

Trumpet plots: Visualizing The Relationship Between Allele Frequency And Effect Size In Genetic Association Studies

Lucia Corte^{1,2}, Lathan Liou¹, Paul F. O'Reilly¹, Judit García-González¹

¹Department of Genetics and Genomic Sciences, Icahn School of Medicine, Mount Sinai, New York City, NY 10029, USA

²Center for Excellence in Youth Education, Icahn School of Medicine, Mount Sinai, New York City, NY 10029, USA

Abstract

Recent advances in genome-wide association study (GWAS) and sequencing studies have shown that the genetic architecture of complex diseases and traits involves a combination of rare and common genetic variants, distributed throughout the genome. One way to better understand this architecture is to visualize genetic associations across a wide range of allele frequencies. However, there is currently no standardized or consistent graphical representation for effectively illustrating these results.

Here we propose a standardized approach for visualizing the effect size of risk variants across the allele frequency spectrum. The proposed plots have a distinctive trumpet shape, with the majority of variants having low frequency and small effects, while a small number of variants have higher frequency and larger effects. These plots, which we call 'trumpet plots', can help to provide new and valuable insights into the genetic basis of traits and diseases, and can help prioritize efforts to discover new risk variants. To demonstrate the utility of trumpet plots in illustrating the relationship between the number of variants, their frequency, and the magnitude of their effects in shaping the genetic architecture of complex diseases and traits, we generated trumpet plots for more than one hundred traits in the UK Biobank. To facilitate their broader use, we have developed an R package 'TrumpetPlots' and R Shiny application, available at <https://juditgg.shinyapps.io/shinytrumpets/>, that allows users to explore these results and submit their own data.

Research Area: Genetics and Genomics

Classifications: Genetics, Bioinformatics, Biomedical Science, Statistics and Probability

STATEMENT OF NEED

Visualizations are powerful tools that have helped the field of genetics to better understand and communicate complex findings. By using visual aids like Manhattan and Volcano plots, genetic variants identified through genome-wide association studies can be more easily pinpointed. With the advancement of genome-wide association and sequencing studies, a mounting number of significant genetic variants, both common and rare, are being discovered. To better understand the relationship between these variants, combining these findings into single visualizations help to observe the relationship between effect size and allele frequency, providing a clearer picture of the genetic architecture of different traits and diseases. However, there is currently no consistent method for illustrating these results. In this paper, we propose a standardized approach for visualizing the effect size of risk variants across the allele frequency spectrum, generate plots for over a hundred traits in the UK Biobank, and provide to the field a R package and R Shiny application to explore their own results.

Availability of supporting source code and requirements

- Project name:
 - R package available in project 'TrumpetPlots'
<https://gitlab.com/JuditGG/trumpetplots>
 - R shiny app and analyses in the UK Biobank available in project '*freq_or_plots*'
https://gitlab.com/JuditGG/freq_or_plots
- Project home page: <https://juditgg.shinyapps.io/shinytrumpets/>
- Operating system(s): Platform independent.
- Programming language: R
- RRID: Not applicable
- License: MIT

Data Availability and implementation

All data used in this manuscript is publicly available. Rare variant associations are available in supplementary data table 2 of the original publication (DOI: <https://doi.org/10.1038/s41586-021-04103-z>), GWAS summary statistics are available in the website <https://www.nealelab.is/uk-biobank/>.

The code is freely available at https://gitlab.com/JuditGG/freq_or_plots (UK Biobank analyses), <https://gitlab.com/JuditGG/trumpetplots> (R package with test data) and <https://juditgg.shinyapps.io/shinytrumpets/> (R Shiny application).

List of abbreviations

GWAS: genome-wide association study

LD: linkage disequilibrium

Log: logarithm or logarithmic

Competing interests

The authors declare no competing interests

Authors' contributions

LC: Data curation, Formal Analysis, Investigation, Writing – original draft.

LL: Software, Validation, Visualization.

PFO: Conceptualization, Funding acquisition, Formal analysis, Supervision, Writing – review & editing.

