Chapter 2
Candidate Gene Association Studies in Stroke

Elizabeth G. Holliday, Christopher J. Oldmeadow, Jane M. Maguire, and John Attia

Introduction

Although some forms of stroke are inherited in a Mendelian manner, being caused by a single genetic locus of strong effect, most strokes are sporadic and, like other complex diseases such as asthma and diabetes mellitus, have a genetic component that is complex and involves multiple loci of modest effect. Controversy surrounds whether these genetic variants are likely to be common (>1–5% prevalence) across most populations, each with very small effect sizes (the so-called common disease-common variant hypothesis), or whether these variants are likely to be rare (<1%), perhaps even unique to families or individuals, with larger effect sizes (the so-called common disease-rare variant hypothesis). In either case, a common approach over the last 10 years has been to test association between common genetic variants and stroke using traditional population-based designs such as case-control and cohort studies. Genetic association studies have steadily increased in popularity;

E.G. Holliday, B.Sc. (Hons), Ph.D. (✉)
Centre for Clinical Epidemiology and Biostatistics,
School of Medicine and Public Health, Faculty of Health, University of Newcastle,
University Drive, Callaghan, NSW 2308, Australia
e-mail: liz.holliday@newcastle.edu.au

C.J. Oldmeadow, Ph.D., BMath (Hon)
School of Medicine and Public Health, Faculty of Health, University of Newcastle,
Newcastle, NSW, Australia

J.M. Maguire, Ph.D., RN, B.A., BNurs (Hons)
School of Nursing and Midwifery, University of Newcastle,
Callaghan, NSW, Australia

J. Attia, B.Sc., M.Sc., M.D., Ph.D.
Department of Medicine, John Hunter Hospital, University of Newcastle,
New Lambton, NSW, Australia

P. Sharma, J.F. Meschia (eds.), *Stroke Genetics*,
DOI 10.1007/978-0-85729-209-4_2, © Springer-Verlag London 2013

the number of published studies has increased from just under 500 in the year 2000 to nearly 8,500 in 2010 [1, 2], with about 95% of these being candidate gene association studies (CGAS). The remainder were genome-wide association studies (GWAS) (see Chap. 4).

The two approaches to testing for disease-associated variants are fundamentally different; the GWAS is a data-mining tool where no knowledge of the possible causative mechanisms is needed, whereas the CGAS is hypothesis driven, where putative functional variants in genes are specified a priori and tested for association with disease risk. In this chapter, we cover the study designs in CGAS, sources of bias, advantages and disadvantages of this method, pooling CGAS using meta-analysis, the uneasy relationship between CGAS and GWAS, and the role of CGAS in elucidating causation using the method of Mendelian randomization. The focus will be mainly methodological; for encyclopaedic reviews of actual genetic variants associated with stroke, readers are referred to other sources [3–5], as well as to the Human Genome Epidemiology (HuGE) Navigator, a continuously updated electronic tool that can help identify genetic association literature for specific diseases and genes [1, 2].

Study Design in Candidate Gene Association Studies

Family-based designs using trios and sibling pairs have been used in the past (see reviews [6, 7]) but, despite some advantages, have recently fallen out of favor due to the difficulty of recruiting siblings and parents for late-onset diseases [8, 9]. Currently, the most popular designs for genetic association studies are case-control and cohort studies. In the former, a particular genetic variant is tested in groups of ascertained cases and controls, and in the latter, a genetic variant is tested on blood drawn at baseline and the disease outcome determined at follow-up. The choice of genetic variant is most commonly a single nucleotide polymorphism (SNP) but can equally be a DNA repeat (which includes minisatellite, microsatellite, variable number tandem repeat, or trinucleotide repeat), an insertion/deletion polymorphism, or a copy number variant (a large genomic region, >1 kb, whose copy number varies from the expected diploid state due to deletion or duplication) (Fig. 2.1).

Genetic association studies are thought to be more robust against confounding than traditional (nongenetic) association studies. For example, recall bias, which can occur when cases overcall previous risk factor exposure knowing they have the outcome of interest, does not influence genetic associations. There is also a lack of temporal bias, which can occur when a risk factor of interest is measured after disease development and the direction of causality is unclear. The constitutional genetic sequence is present at birth and does not change with disease (excepting changes in tumor DNA in the case of cancer).

