Biochemical and Biophysical Research Communications 396 (2010) 143–146

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Genetics of complex disorders

Juha Kere

Karolinska Institutet, Department of Biosciences and Nutrition, 14157 Huddinge, Sweden
Folkhälsan Institutet of Genetics, Helsinki, and Department of Medical Genetics, University of Helsinki, 00014 Helsinki, Finland

ARTICLE INFO

Article history:
Received 31 March 2010

Keywords:
Gene mapping
Linkage disequilibrium
Genome-wide association study

ABSTRACT

The success stories of identifying genes in Mendelian disorders have stimulated research that aims at identifying the genetic determinants in complex disorders, in which both genetics, environment and chance affect the pathogenetic processes. This review summarizes the brief history and lessons learned from genetic analysis of complex disorders and outlines some landscapes ahead for medical research.

1. A short history of rapidly changing analytic methods

The history of modern human genetics research is much the history of the rapidly changing methodologies. The vast size of the human genome appeared at first an impossible task to work with, and certainly to know in full detail. But with the advent of new analytical approaches roughly every 10 years, we have now reached the point where individual genome sequencing is within the reach for many laboratories.

The brief history and some of the key papers to read related to each stage are listed in Table 1. Modern human molecular genetics research owes much to the innovative method that bears the name of its inventor: Southern blotting (and the analogous gel-blotting techniques named with a tongue in the cheek, Northern and Western blotting). More exactly, Southern blotting applied to the analysis of restriction fragments obtained after digestion with a growing list of various bacterial restriction enzymes allowed the analysis of genome position-specific polymorphic markers [1]. The idea that these markers, restriction fragment length polymorphisms (or RFLPs) could be used as anonymous addresses for the various positions in the human genome led to systematic efforts to clone probes and screen them for their polymorphic nature [2]. Ultimately, hundreds of such cloned probes combined with the study of large families were used to derive the first continuous genetic maps of the human genome [3]. Such maps then made it possible to attempt the mapping of any Mendelian trait to a named position in the genome, with first success stories setting the stage of the first wave of disease gene identifications [10].

Southern blotting combined with the use of radioactive probes was not optimal for large-scale projects. The invention of the polymerase chain reaction (PCR) changed molecular genetic analyses rapidly, and a new class of genetic markers was discovered. When RFLPs were caused either by single nucleotide changes altering restriction enzyme recognition sites or larger-scale variations in tandemly arranged DNA repeat units, the new class of microsatellite markers were based on length variation in short sequence repeats [4]. This class of markers, often involving poly-CA dinucleotide repeats, or various tri- or tetranucleotide repeats, could be analyzed more rapidly and easily, and moreover, there appeared to be many more of them in the genome. The second-generation genetic linkage maps of the human genome were then based entirely on PCR-amplified microsatellite markers [5].

The analysis of microsatellite markers still required the separation of differently-sized DNA fragments. Another line of technology was being developed based on the high selectivity of DNA hybridization that could discern sequence differences as small as a single nucleotide. The rather common and simple assay based on the application of short DNA probes on filter papers as dot-blots was refined and miniaturized to make microarrays of DNA probes on glass slides. The first applications of the microarray technology were directed toward measuring levels of different transcripts in RNA samples [11]. When the common variation in DNA sequences between individuals denoted single-nucleotide polymorphisms (SNPs) started to be appreciated, it became possible to make microarrays assaying at first thousands and later hundreds of thousands of SNPs all over the genome that were discovered in the meantime by genome sequencing and other means.

In parallel with these developments, DNA sequencing technologies took rapid leaps. The dideoxynucleotide-based sequencing method invented by Fred Sanger [12] was taken from manually-read radioactively labelled gels to machine-read fluorescently labelled gels and ultimately to matrix-filled capillaries. Massively parallel methods based on immobilized DNA molecules on the surface of microparticles then evolved in various forms to the new

E-mail address: juha.kere@ki.se

0006-291X/$ - see front matter © 2010 Published by Elsevier Inc.
doi:10.1016/j.bbrc.2010.04.013

Table 1
Selected key events in the mapping of disease genes.

Development	Importance	References
DNA blotting	Easy analysis of DNA variation	[1]
Concept of RFLP maps	Vision of human gene mapping	[2]
Continuous linkage maps of the human genome	Genome-wide linkage mapping of human genes	[3]
Microsatellite maps	Mapping of many human disease genes	[4,5]
Extended linkage disequilibrium in the genome	Vision of linkage disequilibrium mapping in complex disorders	[6,7]
The HapMap project	Detailed analysis of linkage disequilibrium in the human genome	[8]
Genome-wide association studies in complex disorders	Hundreds of loci identified	[9]

generation DNA sequencing technologies. Again, one example of a yet newer technology is based on assaying the progressing synthesis of a DNA strand of individual DNA polymerase molecules on nanofabricated devices [13].

