Center for Genomics Research and Innovation

H3ABioNet

Genome-Wide Association Study (GWAS):
Computational Resources & Workflow

Segun Fatumo *PhD*

Associate Professor, Department of Non-Communicable Disease Epidemiology, LSHTM, UK
Group Leader, The African Computational Genomics Group, MRC/UVRI and LSHTM, Uganda

segun.fatumo@lshtm.ac.uk

Intended Learning Outcomes

By the end of the session, students will be able to

- Recognise necessary computation tools for performing genome-wide association study (GWAS)
- Explain the workflow in conducting GWAS
- Discuss Ethical steps necessary for sample collection for genomic study

GWAS: From the beginning to the end

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Genomewide Association Study of Severe Covid-19 with Respiratory Failure

The Severe Covid-19 GWAS Group*

ABSTRACT

BACKGROUND

There is considerable variation in disease behavior among patients infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes coronavirus disease 2019 (Covid-19). Genomewide association analysis may allow for the identification of potential genetic factors involved in the development of Covid-19.

METHODS

We conducted a genomewide association study involving 1980 patients with Covid-19 and severe disease (defined as respiratory failure) at seven hospitals in the Italian and Spanish epicenters of the SARS-CoV-2 pandemic in Europe. After quality control and the exclusion of population outliers, 835 patients and 1255 control participants from Italy and 775 patients and 950 control participants from Spain were included in the final analysis. In total, we analyzed 8,582,968 single-nucleotide polymorphisms and conducted a meta-analysis of the two case–control panels.

RESULTS

We detected cross-replicating associations with rs11385942 at locus 3p21.31 and with rs657152 at locus 9q34.2, which were significant at the genomewide level ($P<5\times10^{-8}$) in the meta-analysis of the two case–control panels (odds ratio, 1.77; 95% confidence interval [CI], 1.48 to 2.11; $P=1.15\times10^{-10}$; and odds ratio, 1.32; 95% CI, 1.20 to 1.47; $P=4.95\times10^{-8}$, respectively). At locus 3p21.31, the association signal spanned the genes *SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6* and *XCR1*. The association signal at locus 9q34.2 coincided with the *ABO* blood group locus; in this cohort, a blood-group–specific analysis showed a higher risk in blood group A than in other blood groups (odds ratio, 1.45; 95% CI, 1.20 to 1.75; $P=1.48\times10^{-4}$) and a protective effect in blood group O as compared with other blood groups (odds ratio, 0.65; 95% CI, 0.53 to 0.79; $P=1.06\times10^{-5}$).

CONCLUSIONS

We identified a 3p21.31 gene cluster as a genetic susceptibility locus in patients with Covid-19 with respiratory failure and confirmed a potential involvement of the ABO blood-group system. (Funded by Stein Erik Hagen and others.)

The authors' full names, academic degrees, and affiliations are listed in the Appendix. Address reprint requests to Dr. Franke at the Institute of Clinical Molecular Biology and University Hospital of Schleswig-Holstein, Christian-Albrechts-University, Rosalind-Franklin-Str. 12, D-24105 Kiel, Germany, or at a.franke@mucosa.de; or to Dr. Karlsen at the Division of Surgery, Inflammatory Diseases, and Transplantation, Oslo University Hospital Rikshospitalet and University of Oslo, Postboks 4950 Nydalen, N-0424 Oslo, Norway, or at t.h.karlsen@medisin.uio.no.

*Dr. Franke serves as an author on behalf of the Covid-19 Host Genetics Initiative; members of the Initiative are listed in Supplementary Appendix 1, available at NEJM.org.

Dr. Ellinghaus and Ms. Degenhardt and Drs. Valenti, Franke, and Karlsen contributed equally to this article.

This article was published on June 17, 2020, at NEJM.org.