However, there are new sources of bias to consider in genetic association studies, and multiple articles have been written about these, both generally (e.g., [11]), and in the context of stroke [12, 13]. This has led to an extension of the guidelines for

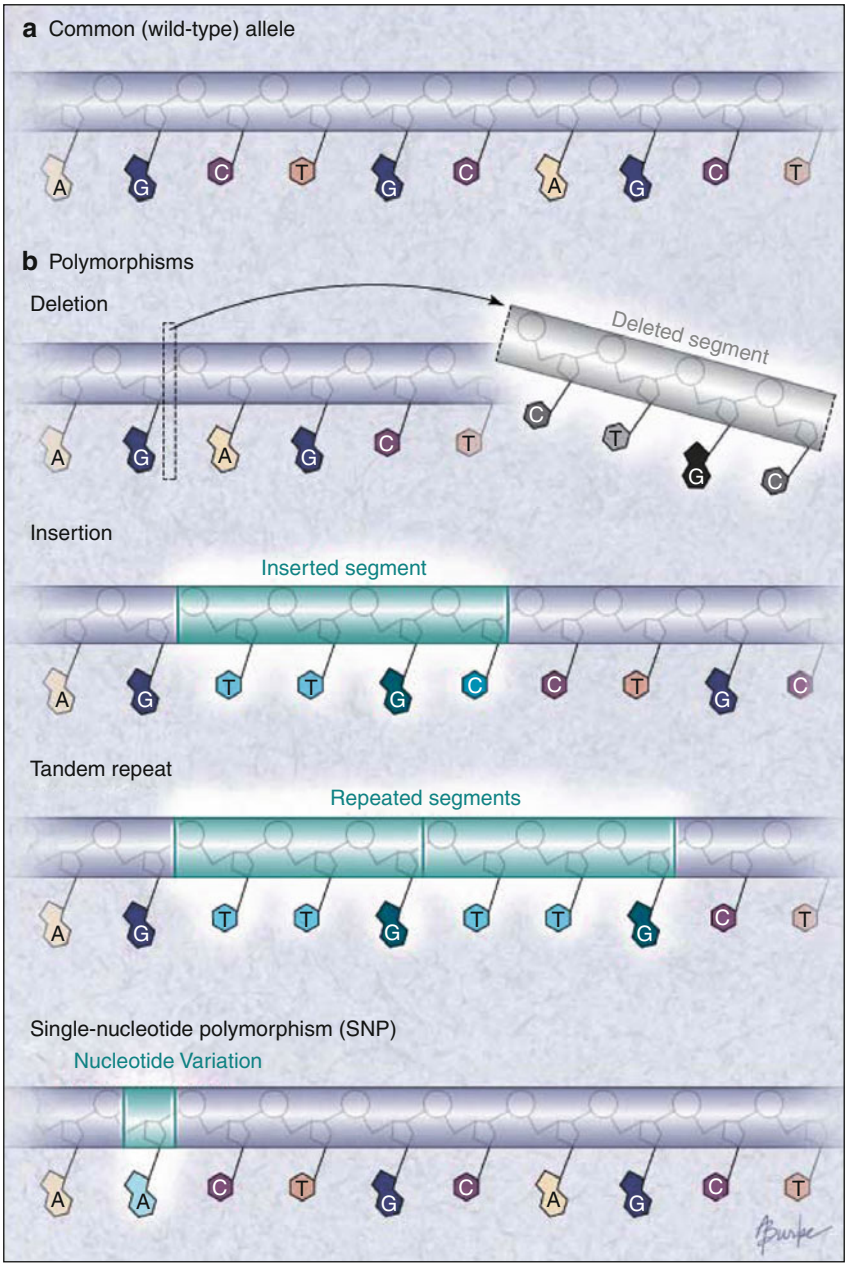

Fig. 2.1 Common (wild-type) allele (**a**) and four types of genetic polymorphisms (**b**). DNA polymorphisms include deletions (in which a DNA sequence is missing compared with the common allele) and insertions (in which a DNA sequence is added compared with the common allele). Repeats may also occur in which the same sequence repeats multiple times. Depending on the size of the repeating unit and the number of repeats, these variants may have different names, such as satellites, microsatellites, minisatellites, or copy number variants. Single nucleotide polymorphisms (*SNPs*), variations at a single base-pair location, are the most common types of polymorphism in the human genome (Reprinted with permission from Attia et al. [10])

reporting observational studies geared specifically to genetic association studies (see STREGA [14]). The recommendations in those guidelines will act as a template for an epidemiological discussion of sources of bias here. Readers looking for more details on the statistical issues in CGAS are referred to an excellent review [15].

Genotyping Error

Although measuring genetic polymorphisms may have the cachet of being exact and absolute, in fact, there is error, just as there is for traditional epidemiological factors such as blood pressure or diet. Multiple articles have outlined the sources and magnitude of error in genotyping assays, which can range up to 30% [16]. Sources of error and associated bias in case and control studies include:

- Case-control sampling differences: controls may be selected from population-based studies, with cases coming from hospital series. In practice, this means that samples are collected in different laboratories, under different conditions, potentially many years apart. Such differences in sample handling may lead to variable DNA quality and variable DNA amplification for cases and controls [17].
- Batch effects: if cases and controls are processed separately in different batches, any errors affecting batch genotyping accuracy can lead to spurious genotype frequency differences between cases and controls [18].
- Cross-platform effects: sometimes cases are genotyped in one laboratory on one type of array, and controls are processed in a completely different laboratory on a different array. Such cross-platform comparisons can lead to extreme batch effects.

