Lessons from studying monogenic disease for common disease

Leena Peltonen1,2,3,*, Markus Perola1,2, Jussi Naukkarinen1,2 and Aarno Palotie3,4

1Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland, 2Department of Medical Genetics and Program for Molecular Medicine, University of Helsinki, Finland, 3The Broad Institute of MIT and Harvard, Boston, MA and 4Finnish Genome Center, Department of Clinical Chemistry and Program for Molecular Medicine, University of Helsinki, Finland

Received March 2, 2006; Revised and Accepted March 9, 2006

The prevailing paradigm for common disease emphasizes the role of common variants predisposing to various rampant health problems, and these genetic risk profiles interact with environmental and life style risk factors triggering the disease process and modifying its progress. However, most of what we know about the molecular background of common diseases is in fact based on what we have learned from rare familial forms of these traits. Further, often the mutations identified in even more rare monogenic diseases sharing some trait components with common diseases have exposed critical new pathways involved in the molecular pathogenesis of common health problems. In this review, we aim to exemplify some of the lessons learned from rare Mendelian forms of diseases that have significantly contributed to our understanding of the genetic background of common diseases and their trait components.

MENDELIAN CASES OF COMPLEX TRAITS

Identification and characterization of different mutations in Mendelian variants of common traits have been extremely successful with improved accuracy of genome-wide analyses provided by the Genome Project. These, usually rare single gene disorders masquerade a multifactorial trait in their clinical phenotype, but careful data collection of family history (Mendelian segregation in a family), and detailed clinical examination (the cases are exceptionally severe or deviate otherwise from the typical disease cases) have revealed the exceptionally strong genetic character of the trait. Typically, linkage and association studies in such families have guided to a gene responsible for the disease in those particular families. Usually the mutations identified have been nearly family-specific and their value in clinical diagnosis has remained modest. Neither have they explained the majority of cases at the population level. However, by studying these ‘human allelic knockouts’ a unique view into the pathophysiology involved has been obtained. Table 1 provides a list of examples for which rare Mendelian forms of common diseases have resulted in the identification of disease genes and subsequent functional studies have evidenced their relevance for common traits. Informative examples among them include familial Alzheimer’s disease (AD), MODY(s) (maturity onset diabetes of the young), familial combined hyperlipidemia (FCHL) and rare familial forms of epilepsies and migraine. These disorders, characterized by their distinct heritability have opened new avenues in further evaluation of the more common forms of dementia, diabetes, dyslipidemias and common forms of epilepsy and migraine.

Familial ADs

If only judged by the number of publications, most genetic studies on AD have considered one genetic polymorphism, the Apolipoprotein E epsilon 2/3/4 variant as a major risk determinant. However, most insight into the molecular pathogenesis and events resulting in characteristic tissue pathology has been produced by studies of the early onset familial forms of the disease, showing classical Mendelian inheritance in exceptional dementia families. In Caucasians, about 1–6% of all AD is early onset (<60 years) and ~60% of this early-onset AD is familial, with 13% inherited in an autosomal dominant manner (1,2). The AD mutations identified in these families exposed genes encoding presenilins and the amyloid precursor protein (APP), the common denominator being that they all alter APP processing in the brain. Together, variations in these three genes cover ~70% of early-onset AD mutations in the French (2). There is most likely considerable population based variance in these frequencies, but all in all, these rare variants probably explain only about 1–2% of AD cases. That said, the accumulated knowledge of the critical metabolic pathway, and the emerging understanding that one
metabolic pathway is shared by all the familial forms of the disease cannot be surpassed in importance when considering the etiology of the complex form of the trait. The understanding of rare Mendelian forms of AD has even paved the way for new hypotheses concerning the etiology and pathogenesis of neurodegenerative diseases more generally. In AD, mutations in the APP or presenilins result in disease by a mechanism that involves the deposition in the brain, of abnormally processed amyloid. This is also seen in other neurodegenerative diseases such as Creutzfeldt–Jakob disease and Parkinson’s disease (3).

