Rare and low-frequency variants and predisposition to complex disease

Patrick K. Albers

Wellcome Trust Centre for Human Genetics
Big Data Institute
Medical Sciences Division

Green Templeton College

University of Oxford

Supervised by

Professor Gil McVean

Professor Mark McCarthy

Submitted in Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy (DPhil)

Hilary 2017

Abstract

Recent advances in high-throughput genomic technologies have enabled the collection of DNA information for thousands of individuals, providing unprecedented opportunities to learn about the genetic architecture of complex disease. One important finding has been that the majority of variants in the human genome are low in frequency or rare. It has been hypothesised that the recent explosive growth of the human population afforded unexpectedly large amounts of rare variants with small but deleterious effects, suggesting that rare variants may play a significant role in the predisposition to complex disease. Moreover, properties specific to rare variants embody a rich source of information relating to their evolutionary history.

In this thesis, I develop several statistical methods to address problems associated with the analysis of rare variants in the context of large cohorts linked to biomedical phenotype data, and to leverage the information they encode. Firstly, one constraint in genome-wide association studies is that lower-frequency variants are not captured by genotyping methods, but instead must be predicted through imputation from a reference panel. I develop a method to improve imputation accuracy by integrating genotype data from multiple reference datasets, which outperformed imputations from separate references in almost all comparisons (mean correlation with masked genotypes $r^2 > 0.9$). In a series of simulated case-control experiments, I demonstrate that this approach (meta-imputation) increases power to identify low-frequency variants of intermediate or high penetrance, improving power by 2.2–3.6%. Secondly, I utilise rare variants as identifiers for recently co-inherited shared haplotypes, as rare variants are likely to have originated recently through mutation, making them highly population-specific. I develop a non-probabilistic method to detect shared haplotype segments that are identical by descent (IBD) from patterns of allele sharing and the detection of recombination breakpoints. I show that the latter can be inferred with higher accuracy at very low allele frequencies ($\leq 0.05\%$, $r^2 > 0.99$) using either haplotype or genotype data. Thirdly, I show that genotype error poses a major problem in the analysis of empirical data, for example as obtained through whole genome sequencing or SNP genotyping, in particular towards lower allele frequencies (false positive rate, FPR = 0.1, at frequency $\leq 0.05\%$). I therefore subsequently propose a novel approach to infer IBD from genotype data using a Hidden Markov Model (HMM) under an empirical error model, which I construct by identifying misclassified genotypes in existing datasets, showing that the HMM is robust in presence of error ($\leq 0.05\%$, $r^2 > 0.98$) while previous methods fail ($r^2 < 0.02$). Finally, the age of a rare allele (time since its creation through mutation) may provide evidence about the selective forces that resulted in its observed frequency, and its impact on fitness. I further develop a novel method to estimate rare allele age, based on the inferred IBD structure of a sample. I demonstrate that allele age can be estimated with high accuracy using the HMM-based approach for IBD detection, even in presence of genotype error (Spearman correlation coefficient $r_S = 0.74$, compared to $r_S = 0.82$ when true IBD data is available). I apply this method to data from the 1000 Genomes Project, showing that there are notable age differences between rare alleles of varying predicted phenotypic consequences.

Acknowledgements

This work would not have been possible without the patience of my supervisors, Gil McVean and Mark McCarthy. If my work will prove to be useful, it is because of their guidance.

I dedicate this thesis to my parents, for their unconditional love and their relentless efforts to support me in my studies. I also want to express my gratitude to my mentors, who have guided me over the years, and those who taught me valuable lessons, during my time as a Bachelor student, during my time as a Master student, and in between; to Nico Michiels (Eberhard Karls Universität Tübingen, Germany) who was the first to spark my interest in statistics, to Christopher Bartlett (James Cook University, Australia) who made my first research project possible, to the people of Vanuatu who will always have a special place in my heart (I was the scientific advisor to a marine protected area), to the people of Laos who taught me to appreciate the small things in life (I was a teacher in a Buddhist temple school), to Mikeal Thollesson (Uppsala Universitet, Sweden) who fostered my enthusiasm for Evolutionary Biology, to Sophie Caillon (Université de Montpellier II, France) who showed me that communication is key, to Scott Edwards (Harvard University, United States) who stopped working on a crucial grant application to write a recommendation letter for my application to the University of Oxford, to Dirk Mentzler (Ludwig-Maximilians Universität Munich, Germany) who taught me to not blindly trust any statistical results. Lastly, I want to thank my friends for their support and their understanding.

Contents

List of Figures

List of Tables

My definition of a scientist is that you can complete the following sentence: 'he or she has shown that . . . '

— E. O. Wilson

Bibliography

1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T., and McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**(7422), 56–65.

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., and Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, **526**(7571), 68–74.

Abecasis, G. R., Cherny, S. S., and Cardon, L. R. (2001). The impact of genotyping error on family-based analysis of quantitative traits. *European Journal of Human Genetics*, **9**(2), 130–134.

Albrechtsen, A., Korneliussen, T. S., Moltke, I., Hansen, T. v. O., Nielsen, F. C., and Nielsen, R. (2009). Relatedness Mapping and Tracts of Relatedness for Genome-Wide Data in the Presence of Linkage Disequilibrium. *Genetic Epidemiology*, **33**(3), 266–274.

Altshuler, D., Daly, M. J., and Lander, E. S. (2008). Genetic Mapping in Human Disease. *Science*, **322**(5903), 881–888.

Altshuler, D. L., Bentley, D. R., Chakravarti, A., Collins, F. S., Donnelly, P., Gabriel, S. B., Knoppers, B. M., Lander, E. S., Lehrach, H., Mardis, E. R., Nickerson, D. A., Peltonen, L., Wang, J., Wilson, R. K., Gibbs, R. A., Deiros, D., Metzker, M., Wheeler, D., Li, J., Jian, M., Li, G., Li, R., Liang, H., Tian, G., Wang, B., Wang, W., Yang, H., Zhang, X., Zheng, H., Ambrogio, L., Bloom, T., Sougnez, C. L., Gormley, N., Humphray, S., Kingsbury, Z., Koko-Gonzales, P., Stone, J., McKernan, K. J., Costa, G. L., Ichikawa, J. K., Lee, C. C., Borodina, T. A., Dahl, A., Davydov, A. N., Marquardt, P., Mertes, F., Nietfeld, W., ROSENSTIEL, P., Schreiber, S., Soldatov, A. V., Timmermann, B., Tolzmann, M., Egholm, M., Affourtit, J., Ashworth, D., Attiya, S., Bachorski, M., Buglione, E., Burke, A., Caprio, A., Celone, C., Clark, S., Conners, D., Gu, L., Guccione, L., Kao, K., Kebbel, A., Knowlton, J., Labrecque, M., McDade, L., Mealmaker, C., Minderman, M., Nawrocki, A., Niazi, F., Pareja, K., Ramenani, R., Riches, D., Song, W., Turcotte, C., Wang, S., Fulton, L., Weinstock, G., Burton, J., Carter, D. M., Churcher, C., Coffey, A., Cox, A., Quail, M., Skelly, T., Stalker, J., Swerdlow, H. P., Turner, D., De Witte, A., Giles, S., Sabo, A., Yu, J., Fang, X., Guo, X., Li, Y., Luo, R., Tai, S., Wu, H., Zheng, H., Zheng, X., Zhou, Y., Wang, J., Huang, W., Indap, A., Kural, D., Lee, W.-P., Stromberg, M. P., Ward, A. N., Lee, C., Mills, R. E., Shi, X., Daly, M. J., DePristo, M. A., Ball, A. D., Browning, B. L., Cibulskis, K., Garimella, K. V., Grossman, S. R., Hanna, M., Jaffe, D. B., Kernytsky, A. M., Li, H., Maguire, J. R., McCarroll, S. A., McKenna, A., Nemesh, J. C., Philippakis, A. A., Poplin, R. E., Price, A., Rivas, M. A., Sabeti, P. C., Schaffner, S. F., Shefler, E., Shlyakhter, I. A., Cooper, D. N., Ball, E. V., Mort, M., Phillips, A. D., Stenson, P. D., Sebat, J., Makarov, V., Ye, K., Yoon, S. C., Clark, A. G., Boyko, A., Degenhardt, J., Gutenkunst, R. N., Kaganovich, M., Keinan, A., Lacroute, P., Ma, X., Reynolds, A., Clarke, L., Cunningham, F., Herrero, J., Keenen, S., Kulesha, E., Leinonen, R., McLaren, W. M., Radhakrishnan, R., Smith, R. E., Zalunin, V., Zheng-Bradley, X., Korbel, J. O., Stütz, A. M., Bauer, M., Keira Cheetham, R., Cox, T., Eberle, M., James, T., Kahn, S., Murray, L., Ye, K., De La Vega, F. M., Fu, Y., Hyland, F. C. L., Manning, J. M., McLaughlin, S. F., Peckham, H. E., Sakarya, O., Sun, Y. A., Tsung, E. F., Batzer, M. A., Konkel, M. K., Walker, J. A., Sudbrak, R., Albrecht, M. W., Amstislavskiy, V. S., Herwig, R., Parkhomchuk, D. V., Agarwala, R., Khouri, H. M., Morgulis, A. O., Paschall, J. E., Phan, L. D., Rotmistrovsky, K. E., Sanders, R. D., Shumway, M. F., Xiao, C., Lunter, G., Marchini, J. L., Moutsianas, L., Myers, S., Tumian, A., Desany, B., Knight, J., Winer, R., Craig, D. W., Beckstrom-Sternberg, S. M., Christoforides, A., Kurdoglu, A. A., Pearson, J. V., Sinari, S. A., Tembe, W. D., Haussler, D., Hinrichs, A. S., Katzman, S. J., Kern, A., Kuhn, R. M., Przeworski, M., Hernandez, R. D., Howie, B., Kelley, J. L., Cord Melton, S., Abecasis, G. R., Li, Y., Anderson, P., Blackwell, T., Chen, W., Cookson, W. O., Ding, J., Min Kang, H., Lathrop, M., Liang, L., and Moffatt, M. F. a. (2010). A map of human genome variation from population-scale sequencing. *Nature*, **467**(7319), 1061–1073.

Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., Rhodes, M., Reich, D. E., and Hirschhorn, J. N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics*, **74**(6), 1111–1120.

Bodmer, W. and Bonilla, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genetics*, **40**(6), 695–701.

Boehnke, M. and Cox, N. J. (1997). Accurate Inference of Relationships in Sib-Pair Linkage Studies. *The American Journal of Human Genetics*, **61**(2), 423–429.

Boomsma, D. I., Wijmenga, C., Slagboom, E. P., Swertz, M. A., Karssen, L. C., Abdellaoui, A., Ye, K., Guryev, V., Vermaat, M., van Dijk, F., Francioli, L. C., Hottenga, J. J., Laros, J. F. J., Li, Q., Li, Y., Cao, H., Chen, R., Du, Y., Li, N., Cao, S., van Setten, J., Menelaou, A., Pulit, S. L., Hehir-Kwa, J. Y., Beekman, M., Elbers, C. C., Byelas, H., de Craen, A. J. M., Deelen, P., Dijkstra, M., den Dunnen, J. T., de Knijff, P., Houwing-Duistermaat, J., Koval, V., Estrada, K., Hofman, A., Kanterakis, A., van Enckevort, D., Mai, H., Kattenberg, M., van Leeuwen, E. M., Neerincx, P. B. T., Oostra, B., Rivadeneira, F., Suchiman, E. H. D., Uitterlinden, A. G., Willemsen, G., Wolffenbuttel, B. H., Wang, J., de Bakker, P. I. W., van Ommen, G.-J., and van Duijn, C. M. (2013). The Genome of the Netherlands: design, and project goals. *European Journal of Human Genetics*, **22**(2), 221–227.

Borodulin, K., Vartiainen, E., Peltonen, M., Jousilahti, P., Juolevi, A., Laatikainen, T., Mannisto, S., Salomaa, V., Sundvall, J., and Puska, P. (2015). Forty-year trends in cardiovascular risk factors in Finland. *European Journal of Public Health*, **25**(3), 539–546.

Botstein, D. and Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics*, **33**(3s), 228–237.

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics*, **32**(3), 314–331.

Browning, B. L. and Browning, S. R. (2011). A fast, powerful method for detecting identity by descent. *American journal of human genetics*, **88**(2), 173–182.

Browning, B. L. and Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*, **194**(2), 459–471.

Browning, B. L. and Browning, S. R. (2016). Genotype Imputation with Millions of Reference Samples. *American journal of human genetics*, **98**(1), 116–126.

Browning, S. R. (2008). Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes. *Genetics*, **178**(4), 2123–2132.

Browning, S. R. and Browning, B. L. (2010). High-Resolution Detection of Identity by Descent in Unrelated Individuals. *The American Journal of Human Genetics*, **86**(4), 526–539.

Browning, S. R. and Browning, B. L. (2012). Identity by Descent Between Distant Relatives: Detection and Applications. *Annual Review of Genetics*, **46**(1), 617–633.

Buetow, K. H. (1991). Influence of aberrant observations on high-resolution linkage analysis outcomes. *The American Journal of Human Genetics*, **49**(5), 985–994.

Burdett, T., Hall, P., Hasting, E., Hindorff, L., Junkins, H., Klemm, A., MacArthur, J., Manolio, T., Morales, J., Parkinson, H., *et al.* (2016). The nhgri-ebi catalog of published genome-wide association studies. *Available at: www.ebi.ac.uk/gwas. Accessed 2017-01-20, version 1.0.*

Bustamante, C. D., Burchard, E. G., and De La Vega, F. M. (2011). Genomics for the world. *Nature*, **475**(7355), 163–165.

Cai, L., Fisher, A. L., Huang, H., and Xie, Z. (2016). CRISPR-mediated genome editing and human diseases. *Genes & Diseases*, **3**(4), 244–251.

Chakravarti, A. (1999). Population genetics - making sense out of sequence. *Nature Genetics*, **21**, 56–60.

Chen, J., Zhang, J.-G., Li, J., Pei, Y.-F., and Deng, H.-W. (2013). On Combining Reference Data to Improve Imputation Accuracy. *PloS one*, **8**(1).

Choi, Y., Wijsman, E. M., and Weir, B. S. (2009). Case-control association testing in the presence of unknown relationships. *Genetic Epidemiology*, **33**(8), 668–678.

Cirulli, E. T. and Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Publishing Group*, **11**(6), 415–425.

Colombo, R. (2007). Dating mutations. *eLS*.

Correns, K. F. J. (1899). Untersuchungen über die Xenien bei *Zea mays*. *Berichte der Deutschen Botanischen Gesellschaft*, **17**, 410–418.

Cotterman, C. W. (1940). *A calculus for statistico-genetics*. Ph.D. thesis, The Ohio State University.

Coventry, A., Bull-Otterson, L. M., Liu, X., Clark, A. G., Maxwell, T. J., Crosby, J., Hixson, J. E., Rea, T. J., Muzny, D. M., Lewis, L. R., Wheeler, D. A., Sabo, A., Lusk, C., Weiss, K. G., Akbar, H., Cree, A., Hawes, A. C., Newsham, I., Varghese, R. T., Villasana, D., Gross, S., Joshi, V., Santibanez, J., Morgan, M., Chang, K., Hale IV, W., Templeton, A. R., Boerwinkle, E., Gibbs, R., and Sing, C. F. (2010). Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nature communications*, **1**(8), 131–6.