Population Stratification

Population stratification refers to a situation where there is an imbalance of subgroups, usually ancestral subgroups, between cases and controls. If an ancestral subgroup is associated with a particular genotype and also has a higher incidence of disease, this can cause confounding; the genotype may appear to be related to disease due to systematic ancestral differences between cases and controls, rather than true association of genotypes with disease. The magnitude of confounding due to population stratification has been debated [19, 20]. On the one hand, these effects are very likely to be small on a theoretical basis [21], and empirically [22], but on the other hand, it is argued that the genetic effects we are trying to detect are also small (odds ratios typically <1.4) and may be obscured. Many methods for adjustment have been developed, including structured association [23], genomic control

(GC) [24, 25], and principal components analysis [26]. Current practice is to either use one of these adjustment methods, restrict cases and controls to a single ethnic group or adjust for self-reported ethnicity.

Hardy-Weinberg Equilibrium

Both genotyping error and population stratification can produce deviations from Hardy-Weinberg equilibrium (HWE) proportions in the control group, hence making HWE tests a useful quality control check. Studies have found deviation from HWE for a larger than expected proportion of published studies, raising the possibility of occult error in many published association studies [27, 28]. The significance level for judging deviation from HWE, however, has been a moving target, with values anywhere between (alpha) $\alpha = 0.05$ and (alpha) $\alpha = 5 \times 10^{-6}$; this is further compounded by the fact that the estimated significance is dependent on sample size (i.e., two studies of different size could have the same HW proportions, but one appears to be consistent with HWE while the other does not). Some have suggested that the measure of HWE should be based on measures of the degree of disequilibrium rather than the *p* value, since the former is independent of sample size [29]. These measures of degree of disequilibrium can then be used to adjust the association analysis [30].

Misspecification of Outcomes

The cases and controls in a CGAS are chosen based on their disease outcome (e.g., stroke vs. no stroke), and any difference in the frequency of a genetic variant is said to be related to that outcome. In practice, however, stroke case and control groups also differ in the frequency of many risk factors for stroke (e.g., diabetes mellitus, cigarette smoking, obesity, and hypertension), and it is also possible that the genetic variant is associated with one of these as a "covert" outcome. In this case, the risk factor is associated with both genetic variant and stroke outcome and acts like a confounder, leading one to say that a variant is associated with stroke when in fact it is related to, for example, diabetes mellitus. This occurred with variants of the *FTO* gene, which were found to be associated with diabetes; however, diabetic cases had higher body mass index (BMI) than controls and when this was adjusted for, association between *FTO* and diabetes disappeared, indicating that the true association was between *FTO* and BMI [31]. The solution to potential confounding problem is to adjust for imbalances in known risk factors or to match cases and controls on potentially confounding variables.

Misspecification can also occur when there is unsuspected heterogeneity of outcomes. Stroke itself is heterogeneous, with many possible differences in the pathophysiologic mechanisms underlying the various subtypes. For example, it is

biologically likely that ischemic and hemorrhagic strokes have fundamentally different aetiologies, and hence, any attempt to find an association between a genetic variant and all strokes could fail. Likewise ischemic stroke may be considered homogenous by one group but heterogeneous by another, who may subdivide this further into large artery, small vessel, and cardioembolic stroke. CGAS have been published for both ischemic stroke as a whole, and for subgroups. Until recently, there was very little standardization in classifying ischemic stroke in CGAS [32].

Multiple Comparisons

The problem of multiple comparisons in CGAS is somewhat hidden in that it is often unclear how many genetic variants had been tested prior to finding and reporting the one that showed an association; when this happens, the likelihood of type I error, that is, incorrectly declaring an association as statistically significant, is increased. Given the difficulty in publishing negative results, it is often the lone positive result that gets published, often using an unadjusted significance threshold of $p < 0.05$. Given the play of chance and the need to have a "strong" result in order to break the publication barrier, this first study also tends to overstate the effect size of the genetic variant, a phenomenon known as the "winner's curse" [33, 34]. Subsequent studies often show lower effect sizes or refute the original association completely.

Advantages and Disadvantages of CGAS

Lack of attention to the five areas enumerated above has led to poor-quality CGAS being published and a perception that CGAS have a poor record of replication [11, 35]. Hence, they have fallen somewhat out of favor. This is unfortunate, given their tremendous efficiency in terms of power. Given the evidence that genetic effects are likely to be small, the candidate gene approach has an advantage over GWAS due to the relative lack of multiple comparisons. For example, imagine a SNP with minor allele frequency of 20% and an additive genetic model where each copy of the risk allele increases stroke risk by 20% (i.e., odds ratio of 1.2), and the baseline risk of stroke in a particular community is 1%. The sample size needed for 80% power at a significance level of $p < 0.05$, if only one genetic variant is tested, is 1,400 people in each group. However, if the genetic variant is one of about one million independent genetic variants tested (as is typical in GWAS), adjustment for multiple testing using a Bonferroni correction requires a significance level of $0.05/10^6$ or 5×10^{-8}. To detect the same association at this adjusted significance level with 80% power would require just over 7,000 people/group, a very sizable increase.