Familial diabetes, MODY

MODY represents a rare familial form of common diabetes and accounts for 2–5% of diabetes mellitus (DM) in most western societies (4), for <5% of DM in childhood and 0.5–1% of non-insulin dependent DM in most Caucasian (European) populations (5). All known forms are characterized by an autosomal dominant inheritance and the age of onset in early adulthood (usually before 25 years of age) (6). Currently six mutated genes for MODY are known, whereas some forms of MODY remain un-characterized at the molecular level (7). Mutations in these six genes account for 80% of European and 20% of Japanese with clinical MODY. The mildest form of MODY is caused by mutations in the glucokinase gene, while five other more severe forms are caused by mutations in genes encoding various transcription factors. Identification of variant MODY genes has provided insight into the development of B-cell dysfunction in the pancreas paving way towards understanding of the molecular processes behind more common forms of DM as well. The fact that all MODY genes are critical for development and function of the pancreatic beta cells has not only produced a new paradigm for the genetic studies of more common forms of diabetes—but also opened the way for understanding the pathophysiology behind different forms of diabetes (7). It became evident that MODY is not a single entity but presents clinical, metabolic and genetic heterogeneity, the phenotype correlating with the causative mutations. Further, some genetically specified forms of MODY are responsive to specific treatments underlining the importance of careful molecular dissection of the trait. (8). These lessons have definitely molded our thinking of the molecular background and its significance for phenotype variations in common forms of diabetes.

Familial dyslipidemias

FCHL is a combination of heterogenous dyslipidemias segregating in families and predisposing affected individuals to early-onset cardiovascular disease. FCHL is considered the most common familial dyslipidemia in Caucasian populations and estimated to be responsible for ~10% of early-onset coronary heart disease. Though the syndrome was described in the early 1970’s among survivors of early myocardial infarcts, the first FCHL-associated gene was described only recently in Finnish FCHL families (9). This gene on 1q encodes the upstream stimulatory factor-1 (USF1), a ubiquitous transcription factor regulating numerous genes participating in lipid metabolism.
and glucose metabolism (10,11) (Fig. 1). The involvement of the allelic variants of \textit{USF1} in dyslipidemias and trait components of the metabolic syndrome has been replicated in different populations and study samples ascertained for these traits (9,12–15). The identified risk alleles do not seem to represent mutations breaking the protein, but rather variants putatively affecting the transcript level of the gene itself. Importantly, allelic variants of \textit{USF1} seem to contribute to cardiovascular disease risk also at the population level (16). The function of \textit{USF1} and especially the number of highly relevant downstream genes regulated by \textit{USF1} makes this gene, initially identified in rare exceptional families with combined hyperlipidemia, an attractive candidate for trait components of the metabolic syndrome.

\textbf{Familial epilepsies and atypical migraines}

Epilepsies and migraine are two paroxysmal CNS disorders in which major efforts have been invested in studies of Mendelian trait variants as models to unravel the pathogenic mechanisms involved in the more complex forms of the trait. Similar to many Mendelian forms of common traits even within exceptional multi-case families, the range of the phenotype is broad: some of the affected family members have an extreme clinical phenotype, some have phenotypes indistinguishable from more common forms of migraine or epilepsy. Thus, although these familial forms represent only a few percentages of the overall prevalence, they have been considered attractive models in identifying the underlying metabolic pathways. Interestingly, they all seem to fall under the umbrella of channelopathies. In monogenic epilepsies alterations in potassium, sodium and chloride channels as well as GABA and acetylcholine receptors have been described (17). In the case of Mendelian forms of migraine, Familial Hemiplegic Migraine (FHM), alterations in calcium and potassium channel genes and in an intracellular Na+/K+ -ATPase (18,19) have been described. The apparent, well documented central role of alterations in ion channels in the Mendelian forms of these two traits has stimulated the hypothesis that susceptibility to more common forms of migraine and idiopathic epilepsy could also, at least partially, be explained by altered functions of neuronal ion channels. In the case of epilepsy, the calcium channel gene \textit{CACNA1H} and the \textit{GABARD} gene have been associated with common forms of idiopathic epilepsy; \textit{CACNA1H} with idiopathic generalized epilepsy and childhood absence epilepsy, \textit{GABARD} with

