Benchmarking Privatization for GWAS Genomic Data Sharing: Perturbed Allele Frequencies via DP vs. LDP-via-GRR Mechanisms

Eylül Sevinç Sevinç, Erkmen Erken

SUMMARY

In this project, we will benchmark three ways to privately share aggregate statistics from GWAS, a publicly available genomic data. Our goal is to release allele frequencies while reducing the risk of membership inference attacks and still preserving downstream utility. We will compare: (1) global DP by adding Laplace noise directly to allele counts per SNP, (2) a correlation-aware global DP method that adds Laplace noise to counts of common patterns within SNP blocks and then derives allele frequencies, and (3) local DP using generalized randomized response (GRR) where each individual reports few randomly chosen SNP locally with randomized response and the server debiases the aggregate to estimate frequencies. We will evaluate privacy with an LR-style membership inference test and evaluate utility by how well the privatized frequencies preserve association signals (chi-square statistics and significant-SNP overlap) compared to the non-private control data.

MOTIVATION AND PROBLEM STATEMENT

Genomic research has seen rapid progress in recent years and has become critical to modern medicine. Human genetic data has had a growing importance in medical and biological research for personalized medicine, drug discovery, the origin of life, and evolution. The genetic data also allows risk prediction by linking genetics variations to disease susceptibility, along with the development of safer therapies.[1]

At the same time, human genomic data is also a topic for many privacy related concerns because the data is very information-rich, difficult to truly anonymize, and highly identifying.[2] Unlike a password, a person can not change their genome if it is exposed. Even when obvious identifiers are removed, the genetic information can often be linked back to an individual by combining it with other publicly available information. Currently, the large scale genomic science depends on broad data sharing across institutions, and even summary statistics that do not directly release the genomic information can leak information about whether an individual is in the database.[3] Consequently, genomic data has a history of privacy attacks, including membership inference attacks from aggregate statistics.[4]

Before we describe our privacy setting and threat model, we will briefly define core genetics terms that will appear later in our project. A **genome-wide association study (GWAS)** compares two groups of people to find genomic positions where genetics variation is statistically associated with a trait.[5] The variants most commonly tested in GWAS are **single nucleotide polymorphisms (SNPs)**, which are single letter DNA differences (like A vs. G, or T vs. C) that are frequent enough in the population to analyze reliably. At any position with a SNP, an **allele** is one of the possible letters/versions, and an **allele frequency** summarizes how common a given allele is within a group of people, which makes it a compact way to publish results without releasing anyone's full genome.[5] In practice, GWAS outputs are often shared as summary statistics, and also a ranked list of the strongest signals. Researchers usually focus on the allele frequencies to drive follow-up experiments and biological interpretation, so protecting these information is the most important thing in privacy-preserving genomic data sharing.[4,5] Prior research has shown that the aggregate genetic releases can enable a likelihood-ratio (LR) style **membership inference attacks**, where an adversary tests whether an individual's genotype is more consistent with being inside the study cohort than outside it.[4] This problem introduces the question of how to release useful GWAS aggregates while minimizing the risk of membership inference and re-identification, and also retaining biological utility.

TECHNICAL APPROACH

Our project is intended to build a reproducible benchmarking pipeline for a GWAS sharing task, inspired by the iDASH privacy challenge.[6] This task concerns the utility-privacy trade-off when sharing aggregate results from human genomic data in GWAS form: privacy-preserving sharing of allele-frequencies.

1) Data resources and preprocessing

Datasets used will be from the Personal Genome Project (PGP)[7] and from HapMap CEU [8]. Data will be divided into three: case, control, and test. Case cohort data will be data of individuals which will be subject to privacy conservation when sharing, whereas the control data will be assumed to be public. We plan to get the case group from PGP, and control + test from HapMap CEU. We plan to use two genomic segments: chromosome 2 with ~311 SNPs spanning 5 Mb and chromosome 10 with ~610 SNPs spanning 1 Mb for evaluating allele-frequency perturbation under LR attacks. To be able to perform perturbation using ***Method 2***, we will preprocess the datasets to obtain haploblocks and their corresponding haplotypes. A **haplotype block** is a stretch of nearby SNPs that are usually inherited together, so haplotypes are common patterns on SNPs observed together.