Cox, D. G. and Kraft, P. (2006). Quantification of the Power of Hardy-Weinberg Equilibrium Testing to Detect Genotyping Error. *Human heredity*, **61**(1), 10–14.

Crow, J. F. (1954). Breeding structure of populations. ii. effective population number. *Statistics and mathematics in biology*, **543**, 556.

Crow, J. F. and Kimura, M. (1970). An introduction to population genetics theory. *An introduction to population genetics theory.*

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., and Lander, E. S. (2001). High-resolution haplotype structure in the human genome. *Nature Genetics*, **29**(2), 229–232.

de Bakker, P. I. W., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., and Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nature Genetics*, **37**(11), 1217–1223.

De Vries, H. M. (1900). Sur la loi de disjonction des hybrides. *Comptes rendus de l'Académie des Sciences*, **130**, 845–847.

Deelen, P., Kanterakis, A., van Dijk, F., Medina-Gomez, C., Kreiner-Møller, E., Rivadeneira, F., Gutierrez-Achury, J., van Enckevort, D., Dijkstra, M., Byelas, H., Genome of Netherlands Consortium, de Bakker, P. I. W., and Swertz, M. A. (2014). Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of The Netherlands'. *European Journal of Human Genetics*, **22**(11), 1321–1326.

Delaneau, O., Coulonges, C., and Zagury, J.-F. (2008). Shape-IT: new rapid and accurate algorithm for haplotype inference. **9**(1), 540.

Delaneau, O., Marchini, J., and Zagury, J.-F. (2011). A linear complexity phasing method for thousands of genomes. *Nature methods*, **9**(2), 179–181.

Delaneau, O., Zagury, J.-F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nature methods*, **10**(1), 5–6.

Devlin, B., Roeder, K., and Wasserman, L. (2001). Genomic control, a new approach to genetic-based association studies. *Theoretical population biology*, **60**(3), 155–166.

Donnelly, K. P. (1983). The probability that related individuals share some section of genome identical by descent. *Theoretical population biology*, **23**(1), 34–63.

Douglas, J. A., Boehnke, M., and Lange, K. (2000). A Multipoint Method for Detecting Genotyping Errors and Mutations in Sibling-Pair Linkage Data. *The American Journal of Human Genetics*, **66**(4), 1287–1297.

Douglas, J. A., Skol, A. D., and Boehnke, M. (2002). Probability of Detection of Genotyping Errors and Mutations as Inheritance Inconsistencies in Nuclear-Family Data. *The American Journal of Human Genetics*, **70**(2), 487–495.

Eberle, M. A., Fritzilas, E., Krusche, P., Kallberg, M., Moore, B. L., Bekritsky, M. A., Iqbal, Z., Chuang, H.-Y., Humphray, S. J., Halpern, A. L., Kruglyak, S., Margulies, E. H., McVean, G., and Bentley, D. R. (2016). A reference data set of 5.4 million phased human variants validated by genetic inheritance from sequencing a three-generation 17-member pedigree. *Genome Research*, **27**(1), 1–9.

Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L., and Järvelä, I. (2002). Identification of a variant associated with adult-type hypolactasia. *Nature genetics*, **30**(2), 233–237.

Ewens, W. J. (2012). *Mathematical Population Genetics 1: Theoretical Introduction*, volume 27. Springer Science & Business Media.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford University Press, Oxford.

Fisher, R. A. (1949). The theory of inbreeding. *The theory of inbreeding*.

Fisher, R. A. (1954). A fuller theory of "junctions" in inbreeding. *Heredity*, **8**(2), 187–197.

Forney, G. D. (1973). The Viterbi Algorithm. In *Proceedings of the IEEE*, pages 268–278.

Frazer, K. A., Murray, S. S., Schork, N. J., and Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics*, **10**(4), 241–251.

Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., Gabriel, S. B., Topol, E. J., Smoller, J. W., Pato, C. N., Pato, M. T., Petryshen, T. L., Kolonel, L. N., Lander, E. S., Sklar, P., Henderson, B., Hirschhorn, J. N., and Altshuler, D. (2004). Assessing the impact of population stratification on genetic association studies. *Nature Genetics*, **36**(4), 388–393.

Fu, W., O'Connor, T. D., Jun, G., Kang, H. M., Abecasis, G., Leal, S. M., Gabriel, S., Altshuler, D., Shendure, J., Nickerson, D. A., Bamshad, M. J., Project, N. E. S., and Akey, J. M. (2012). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*, **493**(7431), 216–220.

Fu, Y. X. (1995). Statistical Properties of Segregating Sites. *Theoretical population biology*, **48**(2), 172–197.

Fu, Y.-X. and Li, W.-H. (1999). Coalescing into the 21st Century: An Overview and Prospects of Coalescent Theory. *Theoretical population biology*, **56**(1), 1–10.

Fuchsberger, C., Flannick, J., Teslovich, T. M., Mahajan, A., Agarwala, V., Gaulton, K. J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D. J., Rivas, M. A., Perry, J. R. B., Sim, X., Blackwell, T. W., Robertson, N. R., Rayner, N. W., Cingolani, P., Locke, A. E., Tajes, J. F., Highland, H. M., Dupuis, J., Chines, P. S., Lindgren, C. M., Hartl, C., Jackson, A. U., Chen, H., Huyghe, J. R., van de Bunt, M., Pearson, R. D., Kumar, A., Müller-Nurasyid, M., Grarup, N., Stringham, H. M., Gamazon, E. R., Lee, J., Chen, Y., Scott, R. A., Below, J. E., Chen, P., Huang, J., Go, M. J., Stitzel, M. L., Pasko, D., Parker, S. C. J., Varga, T. V., Green, T., Beer, N. L., Day-Williams, A. G., Ferreira, T., Fingerlin, T., Horikoshi, M., Hu, C., Huh, I., Ikram, M. K., Kim, B.-J., Kim, Y., Kim, Y. J., Kwon, M.-S., Lee, J., Lee, S., Lin, K.-H., Maxwell, T. J., Nagai, Y., Wang, X., Welch, R. P., Yoon, J., Zhang, W., Barzilai, N., Voight, B. F., Han, B.-G., Jenkinson, C. P., Kuulasmaa, T., Kuusisto, J., Manning, A., Ng, M. C. Y., Palmer, N. D., Balkau, B., áková, A. S., Abboud, H. E., Boeing, H., Giedraitis, V., Prabhakaran, D., Gottesman, O., Scott, J., Carey, J., Kwan, P., Grant, G., Smith, J. D., Neale, B. M., Purcell, S., Butterworth, A. S., Howson, J. M. M., Lee, H. M., Lu, Y., Kwak, S.-H., Zhao, W., Danesh, J., Lam, V. K. L., Park, K. S., Saleheen, D., So, W. Y., Tam, C. H. T., Afzal, U., Aguilar, D., Arya, R., Aung, T., Chan, E., Navarro, C., Cheng, C.-Y., Palli, D., Correa, A., Curran, J. E., Rybin, D., Farook, V. S., Fowler, S. P., Freedman, B. I., Griswold, M., Hale, D. E., Hicks, P. J., Khor, C.-C., Kumar, S., Lehne, B., Thuillier, D., Lim, W. Y., Liu, J., van der Schouw, Y. T., Loh, M., Musani, S. K., Puppala, S., Scott, W. R., Yengo, L., Tan, S.-T., Taylor, H. A., Thameem, F., Wilson, G., Wong, T. Y., Njølstad, P. R., Levy, J. C., Mangino, M., Bonnycastle, L. L., Schwarzmayr, T., Fadista, J., Surdulescu, G. L., Herder, C., Groves, C. J., Wieland, T., Bork-Jensen, J., Brandslund, I., Christensen, C., Koistinen, H. A., Doney, A. S. F., Kinnunen, L., Esko, T., Farmer, A. J., Hakaste, L., Hodgkiss, D., Kravic, J., Lyssenko, V., Hollensted, M., Jørgensen, M. E., Jørgensen, T., Ladenvall, C., Justesen, J. M., Käräjämäki, A., Kriebel, J., Rathmann, W., Lannfelt, L., Lauritzen, T., Narisu, N., Linneberg, A., Melander, O., Milani, L., Neville, M., Orho-Melander, M., Qi, L., Qi, Q., Roden, M., Rolandsson, O., Swift, A., Rosengren, A. H., Stirrups, K., Wood, A. R., Mihailov, E., Blancher, C., Carneiro, M. O., Maguire, J., Poplin, R., Shakir, K., Fennell, T., DePristo, M., de Angelis, M. H., Deloukas, P., Gjesing, A. P., Jun, G., Nilsson, P., Murphy, J., Onofrio, R., Thorand, B., Hansen, T., Meisinger, C., Hu, F. B., Isomaa, B., Karpe, F., Liang, L., Peters, A., Huth, C., O'Rahilly, S. P., Palmer, C. N. A., Pedersen, O., Rauramaa, R., Tuomilehto, J., Salomaa, V., Watanabe, R. M., Syvänen, A.-C., Bergman, R. N., Bharadwaj, D., Bottinger, E. P., Cho, Y. S., Chandak, G. R., Chan, J. C. N., Chia, K. S., Daly, M. J., Ebrahim, S. B., Langenberg, C., Elliott, P., Jablonski, K. A., Lehman, D. M., Jia, W., and Ma, R. (2016). The genetic architecture of type 2 diabetes. *Nature*, **536**(7614), 41–47.

