

## HapMap and Mapping Genes for Cardiovascular Disease

Kiran Musunuru, MD, PhD; Sekar Kathiresan, MD

A key goal of biomedical science is to understand why individuals differ in their susceptibility to disease. Family history is among the established risk factors for most forms of cardiovascular disease, in part because inherited DNA sequence variants play a causal role in disease susceptibility. Consequently, the search for these variants has intensified over the past decade.<sup>1-3</sup> One class of DNA sequence variants takes the form of single nucleotide changes (single nucleotide polymorphisms, or SNPs), usually with two variants or alleles for each SNP.<sup>4</sup> SNPs are scattered throughout the 23 pairs of chromosomes of the human genome, and roughly 11 million common polymorphisms (ie, those >1% frequency) are estimated to exist.<sup>5</sup> A combination of SNP alleles along a chromosome is termed a haplotype.

The International Haplotype Map Project was designed to create a public genome-wide database of common SNPs and, consequently, enable systematic studies of most common SNPs for their potential role in human disease.<sup>6-8</sup> We review the following: (1) the concept of linkage disequilibrium or allelic association, (2) the HapMap project, and (3) several examples of the utility of HapMap data in genetic mapping for cardiovascular disease phenotypes.

### Linkage Disequilibrium: Correlation Among SNPs

Groups of SNPs across the genome are correlated with each other, a phenomenon known as linkage disequilibrium (LD) or allelic association. To understand how LD arises, one needs to recall that during meiosis, recombination occurs at multiple sites between each pair of chromosomes, thus providing for an extra source of genetic variability to pass on to offspring. This is not a random process that occurs with equal probability at every place along a chromosome; rather, there are large stretches of DNA along which there is a very low probability of recombination, punctuated by recombination “hotspots,” where it occurs relatively more often. The consequence is that the stretches of DNA between hotspots tend to stay together—in what are referred to as “haplotype blocks”—as they are passed along from generation to generation.<sup>9,10</sup>

To understand how SNPs arise and become correlated with other SNPs, consider the following hypothetical example (Figure 1). At some time in the remote past, a mutation of a

single nucleotide in a single individual results in a base change from “A” (adenine) to “G” (guanine). Previously, there was no variation at that site in the population, with everybody else having an “A” allele at the position in both copies of the gene (one copy on each of the paired chromosomes). There is an SNP nearby that is a “C” (cytosine) allele 50% of the time and a “T” (thymine) allele the other 50%. It so happens that the A→G mutation arose on a chromosome in which the identity of the nearby SNP is a “C” allele. If the mutation is not so harmful that natural selection would cull it out of the population, it is transmitted to many successive generations; in this example, it spreads through the population until 10% of chromosomes in the population have a “G” allele at the position.

Because the new A/G SNP and the old C/T SNP are close together with no recombination hotspots between them, resulting in essentially no recombination between them in successive generations, all chromosomes with a “G” allele at the first SNP also have a “C” allele at the second SNP. In contrast, chromosomes with an “A” allele at the first SNP have some chance of having a “C” allele at the second SNP, with the others having a “T” allele, reflecting the state of affairs before the origin of the new SNP. The C/T SNP has become correlated with the A/G SNP, and knowledge of the allele at one of the SNPs confers some information about the allele at the other SNP.

SNPs within a haplotype block and, to a lesser extent, SNPs in nearby haplotype blocks tend to remain correlated over time. The degree of correlation or LD can be quantified in two different ways, the calculated values of  $D'$  and  $r^2$ .  $D'$  measures the deviation of haplotype frequencies from linkage equilibrium and  $r^2$  is a measurement of correlation between a pair of variables.  $r^2$  is particularly useful in genetic mapping—when  $r^2=1$  (the maximum value), knowing the genotypes of alleles of one SNP is perfectly predictive of the genotypes of another SNP. (Please see Wang et al<sup>11</sup> for an expanded discussion of these concepts and the mathematical formulations.) Although any haplotype made up of  $n$  SNPs (each with two possible alleles) potentially has  $2^n$  combinations of SNP alleles, far fewer combinations are actually seen in a population because of correlation among the SNPs. In principle, knowledge of the correlation structure among all SNPs in the genome—as represented by a vast array of

From the Cardiovascular Research Center and Cardiology Division, and Center for Human Genetic Research, Massachusetts General Hospital, Boston and Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Mass (K.M., S.K.).

