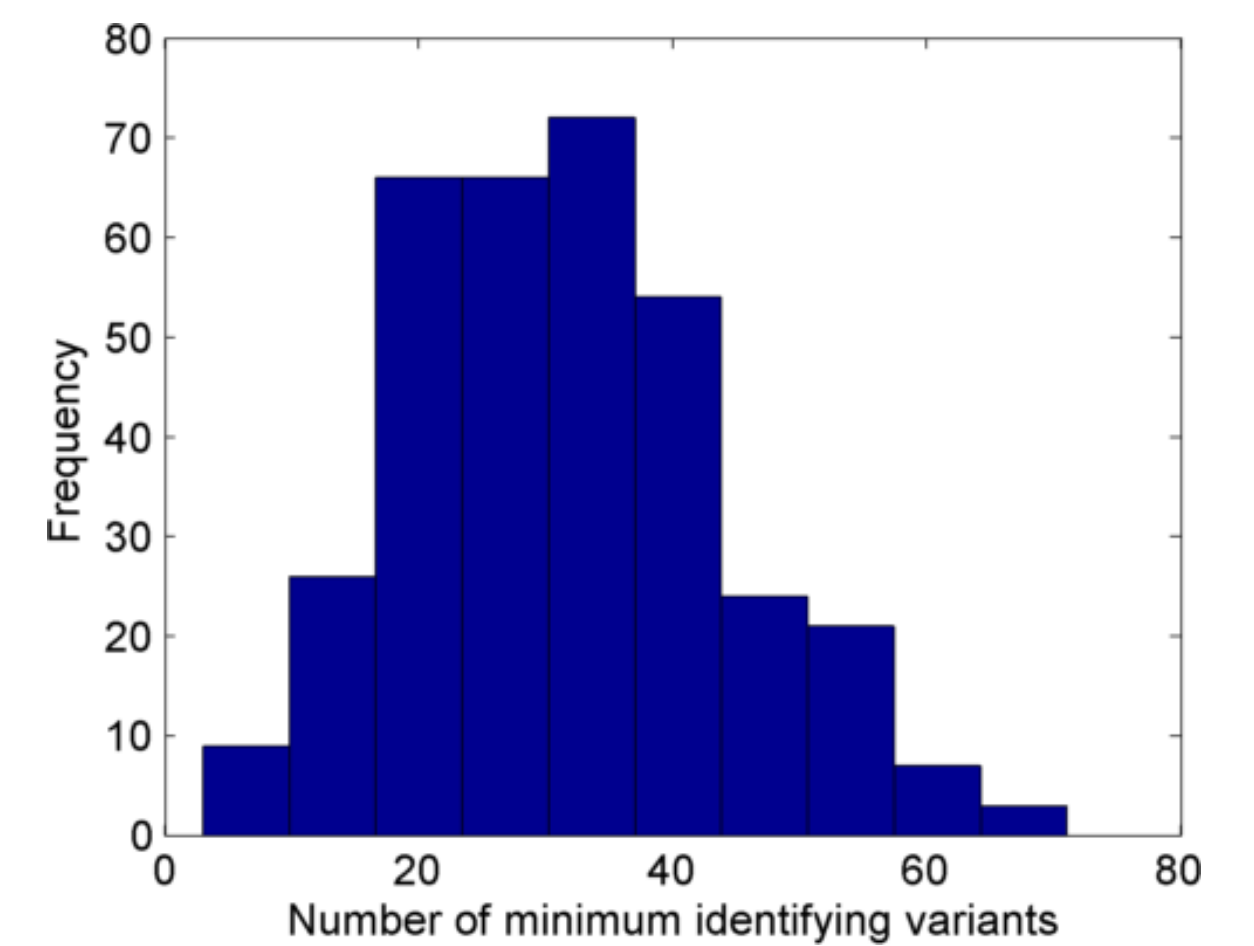
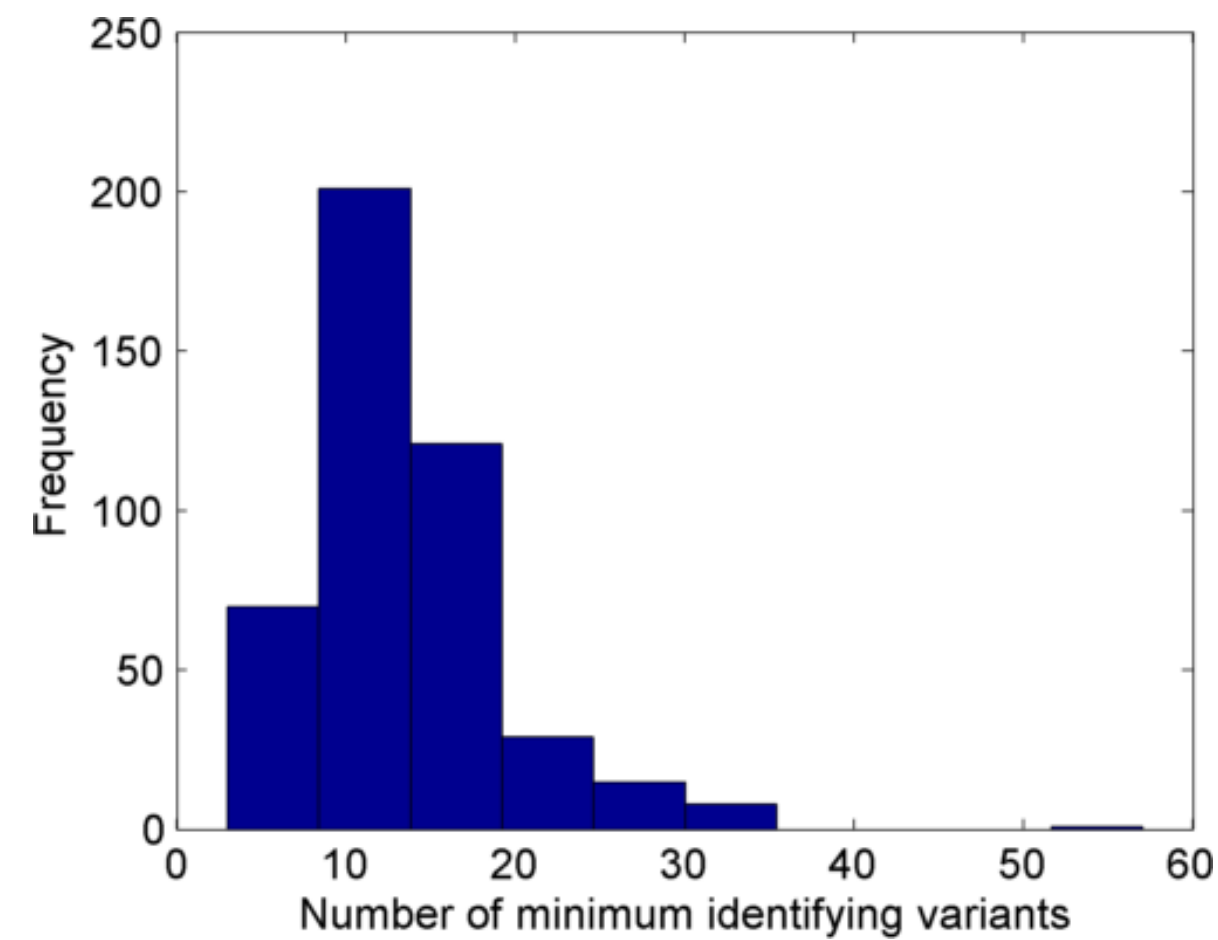