N Engl J Med 2020;383:1522-34.
DOI: 10.1056/NEJMoa2020283

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-04951-w OPEN

Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes

Angli Xue[1], Yang Wu[1], Zhihong Zhu[1], Futao Zhang[1], Kathryn E. Kemper[1], Zhili Zheng[1,2], Loic Yengo[1], Luke R. Lloyd-Jones[1], Julia Sidorenko[1,3], Yeda Wu[1], eQTLGen Consortium#, Allan F. McRae[1,4], Peter M. Visscher[1,4], Jian Zeng[1] & Jian Yang[1,2,4]

Type 2 diabetes (T2D) is a very common disease in humans. Here we conduct a meta-analysis of genome-wide association studies (GWAS) with ~16 million genetic variants in 62,892 T2D cases and 596,424 controls of European ancestry. We identify 139 common and 4 rare variants associated with T2D, 42 of which (39 common and 3 rare variants) are independent of the known variants. Integration of the gene expression data from blood ($n=14,115$ and 2765) with the GWAS results identifies 33 putative functional genes for T2D, 3 of which were targeted by approved drugs. A further integration of DNA methylation ($n=1980$) and epigenomic annotation data highlight 3 genes (*CAMK1D*, *TP53INP1*, and *ATP5G1*) with plausible regulatory mechanisms, whereby a genetic variant exerts an effect on T2D through epigenetic regulation of gene expression. Our study uncovers additional loci, proposes putative genetic regulatory mechanisms for T2D, and provides evidence of purifying selection for T2D-associated variants.

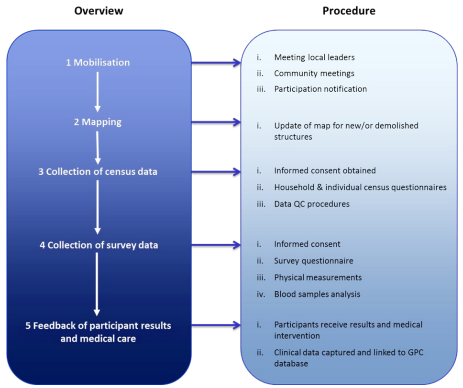
[1] Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia. [2] The Eye Hospital, School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China. [3] Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu 51010, Estonia. [4] Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia. These authors contributed equally: Angli Xue, Yang Wu. These authors jointly supervised this work: Jian Zeng, Jian Yang. #A full list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to J.Z. (email: j.zeng@uq.edu.au) or to J.Y. (email: jian.yang@uq.edu.au)

Computational Resources

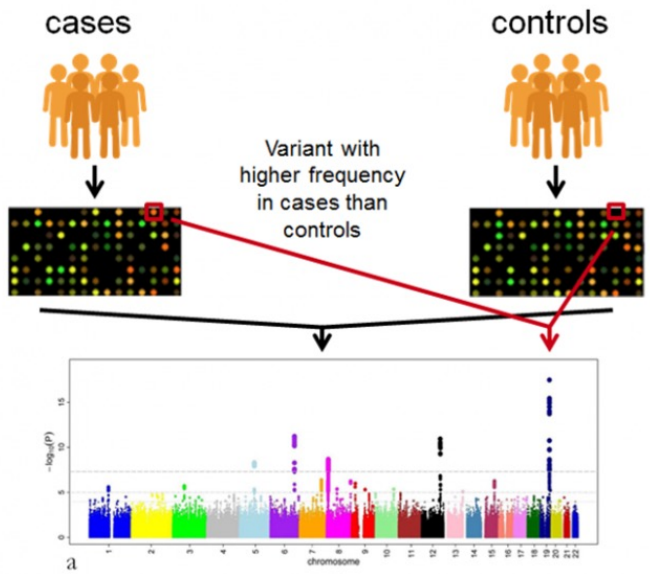
Unix (Bash, AWK, Sed, Grep, etc)

- Tools
  - Genome Studio
  - genCall
  - zCall
- Tools
  - **Plink**
  - EigenSoft
  - R
- Tools
  - **Plink**
  - GEMMA
  - R
- Tools
  - Impute2
  - PBWT
  - Online
- Tools
  - METAL
  - Metasoft
  - MANTRA
- Tools
  - FINEMAP
  - BFMAP
  - SuSiE
- Tools
  - **Plink**
  - PRSice
  - LDpred

