Identifying susceptibility genes of IgA nephropathy: research in progress

Moumita Barua and York Pei

Division of Nephrology, University Health Network and University of Toronto, Toronto, Ontario, Canada

Correspondence and offprint requests to: York Pei; E-mail: york.pei@uhn.on.ca

Keywords: candidate gene association; genome scan; GWAS; IgA nephropathy

Genetic susceptibility of IgA nephropathy (IgAN)

Immunoglobulin A nephropathy (IgAN) is the most common primary glomerular disease worldwide and a significant cause of end-stage renal disease [1]. Although dysregulation of mucosal immunity is a notable feature of this disease, the pathogenesis of IgAN remains poorly understood and its treatment is limited. Familial clustering of patients with IgAN suggests a genetic predisposition [2–5]. Recent genome-wide studies of multiplex families with IgAN have identified two loci with significant linkage on chr. 6q22 [6] and chr. 2q36 [7], and two loci (see the Appendix) with suggestive linkage on chr. 4q26–31 and chr. 17q12–22 [8]. Moreover, the absence of linkage to these loci in a large family with multiple affected members suggests additional genetic heterogeneity (see the Appendix) [5]. Taken together, these data indicate that IgAN is a complex disease likely influenced by multiple genetic and environmental factors.

The identification of susceptibility genes for IgAN has the potential to improve risk prediction and may yield novel insights into the primary pathogenesis of this disease. The latter knowledge is essential for future development of mechanism-based therapeutics. However, none of these putative susceptibility genes for IgAN has been identified thus far. To map these susceptibility genes, a candidate gene or genome-wide approach can be taken [9–12]. The former requires a conscious effort to select and then test polymorphic markers (see the Appendix) from one or more candidate gene loci among ~25 000 genes in the human genome [13] for association or linkage with the disease. In contrast, the latter tests a dense set of equally spaced polymorphic markers across the entire genome for association or linkage with the disease and is agnostic for any underlying assumptions. To date, the candidate gene approach has been used exclusively for sporadic IgAN [1,14–18], while three genome-wide linkage studies have been employed for familial IgAN [6–8].

Candidate gene association studies

The search for susceptibility genes for complex diseases has been fraught with problems [9]. Although many putative susceptibility genes were identified by the candidate gene approach, few were reproducible in subsequent studies. A recent review of the literature showed that only 6/166-reported associations were replicated in follow-up studies [11]. This is likely an overestimate given that there is a publication bias for positive studies. Most of these associations are likely spurious since they were identified from studies of a small sample size that were inadequately powered to detect genetic loci of modest effect that are typical of most complex trait disorders [9,11]. Additionally, failure to control for population stratification (differences in marker allele frequencies between cases and controls due to different ancestral origins) and to correct for multiple comparisons may also lead to spurious associations. Other reasons for the lack of reproducibility include genetic and etiological heterogeneity; variable linkage disequilibrium (LD; see the Appendix) between the tested and causative variant; and inadequate power of the replication studies [9–12].

Genome-wide linkage studies

Genome-wide linkage studies have been successfully employed for mapping Mendelian diseases and are robust against population stratification. However, their utility in complex diseases/traits is limited [12]. First, while the linkage approach has been successful in mapping rare genes with large effects, it is much less powerful for mapping common variants with modest effects, as expected for complex trait disorders. Second, this approach may be limited by the availability of many large families with multiple affected members. Nonetheless, several loci have been identified for familial IgAN, and follow-up mapping using sporadic cases may allow for the identification of the disease gene(s). However, the latter strategy may not work if the disease gene(s) for familial IgAN underpin only a small fraction of sporadic cases. This possibility is raised by a recent study in
which fine-mapping of the \textit{IGAN1} locus with a large patient cohort failed to identify a definitive association signal [18].

**Genome-wide association studies (GWAS)**

The recent completion of the Phase II HapMap project has generated more than 3 million well-validated single nucleotide polymorphisms (SNPs; see the Appendix) designed for mapping common variants of complex traits [19]. Additionally, recent advances in high-throughput genotyping [20] have made it possible to conduct cost-effective studies of dense sets of common SNPs across the genome for associations with complex diseases. A number of different array platforms are now available to rapidly screen up to one million SNPs on a single chip [20]. In many cases, tagging SNPs are used to effectively capture most of the common haplotypes (see the Appendix) found in the human genome [21].

The past 18 months has shown the promise of GWAS, and multiple susceptibility gene loci have been identified under many complex medical conditions such as coronary artery disease, diabetes mellitus, rheumatoid arthritis, Crohn’s disease and various forms of cancer, among others. To date, >150 risk loci have been identified in studies from <60 common diseases and traits [22]. The overwhelming success of these studies has led to surprising new insights into disease pathophysiology and possible therapeutic approaches. Typically, these studies have employed arrays with >100 000 SNPs to assay common genetic variants in large samples of patients and control subjects (>1000/group). They have shown the ability to identify loci of moderate to low risk. Indeed, most of the robust risk loci identified in these conditions are associated with heterozygote odds ratios of 1.2–2.0. Gratifyingly, many associations have been replicated in multiple studies showing the feasibility and robustness of the approach [22].