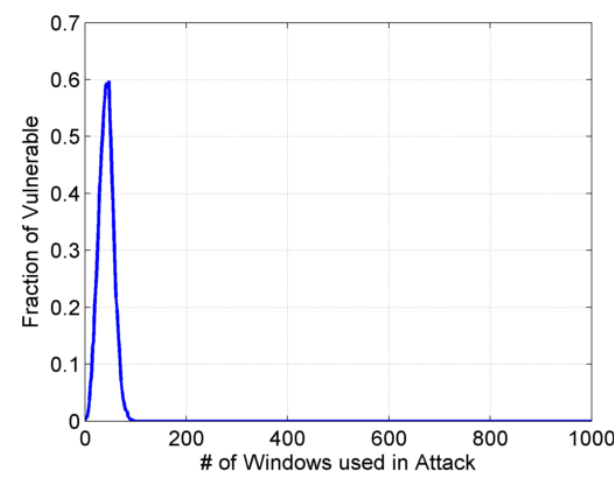
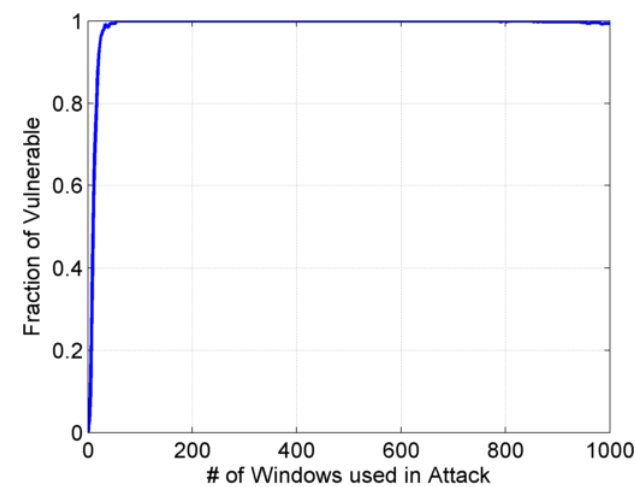
d) Known Indels (1kG MAF>0.01)