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{example.png}
\caption{Example of a gene identified in rare families and it’s potential significance for trait components of common diseases, transcription factor USF1: the expression of numerous genes involved in lipid and glucose metabolism must be dynamic to meet the changing requirements of our physiology. USFs activate and repress the expression of many of these genes according to hormonal and metabolic cues such as insulin and glucose by binding an E-box promoter sequence upstream of its target genes. In the presence of insulin, the USF1/USF2 dimer becomes phosphorylated, precluding its binding the E-box and resulting in decreased expression of the target gene (15). Variation in the function of USF1 has potentially far reaching effects on lipid and glucose metabolism, features relevant to a number of common dyslipidemias and trait components of the metabolic syndrome.
}
\end{figure}
idiopathic generalized epilepsy and generalized epilepsy with febrile seizure plus (17). In the case of common forms of migraine, good evidence for the involvement of these genes in trait susceptibility has not yet been reported, although it remains a strong hypothesis.

**RARE MONOGENIC DISEASES REVEALING CRITICAL PATHWAYS**

**Genes of exceptional obesity**

Obesity, a dramatically increasing plague of global societies has been traditionally considered to be largely environmental or lifestyle dependent. However, the lessons from experimental animals and extremely rare monogenic syndromes in humans have pinpointed the critical importance of brain and adipose tissue-derived hormones and the hormonal crosstalk between fat tissue and the hypothalamus-hypophysis axis. Different monogenic forms of severe obesity have been found to be caused by mutations in specific genes encoding hormones, hormone receptors or their regulatory molecules (20). The pioneer work on leptin, initially identified on the basis of the rodent models and progressing to direct demonstration of the lack of leptin in human patients with syndromic obesity has provided undeniable evidence for the tremendous impact of normal leptin levels on human fat mass and opened new avenues for novel treatment strategies of morbid obesity. Mutations identified in these rare syndromes have not proved to be common among obese individuals at the population level, with a notable exception of one melanocortin receptor gene, MC4R, seem to be important for obesity; up to 4% of obese children are estimated to carry mutations of this gene (20) and some evidence exist that different allelic forms of this gene associated with a higher body mass index at the level of general population (21). Findings in extremely rare forms of obesity, often represented by only a handful of children or families have thus highlighted critical pathways for the control of human fat tissue and its normal metabolism.

**Rare syndromes with hypertension**

Most of our understanding of the molecular pathogenesis of hypertension emerges from a systematic characterization of the mutations in rare monogenic diseases, characterized by high blood pressure. This tedious work has been of paramount importance, because although the environmental and lifestyle risk factors of common hypertension are reasonably well defined (e.g. salt intake, age, gender and body mass index), the primary molecular defects and triggering cascades behind essential hypertension have remained obscure. It is the rare monogenic diseases that have provided an avenue for identification of critical molecules behind abnormally elevated blood pressure. About 20 different single gene disorders that cause either hyper or hypotension have been identified thus far. Markedly, as in AD, all Mendelian forms identified thus far can be considered to represent ‘the final common pathway,’ they all interfere with the normal function of the critical molecules involved in the regulation of salt balance of the body by the kidneys (22).

The epithelial sodium channel (ENaC), functioning in the cortical collection tubule of the nephron plays a crucial role in several forms of monogenic hyper or hypotension. Gain-of-function mutations in the beta and gamma subunits of ENaC cause a rare Liddle’s syndrome with severe hypertension (23). Loss-of-function mutations in any of the three subunits of ENaC cause an autosomal recessive syndrome with hypotension (pseudohypoaldosteronism type I, OMIM 264350). Thus, dysfunctions of all three subunits of this sodium channel result in human diseases, either in hyper or hypotension. Another set of rare syndromes, resulting from mutations in the mineralocorticoid receptor gene (24) affect the activity of this receptor, an important regulator of ENaC activity (25). Such studies of rare Mendelian disorders with dysfunctions of blood pressure regulation have highlighted the crucial importance of a set of molecules functioning in Henle’s loop and regulating salt balance under the control of renin-angiotensin-aldosterone system has advanced our understanding of the molecular pathogenesis of the more common essential hypertension as well as provided new target molecules for pharmaceutical industry (25).

