data)

5. Out of the 35 reference sequence 35-mers that overlap with the site, fewer than 18 can be mapped elsewhere with at most one mismatch

References

[1] The 1000 Genomes Project Consortium. "A map of human genome variation from population-scale sequencing." *Nature* (2010) 1061–1073.

[2] C.A. Anderson, D. Brocklebank, and A.P. Morris. "A comparison of reference panels for imputation in genome-wide association studies." *The European Journal of Human Genetics*, to appear.

[3] C.A. Anderson, F.H. Pettersson, J.C. Barrett, J.J. Zhuang, J. Ragoussis, L.R. Cardon, and A.P. Morris. "Evaluation the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms." *The American Journal of Human Genetics* (2008) 112–119.

[4] S.-M. Ahn, T.-H. Kim, S. Lee, et al. "The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group." *Genome Research* 19(9) (2009) 1622–1629.

[5] S. Bamford, E. Dawson, S. Forbes, J. Clements, R. Pettett, A. Dougan, A. Flanagan, J. Teague, P.A. Futreal, M.R. Stratton, and R. Wooster. "THe COSMIC (Catalogue of Somatic Mutations in Cancer) database and website." *British Journal of Cancer* 91 (2004) 355–358.

[6] J.C. Barrett and L.R. Cardon. "Evaluating the coverage of genome-wide association studies." *Nature Genetics* 38(6) (2006) 659–662.

[7] D.R. Bentley, S. Balasubramanian, H.P. Swerdlow, et al. "Accurate whole genome sequencing using reversible terminator chemistry." *Nature* 456 (2008) 53–59.

[8] T.R. Bhangale, M.J. Rieder, and D.A. Nickerson. "Estimating coverage and power for genetic association studies." *Nature Genetics* 40(7) (2008) 841–843.

[9] S.R. Browning. "Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes." *Genetics* 187 (2008) 2123–2132.

[10] S.R. Browning and B.L. Browning. "High resolution detection of identity by descent in unrelated individuals." *American Journal of Human Genetics* 86 (2010) 1–14.

[11] N.H. Chapman and E.A. Thompson. "A model for the length of tracts of identity by descent in finite random mating populations." *Theoretical Population Biology* 64 (2003) 141–150.

[12] A.G. Clark. "The size distribution of homozygous segments in the human genome." *The American Journal of Human Genetics* 65 (6) (1999) 1489–1492.

[13] D. Curtis, A.E. Vine, and J. Knight. "Study of regions of extended homozygosity provides a powerful method to explore haplotype structure of human populations." *The American Journal of Human Genetics* 72 (2) (2008) 261–278.

[14] R.A. Fisher. "A fuller theory of junctions in inbreeding." *Heredity* 8 (9154) 187–197.

[15] J. Gibson, N.E. Morton, and A. Collins. "Extended tracts of homozygosity in outbred human populations." *Human Molecular Genetics* 15 (5) (2006) 789–795.

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

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

[18] J. Hein, M.H. Schierup, and C. Wiuf. *Gene Genealogogies, Variation and Evolution: A Primer in Coalescent Theory.* Oxford: New York 2005.

[19] J. Hernandez-Sanchez, C.S. Haley, and J.A. Woolliams. "Prediction of IBD based on population history for fine gene mapping." *Genetics, Selection, and Evolution* 38 (2006) 231–252.

[20] W.G. Hill and J. Hernandez-Sanchez. "Prediction of multilocus identity-by-descent." *Genetics* 176 (2007) 2307–2315.

[21] W.G. Hill and A. Robertson. "Linkage disequilibrium in finite populations." *Theoretical and Applied Genetics* 38 (6) (1968) 1432–2242.

[22] A. Hodgkinson and A. Eyre-Walker. "Human triallelic sites: evidence for a new mutational mechanism?" *Genetics* 184 (2010) 233–241.

[23] B.N. Howie, P. Donnelly, and J. Marchini. "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies." *PLOS Genetics* 5 (2009).

[24] L. Huang, Y. Li, A.B. Singleton, J.A. Hardy, G. Abecasis, N.A. Rosenberg, and P. Scheet. "Genotype-imputation accuracy across worldwide human populations." *The American Journal of Human Genetics* 84 (2009) 235–250.

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

[26] I. Ionita-Laza, C. Lange, and N.M Laird. "Estimating the number of unseen variants in the human genome." *Proceedings of the National Academy of Sciences* 106(13) (2009) 5008–5013.

[27] The International HapMap Consortium. "A second generation human haplotype map of over 3.1 million SNPs." *Nature* 447 (2007) 661–678.

[28] E. Jakkula, K. Rehnstrom, T. Varilo, O.P.H. Pietilainen, T. Paunio, N.L. Pedersen, U. deFaire, M.-R. Jarvelin, J. Sahrinen, N. Freimer, S. Ripatti, S. Purcell, A. Collins, M.J. Daly, A. Palotie, and L. Peltonen. "The genome-wide patterns of variation expose significant substructure in a founder population." *The American Journal of Human Genetics* 83 (2008) 1–8.

[29] X. Ke, M.S. Taylor, and L.R. Cardon. "Singleton SNPs in the human genome and implications for genome-wide association studies." *European Journal of Human Genetics* 16 (2008) 506–515.

[30] A. Keinan, J.C. Mullikan, N. Patterson, and D. Reich. "Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans." *Nature Genetics* 39 (2007) 1251–1255.