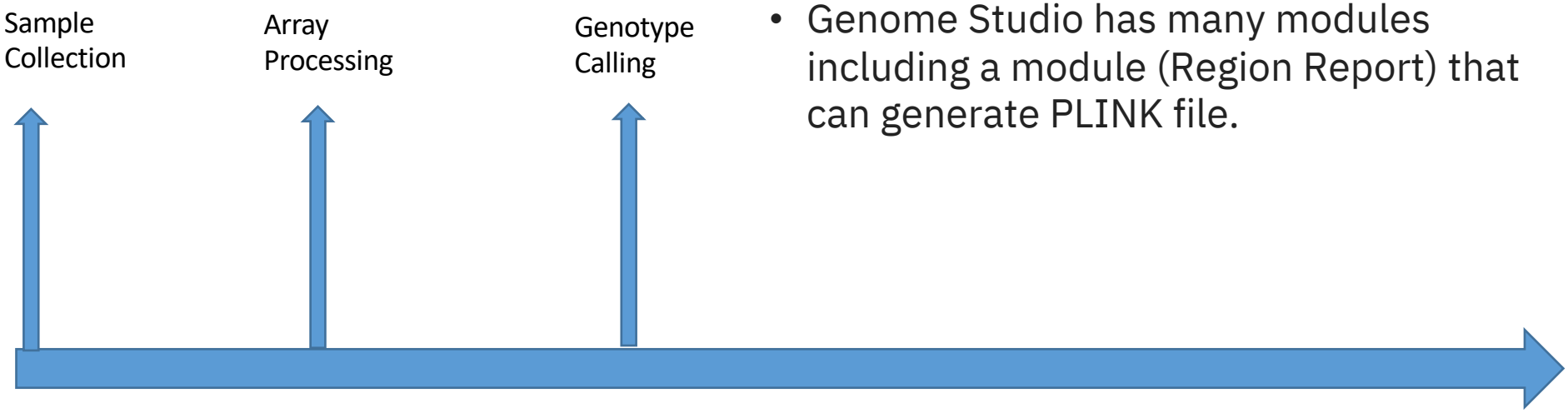
Sample Collection | Array Processing | Genotype Calling | Quality Control | Association Analysis (GWAS) | Genotype Imputation | Meta-analysis &Replication | In-silico Fine-mapping | Polygenic Risk score



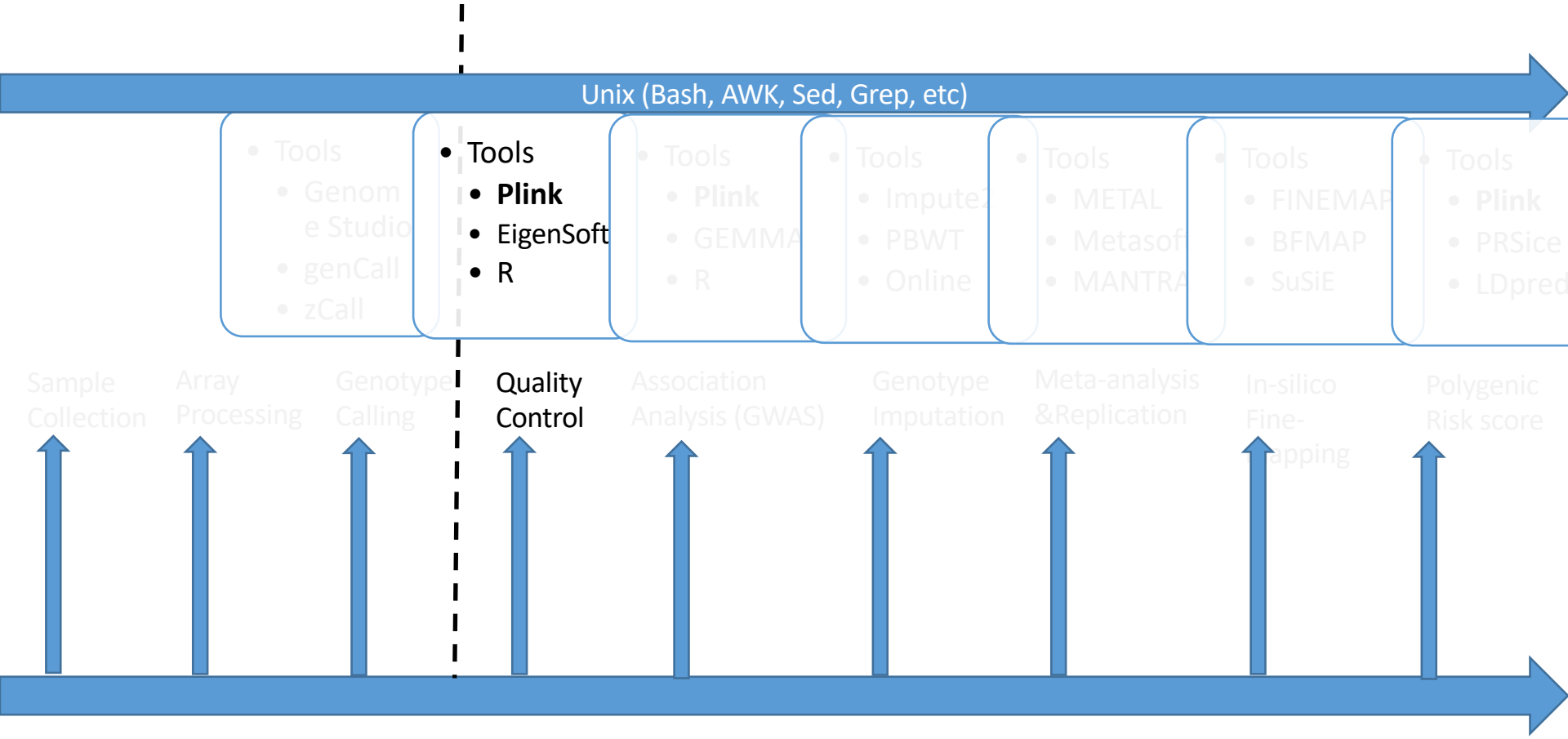
- Training & Mobilisation
- Stakeholder & Community Engagement
- Mapping of household
- Ethics (informed consents,)
- Collection of survey data
- Feedback