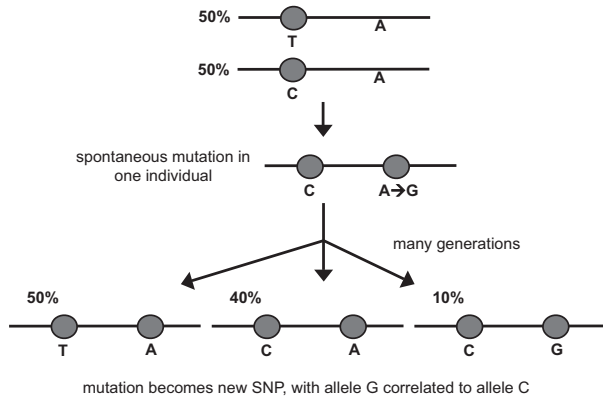
Correspondence to Sekar Kathiresan, MD, Cardiovascular Research Center and Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge St, CPZN 5.252, Boston, MA 02114. E-mail skathiresan@partners.org

(*Circ Cardiovasc Genet.* 2008;1:66-71.)

© 2008 American Heart Association, Inc.

*Circ Cardiovasc Genet* is available at <http://circgenetics.ahajournals.org>

DOI: 10.1161/CIRCGENETICS.108.813675



**Figure 1.** Genesis of a new SNP correlated with an old SNP. Initially there is only 1 SNP (T/C) in the region depicted. A spontaneous mutation in a single individual converts an A nucleotide into a G nucleotide. After many generations, a new A/G polymorphism has emerged, with 10% of the population having the G allele. Because no recombination between the two SNPs has occurred, all chromosomes with the G allele have a C allele at the other SNP. SNP indicates single nucleotide polymorphism.

pair-wise  $D'$  and  $r^2$  values and haplotype combinations—would provide a powerful tool with which to study human genetics and disease.

### The HapMap Project

The International HapMap Project began in October 2002 with the purpose of identifying millions of SNPs throughout the genome, determining the allele frequencies at each SNP, and determining the correlations between SNPs.<sup>6</sup> Drawing on 269 DNA samples from individuals of four different ethnicities—90 residents of Utah in the United States with Northern and Western European ancestry, 90 Yoruba people in Nigeria, 44 Japanese people in Tokyo, and 45 Han Chinese people in Beijing—HapMap has now genotyped more than 3 million SNPs in each of these populations and published the results in a public database.<sup>7,8</sup>

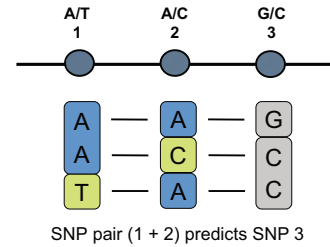
Analyses of this data have yielded a number of important insights into human genetic variation. For example, although the 4 ethnic groups included in the HapMap Project share most SNPs, the allele frequencies at these SNPs can vary widely among the groups. Yoruban individuals appear to have many more rare alleles (frequency  $<5\%$ ) than the other groups, which may reflect the fact that European and Asian populations are “younger” (ie, descended from offshoots of an ancestral African population).<sup>7</sup> Although recombination hotspots are widely distributed across the genome, they are more common near telomeres (the ends) of chromosomes and more rare near the centromeres of chromosomes.<sup>7</sup> SNPs in the vicinity of recombination hotspots have less correlation with surrounding SNPs compared with SNPs at some distance from hotspots.<sup>7</sup>

Although these findings are of biological interest, there are other features of the HapMap data that are particularly useful for the study of human disease.