JGG: Conceptualization, Data curation, Formal Analysis, Investigation, Software, Validation, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Acknowledgements

We thank the participants in the UK Biobank and the scientists involved in the construction of this resource. This work was supported by a grant from the National Institute of Health (R01MH122866) to PFO, by a 2022 NARSAD Young Investigator Grant (Number 30749) by the Brain & Behavior Research Foundation to JGG, and through the computational resources and staff expertise provided by Scientific Computing and the Data Ark (Data Commons) teams at the Icahn School of Medicine at Mount Sinai.

We would like to thank Dr Shea Andrews for helpful discussions on several aspects of the project. We would also like to express our gratitude to the Center for Excellence in Youth Education (CEYE) program for their support and training, which enabled us to carry out this research. Without the invaluable assistance and dedication of CEYE staff, this project would not have been possible.

Authors' information

Lucia Corte

Department of Genetics and Genomic Sciences, Icahn School of Medicine, Mount Sinai, New York City, NY 10029, USA and Center for Excellence in Youth Education, Icahn School of Medicine, Mount Sinai, New York City, NY 10029, USA

Email: lc105@wellesley.edu

ORCID ID: <https://orcid.org/0009-0005-0963-4880>

Lathan Liou

Department of Genetics and Genomic Sciences, Icahn School of Medicine, Mount Sinai, New York City, NY 10029, USA

Email: lathan.liou@icahn.mssm.edu

ORCID ID: <https://orcid.org/0000-0002-8066-5947>

Paul F. O'Reilly

Department of Genetics and Genomic Sciences, Icahn School of Medicine, Mount Sinai, New York City, NY 10029, USA.

email: paul.oreilly@mssm.edu

ORCID ID: <https://orcid.org/0000-0001-7515-0845>

Judit García-González,

Department of Genetics and Genomic Sciences, Icahn School of Medicine, Mount Sinai, New York City, NY 10029, USA.

email: judit.garciagonzalez@mssm.edu,

ORCID ID: <https://orcid.org/0000-0001-6245-740X>

1 **Background**

2 Results visualization is an essential tool for interpreting complex data. By using visual
3 representations such as graphs, charts, and plots, researchers can quickly identify patterns,
4 trends, and outliers that may not be detected in tables of raw data. Visualizations help to gain a
5 more intuitive and comprehensive understanding of the data, especially when dealing with large
6 and complex datasets. Furthermore, visualizing results can facilitate communication of research
7 findings to a broader audience, including non-experts [1]. Visual representations are often more
8 accessible and engaging, making it easier for others to understand and appreciate the
9 significance of the research conducted.

10 In the field of genetics, the use of visualizations has revolutionized the interpretation and
11 communication of research findings. Over the past two decades, visualizations such as
12 Manhattan plots [2], which display the results of genome-wide association studies (GWAS),
13 software packages like haploview [3] that analyze and visualize linkage disequilibrium (LD)
14 patterns of GWAS associated loci, and Volcano plots that assess patterns of differential gene
15 expression [4], have all played crucial roles in illustrating and sharing the key summaries of data
16 that have advanced the field. These and other visualizations [5] have allowed the genetics field
17 to more easily identify candidate causal variants, relevant genes, and potential outliers that may
18 not be apparent in tables of GWAS summary statistics or differential expression results.

19 More recent advances in GWAS and sequencing studies [6]–[8] have resulted in the
20 identification of an increasing number of significant genetic variants, including both common [8]
21 and rare variants [5,6]. Researchers are now starting to combine these findings into single
22 visualizations to observe the relationship between effect size and allele frequency across the full
23 range of significantly associated variants. Given that the number of risk-conferring variants, their
24 frequency in a population, and their effect size can vary across different diseases and traits,
25 using these plots can provide a better and instantaneous understanding of their relative genetic
26 architecture. Recent studies on height [9], schizophrenia [10] and coronary artery disease [11],
27 [12] have already included this full range as the main figure, highlighting the utility of this type of
28 visualization. However, a formal and consistent method for illustrating these results has not yet
29 emerged.