e) De-novo Predicted

f) Known Indels (1kG MAF>0.01)

g) De-novo Predicted

Genotype



Existence of Deletion

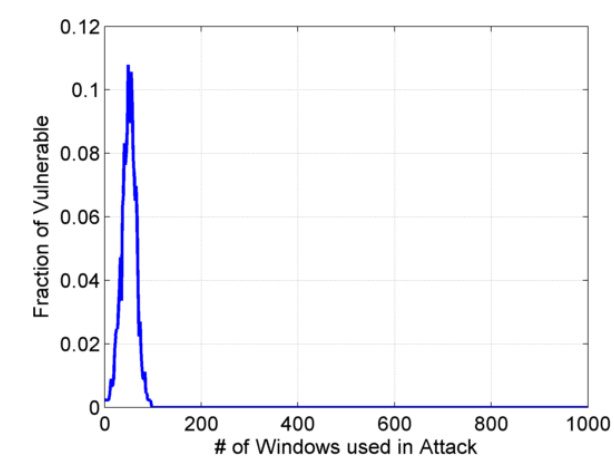
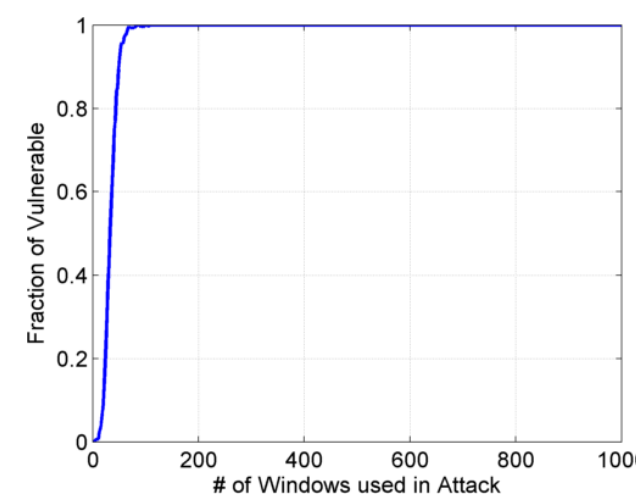
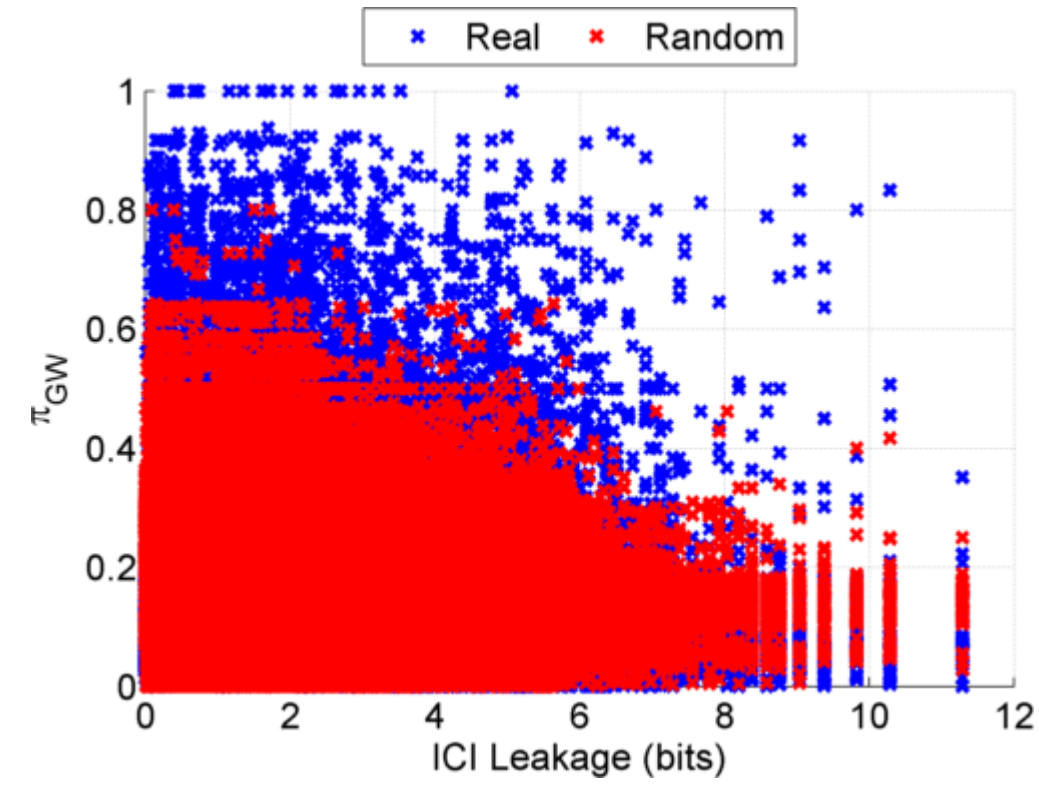


Fig 3

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

a) Deletion Genotyping



Gtex extremity attack: known indels (1kG MAF>0.01)
c) Genotyping d) Existence of Deletion