2) DP-style perturbation for releasing allele frequencies

The goal is to create a perturbed case-group allele frequencies and benchmark the privacy–utility trade-off by evaluating the privacy risk under an LR-based membership inference test, and evaluate the utility by comparing the perturbed case group against the public control group.

We will implement and compare multiple perturbation designs:
Method 1 (Naive SNP-level Laplace perturbation for DP): We plan to treat case allele counts across N number of SNPs as a histogram query and add Laplace noise to each count. The sensitivity of releasing allele counts across N SNPs is 2N (one individual can change up to two alleles per SNV), which can be used in Laplace(0, f /ε) to determine the amount of noise that needs to be added.

Method 2 (Haplotype-block Laplace perturbation for DP): Instead of perturbing each SNV independently, we will use a mechanism that exploits publicly inferred haplotype blocks to preserve correlations. Haploblock structure and common haplotypes can be inferred from public reference data, independent of the sensitive case cohort.

For each block b, we will compute a histogram of case haplotype counts, add Laplace noise to these haplotype counts using a per-block budget ε_b, and will normalize within the block to obtain noisy haplotype frequencies. We will then derive each SNV's perturbed allele frequency in that block as a weighted sum over noisy haplotype frequencies. We plan to allocate the total privacy budget across blocks so that $\sum_b \varepsilon_b = \varepsilon$, using a rule of budget being proportional to the number of haplotypes in each block.

Method 3 (Local Differential Privacy-GRR via per-individual single-SNP reporting): We will implement a local differential privacy mechanism where each case individual reports only one randomly selected SNP (or a small fixed number of SNPs) from the target region. So, for every person, we randomly choose one SNP index from the list of N SNPs in the region. The person looks up their genotype at that SNP and encodes it as a "minor-allele count" with three possible values: 0 (no minor alleles), 1 (one minor allele), or 2 (two minor alleles). The person then applies generalized randomized response locally: they report the true value with a high probability that depends on the privacy budget epsilon, and otherwise they report one of the other two values at random. The person sends only the SNP index they were asked to report and their randomized response; the server never receives their full genotype vector. From these, we can "debias" the aggregated counts using the known reporting probabilities to obtain an unbiased estimate of the true genotype distribution at that SNP. From the estimated genotype distribution, we compute the estimated minor-allele frequency for that SNP (as the average number of minor alleles per person divided by two). This yields a locally private estimate of case-group allele frequencies across the region.

For privacy-risk computation, we will implement the likelihood ratio (LR) test, which computes a single LR score per individual by aggregating evidence across all SNPs (using the individual's genotype xj, the released/perturbed case minor-allele frequencies, and the corresponding control frequencies).

To quantify risk, we compute LR scores for (i) individuals in the case cohort and (ii) individuals in an external test cohort (not included in case or control). We will set an LR threshold based on the test cohort (e.g., so that only ~5% of test individuals exceed it), and will report LR power as the fraction of case individuals whose LR exceeds this threshold; lower power indicates lower membership inference risk.

$$\bar{L} = \sum_{j=1}^{m} (x_j log\frac{\hat{p}_j}{p_j} + (1 - x_j) log\frac{1 - \hat{p}_j}{1 - p_j}),$$

Figure 1: The likelihood ratio (LR) test, where xj is the allele type at SNV site j, m is the total number of SNVs, pj is the minor allele frequency of SNV j in the case group, and pj is the corresponding number for the reference group [9].

Utility computation (GWAS signal recovery) will be done with computing per-SNP chi-square association statistics comparing the case group to the public control group. Utility is measured by comparing the set of significant SNPs obtained using unperturbed case frequencies vs. the set obtained using perturbed case frequencies (both against the same public control).

For standard p-value cutoffs (e.g. 10^{-5}), we will compute overlap-based recovery (e.g., $|S_{true} \cap S_{priv}|$) and report discovery quality via precision/recall-style summaries (false positives vs false negatives) to conclude whether truly significant SNPs remain significant after perturbation.

5) Implementation details:

- **Language:** Python (NumPy/pandas for data handling; SciPy/statsmodels for chi-square tests; matplotlib for plots).
- **DP implementation:** We will implement Laplace noise and exponential-mechanism sampling ourselves using the formal definitions and algorithms as references.