2. Candidate gene studies

All these methodologies were first applied to the study of monogenic or Mendelian disorders. Even though complex disorders seemed a difficult target for study, the early attempts to disentangle the genetic determinants were based on candidate gene approaches. The rationale was to use genetic association to ask whether certain polymorphisms in genes that were thought to be functionally relevant in a given disease would show different allele frequencies in patients as compared to controls. A large number of studies were published, many based on patient and control collections as small as only tens of samples, or at best several hundreds, and numerous weak associations were reported in the literature. Some of the identified genes have withstood the verification by better-powered genetic association studies, but most have never been replicated in another study.

3. Chromosomal markers as pointers to genes in complex disorders

In complex as well as in monogenic disorders, one way to identify possible disease-related genes has been to study specific rare chromosome abnormalities in patients with a monogenic or complex disorder. In monogenic diseases, there are numerous success stories, such as for example the mapping and identification of the Duchenne muscular dystrophy gene or the anhidrotic ectodermal dysplasia gene [14,15]. In complex disorders, however, some of the identified genes have likewise turned out to be relevant even by genetic association tests; the first dyslexia gene DYX1C1 might serve as an example [16].

4. Genetic linkage studies

Genetic linkage studies in monogenic disorders are based on the simple model of Mendelian inheritance, and linkage analysis algorithms based on this model are very powerful for showing genetic linkage. In complex disorders, however, the situation is much less straightforward as a given genotype would not at all fully correlate with the presence of the disease phenotype. In genetic terms, both phenocopies (the occurrence of the same phenotype in genetically completely different individuals), reduced penetrance (the occurrence of the phenotype in only a fraction of those possessing a given genotype) and variable expressivity (the clinical heterogeneity of a given condition in the presence of the same genotype in different individuals) complicate matters. Therefore, new computational approaches were needed to complement those of model-based monogenic linkage studies.

Many successful approaches in complex disorders were based on affected-only strategies. The idea was to identify genomic regions that more often than by chance were shared between individuals affected with a complex disorder [17]. No information content was given to healthy individuals. Even though phenocopies might hamper the identification of shared genomic regions, large enough sample sets should ultimately implicate specifically disease-linked positions with acceptable levels of statistical evidence.

The 1990s were the golden period of genome-wide linkage studies in complex disorders. Accumulated studies, however, started little by little to cast doubt on the replicability of the mapping results. Typically, no two studies of the same disorder reported consistently the same loci, but instead all chromosomes were littered with positive but nonreplicated linkage hits.

Only quite recently has it become apparent what was the problem with these approaches. We now know that many true susceptibility genes in complex disorders have effect sizes so small that genetic linkage, in order to be powerful enough should have been based two or three orders of magnitude larger sample sets. The early attempts to map genes in complex disorders were not wrong in principle, but the studies were simply underpowered to detect the weak genetic effects underlying many complex disorders.

Nevertheless, strong single gene effects in some complex disorders have been abundantly verified already by genetic linkage and later confirmed by genetic association. An illustrative example is psoriasis, where the overwhelming genetic effect of the HLA-C gene or SNPs nearby was detected in virtually all genetic studies [18].

5. The paradigm of positional cloning in complex disorders

In monogenic disorders, the genetic mapping of the causative locus was followed up by narrowing down the DNA segment including the causative gene by genetic association. The idea was to take advantage of founder effects, especially in recessive disorders where a mutated gene might be inherited for generations in a population until it happened to find its way homozygous in an individual, then causing disease. The same principle was applicable to the thinking in complex disorders [19]. A susceptibility allele with reduced penetrance might travel for generations in families and populations with little if any selection, as many complex disorders only manifest after the reproductive age window.

Some success stories exist with this approach. One example is the genetic mapping and positional cloning of NPSR1 as a susceptibility gene for asthma and related disorders [20]. In these studies, we took advantage of a rural isolated subpopulation in Finland hoping to reduce the genetic complexity of an otherwise heterogeneous complex disorder. By a lucky incidence, a particular asthma risk associated variant appeared to be more common there than in many other populations, possibly helping to reach barely sufficient power to map the gene by linkage and association. Later association studies, including well-powered genome-wide association studies have confirmed the disease association of NPSR1 to asthma.

6. The block structure of human chromosomes and the HapMap project

Millions of SNPs across the genome are not particularly helpful to map and identify disease-associated genes in complex disorders,