Genetic prediction of common complex disorders assessed by next generation sequencing and genome wide analysis

BRUNO DALLAPICCOLA(1), RITA MINGARELLI(1), STEFANIA BOCCIA(2, 3)

ABSTRACT
Insight into the biological make-up of complex disorders can improve their diagnosis, lead to the discovery of new targets for therapy, increase awareness of genome-environment interactions in health and disease, and open the door to predictive medicine. More than 1,600 genome-wide association studies (GWASs) have been published, and have identified thousands of polymorphisms associated with more than 250 common diseases or traits. However, for most of the genomic variants identified so far, only inconclusive associations with complex diseases have been reported and for many of them their predictive value reaches the same level as the traditional risk. The limited value of these results is probably due to regulatory elements in 2-3% of the encoding genome, whose function has only recently been partially decrypted. Nevertheless, genomic sequencing is an attractive tool for personalized medicine. During the last few years several commercial ventures have begun marketing GWASs directly to consumers for medical, genealogic, and even recreational purposes. Although these tests show promise for the future, consumers should be aware of the unreliability of most of their results at the present time. The development of methods integrating clinical and genetic data together with a better understanding of the heritability of complex diseases will be necessary in the endevour to progress towards a personalized medicine. In order to achieve maximum benefits from GWASs while keeping the disadvantages to a minimum, guidelines will be necessary to manage the technical advances and to meet the challenges involved in the clinical application of whole genomic sequencing.

Key words: Common complex disorders; Genome wide association studies; Genetic prediction; Genetics; Heritability; Genetic counseling; Direct-to-consumer genetic testing

(1) Bambino Gesù Children’s Hospital, IRCCS, Rome, Italy
(2) Institute of Hygiene, Università Cattolica del Sacro Cuore, Rome, Italy
(3) IRCCS San Raffaele Pisana, Rome, Italy

CORRESPONDING AUTHOR: Stefania Boccia, Institute of Hygiene, Università Cattolica del Sacro Cuore, L.go F. Vito, 1-00168 – Rome, Italy. Tel: +39 06 35001527, Fax: +39 06 3500152. e-mail: sboccia@rm.unicatt.it
DOE: 10.2427/8691

INTRODUCTION
Common complex diseases, unlike single gene disorders, are caused by the interaction of genetic and environmental factors, each one having a small effect, with a few sometimes acting individually as a necessary, although on their own insufficient, trigger for the disease to occur. To study the genetic background of these phenotypes, principles for genetic mapping have been developed, using populations rather than families. Based on the “common disease-
common variant” (CD-CV) hypothesis (1), it is assumed that a vast number of polymorphisms, classically defined as having an allele frequency >1%, are pathogenically related to common complex diseases. Accordingly, testing of all these variants should shed light on the underlying heritability and clearly identify the relevant key susceptibility genes. During recent years, the CD-CV hypothesis has been tested following the development of catalogues of common variants, haplotype maps, genotyping arrays, and innovative and more accurate statistical methods. This, in turn, has opened the door to predictive medicine by decrypting the genetic setting underlying the susceptibility to complex disorders, and instituting measures for preventing diseases or decreasing their impact upon the patient.

Genome-wide association studies (GWAS) involve the analysis of a comprehensive inventory of hundreds of thousands of single-nucleotide polymorphisms (SNPs) in hundreds of thousands of cases and controls from a population, to find the variants associated with a disease or traits. It has been estimated that 10 million common SNPs are transmitted in blocks across generations, and some particular tag SNPs allow to capture the vast majority of SNPs variation within each block (2). Technical advances and quality control now permit to obtain reliable and affordable genotyping of up to 1 million SNPs of a person’s DNA in a single scan (3). Since 2006, 1,628 GWAS have been published, and identified hundreds loci associated with more than 250 common diseases or traits (4). Even if they are sometimes methodologically questionable (5), an additional 2,811 meta-analyses on 2,194 genes investigated with a candidate gene approach have been published since 2000 (6). Overall, only 12% of SNPs associated with complex phenotypes are located in or occur in tight linkage disequilibrium with protein-coding regions of genes, about 40% fall in the intergenic regions and another 40% in noncoding introns, suggesting a role of these latter regions in the regulation of gene expression. A few major findings emerge from this huge amount of data and include: the moderate effect of the majority of common variants at the disease loci, which increase the risk by 10 to 50%, similarly to the effect of many environmental risk factors; the large number of loci influencing most examined traits; the confirmation of the association between complex diseases and some genes earlier implicated using linkage analysis; and the discovery of many new susceptibility alleles.