[31] A. Kong, G. Masson, M.L. Frigge, A. Gylfason, P. Zusmanovich, G. Thorleifsson, P.I. Olason, A. Ingason, S. Steinberg, T. Rafnar, P. Sulem, M. Mouy, F. Jonsson, U. Thorsteinsdottir, D.F. Gudbjartsson, H. Stefansson, and K. Stefansson. "Detection of sharing by descent, long-range phasing and haplotype imputation." *Nature Genetics* 40.9 (2008) 1068–1075.

[32] G. Laval, E. Patin, L.B. Barriero, and L. Quintana-Murci. "Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions." *PLOS ONE* (2010).

[33] A.-L. Leutnegger, B. Prum, E. Genin, C. Verny, A. Lemainque, F. Clerget-Darpoux, and E.A. Thompson. "Estimation of the inbreeding coefficient through use of genomic data." *American Journal of Human Genetics* 73 (2003) 516–523.

[34] A.-L. Leutnegger, A. Labalme, E. Genin, A. Toutain, E. Steichen, F. Clerget-Darpoux, and P. Edery. "Using genomic inbreeding coefficient estimates for homozygosity mapping of rare recessive traits: application to

Taybi-Linder syndrome." *American Journal of Human Genetics* 79 (2006) 62–66.

[35] Q. Li and R. Wu. "A multilocus model for constructing a linkage disequilibrium map in human populations." *Statistical Applications in Genetics and Molecular Biology* 8 (1) Article 18 (2009).

[36] Y. Li, C. Willer, S. Sanna, and G. Abecasis. "Genotype imputation." *Annual Reviews of Genomics and Human Genetics* 10 (2009)387–406.

[37] F. Liu, C.M. van Duijn, and Y.S. Aulchenko. "Ignoring distant genealogic loops leads to false-positives in homozygosity mapping." *Annals of Human Genetics* 70 (2006) 1–6.

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

[39] J. Marchini and B. Howie. "Genotype imputation for genome-wide association studies." *Nature Reviews: Genetics* 11 (2010) 499–511.

[40] G.T. Marth, E. Czabarka, J. Murvai, and S.T. Sherry. "The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations." *Genetics* 166 (2004) 351–372.

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

[42] G.A.T. McVean. "A genealogical intepretation of linkage disequilibrium." *Genetics* 162 (2002) 987–991.

[43] G.A.T. McVean and N.J. Cardin. "Approximating the coalescent with recombination." *Philosophical Transactions of the Royal Society B* 360 (2005) 1387–1393.

[44] P. Mellars. "Going East: New genetic and archaeological perspectives on the modern human colonization of Eurasia." *Science* 11 (2006) 796–800.

[45] T.H.E Meuwissen and M.E. Goddard. "The use of marker haplotypes in animal breeding schemes." *Genetics, Selection, and Evolution* 28 (1996) 161–176.

[46] T.H.E. Meuwissen and M.E. Goddard. "Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci." *Genetics* 155 (2000) 421–430.

[47] T.H.E. Meuwissen and M.E. Goddard. "Prediction of identity by descent probabilities from marker-haplotypes." *Genetics, Selection, and Evolution* 33 (2001) 605–634.

[48] T.H.E. Meuwissen, B.J. Hayes, and M.E. Goddard. "Prediction of total genetic value using genome-wide dense marker maps." *Genetics* 157 (4) (2001) 1819–1829.

[49] T.H.E. Meuwissen and M.E. Goddard. "Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data." *Genetics, Selection, and Evolution* 36 (2004) 261–279.

[50] T. Ohta and M. Kimura. "Linkage disequilibrium between two segretation nucleotide sites under the steady flux of mutations in a finite population." *Genetics* 68 571–580.

[51] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, and M.A.R Ferreira. "PLINK: a tool set for whole-genome association and population-based linkage analyses." *American Journal of Human Genetics* 81 (2007) 559–575.

[52] N.A. Sheehan and T. Egeland. "Adjusting for founder relatedness in a linkage analysis using prior information." *Human Heredity* 65 (2008) 221–231.

[53] C.C.A. Spencer, Z. Su, P. Donnelly, and J. Marchini. "Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip." *PLOS Genetics* 5(5) (2009) e1000477.

[54] P. Stam. "The distribution of the fraction of the genome identical by descent in finite random mating populations." *Genetical Research* 35 (1980) 131–155.

[55] J.A. Sved. "Linkage disequilibrium and homozygosity of chromosome segments in finite populations." *Theoretical Population Biology* 2 (2) (1971) 125–141.

[56] J.D. Terwilliger and T. Hiekkalinna. "An utter refutation of the fundamental theorem of the HapMap." *European Journal of Human Genetics* 14 (2006) 426–437.

[57] E.A. Thompson. "The IBD process along four chromosomes." *Theoretical Population Biology* 73 (3) (2008) 369–373.

[58] L.N. Trefethen. *Numerical Linear Algebra.* SIAM, Philadelphia: 1997.

[59] J. Wakeley. *Coalescent Theory: An Introduction.* Roberts and Company, Greenwood: 2009.

[60] J. Wang, W. Wang, R. Li, et al. "The diploid genome sequence of an Asian individual." *Nature* 456 (2008) 60–65.

[61] B.S. Weir and W.G. Hill. "Effect of mating structure on variation in linkage disequilibrium." *Genetics* 95 477–488.