30 The aim of this work is to introduce an R package and R shiny application to illustrate the
31 distribution of risk variants across a wide range of allele frequencies. We term the resulting plots
32 ‘trumplet plots’, due to their trumpet-like shape. To demonstrate their utility, we generated

33 trumpet plots for over one hundred continuous traits available in the UK Biobank [13], illustrating
34 the distribution of risk variants across an effect allele frequency range between 0.00001 and 1.
35 These plots are available at <https://juditgg.shinyapps.io/shinytrumpets/> and we illustrate a single
36 trumpet plot combining the results of all of these in **Figure 1**.

37 We propose that trumpet plots are valuable representations of genetic associations across the
38 full allele frequency spectrum that can help researchers to better understand the genetic
39 architecture of traits and diseases and potentially aid in study design and the prioritisation of
40 investments to discover new variants that contribute to disease.

41

42 **Methods**

43 In the following sections, we will explain the various decisions we made when creating trumpet
44 plots. These decisions include selecting the appropriate scale to represent allele frequencies,
45 deciding whether to use the full GWAS summary statistics or independent GWAS variants,
46 addressing issues related to the reporting of rare variant association tests, determining whether
47 to include power curves, and considering the effect size sign of the variants included. By
48 carefully considering each of these factors, we aimed to create informative and visually
49 appealing trumpet plots that illustrate the effect size of genetic associations across a wide range
50 of allele frequencies.

51

52 *Using the logarithmic vs linear scale to represent allele frequencies*

53 In the representation of allele frequencies, the range of values can vary greatly between the
54 smallest and largest frequency. When these associations are plotted on a linear scale, the rare
55 variants can be obscured or difficult to distinguish. To address this issue, we recommend using
56 a logarithmic (log) scale (we use log base 10) for allele frequencies in trumpet plots. Compared
57 to a linear scale, the log scale uses increments that represent a relative increase or decrease,
58 rather than a fixed value increase or decrease. The log scale compresses the allele frequencies
59 that are most common, which results in a more even distribution of values across the scale. This
60 scale of visualization facilitates the identification of important patterns and trends.

61

62 *Identification of independent significant variants to enhance the interpretation of trumpet plots*

63 Genetic association studies involve testing up to millions of genetic variants for their association
64 with a particular trait or disease. However, many of these variants are correlated with each other
65 due to their physical proximity on the genome, which is known as linkage disequilibrium (LD).

66 This means that many variants nearby to causal variants often show significant associations
67 with the trait under study, due only to their correlation rather than any biological involvement
68 with the trait.

69
70 Two methods that can be used to identify independent significant variants in a GWAS are
71 clumping and conditional analysis [14], [15]. Clumping involves selecting a subset of
72 independent significant variants by choosing a lead variant for each LD cluster and then
73 discarding all other variants in that cluster. The lead variant is typically the one with the
74 strongest association with the trait of interest. Conditional analysis, on the other hand, involves
75 identifying independent significant variants after performing a joint analysis of multiple variants
76 together. In this approach, the effect of one variant is *conditioned* on the effect of other variants,
77 meaning that the association between the trait of interest and one variant is evaluated after
78 accounting for the effect of other variants. This can either be performed as a joint analysis (e.g.
79 a regression) with multiple variants in one model or else as an iterative process, where the lead
80 variant and each other variant in the region are tested jointly one-by-one. By considering the
81 effects of multiple variants jointly in single models, conditional analysis helps to identify
82 independent signals more accurately than clumping, which relies only on correlations between
83 variants to indirectly infer independent signals.

84
85 Since the correlation between variants can make it challenging to interpret trumpet plots, we
86 recommend plotting only independent significant variants. These are variants that represent
87 distinct genetic associations with the trait of interest.

88 89 *Variant-level associations for low allele frequencies*

90 While GWAS are a valuable tool for detecting common genetic variants associated with
91 complex traits or diseases, they have limited power to identify associations with rare variants.
92 To address this limitation, sequencing studies [16]–[18], such as whole exome sequencing and
93 whole genome sequencing, have been used to detect variant-level associations with low allele
94 frequencies.