The detection of hundreds of loci involved in modulation of the phenotype of complex traits and diseases provides clues to determine the underlying cellular pathways, and in some cases also gives new hints concerning therapeutic approaches. Several results obtained by studies carried out in very recent years support these issues.

In this narrative review, we will describe some of the most relevant results obtained so far, based on a selection of published studies we consider most useful for pinpointing the state of the art of GWAs of complex traits and diseases and their impact on clinical and public health practice.

GWAS and candidate-gene studies

Table 1 summarizes some of the most relevant results of GWAs of common complex disorders published so far. GWASs have disclosed the etiology of age-related macular degeneration (ARMD), a leading cause of blindness affecting millions of aged individuals. Seven loci with common variants of major effect have been found to be associated with this complex disorder, each of which results in an increased risk of disease (7). Two of these variants, found in the complement factor H-related (CFHR) gene, display allele frequencies of 36% and 57% among unaffected persons. All together, these polymorphisms more than double the disease risk in the siblings of persons affected by ARMD, explain about half of its heritability, and are likely to drive the disease risk by a loss-of-function mechanism (8). These results also point to a central role of the complement-mediated inflammation pathway, which appears to be a promising target for the development of new therapies.

About 100 loci related to autoimmune diseases have been found, some of which are shared by several of these disorders, involving fundamental regulatory pathways as well as other disease-specific pathways.

GWASs have so far identified about 100 susceptibility loci for Crohn’s inflammatory bowel disease, accounting for about 20 to 25% of disease heritability (9). Three of these variants, found in NOD2, IL23R, LRRK2 genes, are common, since all but one have allele
frequencies higher than 9% in the examined populations. These major polymorphisms are associated with an increased risk of disease with odds ratios (ORs) in the range of 1.5 to 4, while the remainder confer small risks (ORs in the range of 1.08 to 1.35). An association was established between the disease and variations in genes with previously unrecognized roles in this disorder, affecting processes such as innate immunity, autophagy and interleukin-23 receptor signaling. The elucidation of the pathogenic role of distinct mutations in cellular and animal models is now leading to the development of new therapeutic strategies.

Studies in type 2 diabetes (T2D) have so far identified about 30 pathogenic loci (10). Many polymorphisms found in these patients affect genes involved in insulin secretion rather than insulin resistance. In addition, genes implicated by biochemical analysis in glucose regulation do not appear to be associated with T2D, but rather with fasting glucose levels. All together, these results explain about 20-25% of the disease heritability.

Candidate gene studies identified four non-HLA type 1 diabetes (T1D) risk loci, INS, CTLA4, PTPN22, and IL2RA. The application of genome-wide SNP typing technology to large sample sets has detected more than 40 susceptibility loci, which account for about 60% of the disease heritability, with ORs in the range of 1.1 to 1.3 (11). Genes of possible functional relevance to T1D include GLIS3, previously related to permanent neonatal diabetes with congenital hypothyroidism and other complications, immunoregulatory genes, and, possibly, some members of the calcium-dependent lectin domain family with immune function.

A GWAS enrolling >100,000 individuals of European ancestry identified 95 loci associated with plasma lipid levels, which represent a major risk factor for myocardial infarction (12). On the whole, these variants explain 20-25% of genetic variance of low-density lipoprotein (LDL) and high-density lipoprotein (HDL) levels. Interestingly, these loci include 18 genes previously implicated in some rare mendelian disorders, supporting the usefulness of GWASs for pinpointing genes harboring rare variants. In addition, this study has shown that loci with only moderate susceptibility effects may have major therapeutic implications. This point is illustrated by HMGCR gene, in which a common variant, found in about 40% of the population, induces a very small change in the LDL levels, while the encoded protein is a target of statins, a class of drugs widely used to reduce LDL levels and the risk of myocardial infarction.