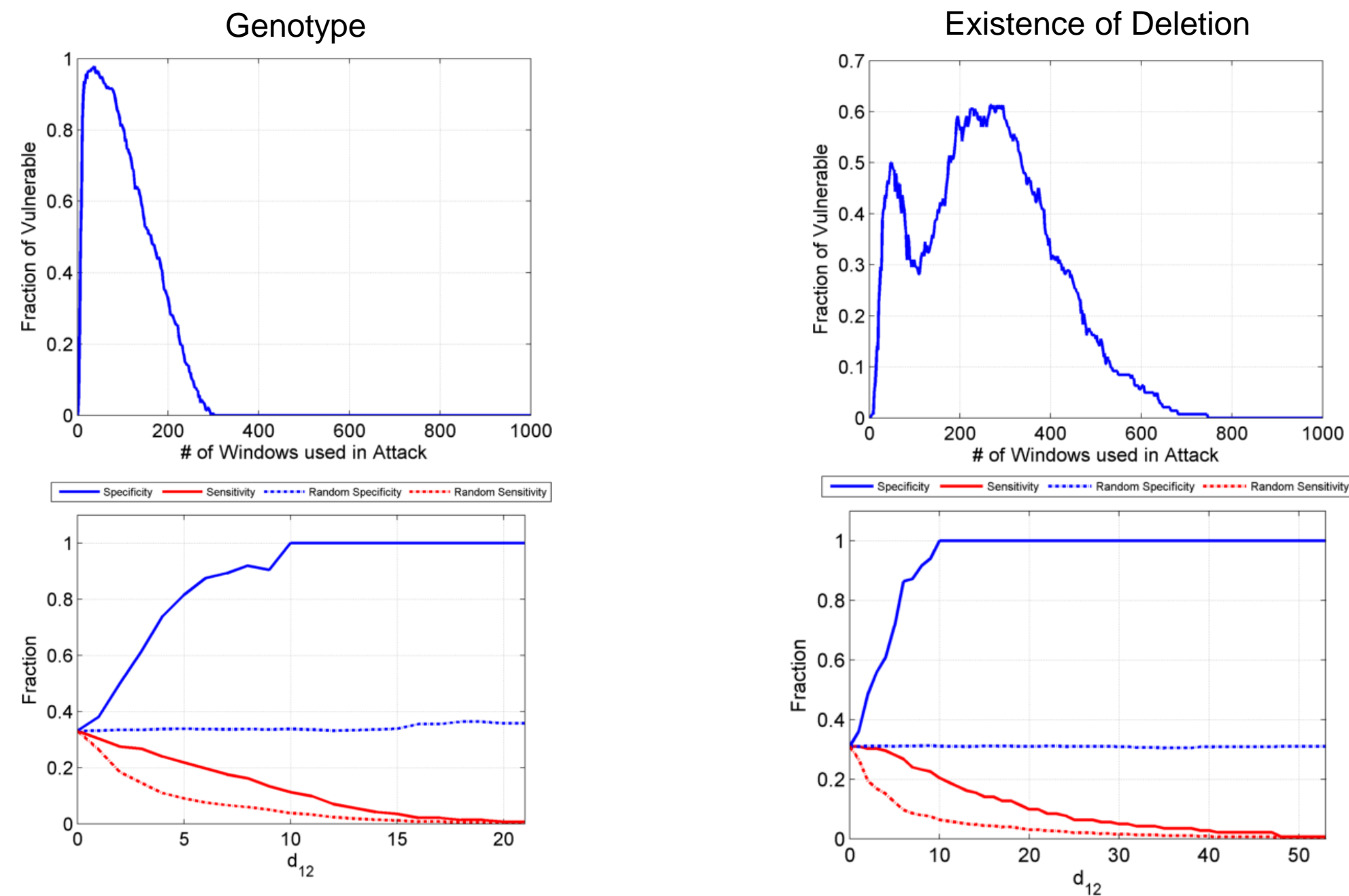
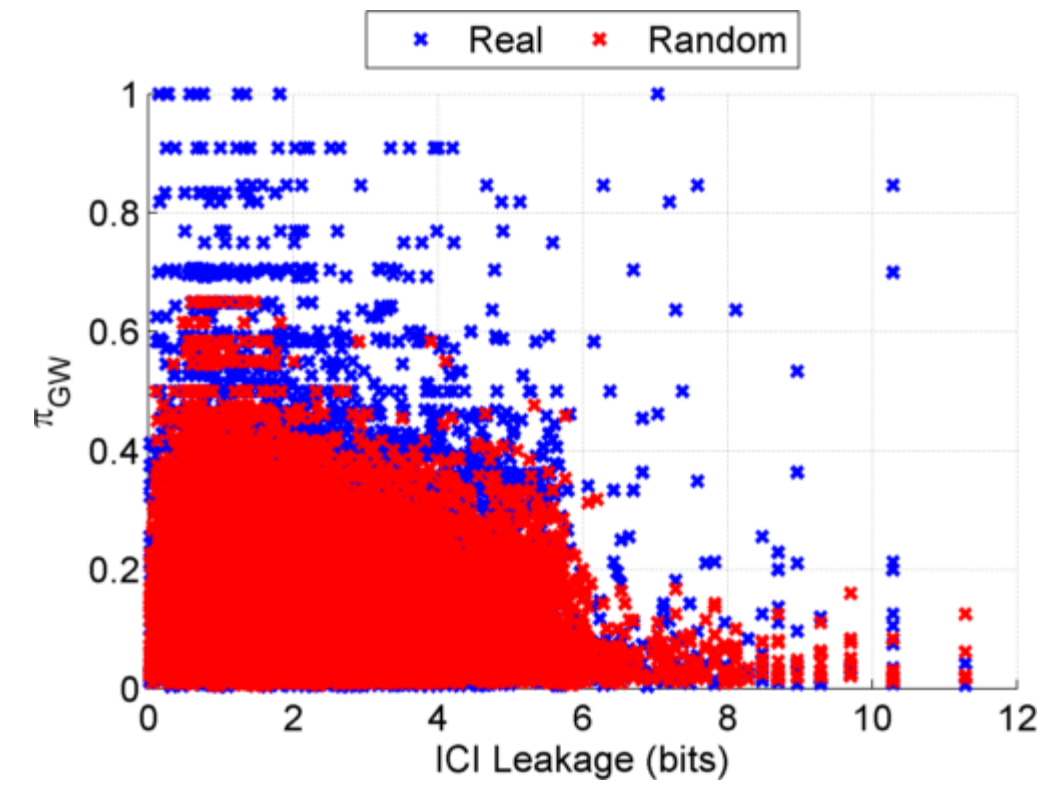
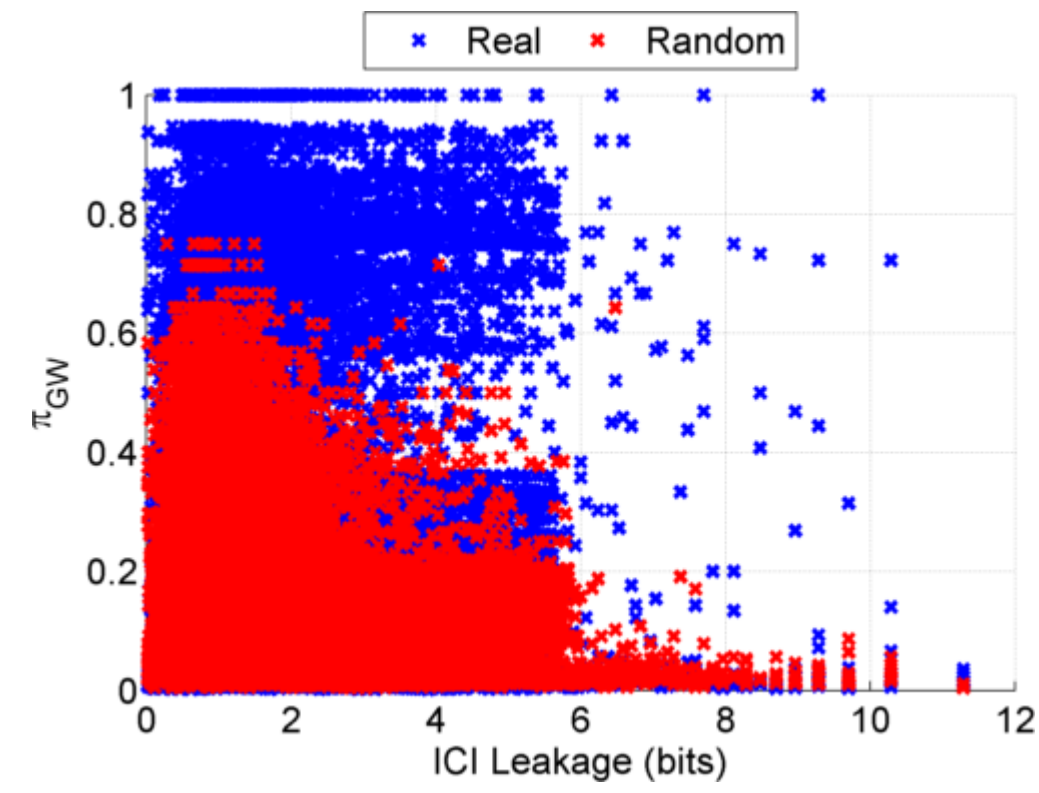