As the number of variants tested in a CGAS is much smaller than for GWAS, there is considerably more scope to fit complex models and tease out the nature of

effects in a reasonable time frame. For instance, if there are K SNPs, then if we do not consider interactions, there are effectively 2^K possibilities of models. There are a variety of computational techniques capable of exploring this large model space for CGAS [36], some that can determine possible Gene×Gene and Gene×Environment interactions.

For smaller numbers of variants, it is also possible to consider the joint effects of markers via haplotype association tests. Haplotypes are a set of alleles on the same strand of DNA that are inherited as a unit, and are thus in strong linkage disequilibrium (LD). By testing haplotypes rather than single markers for association, an increase in power may be gained by reducing the number of comparisons. It is also possible that the effect due to the haplotype will be stronger than the effect observed at constitutive alleles. When investigating the statistical significance of an association due to a haplotype, special algorithms should be employed that account for the LD structure of the haplotype (e.g., permutation procedures).

There are also many examples of CGAS results that are confirmed on meta-analysis but that are not initially detectable on GWAS. For example, an association between chromosome 9p21 variants and ischemic stroke was first identified by CGAS [37, 38] and subsequently confirmed by meta-analysis [39], but has not reached genome-wide significance in GWAS.

The major drawback, however, to the candidate gene approach is the limited knowledge of biochemical pathways and disease aetiology on which to base the choice of candidate. Thus, these studies depend on the ability to predict functional candidate genes and polymorphisms. The GWAS approach is often touted positively as "agnostic" or "hypothesis neutral," in that no prior biological knowledge is needed. Nevertheless, the choice of a candidate gene is increasingly being refined. With the Human Genome Map complete, and the catalogue of variants continually increasing—that is, HapMap [40] and 1000 Genomes [41]—there is more information on which to base the choice of candidate. SNPs can be chosen because they are:

- Nonsynonymous (i.e., lead to an amino acid change, and presumably to a functional effect).
- In a splice site (i.e., they may interfere with RNA splicing of the transcript).
- In a promoter site, upstream of the gene of interest.
- In the same biological pathway as another implicated gene. There are increasingly complete catalogues of biological pathways that can be searched, e.g., KEGG [42].
- In genes with SNPs that have shown evidence of association in a GWAS.
- In genes within loci identified through a linkage study.
- In genes that are differentially expressed in a microarray study.
- In genes shown to have functional effects in in vitro models or in animal models.
- In genes that are good candidates for an intermediate in the disease pathway.

However, it is still difficult to anticipate what SNPs will have a functional effect and hence be "good" candidates; for example, some synonymous SNPs (i.e., those

that do not change the amino acid sequence) have been shown to have functional effects via an effect on mRNA stability [43].

Meta-analysis of CGAS

One of the best ways to confirm a CGAS is to perform a meta-analysis, a way of mathematically pooling results from individual CGAS. Detailed methodology for doing this has been published by multiple groups [44–46], and the main issues will be highlighted here.

Genetic Model for Pooling

Methodology for pooling results in traditional epidemiological studies is well established, but this has focused mainly on pooling contrasts between two groups. The main difference with CGAS is that there are, at minimum, three genotype groups (AA, Aa, and aa) and potentially more if testing a DNA repeat variant. One approach is to collapse the three groups to two by assuming a dominant or recessive genetic model. However, this may be an inaccurate assumption; there is often little evidence to indicate the genetic model operating in a disease association. Theoretical work has suggested that in the absence of any evidence about genetic models, an additive genetic model retains the most power with the least error, even when the true underlying genetic model is dominant or recessive [47]. In an additive model, risk increases with each copy of the risk allele. Alternatively, one can look at the two pairwise comparisons between the three genotype groups and, by dividing them, obtain a parameter (lambda) that can indicate the genetic model; this is an empirical approach that lets the empirical data indicate which genetic model might be operating [48]. Interestingly, data for genetic associations in diabetes mellitus support the concept that most genetic variants act via an additive model [49].

Statistical Methods for Pooling

The usual fixed effects or random effects models for pooling traditional epidemiological data are also applicable for genetic data. The fixed effects model does not make allowance for any between-study variation, whereas the random effects model does; in essence, it allows for the fact that the genetic effect may not be identical in all studies. This between-study variation can be quantified by metrics of heterogeneity, which are important to calculate. Most commonly, heterogeneity is measured using a Breslow-Day Q test (with p value), or I^2, where 25%, 50%, and 75% correspond to benchmarks for low, moderate, and high heterogeneity [50].