95
96 To further enhance the statistical power of rare-variant associations, a commonly used strategy
97 is to aggregate the rare variants detected into functional genetic units, such as genes, and
98 perform collective variation analysis (e.g. gene burden tests [19]–[21]). However, we found that
99 the reporting of rare variant association analyses varies across studies. Some studies report

100 results at the variant level, while others only report results of the functional unit in which rare
101 variants were aggregated. This makes comparing results across independent studies, and
102 determining the functional significance of specific variants, challenging.

103
104 We therefore encourage the reporting of results at the variant level, so that they can be included
105 in visualizations of allele frequency in relation to effect size, such as trumpet plots, aiding study
106 comparisons and variant interpretation. Although the analysis of genetic variants at low
107 frequency is expected to improve with the availability of biobank-scale samples and the
108 development of new methods to reduce biases in association tests, caution should still be
109 exercised when inspecting associations with rare variants, as these can suffer from instability
110 and low power, particularly in the relation to binary traits.

111
112 Statistical power considerations

113 GWAS require careful consideration of statistical power [22], which depends on various factors,
114 including the allele frequency and effect size of variants, represented by the x- and y-axis of
115 trumpet plots, respectively. Common variants, usually defined as having allele frequency greater
116 than 1%, tend to have higher power in association studies because common causal variants are
117 more likely to be present in the sample (either genotyped or imputed), and because their
118 relatively balanced number of alleles is akin to having a larger sample size. Variants with larger
119 effect sizes have higher power because their effects are further from the null hypothesis of zero
120 effect.

121
122 We therefore recommend incorporating power curves into trumpet plots, since they visually
123 represent the statistical power across the allele frequency spectrum for a given sample size and
124 effect size [23]. Moreover, power curves can aid in identifying parts of the association testing
125 space in which the power to detect significant associations is low.

126
127 Two alternative approaches to illustrate the joint distribution of allele frequency and effect size

128 One approach to illustrate the relationship between allele frequency and effect size is to plot
129 only positive effects i.e the allele effect for each variant that increases the value of the
130 phenotype. In this case, the effect sizes are always positive, and both the allele and sign of the
131 association regression coefficients (betas or Odds Ratios, ORs) need to be flipped (to the other
132 allele) if they are reported as negative to ensure that the effect size is greater than zero. If this
133 'flipping' is required, then the allele frequency of the other allele should be reported, which will

134 be 1 minus the original allele frequency. In this case, the allele frequencies of the plot range
135 from 0 to 1.

