

Aggregation

Re-identification of aggregated data is possible

Statistics from allele frequencies can be used to identify genetic trait
participants [1]

Presence of an individual in a group can be determined by using allele frequencies and his DNA profile [2]

[1] R. Wang et al. ‘Learning Your Identity and Disease from Research Papers: Information Leaks in Genome Wide Association Study.’

ACM CCS, 2009

[2] N. Homer et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genetics, 4, Aug. 2008

1

3

Aggregation

Re-identification of aggregated data is possible

Statistics from allele frequencies can be used to identify genetic trial participants [1]

Presence of an individual in a group can be determined by using allele frequencies and his DNA profile [2]

[1] R. Wang et al. “Learning Your Identity and Disease from Research Papers: Information Leaks in Genome Wide Association Study.” ACM CCS, 2009

[2] N. Homer et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genetics, 4, Aug. 2008

Enter Synthetic Genomic Data

Recombination model (Recomb)*

Restricted Boltzmann Machines (RBM)+

Generative Adversarial Networks (GAN)+

Wasserstein GAN (WGAN)^

Recombination RBM (Rec-RBM), new

Recombination GAN (Rec-GAN), new