Genome of the Netherlands Consortium (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature Genetics*, **46**(8), 818–825.

Gibson, G. (2012). Rare and common variants: twenty arguments. *Nature Reviews Genetics*, **13**(2), 135–145.

Gordon, D., Heath, S. C., Liu, X., and Ott, J. (2001). A Transmission/Disequilibrium Test That Allows for Genotyping Errors in the Analysis of Single-Nucleotide Polymorphism Data. *The American Journal of Human Genetics*, **69**(2), 371–380.

Gordon, D., Finch, S. J., Nothnagel, M., and Ott, J. u. r. (2002). Power and Sample Size Calculations for Case-Control Genetic Association Tests when Errors Are Present: Application to Single Nucleotide Polymorphisms. *Human heredity*, **54**(1), 22–33.

Gore, A., Li, Z., Fung, H.-L., Young, J. E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M. A., Kiskinis, E., Lee, J.-H., Loh, Y.-H., Manos, P. D., Montserrat, N., Panopoulos, A. D., Ruiz, S., Wilbert, M. L., Yu, J., Kirkness, E. F., Izpisua Belmonte, J. C., Rossi, D. J., Thomson, J. A., Eggan, K., Daley, G. Q., Goldstein, L. S. B., and Zhang, K. (2011). Somatic coding mutations in human induced pluripotent stem cells. *Nature*, **471**(7336), 63–67.

Griffiths, R. C. (1991). The Two-Locus Ancestral Graph. In *Selected Proceedings of the Sheffield Symposium on Applied Probability*, pages 100–117. Institute of Mathematical Statistics, Hayward, CA.

Griffiths, R. C. and Marjoram, P. (1996). Ancestral inference from samples of DNA sequences with recombination. *Journal of Computational Biology*, **3**(4), 479–502.

Griffiths, R. C. and Marjoram, P. (1997a). An ancestral recombination graph. *Institute for Mathematics and its Applications*, **87**, 257.

Griffiths, R. C. and Marjoram, P. (1997b). Progress in population genetics and human evolution.

Griffiths, R. C. and Tavaré, S. (1998). The age of a mutation in a general coalescent tree. *Communications in Statistics. Stochastic Models*, **14**(1-2), 273–295.

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L., Friedman, J. M., and Pe'er, I. (2009). Whole population, genome-wide mapping of hidden relatedness. *Genome Research*, **19**(2), 318–326.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. (2009). Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genetics*, **5**(10), e1000695–11.

Haldane, J. (1919). The combination of linkage values and the calculation of distances between the loci of linked factors. *J Genet*, **8**(29), 299–309.

Hardy, G. H. (1908). Mendelian Proportions in a Mixed Population. *Science*, **28**(706), 49–50.

Harris, K. and Nielsen, R. (2013). Inferring Demographic History from a Spectrum of Shared Haplotype Lengths. *PLoS Genetics*, **9**(6).

Hayes, B. J., Visscher, P. M., McPartlan, H. C., and Goddard, M. E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, **13**(4), 635–643.

Hein, J., Schierup, M., and Wiuf, C. (2004). *Gene genealogies, variation and evolution: a primer in coalescent theory*. Oxford University Press, Oxford.

Henn, B. M., Gutenkunst, R. N., Indap, A. R., Marth, G. T., Clark, A. G., Yu, F., Gibbs, R. A., 1000 Genomes Project, Durbin, R. M., Flicek, P., Gabriel, S. B., Lander, E. S., Wheeler, D., Cibulskis, K., Fennell, T. J., Jaffe, D. B., Shefler, E., Egholm, M., Fulton, R., Bainbridge, M., Challis, D., Sabo, A., Wang, J., Lee, C., Korn, J. M., Sudbrak, R., Auton, A., Iqbal, Z., Desany, B., Dooling, D., Hurles, M. E., MacArthur, D. G., Abyzov, A., Zhang, Z., Garrison, E. P., Banks, E., Handsaker, R. E., Hartl, C., De La Vega, F. M., Alkan, C., Snyder, M., Muzny, D., Reid, J., Quinlan, A. R., Stewart, C., Wu, J., Gravel, S., Sherry, S. T., McVean, G. A., Abecasis, G. R., Koboldt, D. C., Palotie, A., Bustamante, C. D., Schafer, A. J., and Brooks, L. D. (2011). Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences*, **108**(29), 11983–11988.

Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands of genomes. *G3 (Bethesda, Md.)*, **1**(6), 457–470.

Howie, B. N., Donnelly, P., and Marchini, J. (2009). A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genetics*, **5**(6), e1000529.

Hudson, R. (1990). Gene genealogies and the coalescent process. *Oxford surveys in evolutionary biology*, **7**(1), 44.

Hudson, R. and Kaplan, N. L. (1985). Statistical Properties of the Number of Recombination Events in the History of a Sample of Dna-Sequences. *Genetics*, **111**(1), 147–164.

Hudson, R. R. (1983). Properties of a neutral allele model with intragenic recombination. *Theoretical population biology*, **23**(2), 183–201.

Hudson, R. R. (2001). Two-locus sampling distributions and their application. *Genetics*, **159**(4), 1805–1817.

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**(2), 337–338.

International HapMap 3 Consortium, Schaffner, S. F., Peltonen, L., Dermitzakis, E., Bonnen, P. E., Altshuler, D. M., de Bakker, P. I. W., Deloukas, P., Gabriel, S. B., Gwilliam, R., Hunt, S., Inouye, M., Jia, X., Palotie, A., Parkin, M., Whittaker, P., Yu, F., Chang, K., Hawes, A., Lewis, L. R., Ren, Y., Wheeler, D., Gibbs, R. A., Muzny, D. M., Barnes, C., Darvishi, K., Hurles, M., Korn, J. M., Kristiansson, K., Lee, C., McCarrol, S. A., Nemesh, J., Keinan, A., Montgomery, S. B., Pollack, S., Soranzo, N., Gonzaga-Jauregui, C., Anttila, V., Brodeur, W., Daly, M. J., Leslie, S., McVean, G., Moutsianas, L., Nguyen, H., Zhang, Q., Ghori, M. J. R., McGinnis, R., McLaren, W., Price, A. L., Takeuchi, F., Grossman, S. R., Shlyakhter, I., Hostetter, E. B., Sabeti, P. C., Adebamowo, C. A., Foster, M. W., Gordon, D. R., Licinio, J., Manca, M. C., Marshall, P. A., Matsuda, I., Ngare, D., Wang, V. O., Reddy, D., Rotimi, C. N., Royal, C. D., Sharp, R. R., Zeng, C., Brooks, L. D., and McEwen, J. E. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**(7311), 52–58.