### Uses of HapMap in Genetic Mapping

#### Coverage of the Genome

The large database of genome-wide SNPs provided by HapMap has allowed efficient design of genetic association



**Figure 2.** A 2-marker SNP set tags a third SNP. In this example, only SNPs 1 and 2 have been directly genotyped. Because HapMap has only 3 possible haplotypes for these SNPs (A-A-G, A-C-C, T-A-C), in all cases the identity of SNP 3 can be inferred from a multimarker test comprising SNPs 1 and 2. Note that neither SNP 1 nor SNP 2 alone can predict SNP 3. SNP indicates single nucleotide polymorphism.

studies. A comprehensive test of common SNPs would theoretically involve the genotyping of all 11 million common SNPs in patients with disease and individuals free of disease. However, the correlation structure among SNPs provided by HapMap allows investigators to genotype far fewer SNPs while still retaining statistical power to find regions of the genome associated with disease. Because a given SNP may be in LD with another SNP in the same region, knowledge of the genotype of the first SNP of the pair may be sufficient to infer the genotype of the other SNP, thereby acting as a “tagging” SNP for the other SNP. In this way, a single SNP can potentially serve to “tag” a number of other SNPs. A judiciously chosen panel of approximately 300,000 to 500,000 HapMap SNPs is sufficient to capture the information content of the full 3 million SNPs in HapMap individuals of European or Asian descent, whereas a panel of approximately 1.1 million SNPs is required in Yoruban individuals.<sup>8</sup> Furthermore, panels of tagging SNPs chosen for each HapMap ethnicity have been shown to provide similar power for non-HapMap study populations of the same ethnicity.<sup>12</sup> Greater than 60% coverage of the genome is provided by commercially available SNP “arrays” or “chips” that can interrogate several hundred thousand SNPs in a single experiment<sup>13,14</sup>; successive generations of these chips that interrogate upward of a million SNPs will provide even better coverage, resulting in increased statistical power to find disease associations.

### Multimarker Tests and Imputation

Increased statistical power can also be achieved by using multimarker tests, in which haplotypes of correlated SNPs are used to tag other SNPs. This is possible because the HapMap database reveals which haplotypes are found in populations. For example, for a set of 3 SNPs for which each SNP has 2 possible alleles, there are 8 possible haplotype combinations, but only a few haplotypes may be seen in HapMap. Thus knowledge of the identity of the first SNP or the second SNP alone may not be sufficient to infer the identity of the third SNP, but the combination of the first and second SNPs may predict the third SNP (Figure 2). When used for tagging in this fashion, 2-marker SNP sets have been shown to significantly improve genome coverage by SNP chips—in the case of the Affymetrix 500K Mapping Array Set, from 66% to 78%.<sup>14</sup>

This process of using genotyped SNPs to infer the identities of additional SNPs, without the need for further genotyping, is termed imputation. A validation study in which imputation was performed to predict the identities of SNPs that had also been directly genotyped found greater than 98% agreement between the results in individuals of European ancestry.<sup>15</sup>

Imputation is particularly useful when combining genome-wide data sets that were obtained with different SNP genotyping platforms. For example, in a recent meta-analysis of 3 genome-wide association studies with lipid traits, 2 of the studies were performed using the Affymetrix 500K Mapping Array Set, with the third using the Illumina HumanHap300 BeadChip.<sup>16,17</sup> Although there was only a small overlap of SNPs directly genotyped by the 2 platforms ( $\approx 45\,000$  SNPs), imputation using the haplotypes in the HapMap database generated a greatly enlarged set of genotyped and imputed SNPs ( $\approx 2.2$  million) for all individuals in the 3 studies.<sup>16,17</sup> Combining information in this way enabled the discovery of 8 new gene regions related to low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and/or triglycerides.<sup>16,17</sup>