Fig 4

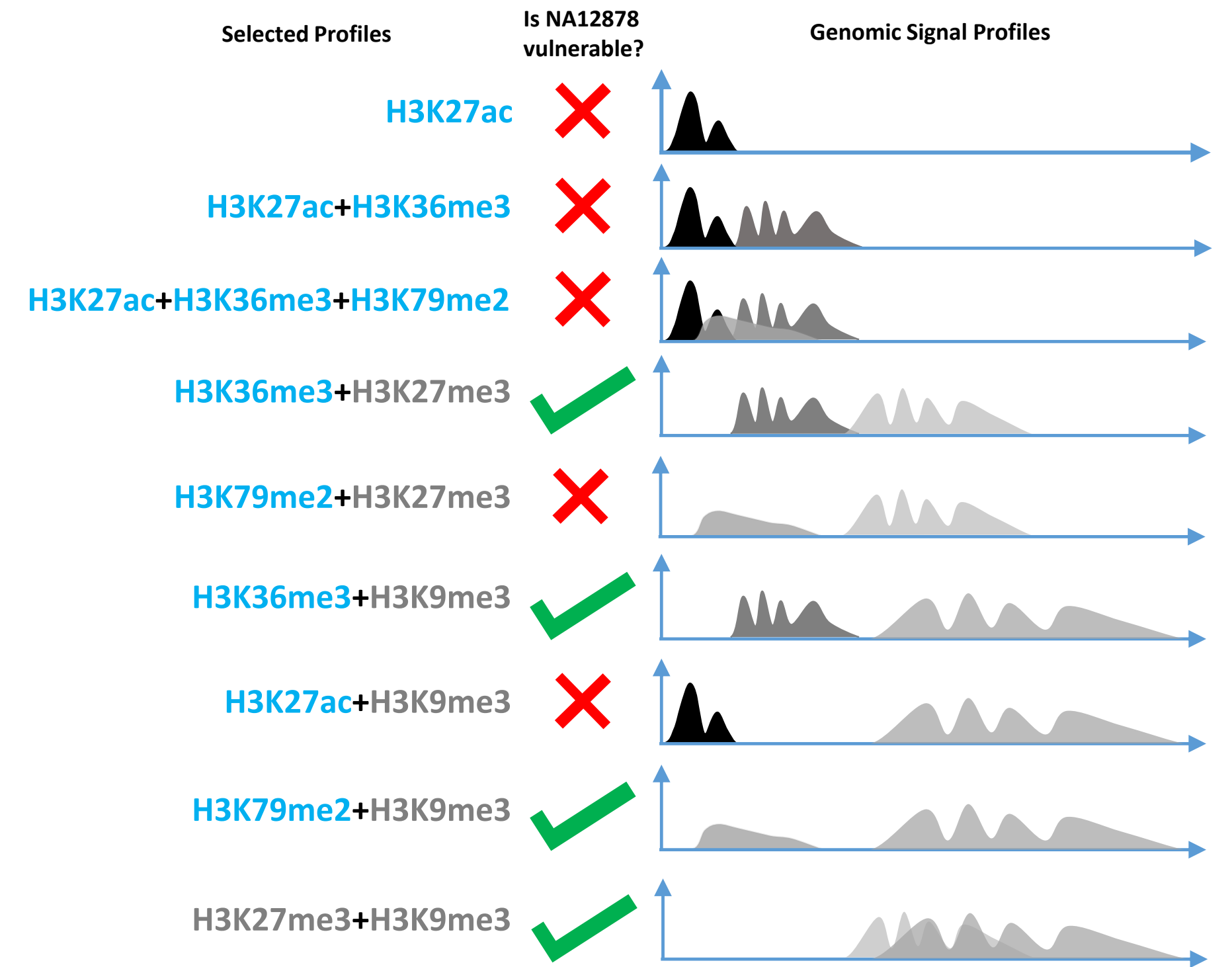
a) Genotyping large deletions

Kasowski

Kilpinen



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



c,d) Extremity Attack Accuracy:

c) Genotyping Only

d) Joint Discovery and Genotyping

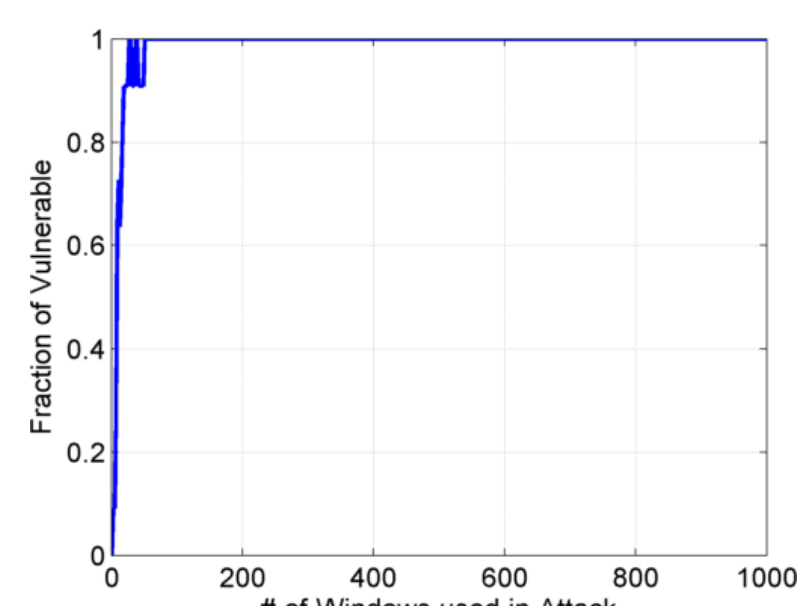
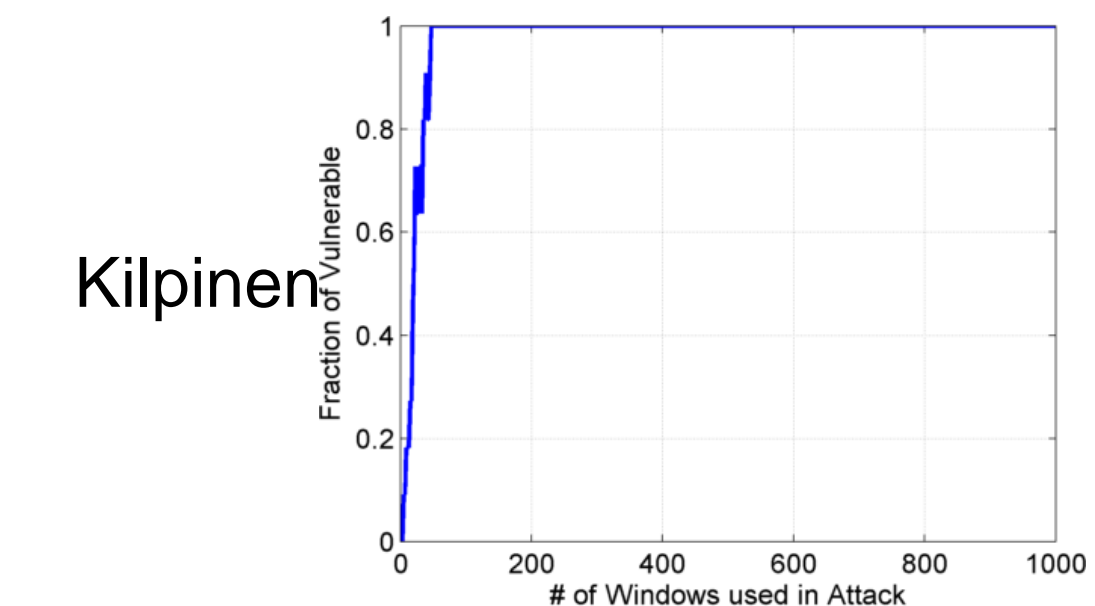