DELIVERABLES

- Project report that describes the problem, methods implemented, DP/LDP via GRR mechanism, and experimental setup.
- Reproducible source code (Python) (data parsing → DP/LDP via GRR mechanism → evaluation → plots) with a clear README.
- DP/LDP via GRR mechanism implementations to make "allele frequencies" differentially private and locally differentially private.
- Evaluation plots and figures.

TIMELINE

Task:	**Dates:**
Data download + preprocessing pipeline	29.12.25-04.01.26, by Eylül, Erkmen
Implementing DP perturbation for allele frequencies using method 1	29.12.25-04.01.26, by Eylül, Erkmen
Implementing DP perturbation for allele frequencies using method 2	05.01.26-11.01.26, by Eylül, Erkmen
Implementing LDP-via-GRR for allele frequencies using method 3	12.01.26 - 18.01.26, by Eylül, Erkmen
Evaluation of utility and privacy + report writing + figures	19.01.26 - Deadline, by Eylül, Erkmen

REFERENCES

[1] Lin, Z., Owen, A. B., & Altman, R. B. (2004). Genomic research and human subject privacy. *Science*, *305*(5681), 183-183.

[2] Naveed, M., Ayday, E., Clayton, E. W., Fellay, J., Gunter, C. A., Hubaux, J. P., ... & Wang, X. (2015). Privacy in the genomic era. *ACM Computing Surveys (CSUR)*, *48*(1), 1-44.

[3] Bonomi, L., Huang, Y., & Ohno-Machado, L. (2020). Privacy challenges and research opportunities for genomic data sharing. *Nature genetics*, *52*(7), 646-654.

[4] Chen, J., Wang, W. H., & Shi, X. (2020). Differential privacy protection against membership inference attack on machine learning for genomic data. In *BIOCOMPUTING 2021: Proceedings of the Pacific Symposium* (pp. 26-37).

[5] Johnson, A., & Shmatikov, V. (2013, August). Privacy-preserving data exploration in genome-wide association studies. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1079-1087).

[6] Jiang, X., Zhao, Y., Wang, X., Malin, B., Wang, S., Ohno-Machado, L., & Tang, H. (2014). A community assessment of privacy preserving techniques for human genomes. *BMC medical informatics and decision making*, *14*(Suppl 1), S1.

[7] Personal Genome Project. (n.d.). *Personal Genome Project* [Data set]. Retrieved December 25, 2025, from https://www.personalgenomes.org

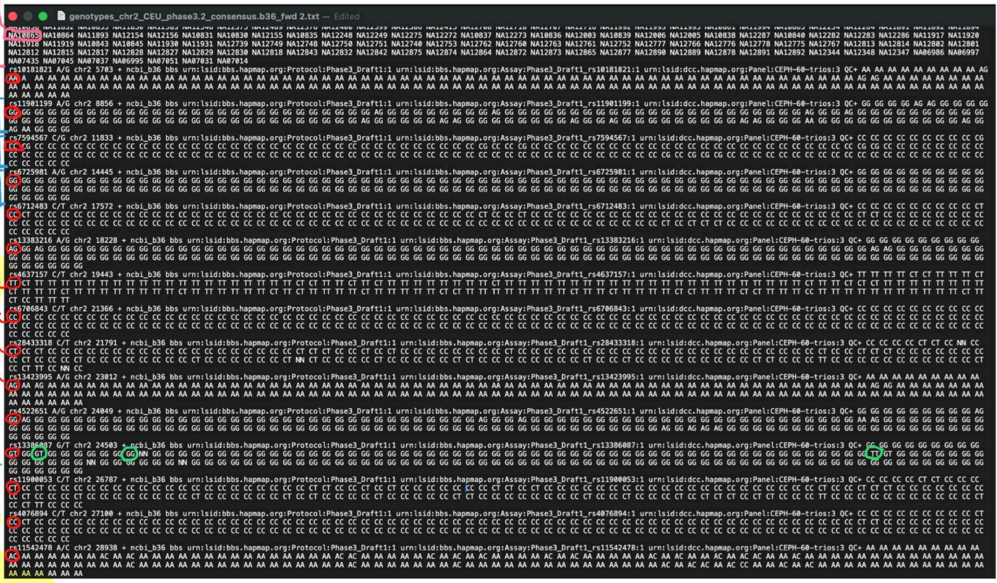
[8] National Center for Biotechnology Information. (n.d.). *HapMap data (FTP directory)* [Data set]. Retrieved December 25, 2025, from https://ftp.ncbi.nlm.nih.gov/hapmap/