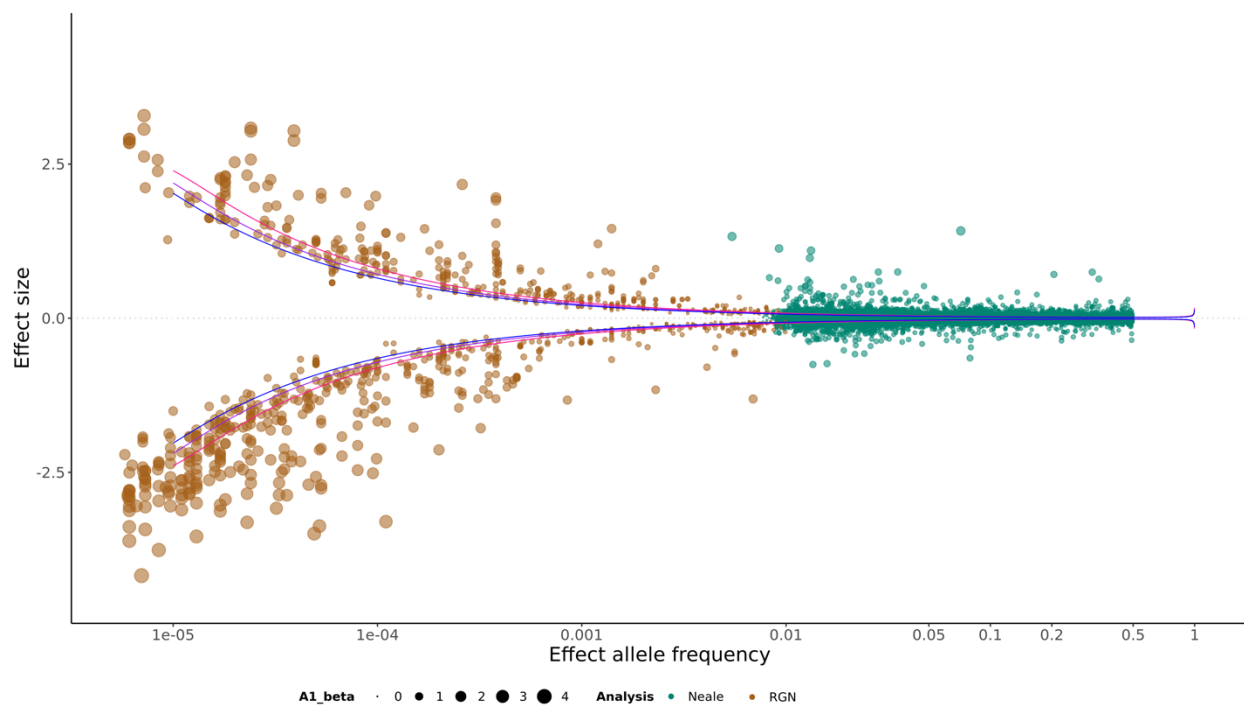
136
137 The other approach, which we recommend, allows for both positive (risk allele in the context of
138 disease phenotypes) and negative (protective allele in the context of disease phenotypes) effect
139 sizes and always corresponds to the minor allele. In this case, the effect size of the allele can
140 have either a positive or negative value, and the allele frequencies of the plot range from 0 to
141 0.5.

142
143 *Practical example: Generating trumpet plots for 129 traits in the UK Biobank*

144 We examined all continuous UK Biobank traits with available GWAS analyses performed by
145 Benjamin Neale's group (<https://www.nealelab.is/uk-biobank/>) and searched whether rare
146 variant associations were available for the same trait (by UK Biobank Field ID) in the exome
147 sequencing analysis performed by the Regeneron team [13].

148
149 Common variant associations were extracted from the Neale's group GWAS summary statistics.
150 For each GWAS, we extracted the independent variants using COJO GCTA (--cojo-slc
151 command), and a random subset of 4,000 unrelated individuals with European ancestry from
152 the UK Biobank as LD reference panel. We selected independent variants with minor allele
153 frequency >0.01 and association P-value < 5×10^{-8} within a 100Kb window.

154
155 Rare variants association results were extracted from the supplementary data table (SD2) of the
156 Regeneron study [13]. This study reports results for both burden tests (which typically
157 aggregated variants and indels) and individual rare variant level tests. To ensure that effect
158 sizes reported in our analyses corresponded to individual rare variants, we extracted only
159 results for 'singleton variants' with predicted loss of function - including stop-gain, frameshift,
160 stop-lost, start-lost and essential splice variants - and deleterious missense variants. Interactive
161 plots illustrating the relationship between allele frequency (x-axis) and odds ratio (y-axis) were
162 plotted using our R package TrumpetPlots (<https://gitlab.com/JuditGG/trumpetplots>), which
163 uses the R packages data.table, ggplot and ggplotly (**Figure 1**).



164
165 **Figure 1.** Distribution of allele frequencies and effect sizes for genetic associations across 129
166 continuous traits from the UK Biobank. Power curves for statistical power of 0.5 (blue), 0.7 (purple)
167 and 0.9 (pink) were constructed for the median GWAS sample size (N=351,550).

168
169 R Shiny application

170 We developed a user-friendly web application called *shinytrumpets* to visualize trumpet plots for
171 our UK Biobank results, as well as any other genetic association results that can be uploaded
172 by the user. With *shinytrumpets*, researchers with no knowledge of R programming can easily
173 upload and visualize their own datasets.