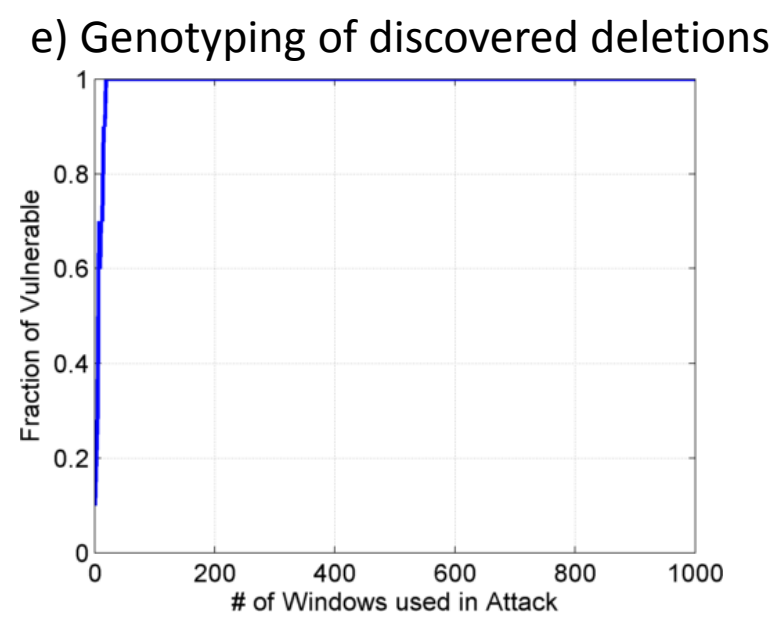
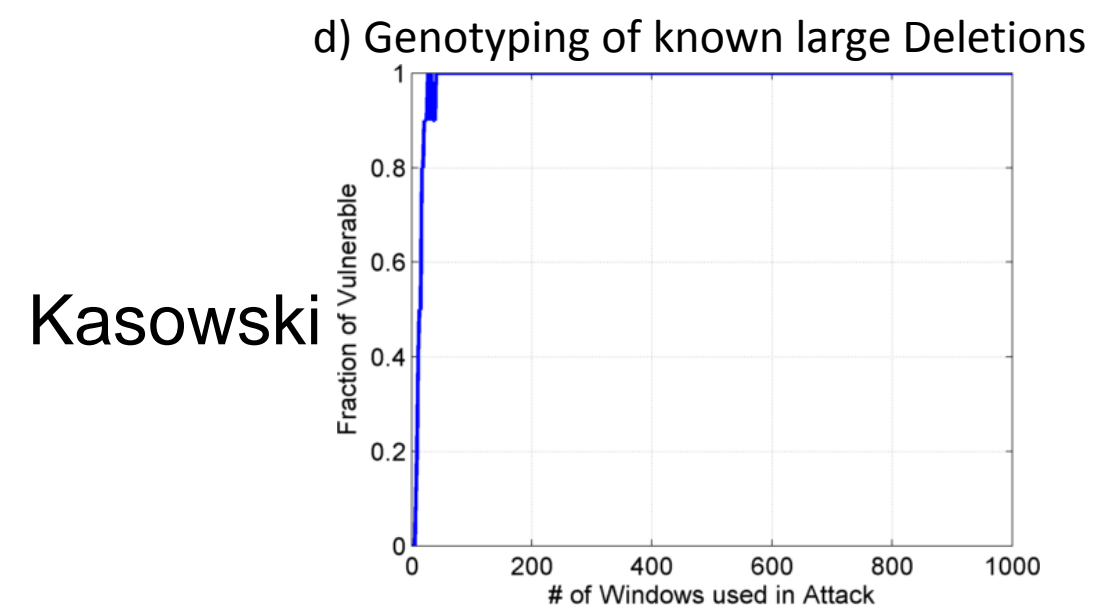
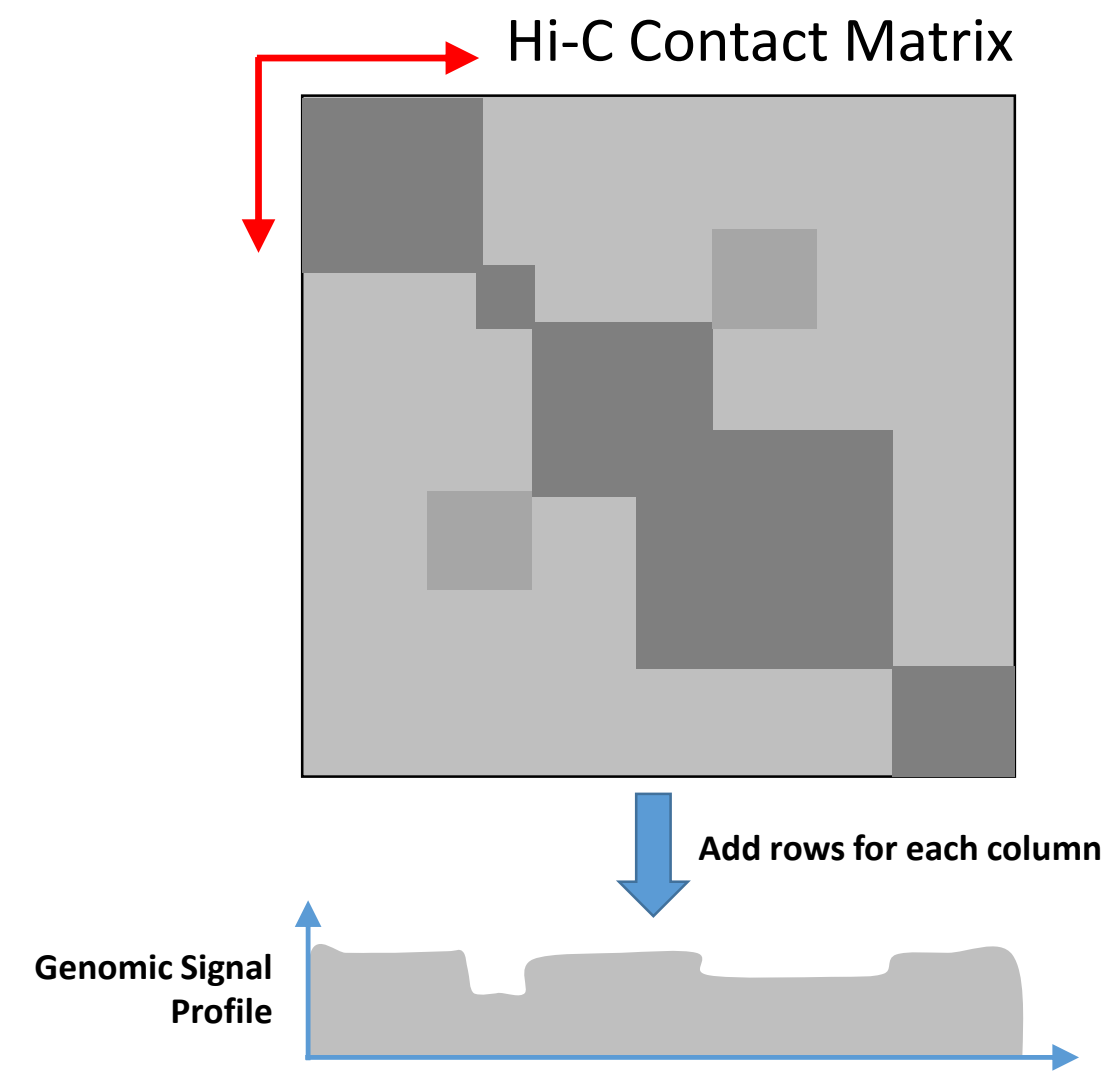