The largest recent GWASs have further explained the heritability of many common disorders. An illustrative example was the discovery of three loci which modulate the erythroid development, together explaining >25% of the genetic variance of fetal hemoglobin (HbF) levels, associated with reduced severity in β-thalassemia and sickle cell anemia (13). This result has raised hopes of treating these diseases by increasing the HbF expression. Another example is height, a heritable polygenic trait, for which more than 180 loci implicated

<table>
<thead>
<tr>
<th>PHENOTYPE</th>
<th>NUMBER OF LOCI</th>
<th>PERCENTAGE OF EXPLAINED HERITABILITY*</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes type 1</td>
<td>41</td>
<td>~60</td>
<td>11</td>
</tr>
<tr>
<td>HbF levels</td>
<td>3</td>
<td>~50</td>
<td>12</td>
</tr>
<tr>
<td>Macular degeneration</td>
<td>7</td>
<td>~50</td>
<td>7, 8</td>
</tr>
<tr>
<td>Diabetes type 2</td>
<td>39</td>
<td>20-25</td>
<td>10</td>
</tr>
<tr>
<td>Crohn's disease</td>
<td>99</td>
<td>20-25</td>
<td>9</td>
</tr>
<tr>
<td>LDL and HDL levels</td>
<td>95</td>
<td>20-25</td>
<td>13</td>
</tr>
<tr>
<td>Stature</td>
<td>180</td>
<td>~12</td>
<td>13</td>
</tr>
</tbody>
</table>

* Calculated by dividing the phenotypic variance explained by polymorphisms at loci identified by GWA by the total heritability calculated on epidemiologic parameters.
in several biological pathways, including TGF-β signaling, have been identified (14).

Several GWASs have shown that some susceptibility loci can be shared by complex phenotypes considered to be unrelated (15). Illustrative examples pointing to common etiologic pathways in disparate conditions include Crohn’s and Parkinson’s disease (LRRK2 gene), childhood asthma and Crohn’s disease (ORMDL3 gene), T2D, melanoma, and coronary disease (CDKN2A, CDKN2B genes), prostate cancer, and height (JAZF1 gene). The potential for using drugs that are effective in the treatment of one condition for the treatment of another remains to be established.

In general, the SNPs-disease association has a modest effect size, having a median OR per copy of the risk allele of 1.33, with several variants carrying ORs above 3.00 and a few exceeding 12.00 (16). Therefore, most common gene variants implicated by such studies are responsible for only a small fraction of the predicted genetic variation. However, small effect sizes do not necessarily mean that a gene variant is of no interest or use. Effect size is directly related to the real variant’s functional role, which can only result in slight changes in gene expression or in protein function. Accordingly, the gene pathway critical for a particular condition or pharmacological action on a given protein may produce a much larger effect in the disease control (17). Interestingly, some polymorphisms with ORs of less than 1.45 provide insight into the pathophysiology of the disease and uncover new targets for therapy. This is shown by PPARG gene associated with T2D, whose protein product is recognized as the receptor for the thiazolidinedione class of insulin sensitizers. In addition, KCNJ11 gene, whose variants are associated with diabetes with an OR of 1.2, codes for the sulphonylurea receptor, a major target for drug therapy of this disease (18). Similarly, IL12B polymorphisms, associated with psoriasis, encode proteins that are targets for anti-p40 antibodies (19).