### Interpreting Association Results

The HapMap database facilitates the interpretation of a genetic association result and can help arrive at an “associated” or “critical” interval, a region of the genome likely to contain the causal polymorphism. Given an index SNP with definitive statistical evidence for association with a trait or disease of interest, one can refer to the HapMap database and use the correlation structure to identify other SNPs in LD and thereby define the region in which to look for the causal variant. For example, several genome-wide association studies have highlighted an association of common noncoding SNPs on chromosome 9p21 with coronary artery disease or myocardial infarction.<sup>18–20</sup> Given the public HapMap resource and such an association result, investigators are readily able to evaluate the patterns on SNP correlation around the index SNP(s) and delimit the region of association. Using data derived from HapMap, Schunkert et al described the correlation structure for SNPs on 9p21 (Figure 3).<sup>21</sup> SNPs spanning a distance of  $\approx 60$  kilobases are correlated with one of the index SNPs (rs13330499) with  $r^2$  of at least 0.5. The search for a causal variant for coronary artery disease has now been narrowed from the entire genome to a small span of DNA sequence.

Another such example involves genetic variation on chromosome 1p13 associated with both low-density lipoprotein cholesterol<sup>16,17,22,23</sup> and coronary artery disease,<sup>20</sup> with multiple genome-wide association studies identifying rs599839 as an index SNP for these phenotypes. On interrogation of this SNP in HapMap, it is evident that the set of SNPs in strong LD with rs599839 span a region roughly 100 kilobases in size. In this region lie at least 4 genes—*CELSR2*, *PSCRC1*, *MYBPHL*, and *SORT1*—and any of these may represent the gene influencing both low-density lipoprotein cholesterol and coronary artery disease. These genes may now be prioritized for the next set of studies (ie, deep sequencing of the

100-kilobase region in humans and manipulation of these 4 positional candidate genes in cells or mice).

HapMap data may also facilitate “fine mapping” of an initial association result. In fine mapping, additional SNPs (beyond the index SNP) within an associated interval are tested to see if they provide stronger evidence for association. As an example, genome-wide association mapping for triglyceride levels identified an SNP in the glucokinase regulatory protein gene (*GCKR*) as being highly associated with triglyceride levels.<sup>24,25</sup> The index *GCKR* SNP was intronic (rs780094) and the associated interval spanned  $\approx 400$  kilobases and contained 17 genes. To fine-map across the associated interval, an additional 120 SNPs were selected from HapMap to tag the associated interval. With fine mapping, a common missense SNP in *GCKR* (rs1260326) that changes the amino acid 446 of the protein from proline to leucine emerged as the strongest association signal.<sup>25</sup> These results now raise the next testable hypothesis, that the coding variant affects the function of *GCKR* (possibly by altering binding to glucokinase) and thereby alters triglyceride and glucose levels.