International HapMap Consortium (2003). The International HapMap Project. *Nature*, **426**(6968), 789–796.

International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M. M. Y., Tsui, S. K. W., Xue, H., Wong, J. T.-F., Galver, L. M., Fan, J.-B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.-F., Phillips, M. S., Roumy, S., Sallée, C., Verner, A., Hudson, T. J., Kwok, P.-Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.-C., Mak, W., Song, Y. Q., Tam, P. K. H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., de Bakker, P. I. W., Barrett, J., Chretien, Y. R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Sabeti, P., Saxena, R., Schaffner, S. F., Sham, P. C., Varilly, P., Altshuler, D., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimmah, C., Royal, C. D. M., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Yakub, I., Onofrio, R. C., Richter, D. J., Birren, B. W., Daly, M. J., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., McLay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'Archevêque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., McEwen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R., and Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**(7164), 851–861.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**(6822), 860–921.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*, **431**(7011), 931–945.

Kaiser, J. (2008). A plan to capture human diversity in 1000 genomes (Science (395)). *Science*, **319**(5868), 1336.

Kang, S. J., Gordon, D., and Finch, S. J. (2004). What SNP genotyping errors are most costly for genetic association studies? *Genetic Epidemiology*, **26**(2), 132–141.

Keinan, A. and Clark, A. G. (2012). Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*, **336**(6082), 740–743.

Kelleher, J., Etheridge, A. M., and McVean, G. (2016). Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLoS computational biology*, **12**(5), e1004842–22.

Kimura, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, **61**(4), 893–903.

Kimura, M. and Ota, T. (1973). The age of a neutral mutant persisting in a finite population. *Genetics*, **75**(1), 199–212.

Kingman, J. F. C. (1982a). Exchangeability and the evolution of large populations. *Exchangeability in Probability and Statistics*, pages 97—-112.

Kingman, J. F. C. (1982b). On the genealogy of large populations. *Journal of Applied Probability*, **19**(A), 27–43.

Kingman, J. F. C. (1982c). The coalescent. *Stochastic processes and their applications*, **13**(3), 235–248.

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J.-Y., Sackler, R. S., Haynes, C., Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C., and Hoh, J. (2005). Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science*, **308**(5720), 385–389.

Kong, A., Masson, G., Frigge, M. L., Gylfason, A., Zusmanovich, P., Thorleifsson, G., Olason, P. I., Ingason, A., Steinberg, S., Rafnar, T., Sulem, P., Mouy, M., Jonsson, F., Thorsteinsdottir, U., Gudbjartsson, D. F., Stefansson, H., and Stefansson, K. (2008). Detection of sharing by descent, long-range phasing and haplotype imputation. *Nature Genetics*, **40**(9), 1068–1075.

Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics*, **22**(2), 139–144.

Kryukov, G. V., Pennacchio, L. A., and Sunyaev, S. R. (2007). Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *The American Journal of Human Genetics*, **80**(4), 727–739.

Labuda, M., Labuda, D., Korab-Laskowska, M., Cole, D. E., Zietkiewicz, E., Weissenbach, J., Popowska, E., Pronicka, E., Root, A. W., and Glorieux, F. H. (1996). Linkage disequilibrium analysis in young populations: pseudo-vitamin d-deficiency rickets and the founder effect in french canadians. *American journal of human genetics*, **59**(3), 633.

Lander, E. S. (1996). The new genomics: Global views of biology. *Science*, **274**(5287), 536–539.

Lawson, D. J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, **8**(1), e1002453.

Leutenegger, A.-L., Prum, B., Génin, E., Verny, C., Lemainque, A., Clerget-Darpoux, F., and Thompson, E. A. (2003). Estimation of the Inbreeding Coefficient through Use of Genomic Data. *American journal of human genetics*, **73**(3), 516–523.

Li, H. and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, **475**(7357), 493–U84.

Li, N. and Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, **165**(4), 2213–2233.

Li, W. H. (1975). The first arrival time and mean age of a deleterious mutant gene in a finite population. *The American Journal of Human Genetics*, **27**(3), 274–286.

Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype imputation. *Annual review of genomics and human genetics*, **10**, 387–406.

Li, Y., Willer, C. J., Ding, J., Scheet, P., and Abecasis, G. R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology*, **34**(8), 816–834.

Lim, E. T., Würtz, P., Havulinna, A. S., Palta, P., Tukiainen, T., Rehnström, K., Esko, T., Mägi, R., Inouye, M., Lappalainen, T., Chan, Y., Salem, R. M., Lek, M., Flannick, J., Sim, X., Manning, A., Ladenvall, C., Bumpstead, S., Hämäläinen, E., Aalto, K., Maksimow, M., Salmi, M., Blankenberg, S., Ardissino, D., Shah, S., Horne, B., McPherson, R., Hovingh, G. K., Reilly, M. P., Watkins, H., Goel, A., Farrall, M., Girelli, D., Reiner, A. P., Stitziel, N. O., Kathiresan, S., Gabriel, S., Barrett, J. C., Lehtimäki, T., Laakso, M., Groop, L., Kaprio, J., Perola, M., McCarthy, M. I., Boehnke, M., Altshuler, D. M., Lindgren, C. M., Hirschhorn, J. N., Metspalu, A., Freimer, N. B., Zeller, T., Jalkanen, S., Koskinen, S., Raitakari, O., Durbin, R., MacArthur, D. G., Salomaa, V., Ripatti, S., Daly, M. J., Palotie, A., and for the Sequencing Initiative Suomi (SISu) Project (2014). Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population. *PLoS Genetics*, **10**(7), e1004494–12.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary mathematics*, **80**(1), 221–239.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of Next-Generation Sequencing Systems. *Journal of biomedicine & biotechnology*, **2012**(7), 1–11.

Loh, P.-R., Palamara, P. F., and Price, A. L. (2016a). Fast and accurate long-range phasing in a uk biobank cohort. *Nature genetics*.

Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G. R., Durbin, R., and Price, A. L. (2016b). Reference-based phasing using the Haplotype Reference Consortium panel. Technical report.

Malécot, G. (1948). Mathematics of heredity. *Les mathematiques de l'heredite.*

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F. C., McCarroll, S. A., and Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, **461**(7265), 747–753.

Marchini, J. and Howie, B. (2010). Genotype imputation for genome-wide association studies. **11**(7), 499–511.

Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics*, **39**(7), 906–913.

Mardis, E. R. (2006). Anticipating the 1,000 dollar genome. *Genome biology*, **7**(7), 112.

Mardis, E. R. (2017). DNA sequencing technologies: 2006–2016. *Nature Protocols*, **12**(2), 213–218.

Marjoram, P. and Wall, J. D. (2006). Fast" coalescent" simulation. *BMC genetics*, **7**(1), 16.

Marth, G. T., Yu, F., Indap, A. R., Garimella, K., Gravel, S., Leong, W. F., Tyler-Smith, C., Bainbridge, M., Blackwell, T., Zheng-Bradley, X., Chen, Y., Challis, D., Clarke, L., Ball, E. V., Cibulskis, K., Cooper, D. N., Fulton, B., Hartl, C., Koboldt, D., Muzny, D., Smith, R., Sougnez, C., Stewart, C., Ward, A., Yu, J., Xue, Y., Altshuler, D., Bustamante, C. D., Clark, A. G., Daly, M., DePristo, M., Flicek, P., Gabriel, S., Mardis, E., Palotie, A., Gibbs, R., and 1000 Genomes Project (2011). The functional spectrum of low-frequency coding variation. *Genome biology*, **12**(9), R84.