Fig 5

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

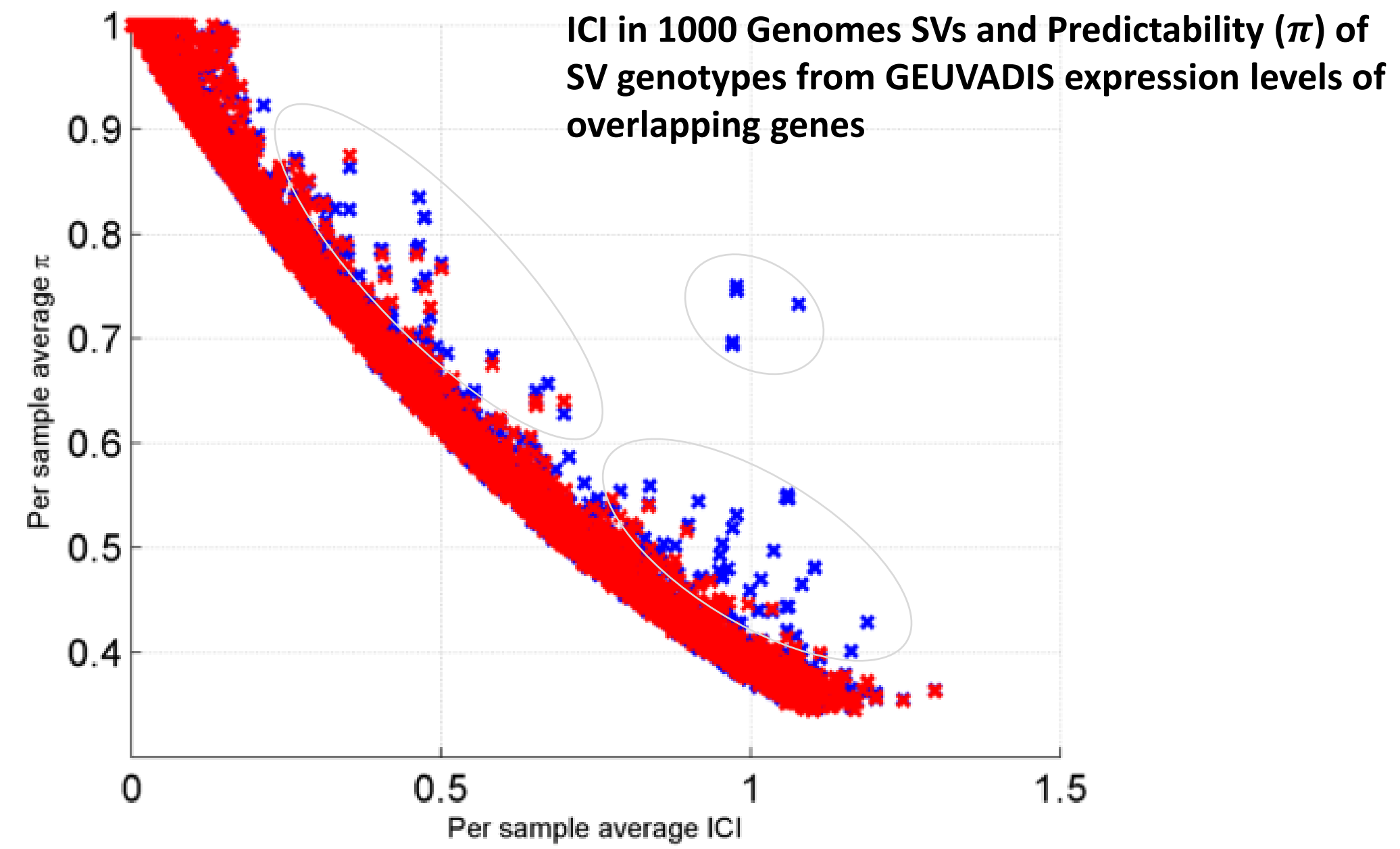
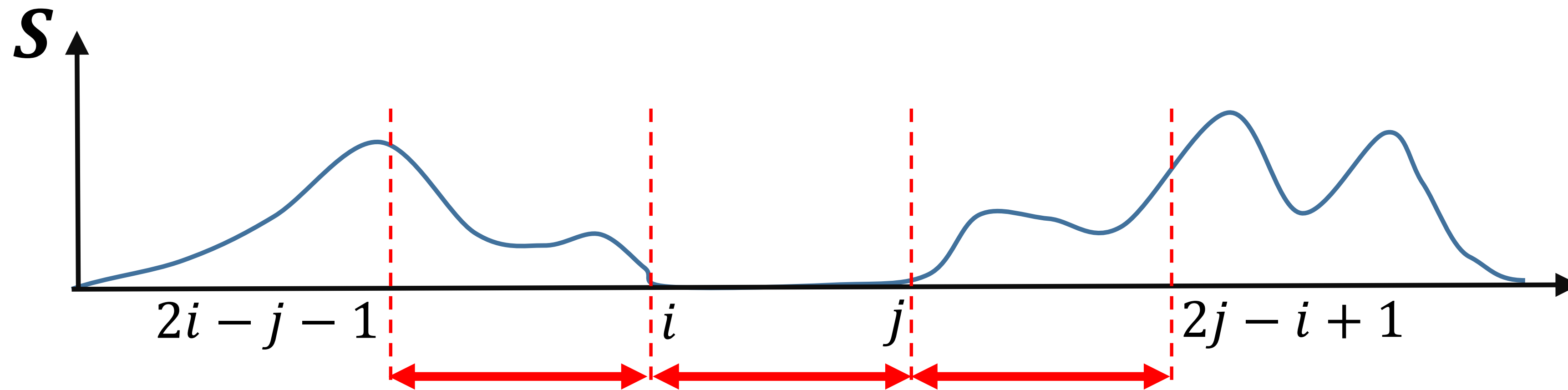
b) π_{SW} vs ICI for 1kG Structural Variants

Fig S2 Features for genotyping deletions



Left
Neighbor

Deletion

Right
Neighbor

$$\bar{S}[2i-j-1, i-1]$$

$$\bar{S}[i, j]$$

$$\bar{S}[j+1, 2j-i+1]$$

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

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

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

Fig S3
Allele Frequency Distribution of Large and
Small Deletions

Fig S4

a)