### Limitations of HapMap

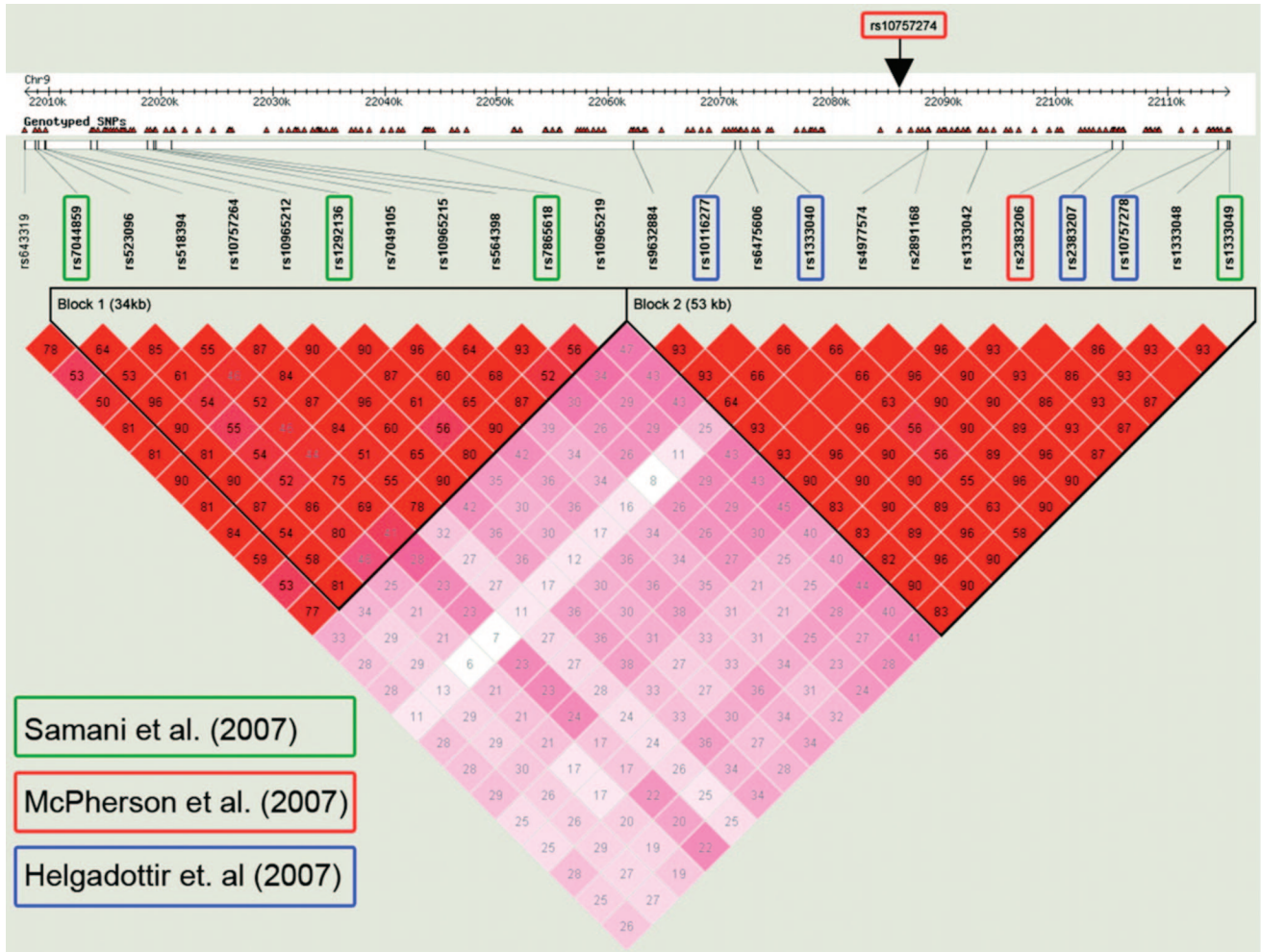
A major limitation of the HapMap project is that low-frequency SNPs (ie, with minor allele frequencies between 0.5% and 5%) are incompletely captured in the database. Rare SNPs ( $<0.5\%$  frequency) are even more underrepresented. As it is likely that an important fraction of disease-causing variants are of low frequency or rare, these will be difficult to identify through the use of tagging SNPs selected from the HapMap database.

An additional limitation is that genotypes are only available for individuals from 4 ethnic groups (European descent in Utah, Yoruban, Japanese, and Han Chinese) at the time of the second phase of HapMap. Although it has been shown that the correlation structures in each of these groups remains valid in other cohorts of the same ethnicity,<sup>12</sup> this may not hold true for ethnicities not represented in HapMap.

Both of these shortcomings are to be squarely addressed by new projects that are now underway. The third phase of HapMap will include genotyping of SNPs in individuals of additional ethnicities beyond the original 4 and thus will extend the utility of HapMap to a wider variety of populations under study worldwide.<sup>8</sup> On an even larger scale, the 1000 Genomes Project, launched in January 2008, aims to fully sequence the genomes of at least 1000 individuals from 11 ethnic/regional groups (including individuals from the original HapMap Project).<sup>26</sup> This effort will markedly increase the number of low-frequency SNPs available for study and, with integration into the existing HapMap database, allow for an extension of the correlation structure to these low-frequency SNPs.