Where pooled genetic estimates are heterogeneous, it is good practice to explore and understand the source of heterogeneity rather than just continuing with "brute force" pooling. This heterogeneity may indicate genotyping error, misclassification of outcomes, or, indeed, different interacting environmental variables between studies [51]. It appears unlikely that there are true differences in a genetic effect between ethnic groups; a survey of identical association studies between different ethnic groups overwhelmingly showed that although allelic frequencies vary widely across ethnic groups, if an allele is present, it has the same magnitude of effect across ethnic groups [52]. The current practice of pooling using random effects when heterogeneity is present does not necessarily increase power and in fact can result in power "deserts," where adding data sets produces no increase in power; using fixed effects in this situation is not a solution either, as it leads to many false-positive signals [53, 54].

Because of the winner's curse discussed above, it is also good practice to pool CGAS omitting the first published study, to ensure that it is not just the initial study that is driving the significance of the pooled estimate. It is also good practice to provide a "funnel" plot [55]; this graph plots a measure of sample size (usually variance) on the *y*-axis with the magnitude of effect on the x-axis. Small studies should show wide variation in odds ratios, and this variation should decrease as the sample size increases, leading to an inverted funnel shape. If small studies with negative results have not been able to break the publication barrier, part of the funnel will be missing, a potential indication of publication bias. Although the funnel plot may be judged subjectively, formal statistical tests (e.g., Egger test) can also be used which yield *p* values judged in the usual way [56, 57].

Graphical displays of the data using Peto plots are a good way to show the effect sizes from different studies and the variation between them; an example of this plot is shown for the Thr/Met rs6065 polymorphism of glycoprotein 1bα (alpha) and ischemic stroke from Maguire et al. (Fig. 2.2).

The Flip/Flop Phenomenon

A unique source of heterogeneity in genetic association studies is flip/flop, a situation in which the direction of the genetic effect alternates between the different CGAS being pooled [59]. This is thought to occur when the genetic variant studied is simply a marker for the true causal variant and is in linkage disequilibrium with it. The different LD structure between the study populations means that sometimes the causal variant is in LD with one allele of the marker SNP and sometimes the other allele, leading to a "flip" in the direction of the effect between study populations. Simulation studies indicate that the degree of LD required for this to occur is consistent with levels observed in the human genome [60]. In such cases, using usual pooling methods (i.e., pooling effect sizes and directions) will lead to a spuriously negative summary estimate; an alternative is to pool *p* values (e.g., using the original method of Fisher, as in [61]).

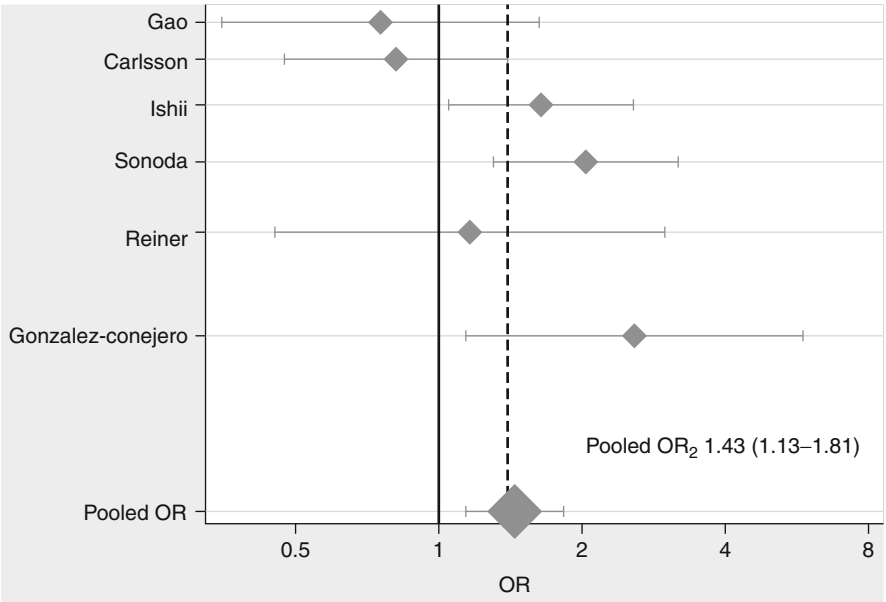

Fig. 2.2 This shows the odds ratio for the Met/Thr variant versus the Thr/Thr variant in glycoprotein GP1bα and ischemic stroke. In this case, significant heterogeneity was present ($I^2 = 58\%$, p value = 0.04) (Reprinted with permission from Maguire et al. [58])

Surveys of quality of CGAS meta-analyses have been performed and indicate that there is much room for improvement; suggestions for guidelines have also been proposed [62].