174
175 If a user uploads their own results, *shinytrumpets* prompts them to upload the input data files
176 and specify the sample size used for the study, such as the GWAS sample size. This
177 information is used to perform power calculations for the visualization. *Shinytrumpets* offers an
178 intuitive interface for users to explore and download trumpet plots.

179
180 Discussion

181 Visual representations of genetics and genomics results, such as Manhattan [2], Q-Q plots [2],
182 haploview [3] or Volcano plots [4], have been helpful in interpreting research findings and

183 identifying patterns, trends, and outliers that may not be easily apparent in tables of raw data.
184 These visualizations have revolutionized the interpretation and communication of research
185 findings relating to the identification of GWAS associated loci, putatively causal genes, and
186 potential outliers.

187
188 In this manuscript, we introduce a new R package and shiny application to illustrate the
189 distribution of risk variants across a wide range of allele frequencies, which we coin 'trumpet
190 plots'. We illustrate the distribution of variant effect sizes across the allele frequency range (from
191 0.00001 to 1) for over 100 continuous traits available in the UK Biobank, and propose that these
192 plots are valuable representations of genetic associations that can help researchers better
193 understand the genetic architecture of traits and diseases and prioritize certain study designs
194 (e.g. sequencing or GWAS) to discover new variants that contribute to disease.

195
196 One important consideration when interpreting the trumpet plots we constructed for the UK
197 Biobank is that they only represent individuals of European ancestry. The relationship between
198 effect size and allele frequency can be affected by population genetic differences [24], [25] and
199 as such, one interesting application of trumpet plots could be to compare the joint distribution of
200 allele frequencies and effect sizes across different ancestries to identify similarities and
201 differences for further investigation. Insights about the similarities and differences across
202 populations in the relationship between effect size and allele frequency could have important
203 implications for disease risk prediction and prevention strategies.

204
205 In conclusion, we emphasize the significance of data visualization in the genetics field and
206 present a novel R package and shiny application for visualizing the relationship between allele
207 frequency and effect size in association studies. We hope that the proposed 'trumpet plots' will
208 provide a valuable representation of genetic associations and will enhance the interpretation of
209 the association results across the allele frequency spectrum.

210

References

- [1] C. O. Iyegbe and P. F. O'Reilly, "Genetic origins of schizophrenia find common ground," *Nature*, vol. 604, no. 7906, pp. 433–435, Apr. 2022, doi: 10.1038/d41586-022-00773-5.
- [2] S. D. Turner, "qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots," *Journal of Open Source Software*, vol. 3, no. 25, p. 731, May 2018, doi: 10.21105/joss.00731.
- [3] J. C. Barrett, B. Fry, J. Maller, and M. J. Daly, "Haploview: analysis and visualization of LD and haplotype maps," *Bioinformatics*, vol. 21, no. 2, pp. 263–265, Jan. 2005, doi: 10.1093/bioinformatics/bth457.
- [4] W. Li, "Volcano plots in analyzing differential expression with mRNA microarrays," *J. Bioinform. Comput. Biol.*, vol. 10, no. 06, p. 1231003, Dec. 2012, doi: 10.1142/S0219720012310038.
- [5] A. P. Boughton *et al.*, "LocusZoom.js: interactive and embeddable visualization of genetic association study results," *Bioinformatics*, vol. 37, no. 18, pp. 3017–3018, Sep. 2021, doi: 10.1093/bioinformatics/btab186.
- [6] W. Zhou *et al.*, "SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests," *Nat Genet*, vol. 54, no. 10, pp. 1466–1469, Oct. 2022, doi: 10.1038/s41588-022-01178-w.
- [7] J. Mbatchou *et al.*, "Computationally efficient whole-genome regression for quantitative and binary traits," *Nat Genet*, vol. 53, no. 7, Art. no. 7, Jul. 2021, doi: 10.1038/s41588-021-00870-7.
- [8] C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee, "Second-generation PLINK: rising to the challenge of larger and richer datasets," *Gigascience*, vol. 4, p. 7, 2015, doi: 10.1186/s13742-015-0047-8.
- [9] L. Yengo *et al.*, "A saturated map of common genetic variants associated with human height," *Nature*, vol. 610, no. 7933, Art. no. 7933, Oct. 2022, doi: 10.1038/s41586-022-05275-y.
- [10] V. Trubetskoy *et al.*, "Mapping genomic loci implicates genes and synaptic biology in schizophrenia," *Nature*, vol. 604, no. 7906, Art. no. 7906, Apr. 2022, doi: 10.1038/s41586-022-04434-5.
- [11] K. G. Aragam *et al.*, "Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants," *Nat Genet*, vol. 54, no. 12, Art. no. 12, Dec. 2022, doi: 10.1038/s41588-022-01233-6.