### Conclusion

HapMap is a public resource that has critically enabled genome-wide association mapping using common DNA sequence variants. These genetic mapping studies have proven useful in identifying novel contributors to cardiovascular traits including myocardial infarction,<sup>18–20</sup> atrial fibrillation,<sup>27</sup>



**Figure 3.** Correlation structure at the 9p21 locus associated with myocardial infarction. Displayed are the linkage disequilibrium relations (as defined by the  $r^2$  metric) between pairs of SNPs in the region, with each square representing the pair-wise strength and significance of correlation, with red indicating strong correlation (high  $r^2$  value) and white indicating weak correlation (low  $r^2$  value). The index SNPs with the strongest association evidence from 3 genome-wide association studies,<sup>18–20</sup> each of which used a different SNP genotyping platform, are indicated with boxes; the other SNPs were identified from the HapMap database. Reprinted from Schunkert et al<sup>21</sup> with permission from the American Heart Association. Copyright 2008 American Heart Association.

lipid levels,<sup>16,17,22,23</sup> diabetes mellitus,<sup>24,28,29</sup> statin-induced myopathy,<sup>30</sup> electrocardiographic QT interval,<sup>31</sup> and abdominal aortic aneurysm.<sup>32</sup> Further application of tools such as HapMap should clarify the full spectrum of DNA sequence differences that confer susceptibility to cardiovascular disease.

### Acknowledgments

The authors thank Dr Mark J. Daly, who provided the illustration on which Figure 2 is based.

### Sources of Funding

Dr Kathiresan is supported by a Doris Duke Charitable Foundation Clinical Scientist Development Award, a charitable gift from the Fannie E. Rippel Foundation, the Donovan Family Foundation, and a K23 career development award from the United States National Institutes of Health. Dr Musunuru is supported by a T32 grant in Cell and Molecular Training for Cardiovascular Biology from the National Institutes of Health.

### Disclosures

Dr Musunuru has received consulting fees from Alnylam Pharmaceuticals and honoraria from the American College of Cardiology

Foundation within the last year. Dr Kathiresan reports no potential conflicts.

### References

- Lloyd-Jones DM, Nam BH, D'Agostino RBS, Levy D, Murabito JM, Wang TJ, Wilson PW, O'Donnell CJ. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA*. 2004;291:2204–2211.
- Fox CS, Parise H, D'Agostino RBS, Lloyd-Jones DM, Vasan RS, Wang TJ, Levy D, Wolf PA, Benjamin EJ. Parental atrial fibrillation as a risk factor for atrial fibrillation in offspring. *JAMA*. 2004;291:2851–2855.
- Lee DS, Pencina MJ, Benjamin EJ, Wang TJ, Levy D, O'Donnell CJ, Nam BH, Larson MG, D'Agostino RB, Vasan RS. Association of parental heart failure with risk of heart failure in offspring. *N Engl J Med*. 2006;355:138–147.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D, International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 2001;409:928–933.

5. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet.* 2001;27:234–236.
6. The International HapMap Consortium. The International HapMap Project. *Nature.* 2003;426:789–796.
7. International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005;437:1299–1320.
8. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, 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 SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, 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, Wayne MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, LeBoeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verne A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermizakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thrisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Anigwu T, Marshall PA, Nkwodimma C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niihara N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449:851–861.
9. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet.* 2001; 29:229–232.
10. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science.* 2002; 296:2225–2229.
11. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet.* 2005;6: 109–118.
12. de Bakker PI, Burtt NP, Graham RR, Guiducci C, Yelensky R, Drake JA, Bersaglieri T, Penney KL, Butler J, Young S, Onofrio RC, Lyon HN, Stram DO, Haiman CA, Freedman ML, Zhu X, Cooper R, Groop L, Kolonel LN, Henderson BE, Daly MJ, Hirschhorn JN, Altshuler D. Transferability of tag SNPs in genetic association studies in multiple populations. *Nat Genet.* 2006;38:1298–1303.
13. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. *Nat Genet.* 2006;38:659–662.
14. Pe'er I, de Bakker PI, Maller J, Yelensky R, Altshuler D, Daly MJ. Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet.* 2006;38:663–667.
15. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;39:906–913.
16. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, Strait J, Duren WL, Maschio A, Busonero F, Mulas A, Albai G, Swift AJ, Morken MA, Narisu N, Bennett D, Parish S, Shen H, Galan P, Meneton P, Hercberg S, Zelenika D, Chen WM, Li Y, Scott LJ, Scheet PA, Sundvall J, Watanabe RM, Nagaraja R, Ebrahim S, Lawlor DA, Ben-Shlomo Y, Davey-Smith G, Shuldiner AR, Collins R, Bergman RN, Uda M, Tuomilehto J, Cao A, Collins FS, Lakatta E, Lathrop GM, Boehnke M, Schlessinger D, Mohlke KL, Abecasis GR. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet.* 2008;40:161–169.
17. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen MR, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altshuler DM, Orho-Melander M. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet.* 2008;40:189–197.
18. McPherson R, Pertsemidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science.* 2007;316:1488–1491.
19. Helgadóttir A, Thorleifsson G, Manolescu A, Gretarsdóttir S, Blondal T, Jonasdóttir A, Jonasdóttir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdóttir S, Jonsdóttir T, Palsson S, Einarsson T, Gunnarsdóttir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgerirsson G, Thorsteinsdóttir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science.* 2007;316:1491–1493.
20. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, König IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwendijk W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H, WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. *N Engl J Med.* 2007;357:443–453.
21. Schunkert H, Gotz A, Braund P, McGinnis R, Tregouet DA, Mangino M, Linsel-Nitschke P, Cambien F, Hengstenberg C, Stark K, Blankenberg S, Tiret L, Ducimetiere P, Keniry A, Ghorji MJ, Schreiber S, El Mokhtari NE, Hall AS, Dixon RJ, Goodall AH, Liptau H, Pollard H, Schwarz DF, Hothorn LA, Wichmann HE, König IR, Fischer M, Meisinger C, Ouwendijk W, Deloukas P, Thompson JR, Erdmann J, Ziegler A, Samani NJ, Cardiogenics Consortium. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation.* 2008;117:1675–1684.
22. Wallace C, Newhouse SJ, Braund P, Zhang F, Tobin M, Falchi M, Ahmadi K, Dobson RJ, Marcano AC, Hajat C, Burton P, Deloukas P, Brown M, Connell JM, Dominiczak A, Lathrop GM, Webster J, Farrall M, Spector T, Samani NJ, Caulfield MJ, Munroe PB. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am J Hum Genet.* 2008;82: 139–149.
23. Sandhu MS, Waterworth DM, Debenham SL, Wheeler E, Papadakis K, Zhao JH, Song K, Yuan X, Johnson T, Ashford S, Inouye M, Luben R, Sims M, Hadley D, McArdle W, Barter P, Kesaniemi YA, Mahley RW, McPherson R, Grundy SM, Wellcome Trust Case Control Consortium, Bingham SA, Khaw KT, Loos RJ, Waeber G, Barroso I, Strachan DP, Deloukas P, Vollenweider P, Wareham NJ, Mooser V. LDL-cholesterol concentrations: a genome-wide association study. *Lancet.* 2008;371: 483–491.
24. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlsson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirm GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science.* 2007;316:1331–1336.

25. Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, Roos C, Tewhey R, Rieder MJ, Hall J, Abecasis G, Tai ES, Welch C, Arnett DK, Lyssenko V, Lindholm E, Saxena R, de Bakker PI, Burt N, Voight BF, Hirschhorn JN, Tucker KL, Hedner T, Tuomi T, Isomaa B, Eriksson KF, Taskinen MR, Wahlstrand B, Hughes TE, Parnell LD, Lai CQ, Berglund G, Peltonen L, Vartiainen E, Jousilahti P, Havulinna AS, Salomaa V, Nilsson P, Groop L, Altshuler D, Ordovas JM, Kathiresan S. A common missense variant in the glucokinase regulatory protein gene (GCKR) is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. *Diabetes*. 2008 Aug 4 [Epub ahead of print].
26. 1000 Genomes Project. Meeting report: a workshop to plan a deep catalog of human genetic variation. Available at: [http://www.1000genomes.org/bcms/1000\\_genomes/Documents/1000Genomes-MeetingReport.pdf](http://www.1000genomes.org/bcms/1000_genomes/Documents/1000Genomes-MeetingReport.pdf). Accessed August 6, 2008.
27. Gudbjartsson DF, Arnar DO, Helgadóttir A, Gretarsdóttir S, Holm H, Sigurdsson A, Jonasdóttir A, Baker A, Thorleifsson G, Kristjánsson K, Pálsson A, Blöndal T, Sulem P, Backman VM, Hardarson GA, Palsdóttir E, Helgason A, Sigurjónsdóttir R, Sverrisson JT, Kostulas K, Ng MC, Baum L, So WY, Wong KS, Chan JC, Furie KL, Greenberg SM, Sale M, Kelly P, MacRae CA, Smith EE, Rosand J, Hillert J, Ma RC, Ellinor PT, Thorgeirsson G, Gulcher JR, Kong A, Thorsteinsdóttir U, Stefánsson K. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature*. 2007;448:353–357.
28. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*. 2007;316:1341–1345.
29. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, Wellcome Trust Case Control Consortium (WTCCC), McCarthy MI, Hattersley AT. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*. 2007;316:1336–1341.
30. SEARCH Collaborative Group, Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, Gut I, Lathrop M, Collins R. SLCO1B1 variants and statin-induced myopathy: a genomewide study. *N Engl J Med*. 2008;359:789–799.
31. Arking DE, Pfeufer A, Post W, Kao WH, Newton-Cheh C, Ikeda M, West K, Kashuk C, Akyol M, Perz S, Jalilzadeh S, Illig T, Gieger C, Guo CY, Larson MG, Wichmann HE, Marban E, O'Donnell CJ, Hirschhorn JN, Kaab S, Spooner PM, Meitinger T, Chakravarti A. A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat Genet*. 2006;38:644–651.
32. Helgadóttir A, Thorleifsson G, Magnusson KP, Gretarsdóttir S, Steinthorsdóttir V, Manolescu A, Jones GT, Rinkel GJ, Blankensteijn JD, Ronkainen A, Jaaskelainen JE, Kyo Y, Lenk GM, Sakalihasan N, Kostulas K, Gottsater A, Flex A, Stefánsson H, Hansen T, Andersen G, Weinsheimer S, Borch-Johnsen K, Jorgensen T, Shah SH, Quyyumi AA, Granger CB, Reilly MP, Austin H, Levey AI, Vaccarino V, Palsdóttir E, Walters GB, Jonsdóttir T, Snorraddóttir S, Magnúsdóttir D, Gudmundsson G, Ferrell RE, Sveinbjörnsdóttir S, Hernesniemi J, Niemela M, Limet R, Andersen K, Sigurdsson G, Benediktsson R, Verhoeven EL, Teijink JA, Grobbee DE, Rader DJ, Collier DA, Pedersen O, Pola R, Hillert J, Lindblad B, Valdimarsson EM, Magnadóttir HB, Wijmenga C, Tromp G, Baas AF, Ruigrok YM, van Rij AM, Kuivaniemi H, Powell JT, Mattiasson SE, Gulcher JR, Thorgeirsson G, Kong A, Thorsteinsdóttir U, Stefánsson K. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet*. 2008;40:217–224.

---

KEY WORDS: cardiovascular diseases ■ genes ■ mapping

## HapMap and Mapping Genes for Cardiovascular Disease

Kiran Musunuru and Sekar Kathiresan

*Circ Cardiovasc Genet.* 2008;1:66-71

doi: 10.1161/CIRCGENETICS.108.813675

*Circulation: Cardiovascular Genetics* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Copyright © 2008 American Heart Association, Inc. All rights reserved.

Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://circgenetics.ahajournals.org/content/1/1/66>

**Permissions:** Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Circulation: Cardiovascular Genetics* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the [Permissions and Rights Question and Answer](#) document.

**Reprints:** Information about reprints can be found online at:  
<http://www.lww.com/reprints>

**Subscriptions:** Information about subscribing to *Circulation: Cardiovascular Genetics* is online at:  
<http://circgenetics.ahajournals.org/subscriptions/>