**Do TGF-β1 gene variants predispose to IgAN?**

In this issue of \textit{Nephrol Dial Transplant}, Vuong et al. present a candidate gene study that examined the role of TGF-β1 gene variants on the susceptibility of IgAN [23].

The choice of this candidate gene was based on the notion that TGF-β1 may increase isotype switching of IgM to IgA and IgA production in B-lymphocytes. The authors compared the genotypes of five SNPs from the \textit{TGF-β1} locus between 212 Caucasian patients with biopsy-proven IgAN and 477 age- and ethnically-matched controls from Sweden. In a subgroup analysis, they found that three SNPs (rs2241715, rs1982073 and rs1800469) were associated with IgAN in male (\(P < 0.02\)) but not female patients. Furthermore, the authors performed a meta-analysis including their study with two other studies of Japanese and Italian patients, respectively, that passed their selection criteria. In the meta-analysis, 640 patients were compared with 880 controls, and the author found a protective effect (odds ratio \(\sim 0.8\)) for IgAN in two SNPs, rs1800469 and rs1982073, which are in complete LD with each other.

The strengths of this study include a putative homogenous population of patients and controls from Sweden, good quality genotyping, control for multiple comparisons of single markers and haplotypes by permutation, and the use of meta-analysis to support the observed associations. On the other hand, without genomic control measures [22–26], it is not possible to eliminate population stratification as a potential confounder. Additionally, the sample size of the current study was modest for a complex trait disorder, and the observed associations were only detectable in a subgroup of affected males. Finally, while the meta-analysis supports the observed associations, positive publication bias can skew the results of this analysis. In this regard, we have recently published an association study in 732 Caucasian patients with IgAN and 503 local control subjects from Canada, France and Finland, using a customized bead array with 1536 SNPs [18]. We used EIGENSTRAT (see the Appendix) to correct for population stratification. Using logistic regression to adjust for age, gender, study site and population structure, we found that IL5RA and TNFRSF6B gene variants were associated with sporadic IgAN. In the same study, we also examined five tagging SNPs at the \textit{TGF-β1} locus, including rs1800469, but did not find any association with IgAN (Table 1). If these negative data were published and included in the meta-analysis, the conclusion would have been very different. In the same study, we also examined multiple SNPs for tumour necrosis factor α (TNFα), polymeric immunoglobulin receptor (CD89) and

| SNP ID | Chr. | Physical location | Logit regression Marker quality |
|--------|------|------------------|-----------------------------|-----------------------------|-----------------------------|
| rs4803454 | 19 | 46524236 | 0.938 | 0.984 | 0.140 | 0.0112 | 0.8 |
| rs10417924 | 19 | 46525007 | 0.985 | 0.701 | 0.450 | 0.641 | 0.0168 | 0.75 |
| rs8179181 | 19 | 46530046 | 0.833 | 0.114 | 0.201 | 0.128 | 0.0112 | 0.83 |
| rs8110090 | 19 | 46537712 | 0.675 | 0.123 | 0.829 | 0.456 | 0.0144 | 0.77 |
| rs1800469 | 19 | 46552136 | 0.987 | 0.779 | 0.719 | 0.386 | 0.0144 | 0.81 |

*Logistic regression was used to adjust for the effects of age, gender, study site and first two principal components (PC1 and PC2) from EIGENSTRAT analysis.*
the selectin gene cluster but were unable to replicate any of the previously reported associations [14–17]. Our study had sufficient power to detect a susceptibility locus with a genotype relative risk (GRR) of 0.5 over allele frequencies of 0.2–0.5 under a multiplicative model [27]. However, even a sample size of 3000 patients and 3000 controls would not be sufficiently powered to detect a multiplicative GRR of 0.8 at a risk allele frequency of 0.3, as detected in the study by Vuong et al.

Looking into the crystal ball—where do we go from now?

Minimizing phenotypic heterogeneity, robust quality control for genotyping, application of genomic controls for population stratification and the use of large sample sets for initial detection and follow-up replication of associations appear to be the key parameters for the successful association studies [10–12]. Despite the immense challenge, we believe that properly powered GWAS will allow for the identification of common susceptibility gene variants for IgAN with modest effects. This approach, however, requires a very substantial patient resource and will therefore need the collective effort of the IgAN research community—together, ‘yes, we can!’.

Glossary

Linkage disequilibrium: The nonrandom association of alleles at two or more polymorphic sites on the same chromosome.

Haplotype: A combination of alleles from multiple loci on the same chromosome, all of which are co-segregated together.

EIGENSTRAT: Statistical analysis that detects and corrects for ancestral genetic background differences between cases and controls in genetic association studies.

Conflict of interest statement. None declared.

(See related article by M. T. Vuong et al. Genetic variation in the transforming growth factor-β1 gene is associated with susceptibility to IgA nephropathy. Nephrol Dial Transplant 2009; 24: 3061–3067.)

References

27. Iles M. What can genome-wide association studies tell us about the genetics of common disease? PLoS Genetics 2008; 4: 0001–0002

Received for publication: 8.5.09; Accepted in revised form: 6.7.09