Despite the promising data mentioned above, the association with complex diseases of most of the genomic variants identified so far in GWASs have shown to be inconclusive, and their functional significance remains elusive. Among the problems and drawbacks of many of these studies, one should include the need for replicating results by independent researchers as well as the need for validation by prospective studies, which typically takes years. A major issue is that many genomic variants have a very modest association with diseases and this raises the question about their use in clinical practice. An illustrative example is the association between 12 genomic variants conferring a 1.6 times greater risk of heart attack in 10% of people of European ancestry, over the population risk of the same origin, independent of other risk factors. The 10% of individuals have a lifetime risk for heart attack of about 78%, compared to the population risk of 49% for >40 year-old men, but this specific risk cannot be controlled by any known intervention (20). Empiric risks are available for many common diseases, based simply on family history, independent of genomic testing. Simulation studies have shown that the predictive value of genomic profiling may attain the same level as the traditional risk in predicting cardiovascular disease, being no higher or high enough for predictive diagnosis (21). Implementation of genetic risk prediction in health care, however, is just at its beginning (OR still in its infancy), since it requires a series of studies encompassing all phases of translational research, starting with a comprehensive evaluation of genetic risk prediction. In this sense, it is crucial both to improve and standardize the quality of genetic risk prediction studies, and to enhance the quality of their reporting. The recently published Strengthening the Reporting of Genetic Risk Prediction Studies (GRIPS) statement aims to enhance the transparency of study reporting and thereby improve the synthesis and application of information from multiple studies that may differ in design, conduct, or analysis (22).

Direct-to-consumer (DTC) genetic testing

The ability of genomic scans to screen for many conditions at once and to assess the individual risk of diseases is an attractive promise of personalized medicine, that can be broadly defined as a customisation of healthcare that accommodates individual differences as far as possible at all stages in the process, from prevention, through diagnosis and treatment, to post-treatment follow-up. However, unlike single gene disorders, genetic testing of common complex diseases remains a complicated issue. Based on the current level of knowledge, most sequence information gained
through whole-genomic sequencing analyses are still of unknown meaning and importance. In addition, all the issues listed below should be carefully considered when approaching whole-genome sequencing: the content of the pre-test information to be provided to the subject; the procedure used; the associated benefits; the risks and limitations; the need for skilled interpretation of test results; the post-test information to be successfully conveyed to the subject. Accordingly, the informed consent for this procedure is more complex than for existing genetic testing (23). Whole-genome sequencing shows that every individual is ‘genetically imperfect’, in the sense that he or she has an above-average risk for some disorder or for having children with a genetic disease. A typical person may have approximately 150 rare coding variants affecting approximately 1% of his or her genes (24). Therefore, awareness of genetic risks resulting from large scale implementation of these tests could have some negative social consequences, in terms of stigma, employment, insurance or other. Finally, the difficulty for the patient to understand genetics and application of the concept of risk probability should be also considered.

These caveats deserve major attention because, since 2007, several commercial ventures are providing GWASs directly to consumers for medical, genealogic, and even recreational purposes. A critical appraisal of the scientific basis of commercial genomic testing used to assess health risks and personalized health interventions, based on predictive genomic profiling offered by seven companies, pointed out that there was “insufficient scientific evidence to conclude that genomic profiles are useful in measuring genetic risks for common diseases or in developing personalized diet and lifestyle recommendations for disease prevention” (25). More recently, the United States Government Accountability Office (GAO) has evaluated the results of direct-to-consumer (DTC) genetic testing on five donors, including 15 tests, purchased from four companies. The GAO’s fictitious consumers received test results that were misleading and contradictory across the four companies and of little or no practical use. In addition, DNA-based disease predictions conflicted with the actual medical conditions, and three companies failed to provide expert advice. These observations suggested that although these tests show promise for the future, the consumers should be aware of the unreliability of these results at the present time (26).