- Typically cases and control for a disease (Can also be quantitative eg height, etc)
- The individuals in the study are genotype on a commercial array
- Specialized tool eg Genome Studio can be for genotyping analysis of the microarray data.
- Genome Studio has many modules including a module (Region Report) that can generate PLINK file.



Quality Control

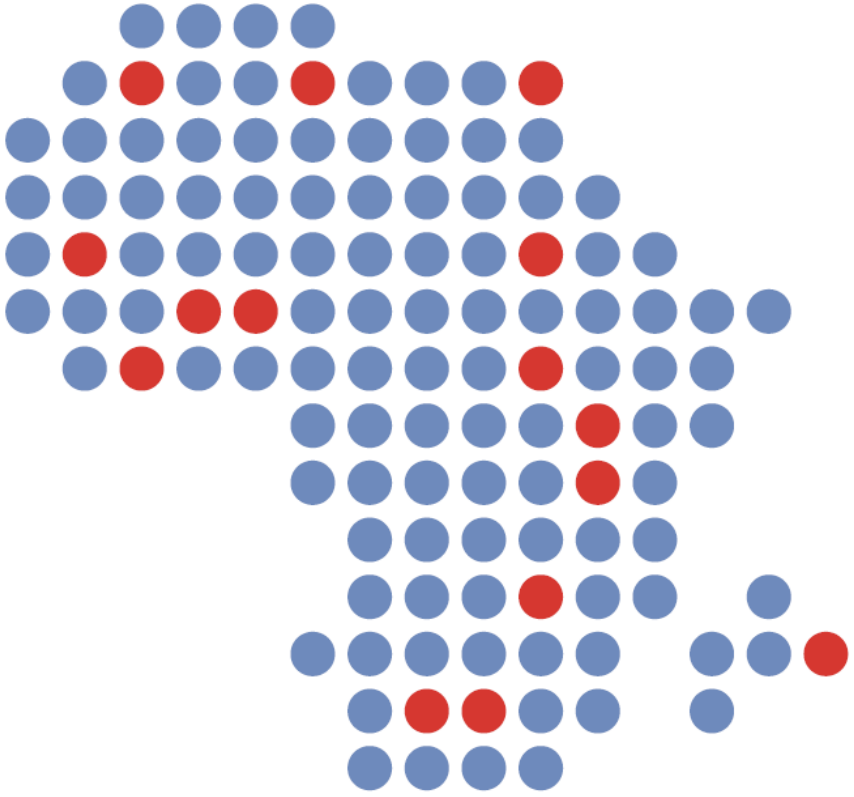


Details in the Next Video

Further Reading

Manduchi, Elisabetta, Patryk R. Orzechowski, Marylyn D. Ritchie, and Jason H. Moore. "Exploration of a diversity of computational and statistical measures of association for genome-wide genetic studies." *BioData mining* 12, no. 1 (2019): 1-16.

Mathaiyan, Jayanthi, Adithan Chandrasekaran, and Sanish Davis. "Ethics of genomic research." *Perspectives in clinical research* 4, no. 1 (2013): 100.



H3ABioNet