Maruyama, T. (1974). The age of an allele in a finite population. *Genetical research*, **23**(2), 137–143.

Mathieson, I. and McVean, G. (2012). Differential confounding of rare and common variants in spatially structured populations. *Nature Genetics*, **44**(3), 243–U29.

Mathieson, I. and McVean, G. (2014). Demography and the Age of Rare Variants. *PLoS Genetics*, **10**(8), e1004528.

Maynard Smith, J. (1989). *Evolutionary genetics.* Oxford University Press.

McCarroll, S. A., Kuruvilla, F. G., Korn, J. M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M. H., de Bakker, P. I. W., Maller, J. B., Kirby, A., Elliott, A. L., Parkin, M., Hubbell, E., Webster, T., Mei, R., Veitch, J., Collins, P. J., Handsaker, R., Lincoln, S., Nizzari, M., Blume, J., Jones, K. W., Rava, R., Daly, M. J., Gabriel, S. B., and Altshuler, D. (2008). Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genetics*, **40**(10), 1166–1174.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., and Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Publishing Group*, **9**(5), 356–369.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., Veldink, J., Peters, U., Pato, C., van Duijn, C. M., Gillies, C. E., Gandin, I., Mezzavilla, M., Gilly, A., Cocca, M., Traglia, M., Angius, A., Barrett, J. C., Boomsma, D., Branham, K., Breen, G., Brummett, C. M., Busonero, F., Campbell, H., Chan, A., Chen, S., Chew, E., Collins, F. S., Corbin, L. J., Smith, G. D., Dedoussis, G., Dorr, M., Farmaki, A.-E., Ferrucci, L., Forer, L., Fraser, R. M., Gabriel, S., Levy, S., Groop, L., Harrison, T., Hattersley, A., Holmen, O. L., Hveem, K., Kretzler, M., Lee, J. C., McGue, M., Meitinger, T., Melzer, D., Min, J. L., Mohlke, K. L., Vincent, J. B., Nauck, M., Nickerson, D., Palotie, A., Pato, M., Pirastu, N., McInnis, M., Richards, J. B., Sala, C., Salomaa, V., Schlessinger, D., Schoenherr, S., Slagboom, P. E., Small, K., Spector, T., Stambolian, D., Tuke, M., Tuomilehto, J., Van den Berg, L. H., Van Rheenen, W., Volker, U., Wijmenga, C., Toniolo, D., Zeggini, E., Gasparini, P., Sampson, M. G., Wilson, J. F., Frayling, T., de Bakker, P. I. W., Swertz, M. A., McCarroll, S., Kooperberg, C., Dekker, A., Altshuler, D., Willer, C., Iacono, W., Ripatti, S., Soranzo, N., Walter, K., Swaroop, A., Cucca, F., Anderson, C. A., Myers, R. M., Boehnke, M., McCarthy, M. I., Durbin, R., Abecasis, G., and Marchini, J. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, **48**(10), 1279–1283.

McClellan, J. and King, M.-C. (2010). Genetic Heterogeneity in Human Disease. *Cell*, **141**(2), 210–217.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A. K., Farrington, S. M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S. H., Dunlop, M. G., Wright, A. F., Campbell, H., and Wilson, J. F. (2008). Runs of Homozygosity in European Populations. *The American Journal of Human Genetics*, **83**(3), 359–372.

McVean, G., Awadalla, P., and Fearnhead, P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*.

McVean, G. A. T. and Cardin, N. J. (2005). Approximating the coalescent with recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**(1459), 1387–1393.

McVean, G. A. T., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R., and Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. *Science*, **304**(5670), 581–584.

Mendel, G. (1866). Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, **4**, 3–47.

Metzker, M. L. (2009). Sequencing technologies — the next generation. *Nature Publishing Group*, **11**(1), 31–46.

Milligan, B. G. (2003). Maximum-Likelihood Estimation of Relatedness. *Genetics*, **163**(3), 1153–1167.

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, **38**(8).

Morgan, T. H. (1911). Random segregation versus coupling in Mendelian inheritance. *Science*, **34**(873), 384–384.

Morral, N., Bertranpetit, J., Estivill, X., and Nunes, V. (1994). The origin of the major cystic fibrosis mutation (DF508) in European populations. *Nature*.

Morris, A. and Cardon, L. (2007). Whole genome association. *Handbook of Statistical Genetics, Third Edition*, pages 1238–1263.

Moskvina, V. and Schmidt, K. M. (2006). Susceptibility of Biallelic Haplotype and Genotype Frequencies to Genotyping Error. *Biometrics*, **62**(4), 1116–1123.

Moskvina, V., Craddock, N., Holmans, P., Owen, M., and O'Donovan, M. (2005). Minor genotyping error can result in substantial elevation in type i error rate in haplotype based case control analysis. In *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*, volume 138, pages 19–19.

Moutsianas, L., Agarwala, V., Fuchsberger, C., Flannick, J., Rivas, M. A., Gaulton, K. J., Albers, P. K., GoT2D Consortium, McVean, G., Boehnke, M., Altshuler, D., and McCarthy, M. I. (2015). The Power of Gene-Based Rare Variant Methods to Detect Disease-Associated Variation and Test Hypotheses About Complex Disease. *PLoS Genetics*, **11**(4), e1005165–24.

Myers, S., Bottolo, L., Freeman, C., McVean, G., and Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science*.

Naidoo, N., Pawitan, Y., Soong, R., Cooper, D. N., and Ku, C.-S. (2011). Human genetics and genomics a decade after the release of the draft sequence of the human genome. *Human Genomics*, **5**(6), 577–622.

Neuhauser, C. (2001). Mathematical models in population genetics. *Handbook of statistical genetics*.

Nordborg, M. (2001). Coalescent theory. *Handbook of statistical genetics*.

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J. E., Rudan, I., McQuillan, R., Fraser, R. M., Campbell, H., Polasek, O., Asiki, G., Ekoru, K., Hayward, C., Wright, A. F., Vitart, V., Navarro, P., Zagury, J.-F., Wilson, J. F., Toniolo, D., Gasparini, P., Soranzo, N., Sandhu, M. S., and Marchini, J. (2014). A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genetics*, **10**(4), e1004234–21.

Ott, J. (1999). *Analysis of human genetic linkage*. JHU Press.

Pajunen, P., Rissanen, H., Härkänen, T., Jula, A., Reunanen, A., and Salomaa, V. (2010). The metabolic syndrome as a predictor of incident diabetes and cardiovascular events in the Health 2000 Study. *Diabetes and Metabolism*, **36**(5), 395–401.

Palamara, P. F. and Pe'er, I. (2013). Inference of historical migration rates via haplotype sharing. *Bioinformatics*, **29**(13), i180–i188.

Palamara, P. F., Lencz, T., Darvasi, A., and Pe'er, I. (2012). Length Distributions of Identity by Descent Reveal Fine-Scale Demographic History. *The American Journal of Human Genetics*, **91**(5), 809–822.

Palin, K., Campbell, H., Wright, A. F., Wilson, J. F., and Durbin, R. (2011). Identity-by-descent-based phasing and imputation in founder populations using graphical models. *Genetic Epidemiology*, **35**(8), 853–860.

Papoulis, A. and Pillai, S. U. (2002). *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.

Pasaniuc, B., Zaitlen, N., Shi, H., Bhatia, G., Gusev, A., Pickrell, J., Hirschhorn, J., Strachan, D. P., Patterson, N., and Price, A. L. (2014). Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics (Oxford, England)*, **30**(20), 2906–2914.

Pauli, F., Racugno, W., and Ventura, L. (2011). Bayesian composite marginal likelihoods. *Statistica Sinica*, pages 149–164.

Pe'er, I., de Bakker, P. I. W., Maller, J., Yelensky, R., Altshuler, D., and Daly, M. J. (2006). Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nature Genetics*, **38**(6), 663–667.