Relationship Between CGAS and GWAS

As discussed above, due to limited understanding of relevant biological pathways from which to select CGAS candidate genes and the "agnostic" approach of GWAS, the latter has been seen as methodologically superior. Gradual reductions in the cost of microarrays for GWAS have also bolstered their popularity. Nevertheless, GWAS remains an expensive option for genetic association studies, and given the greater efficiency of CGAS, the question is perhaps not whether one should supplant the other but how they may complement one another. GWAS may indicate a genetic locus that had hitherto not been suspected; for example, complement factor H had not been suspected in the aetiology of macular degeneration (AMD) until it was identified by GWAS [63]. Once this locus was identified, CGAS of other complement factors were performed, leading to the identification of complement factors C2 and C3. Associations of these loci with AMD have now been confirmed by CGAS pooling and meta-analyses [64, 65]; however, the associations are too weak to have been definitively identified by GWAS at accepted significance levels in available sample sizes.

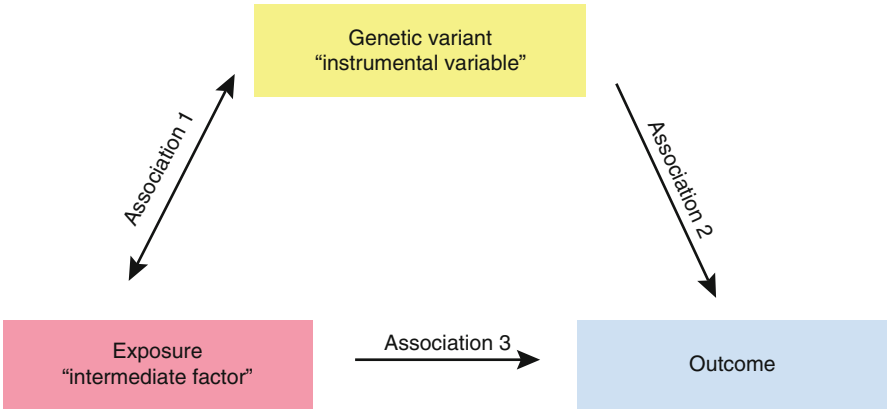

Fig. 2.3 Mendelian randomization (Adapted from Thanassoulis et al. [66])

It is also increasingly recognized that associations identified via GWAS are largely indirect, as tested SNPs are unlikely to be causative but simply correlated with functional variants due to linkage disequilibrium (LD). Hence, considerable subsequent fine mapping and sequencing is typically required to isolate the actual causative variants, which may not be a SNP but alternatively an insertion, deletion, or copy number variant (CNV). It has been suggested that CGAS within loci identified by GWAS could be a component of the arsenal for such fine mapping [40].

CGAS and Mendelian Randomization

The underlying idea of Mendelian randomization is to use the random allocation of alleles at a particular gene at meiosis during conception to act as a randomized controlled trial. Since randomization avoids confounding, the random allocation of a gene allows an unconfounded assessment of the relationship between an environmental exposure (that is dependent on the gene) and a disease outcome (Fig. 2.3). The idea was first suggested by Katan [41]; the term "Mendelian randomization" was originally coined by Gray and Wheatley [42] but was popularized by Davey Smith and Ebrahim [2]. One example given in the latter paper related to homocysteine and heart disease. Given that homocysteine may be influenced by diverse factors such as smoking, diet, and exercise, the relationship with heart disease was controversial. Examining the relationship between variations in the MTHFR gene (which influences homocysteine) and heart disease led them to conclude that elevated levels of homocysteine did in fact increase the risk of heart disease and that elevations were not just indicators of high exposure to other risk factors. However, the Mendelian randomization approach can fail when there is pleiotropy (i.e., a gene has multiple biological effects that can influence an outcome) or when there is

"canalization," a term referring to the homeostatic mechanisms of the human body that seek to restore a phenotype toward normal values. Indeed, randomized controlled trials of folate and vitamin B, which lowered homocysteine levels, did not influence cardiovascular event rates [67], indicating that the Mendelian randomization approach is not foolproof.

Conclusion

We conclude that CGAS, whether using case-control or cohort approaches, are still an efficient study design but that due attention is needed to key methodological issues including genotyping error, population stratification, Hardy-Weinberg equilibrium, specification of outcomes, and multiple comparisons. Replication of CGAS is essential, and meta-analysis of individual studies is an excellent way to increase power and explore the consistency of the signal across multiple populations. We anticipate that CGAS and GWAS will each find their niche in the genetic armamentarium and settle into a complementary and symbiotic existence.