- [12] S. Koyama *et al.*, “Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease,” *Nat Genet*, vol. 52, no. 11, Art. no. 11, Nov. 2020, doi: 10.1038/s41588-020-0705-3.
- [13] J. D. Backman *et al.*, “Exome sequencing and analysis of 454,787 UK Biobank participants,” *Nature*, vol. 599, no. 7886, Art. no. 7886, Nov. 2021, doi: 10.1038/s41586-021-04103-z.
- [14] J. Yang *et al.*, “Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits,” *Nat Genet*, vol. 44, no. 4, pp. 369–375, S1-3, Mar. 2012, doi: 10.1038/ng.2213.
- [15] J. Yang, S. H. Lee, M. E. Goddard, and P. M. Visscher, “GCTA: A Tool for Genome-wide Complex Trait Analysis,” *Am J Hum Genet*, vol. 88, no. 1, pp. 76–82, Jan. 2011, doi: 10.1016/j.ajhg.2010.11.011.
- [16] L. Leitsalu *et al.*, “Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu,” *International Journal of Epidemiology*, vol. 44, no. 4, pp. 1137–1147, Aug. 2015, doi: 10.1093/ije/dyt268.
- [17] J. M. Gaziano *et al.*, “Million Veteran Program: A mega-biobank to study genetic influences on health and disease,” *J Clin Epidemiol*, vol. 70, pp. 214–223, Feb. 2016, doi: 10.1016/j.jclinepi.2015.09.016.
- [18] C. Turnbull *et al.*, “The 100 000 Genomes Project: bringing whole genome sequencing to the NHS,” *BMJ*, vol. 361, p. k1687, Apr. 2018, doi: 10.1136/bmj.k1687.
- [19] S. Lee *et al.*, “Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies,” *Am J Hum Genet*, vol. 91, no. 2, pp. 224–237, Aug. 2012, doi: 10.1016/j.ajhg.2012.06.007.
- [20] S. Feng, D. Liu, X. Zhan, M. K. Wing, and G. R. Abecasis, “RAREMETAL: fast and powerful meta-analysis for rare variants,” *Bioinformatics*, vol. 30, no. 19, pp. 2828–2829, Oct. 2014, doi: 10.1093/bioinformatics/btu367.
- [21] B. M. Neale *et al.*, “Testing for an Unusual Distribution of Rare Variants,” *PLOS Genetics*, vol. 7, no. 3, p. e1001322, Mar. 2011, doi: 10.1371/journal.pgen.1001322.
- [22] E. P. Hong and J. W. Park, “Sample Size and Statistical Power Calculation in Genetic Association Studies,” *Genomics Inf*, vol. 10, no. 2, pp. 117–122, Jun. 2012, doi: 10.5808/GI.2012.10.2.117.
- [23] P. C. Sham and S. M. Purcell, “Statistical power and significance testing in large-scale genetic studies,” *Nat Rev Genet*, vol. 15, no. 5, Art. no. 5, May 2014, doi: 10.1038/nrg3706.

- [24] H. Shi *et al.*, “Population-specific causal disease effect sizes in functionally important regions impacted by selection,” *Nat Commun*, vol. 12, no. 1, Art. no. 1, Feb. 2021, doi: 10.1038/s41467-021-21286-1.
- [25] A. R. Martin *et al.*, “Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations,” *Am J Hum Genet*, vol. 100, no. 4, pp. 635–649, Apr. 2017, doi: 10.1016/j.ajhg.2017.03.004.