[9] Sankararaman S, Obozinski G, Jordan MI, et al: Genomic privacy and limits of individual detection in a pool. Nat Genet 2009, 41:965-7[http://dx.doi.org/10.1038/ng.436], (accessed 18 Apr2014).

APPENDIX

Appendix A: Screenshot of SNP Count Data with Explanations

Each of these is a different person



genotypes_chr2_CEU_phase3.2_consensus.b36_fwd 2.txt — Edited

```
NA10865 NA10864 NA11893 NA12154 NA12156 NA10831 NA10830 NA12155 NA10835 NA12248 NA12249 NA12275 NA12272 NA10837 NA12273 NA10836 NA12003 NA10839 NA12006 NA12005 NA10838 NA12287 NA10840 NA12282 NA12283 NA12286 NA11917 NA11920
NA11918 NA11919 NA10843 NA10845 NA11930 NA11931 NA12739 NA12749 NA12748 NA12750 NA12751 NA12740 NA12753 NA12762 NA12760 NA12763 NA12761 NA12752 NA12777 NA12766 NA12776 NA12778 NA12775 NA12767 NA12813 NA12814 NA12802 NA12801
NA12812 NA12815 NA12817 NA12828 NA12827 NA12829 NA12830 NA12818 NA12843 NA12832 NA12842 NA12875 NA12874 NA12864 NA12872 NA12873 NA12865 NA12877 NA12890 NA12889 NA12878 NA12891 NA12892 NA12344 NA12348 NA12347 NA06986 NA06997
NA07435 NA07045 NA07037 NA06995 NA07051 NA07031 NA07014
 rs10181821 A/G chr2 5703 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs10181821:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ AA AA AA AA AA AA AA AA AA AG
AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AG AG AA AA AA AA AA AA AA AA
AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AA AA
 rs11901199 A/G chr2 8856 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs11901199:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ GG GG GG GG AG AG GG GG GG GG
GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG AG GG GG GG GG GG GG GG GG GG AG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG AG GG GG AG GG GG GG GG GG GG GG GG GG GG
GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG AG GG GG GG GG GG AG AG GG GG GG GG GG GG GG GG AG GG GG AG GG GG GG GG GG GG GG GG AG AG GG GG GG GG AA GG GG GG GG GG GG GG GG GG GG GG AG GG
AG AA GG GG GG
 rs7594567 C/G chr2 11833 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs7594567:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ CC CC CC CC CC CC CC CC CC CG
CC CG CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CG CC CC CG CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CG CG CC CC CC CC CC CC CC CC
CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CG CC CG CG CG CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
CC CC CC CC CC
 rs6725981 A/G chr2 14445 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs6725981:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ GG GG GG GG GG GG GG GG GG GG
GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG
GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG
GG GG GG GG GG
 rs6712483 C/T chr2 17572 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs6712483:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ CC CC CC CC CC CC CC CC CC CT
CC CT CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CT CC CC CT CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CT CT CC CC CC CC CC CC CC CC
CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CT CC CT CT CT CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
CC CC CC CC CC
 rs13383216 A/G chr2 18228 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs13383216:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ GG GG GG GG GG GG GG GG GG
AG GG AG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG AG AG GG GG GG GG GG GG GG GG
GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG
GG GG GG GG GG GG
 rs4637157 C/T chr2 19443 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs4637157:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ TT TT TT TT CT CT TT TT TT CT
TT CT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT CT CT TT TT CT TT CT TT TT TT TT TT TT TT TT CT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT TT CT TT TT CT CT CT TT TT TT TT TT TT TT TT
CT TT TT TT CT TT TT TT TT TT TT TT TT CT TT TT TT TT TT TT CT TT CT TT TT TT TT CT TT TT TT TT TT CT CT TT TT TT TT TT TT TT TT CT TT TT CT TT TT TT TT CT CT TT TT TT CT CT TT TT CC TT TT TT TT TT TT TT TT TT TT TT TT CT TT