References

1. Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. Nat Genet. 2008;40(2):124–5.
2. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1–22.
3. Bersano A, Ballabio E, Bresolin N, Candelise L. Genetic polymorphisms for the study of multifactorial stroke. Hum Mutat. 2008;29(6):776–95.
4. Casas JP, Hingorani AD, Bautista LE, Sharma P. Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. Arch Neurol. 2004;61(11):1652–61.
5. Stankovi S, Majki-Singh N. Advances in the genetic basis of ischemic stroke. J Med Biochem. 2008;27(2):123–34.
6. Laird NM, Lange C. Family-based designs in the age of large-scale gene-association studies. Nat Rev Genet. 2006;7(5):385–94.
7. Laird NM, Lange C. Family-based methods for linkage and association analysis. Adv Genet. 2008;60:219–52.
8. Meschia JF, Kissela BM, Brott TG, Brown Jr RD, Worrall BB, Beck J, et al. The siblings with ischemic stroke study (SWISS): a progress report. Clin Med Res. 2006;4(1):12–21.
9. Meschia JF, Nalls M, Matarin M, Brott TG, Brown Jr RD, Hardy J, et al. Siblings with ischemic stroke study: results of a genome-wide scan for stroke loci. Stroke. 2011;42(10):2726–32.
10. Attia J, Ioannidis JPA, Thakkinstian A, McEvoy M, Scott RJ, Minelli C, et al. How to use an article about genetic association: A: background concepts. JAMA. 2009;301:74–81.
11. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. Lancet. 2003;361(9360):865–72.
12. Dichgans M, Markus HS. Genetic association studies in stroke: methodological issues and proposed standard criteria. Stroke. 2005;36(9):2027–31.

13. Pruissen DM, Kappelle LJ, Rosendaal FR, Algra A. Genetic association studies in ischaemic stroke: replication failure and prospects. Cerebrovasc Dis. 2009;27(3):290–4.
14. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, et al. STrengthening the REporting of Genetic Association studies (STREGA)–an extension of the STROBE statement. Eur J Clin Invest. 2009;39(4):247–66.
15. Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet. 2006;7(10):781–91.
16. Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors: causes, consequences and solutions. Nat Rev Genet. 2005;6(11):847–59.
17. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet. 2005;37(11):1243–6.
18. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010;11(10):733–9.
19. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev. 2002;11(6):505–12.
20. Wacholder S, Rothman N, Caporaso N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol Biomarkers Prev. 2002;11(6):513–20.
21. Khlat M, Cazes MH, Genin E, Guiguet M. Robustness of case-control studies of genetic factors to population stratification: magnitude of bias and type I error. Cancer Epidemiol Biomarkers Prev. 2004;13(10):1660–4.
22. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst. 2000;92(14):1151–8.
23. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. Am J Hum Genet. 2000;67(1):170–81.
24. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55(4): 997–1004.
25. Bacanu SA, Devlin B, Roeder K. Association studies for quantitative traits in structured populations. Genet Epidemiol. 2002;22(1):78–93.
26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904–9.
27. Bardoczy Z, Gyorffy B, Kocsis I, Vasarhelyi B. Re-calculated Hardy-Weinberg values in papers published in atherosclerosis between 1995 and 2003. Atherosclerosis. 2004;173(1): 141–3.
28. Sen S, Burmeister M. Hardy-Weinberg analysis of a large set of published association studies reveals genotyping error and a deficit of heterozygotes across multiple loci. Hum Genomics. 2008;3(1):36–52.
29. Attia J, Thakkinstian A, McElduff P, Milne E, Dawson S, Scott RJ, et al. Detecting genotyping error using measures of degree of Hardy-Weinberg disequilibrium. Stat Appl Genet Mol Biol. 2010;9(1):Article 5.
30. Trikalinos TA, Salanti G, Khoury MJ, Ioannidis JP. Impact of violations and deviations in Hardy-Weinberg equilibrium on postulated gene-disease associations. Am J Epidemiol. 2006;163(4):300–9.
31. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007;316(5826):889–94.
32. Meschia JF. Addressing the heterogeneity of the ischemic stroke phenotype in human genetics research. Stroke. 2002;33(12):2770–4.
33. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet. 2001;29(3):306–9.

34. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet. 2003;33(2):177–82.
35. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al. Replicating genotype-phenotype associations. Nature. 2007;447(7145):655–60.
36. Lin E, Hsu SY. A Bayesian approach to gene-gene and gene-environment interactions in chronic fatigue syndrome. Pharmacogenomics. 2009;10(1):35–42.
37. Smith JG, Melander O, Lovkvist H, Hedblad B, Engstrom G, Nilsson P, et al. Common genetic variants on chromosome 9p21 confers risk of ischemic stroke: a large-scale genetic association study. Circ Cardiovasc Genet. 2009;2(2):159–64.
38. Gschwendtner A, Bevan S, Cole JW, Plourde A, Matarin M, Ross-Adams H, et al. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. Ann Neurol. 2009;65(5): 531–9.
39. Anderson CD, Biffi A, Rost NS, Cortellini L, Furie KL, Rosand J. Chromosome 9p21 in ischemic stroke: population structure and meta-analysis. Stroke. 2010;41(6):1123–31.
40. On beyond GWAS. Nat Genet. 2010;42(7):551.
41. Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. Lancet. 1986;1(8479): 507–8.
42. Gray R, Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. Bone Marrow Transplant. 1991;7 Suppl 3:9–12.
43. Capon F, Allen MH, Ameen M, Burden AD, Tillman D, Barker JN, et al. A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups. Hum Mol Genet. 2004;13(20):2361–8.
44. Thakkinstian A, McElduff P, D'Este C, Duffy D, Attia J. A method for meta-analysis of molecular association studies. Stat Med. 2005;24(9):1291–306.
45. Trikalinos TA, Salanti G, Zintzaras E, Ioannidis JP. Meta-analysis methods. Adv Genet. 2008;60:311–34.
46. Zintzaras E, Lau J. Synthesis of genetic association studies for pertinent gene-disease associations requires appropriate methodological and statistical approaches. J Clin Epidemiol. 2008;61(7):634–45.
47. Thakkinstian A, Thompson JR, Minelli C, Attia J. Choosing between per-genotype, per-allele, and trend approaches for initial detection of gene-disease association. J Appl Stat. 2009;36(6):633–46.
48. Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J. The choice of a genetic model in the meta-analysis of molecular association studies. Int J Epidemiol. 2005;34(6):1319–28.
49. Salanti G, Southam L, Altshuler D, Ardlie K, Barroso I, Boehnke M, et al. Underlying genetic models of inheritance in established type 2 diabetes associations. Am J Epidemiol. 2009;170(5):537–45.
50. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539–58.
51. Greene CS, Penrod NM, Williams SM, Moore JH. Failure to replicate a genetic association may provide important clues about genetic architecture. PLoS One. 2009;4(6):e5639.

52. Ioannidis JP, Ntzani EE, Trikalinos TA. 'Racial' differences in genetic effects for complex diseases. Nat Genet. 2004;36(12):1312–8.
53. Moonesinghe R, Khoury MJ, Liu T, Ioannidis JP. Required sample size and nonreplicability thresholds for heterogeneous genetic associations. Proc Natl Acad Sci USA. 2008;105(2): 617–22.
54. Pereira TV, Patsopoulos NA, Salanti G, Ioannidis JP. Discovery properties of genome-wide association signals from cumulatively combined data sets. Am J Epidemiol. 2009; 170(10):1197–206.
55. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629–34.
56. Chapter 7: publication bias. In: Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F, editors. Methods for meta-analysis in medical research. West Sussex: Wiley; 2000.
57. Sterne JAC, Egger M, Smith GD. Investigating and dealing with publication and other biases. In: Egger M, Smith GD, Altman DG, editors. Systematic reviews in health care. 2nd ed. London: BMJ Books; 2001.
58. Maguire J, Sturm J, Levi C, Thakkinstian A, Whyte S, Attia J. Polymorphisms in platelet glycoprotein 1bα and factor VII and risk of ischemic stroke. A meta-analysis. Stroke. 2008;39:1710–6.
59. Lin PI, Vance JM, Pericak-Vance MA, Martin ER. No gene is an island: the flip-flop phenomenon. Am J Hum Genet. 2007;80(3):531–8.
60. Clarke GM, Cardon LR. Aspects of observing and claiming allele flips in association studies. Genet Epidemiol. 2010;34(3):266–74.
61. Maguire JM, Thakkinstian A, Sturm J, Levi C, Lincz L, Parsons M, et al. Polymorphisms in platelet glycoprotein 1balpha and factor VII and risk of ischemic stroke: a meta-analysis. Stroke. 2008;39(6):1710–6.
62. Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J. The quality of meta-analyses of genetic association studies: a review with recommendations. Am J Epidemiol. 2009;170(11):1333–43.
63. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308(5720):385–9.
64. Gold B, Merriam JE, Zernant J, Hancox LS, Taiber AJ, Gehrs K, et al. Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. Nat Genet. 2006;38(4):458–62.
65. Thakkinstian A, McKay GJ, McEvoy M, Chakravarthy U, Chakrabarti S, Silvestri G, et al. Systematic review and meta-analysis of the association between complement component 3 and age-related macular degeneration: a HUGE review and meta-analysis. Am J Epidemiol. 2011;173:1365–79.
66. Thanassoulis G, O'Donnell CJ. Mendelian randomisation. UptoDate. Wolters Kluwer Health.
67. Armitage JM, Bowman L, Clarke RJ, Wallendszus K, Bulbulia R, Rahimi K, et al. Effects of homocysteine-lowering with folic acid plus vitamin B12 vs placebo on mortality and major morbidity in myocardial infarction survivors: a randomized trial. JAMA. 2010;303(24): 2486–94.

http://www.springer.com/978-0-85729-208-7

Stroke Genetics
Sharma, P.; Meschia, J.F. (Eds.)
2013, XX, 320 p., Hardcover
ISBN: 978-0-85729-208-7