Pennisi, E. (2007). Human Genetic Variation. *Science*, **318**(5858), 1842–1843.

Powell, J. E., Visscher, P. M., and Goddard, M. E. (2010). Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Publishing Group*, **11**(11), 800–805.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, **38**(8), 904–909.

Pritchard, J. K. (2001). Are rare variants responsible for susceptibility to complex diseases? *The American Journal of Human Genetics*, **69**(1), 124–137.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., and Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**(3), 559–575.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286. Alcatel-Lucent Bell Labs, Murray, United States.

Rasmussen, M. D., Hubisz, M. J., Gronau, I., and Siepel, A. (2014). Genome-Wide Inference of Ancestral Recombination Graphs. *PLoS Genetics*, **10**(5), e1004342–27.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R., and Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature*, **411**(6834), 199–204.

Risch, N. and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, **273**(5281), 1516–1517.

Risch, N., de Leon, D., Ozelius, L., Kramer, P., Almasy, L., Singer, B., Fahn, S., Breakefield, X., and Bressman, S. (1995). Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nature Genetics*, **9**(2), 152–159.

Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature*, **405**(6788), 847–856.

Roach, J. C., Glusman, G., Smit, A. F. A., Huff, C. D., Hubley, R., Shannon, P. T., Rowen, L., Pant, K. P., Goodman, N., Bamshad, M., Shendure, J., Drmanac, R., Jorde, L. B., Hood, L., and Galas, D. J. (2010). Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing. *Science*, **328**(5978), 636–639.

Roshyara, N. R. and Scholz, M. (2015). Impact of genetic similarity on imputation accuracy. *BMC Genetics*, **16**(1), 1–16.

Rousset, F. (2002). Inbreeding and relatedness coefficients: What do they measure? *Heredity*, **88**(5), 371–380.

Sanger, F. and Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology*, **94**(3), 441–448.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**(12), 5463–5467.

Scally, A. and Durbin, R. (2012). Revising the human mutation rate: implications for understanding human evolution. *Nature Reviews Genetics*, **13**(10), 745–753.

Schiffels, S. and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature genetics*, **46**(8), 919–925.

Schork, N. J., Murray, S. S., Frazer, K. A., and Topol, E. J. (2009). Common vs. rare allele hypotheses for complex diseases. *Current opinion in genetics & development*, **19**(3), 212–219.

Schroff, M. H. (2016). *Genealogical properties of rare variation and their implications for demographic inference*. Ph.D. thesis, University of Oxford.

Serre, J. L., Simon-Bouy, B., Mornet, E., Jaume-Roig, B., Balassopoulou, A., Schwartz, M., Taillandier, A., Boue, J., and Boue, A. (1990). Studies of RFLP closely linked to the cystic fibrosis locus throughout Europe lead to new considerations in populations genetics. *Human genetics*, **84**(5), 449–454.

Shields, D. C., Collins, A., Buetow, K. H., and Morton, N. E. (1991). Error filtration, interference, and the human linkage map. *Proceedings of the National Academy of Sciences of the United States of America*, **88**(15), 6501–6505.

Simons, Y. B., Turchin, M. C., Pritchard, J. K., and Sella, G. (2014). The deleterious mutation load is insensitive to recent population history. *Nature Publishing Group*, **46**(3), 220–224.

Slack, J. (2014). *Genes: A Very Short Introduction*. Oxford University Press, Oxford.

Slatkin, M. (2000). Allele age and a test for selection on rare alleles. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **355**(1403), 1663–1668.

Slatkin, M. (2008a). Inbreeding coefficients and coalescence times. *doi.org*, pages 1–9.

Slatkin, M. (2008b). Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, **9**(6), 477–485.

Slatkin, M. and Bertorelle, G. (2001). The use of intraallelic variability for testing neutrality and estimating population growth rate. *Genetics*, **158**(2), 865–874.

Slatkin, M. and Rannala, B. (2000). Estimating allele age. *Annual review of genomics and human genetics*, **1**(1), 225–249.

Sobel, E., Papp, J. C., and Lange, K. (2002). Detection and integration of genotyping errors in statistical genetics. *The American Journal of Human Genetics*, **70**(2), 496–508.

Stam, P. (1980). The distribution of the fraction of the genome identical by descent in finite random mating populations. *Genetical research*, **35**(02), 131.

Stone, M. (1961). The Opinion Pool. *The Annals of Mathematical Statistics*, **32**(4), 1339–1342.

Sturtevant, A. H. (1913). The linear arrangement of six sex-linked factors in drosophila, as shown by their mode of association. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, **14**(1), 43–59.

Su, Z., Marchini, J., and Donnelly, P. (2011). HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics*, **27**(16), 2304–2305.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**(2), 437–460.

Tajima, F. (1993). Measurement of dna polymorphism. *Mechanisms of molecular evolution*, pages 37–59.

Takahata, N. (1993). Allelic genealogy and human evolution. *Molecular Biology and Evolution*, **10**(1), 2–22.

Tavaré, S. (2004). Part I: Ancestral inference in population genetics. In *Lectures on probability theory and statistics*, pages 1–188. Springer.

Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, **145**(2), 505–518.

Tennessen, J. A., Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., Kang, H. M., Jordan, D., Leal, S. M., Gabriel, S., Rieder, M. J., Abecasis, G., Altshuler, D., Nickerson, D. A., Boerwinkle, E., Sunyaev, S., Bustamante, C. D., Bamshad, M. J., Akey, J. M., GO, B., GO, S., and Project, N. E. S. (2012). Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science*, **337**(6090), 64–69.

Thompson, E. A. (1974). Gene Identities and Multiple Relationships. *Biometrics*, **30**(4), 667.

Thompson, E. A. (1975). The estimation of pairwise relationships. *Annals of human genetics*, **39**(2), 173–188.

Thompson, E. A. (1976). Estimation of age and rate of increase of rare variants. *The American Journal of Human Genetics*, **28**(5), 442–452.

Thompson, E. A. (2008). The IBD process along four chromosomes. *Theoretical population biology*, **73**(3), 369–373.

Thompson, E. A. (2013). Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics*, **194**(2), 301–326.

Tschermak, E. (1900). Über künstliche Kreuzung bei *Pisum sativum*. *Berichte der Deutschen Botanischen Gesellschaft*, **18**, 232–239.

UK10K Consortium, Walter, K., Crooks, L., Memari, Y., Perry, J. R. B., Xu, C., Futema, M., Lawson, D., Iotchkova, V., Schiffels, S., Hendricks, A. E., Li, R., Floyd, J., Wain, L. V., Humphries, S. E., Barrett, J. C., Plagnol, V., Richards, J. B., Greenwood, C. M. T., Timpson, N. J., Soranzo, N., Danecek, P., Barroso, I., McCarthy, S., Tachmazidou, I., Durbin, R., Hurles, M. E., Kennedy, K., Palotie, A., Zeggini, E., Cocca, M., Huang, J., and Min, J. L. (2015). The UK10K project identifies rare variants in health and disease. *Nature*, **526**(7571), 82–90.

Vartiainen, E., Laatikainen, T., Peltonen, M., Juolevi, A., Mannisto, S., Sundvall, J., Jousilahti, P., Salomaa, V., Valsta, L., and Puska, P. (2010). Thirty-five-year trends in cardiovascular risk factors in Finland. *International Journal of Epidemiology*, **39**(2), 504–518.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Miklos, G. L. G., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R.-R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z. Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S. C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Angela Center, Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.-H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.-H., Coyne, M., Dahlke, C., Mays, A. D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., and Majoros... (2001). The Sequence of the Human Genome. *Science*, **291**(5507), 1304–1351.

Viterbi, A. J. (1967). Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. *IEEE Transactions on Information Theory*, **13**(2), 260–269.

Voight, B. F. and Pritchard, J. K. (2005). Confounding from Cryptic Relatedness in Case-Control Association Studies. *PLoS Genetics*, **1**(3), e32–10.