Three key policy documents were published in Europe in 2010. The European Society of Human Genetics published a document on DTC genetic testing that highlighted the importance of right to information; quality of test performed; clinical usefulness of the test provided; the need for individualized medical supervision; the provision of pre-test information and genetic counseling; follow-up and support in the interpretation of results and their psychosocial impact, the protection of persons not able to consent; respect for privacy and confidentiality; the storing of samples; their propriety and respect for ethical principles in research (27). The UK Human Genetics Commission, a government advisory body, launched its “Common Framework of Principles for DTC Genetic Testing Services”, to guide the development of the EU codes of practice, by taking into account different existing regulatory structures (28). This framework for voluntary regulation covers a broad range of issues for DTC genetic testing that embraces the basic elements of consent, data protection, truth in marketing, scientific rigour and balanced interpretation. Thirdly, a comprehensive report giving a detailed list of recommendations on DTC genetic testing directed towards policy makers and stakeholders has just been released (29, 30). This presents the results of an expert working group’s activity within a project initiated by EASAC (European Academies Science Advisory Council) and FEAM (Federation of the European Academies of Medicine), with support from the IAP (global network of the world’s science academia) and which aims to review the scientific evidence already available on DTC genetic testing, to assess the regulatory developments underway, and to ascertain the principles that should underpin the options for action by public policy-makers. 

Fulfilling the promise of genomics in improving citizens’ health, however, inevitably requires a public health perspective (31, 32). We cannot ignore the potential for increasing health care expenditure, especially in countries where the health care system is publicly funded. Even though the rapidly evolving technologies are dramatically decreasing the cost of whole-genome scanning, overall the cost of interventions related to more personalized (preventive) treatments are expected to increase (33). A prominent US health insurer published in March 2012 a
white paper reporting an expected increase in spending for genetic testing of around $15-25 billion in 2011, compared with $5 billion in 2010 (33). Furthermore, the paper reported the results of a survey among >1,000 physicians and >1,000 citizens which showed that 63% of physicians and 77% of customers agreed with the statements that “Genetic testing gives me the ability to diagnose conditions that would otherwise be unknown”, and “Genetic testing allows for more personalized medical decisions”, respectively. This clearly shows, at least in the US, that the use of whole-genome scanning is likely to increase in the near future. On the other hand, a plethora of publications underlines the need for health literacy on genomics for both physicians and citizens (34). In this regard, Italy is at the forefront, as the Ministry of Health recently funded dedicated projects to set up training courses on translational genomics aimed at the potential prescribers of genomic tests, namely general practitioners, public health specialists of community prevention services, oncologists, gynecologists and neurologists (35). Italy is also the first country in Europe where the Ministry of Health will publish a plan of action for genomics and predictive medicine within the framework of the 2010-2012 National Prevention Plan (36), based on a white paper published by an expert working group (37). From this considerations, it is evident the key role of public health in dissecting the promise and pitfalls of genomic advancements for the benefit of the general population.

CONCLUSIONS

Although whole-genome sequencing is a nascent technique, availability of genomic information is becoming increasingly widespread. Estimation of post-test probability of disease, by applying likelihood ratios derived from integration of multiple common variants to age-appropriate and sex-appropriate pre-test probabilities, provides the proof of principle that clinically meaningful information can be obtained about disease risk and response to drugs in individuals using whole-genome sequencing data (38). Development of methods integrating genetic and clinical data is needed and represents a large step towards personalized medicine. However, in order to obtain maximum benefit from GWASs and to keep disadvantages to a minimum, the establishment of guidelines will be necessary to manage the technical advances and to meet the challenges involved in the clinical application of whole genomic sequencing. It will also be essential to create the necessary infrastructure to drive health benefits in the future, and to increase health literacy among physicians and citizens.

FUNDING: this study was supported in part by a grant from the Italian Ministry of Health (Ricerca Corrente 2012), and in part by a grant from the Centro nazionale per la prevenzione e il Controllo delle Malattie (CCM) within the Italian Ministry of Health (project titled “Predictive genomic tests: census in some Italian regions for the establishment of a register of supply, and promotion of training for physicians”).

References
(7) Maller J, George S, Purcell S, et al. Common variation in three genes, including a noncoding variant in
CFH, strongly influences risk of age-related macular degeneration. Nat Genet 2006; 38: 1055-9
(34) Sørensen K, Brand H. Health literacy: the essential catalyst for the responsible and effective translation of genome-based information for the benefit of population health. Public Health Genomics 2011; 14(4-5): 195-200