CT CC TT TT TT
 rs6706843 C/T chr2 21366 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs6706843:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ CC CC CC CC CC CC CC CC CC CC
CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
CC CC CC CC CC
 rs28433318 C/T chr2 21791 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs28433318:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ CC CC CC CC CT CT CC NN CC
CT CC CT CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CT CT CT CC CC CT CC CT CC CC CC CC CC CC CC CC CT CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CT CC CC CT CT CT CC CC CC CC CC CC CC
CC CT CC CC CC CT CC CC CC CC CC CC CC CC CT CC CC CC CC CC CC CT NN CT CC CC CC CC CT CC CC CC CC CC CT CT CC CC CC CC CC CC CC CC CT CC CC CT CC CC CC CC CC CC CC CC CC CT CT CC CC CC TT CC CC CC CC CC CC CC CC CC CC CC CT
CC CT TT CC NN CC
 rs13423995 A/G chr2 23012 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs13423995:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ AA AA AA AA AA AA AA AA AA
AG AA AG AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AG AG AA AA AA AA AA AA AA AA
AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AA AA AA
 rs4522651 A/G chr2 24049 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs4522651:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ GG GG GG GG GG GG GG GG GG AG
GG AG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG AG GG GG AG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG AA AG GG GG GG GG GG GG GG GG
GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG AG GG AG AG AG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG
GG GG GG GG GG
 rs13386087 G/T chr2 24503 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs13386087:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ GG GG GG GG GG GG GG GG GG
GT GG GT GG GG GG GG GG GG GG NN GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG TT GT GG GG GG GG GG GG GG
GG GG GG GG GG GG NN GG GG GG GG GG GG NN GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG GG
GG GG GG GG GG GG
 rs11900053 C/T chr2 26787 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs11900053:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ CC CC CC CC CT CT CC CC CC
CT CC CT CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CT CT CC CC CT CC CT CC CC CC CC CC CC CC CT CT CC CT CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CT CC CC CT CT CT CC CC CC CC CC CC CC
CC CT CC CC CC CT CC CC CC CC CC CC CC CC CT CC CC CC CC CC CC CT CC CT CC CC CC CC CT CC CC CC CC CC CT CT CC CC CC CC CC CC CC CC CT CC CC CT CC CC CC CT CC CT CT CT CC CT CT CC CC CC TT CC CC CC CC CC CC CC CC CC CC CC CT
CC CT TT CC CC CC
 rs4076894 C/T chr2 27100 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs4076894:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ CC CC CC CC CC CC CC CC CC CT
CC CT CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CT CT CC CC CC CC CC CC CC CC
CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
CC CC CC CC CC
 rs11542478 A/C chr2 28938 + ncbi_b36 bbs urn:lsid:bbs.hapmap.org:Protocol:Phase3_Draft1:1 urn:lsid:bbs.hapmap.org:Assay:Phase3_Draft1_rs11542478:1 urn:lsid:dcc.hapmap.org:Panel:CEPH-60-trios:3 QC+ AA AA AA AA AA AA AA AA AA
AC AA AA AA AA AA AA AC AA AC AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AC AC AA AA AA AA AA AC AA AC AA AC AA AA AA AA AC AA AC AA AA AA AA AA AC AA AC AA AA AC AA AC AC AA AA AC AA AA AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AA AA AA AA AA AA AC AA AA AA AA AA AA AA AA AA AA AA AA AA AA AC AA AA AA AA AA AA AA AA AA AC AA AA AA AA AA AA AA AA AA AA AA AA AA AA AC AA AC AC AA AC CC AA AA AA AC AA AC AA AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AA AA AA
```

These are different SNPs

These are alleles of person 1 in different SNPs.

Each person contributes to the allele frequencies of each SNP.

For example, in this SNP, most people have GG. But, GT and TT are also observed. So each allele has a different frequency. GG has the highest frequency.

A group consisting of many SNPs is a haploblock.