Wakeley, J. (2008). *Coalescent Theory: An Introduction*. W. H. Freeman.

Wakeley, J. and Wilton, P. (2016). Coalescent and models of identity by descent. In R. M. Kliman, editor, *Encyclopedia of Evolutionary Biology*, pages 287 – 292. Academic Press, Oxford.

Wang, J. (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**(1459), 1395–1409.

Watson, J. D. and Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**(4356), 737–738.

Watterson, G. (1996). Motoo Kimura's Use of Diffusion Theory in Population Genetics. *Theoretical population biology*, **49**(2), 154–188.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical population biology*, **7**(2), 256–276.

Watterson, G. A. (1976). Reversibility and the age of an allele. I. Moran's infinitely many neutral alleles model. *Theoretical Population Biology*, **10**(3), 239–253.

Weinberg, W. (1908). Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg.*, **64**, 368–382.

Weissenbach, J. (1993). A second generation linkage map of the human genome based on highly informative microsatellite loci. *Gene*, **135**(1-2), 275–278.

Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.-J., Makhijani, V., Roth, G. T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C. L., Irzyk, G. P., Lupski, J. R., Chinault, C., Song, X.-z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D. M., Margulies, M., Weinstock, G. M., Gibbs, R. A., and Rothberg, J. M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **452**(7189), 872–876.

Winkler, H. (1920). *Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreich*. Verlag G. Fischer, Jena.

Wiuf, C. and Hein, J. (1997). On the number of ancestors to a DNA sequence. *Genetics*, **147**(3), 1459–1468.

Wright, S. (1921). Systems of Mating. I. the Biometric Relations between Parent and Offspring. *Genetics*, **6**(2), 111–123.

Wright, S. (1922). Coefficients of Inbreeding and Relationship. *The American Naturalist*, **56**(645), 330–338.

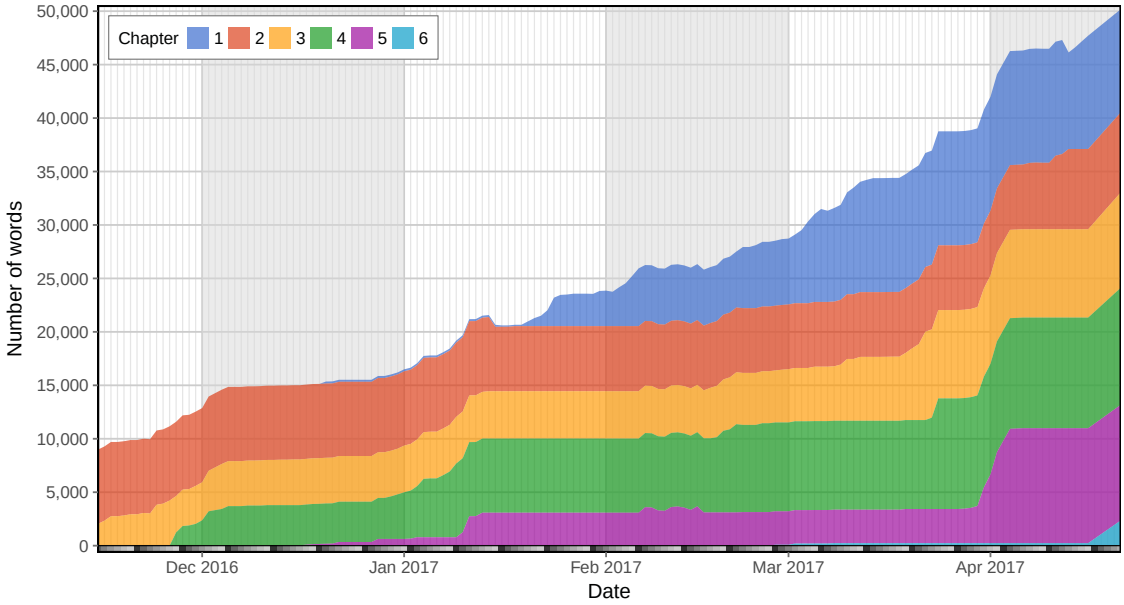
Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, **16**(2), 97–159.

Yu, N., Zhao, Z., Fu, Y. X., Sambuughin, N., Ramsay, M., Jenkins, T., Leskinen, E., Patthy, L., Jorde, L. B., Kuromori, T., and Li, W. H. (2001). Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. *Molecular Biology and Evolution*, **18**(2), 214–222.

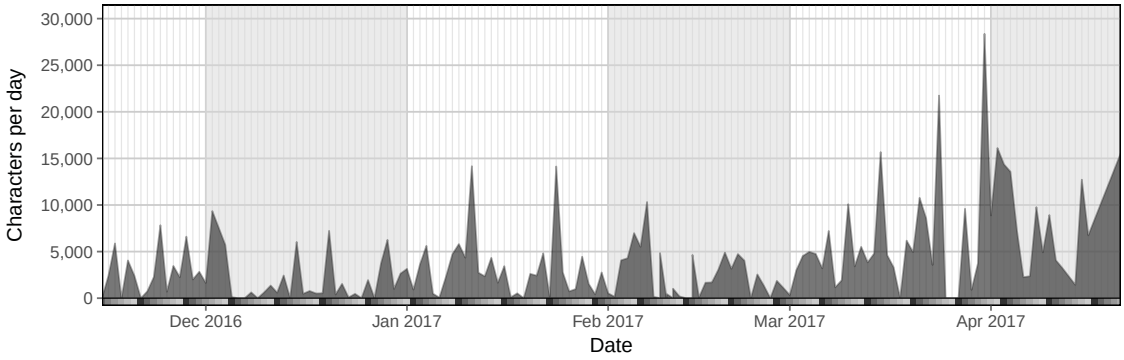
Zuk, O., Schaffner, S. F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M. J., Neale, B. M., Sunyaev, S. R., and Lander, E. S. (2014). Searching for missing heritability: designing rare variant association studies. *Proceedings of the National Academy of Sciences*, **111**(4), E455–64.

Remember kids, the only difference between
screwing around and science
is writing it down.

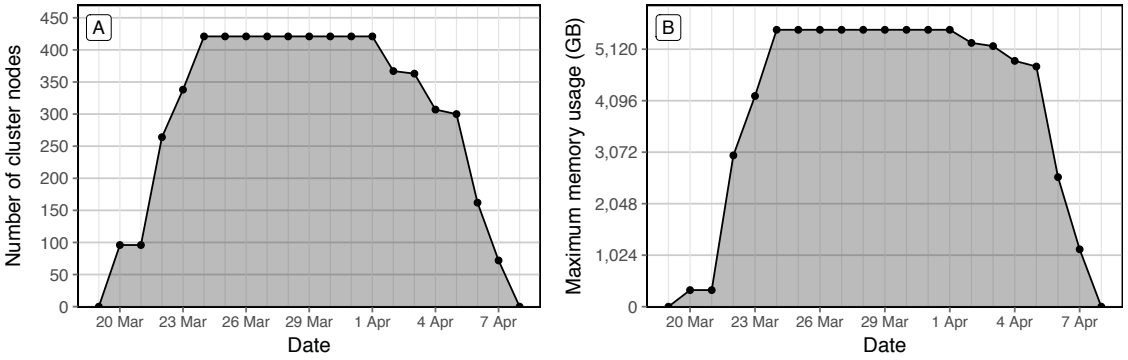
— Adam Savage



Supplementary Figure 1: Word count over time during thesis writing period. Shown for the time since I automatically generated daily backups and until the submission of this thesis.



Supplementary Figure 2: Number of characters written per day. Note that all characters in each LaTeX file were counted.



Supplementary Figure 3: Computer cluster usage one month before the submission date of this thesis. Indicated by the **(A)** number of nodes used and **(B)** daily maximum of computer memory on the cluster of the Wellcome Trust Centre for Human Genetics.