**Genes behind monogenic autoimmunity**

Concerning the complex background of human autoimmune disorders, rare monogenic diseases have provided critical clues of the molecular pathogenesis involved. The investigation of these diseases have identified molecules of the immune system that would have remained uncharacterized without the detailed clinical description of exceptional families that served as a catalyst for further molecular work. A good example is APECED, a rare monogenic autoimmune disease, enriched in Finland and in some other isolated populations, such as the Iranian Jews. Isolation of the mutated gene Aire (autoimmune regulator) by classical positional cloning in Finnish families and subsequent production of the knock out mouse model resulted in the identification of a novel critical molecule that regulates the ectopic expression of a subset of autoantigens, characteristic for peripheral tissues, during the fetal period in thymus (26). Further studies have implied that this process, critical for the so-called central tolerance, is in the crucial position for the development of late-onset common autoimmune diseases like type I diabetes and have provided new avenues for production of accurate animal models and future therapy trials for these common traits (27).

**SAME GENE, DIFFERENT MUTATIONS IN RARE AND COMMON FORMS OF DISEASE**

Characteristic of early onset Mendelian traits with severe phenotypes, the causative mutations generally result in amino acid alterations and truncated protein products and their consequences for molecular functions are relatively easy to predict. Most probably the genetic predisposition for late-onset common traits is more often associated with variants in non-coding areas, likely affecting the regulation of transcription, tissue specific splicing or with variants that carry their effect by some other, more subtle mechanisms during the life span of an individual. Although the
number of well established variants of specific genes associated with any common trait is still limited, they seem to support an overall hypothesis that a large fraction of common predisposing variants do not alter the primary protein structure.

A good example of the differing effects of different type of mutations of the same gene is provided by the case of two forms of lactose intolerance. Homozygote nonsense or truncating mutations breaking the synthesized polypeptide and resulting in complete absence of the intestinal lactase enzyme were recently described to cause congenital lactase deficiency—a severe gastrointestinal disorder characterized by watery diarrhea in infants fed with breast milk or other lactose-containing formulas (28). Interestingly, another variant located 14 kb upstream from the lactase gene has been established resulting in an adult type lactose intolerance (lactase non-persistence), causing much milder late-onset symptoms and being common in most global populations (29). This variant interferes with a cis regulatory element capable of enhancing transcriptional activation of the lactase promoter, which is characteristically down-regulated in humans after weaning (30). The lesson obtained might hold also for other genes in which mutations resulting in a broken polypeptide produce a rare early-onset trait, whereas causative variants behind common late-onset trait might well be identified in non-coding and regulatory regions of the same gene, posing a much harder challenge for researchers tracing those variants.

A somewhat different example involves variants in the ATP-binding cassette transporter A1 (ABCA1), associated with near absence of plasma HDL-cholesterol (HDL-C). Using linkage and positional candidate gene strategies, mutations in the ABCA1 gene were demonstrated to cause Tangier disease—a rare autosomal recessive disease of low HDL-C, leading to the severe deposition of cholesteryl esters in the reticulo-endothelial system in organs like spleen (31,32). Tangier patients have truncating mutations in both alleles of the ABCA1 gene. Interestingly, heterozygote family members showed decreased HDL-C plasma levels without other symptoms of Tangier disease. This stimulated a hypothesis that variants of ABCA1 and other genes in the same pathway might be associated with HDL-C plasma levels also in the general population. There is still a lack of studies that would provide a comprehensive coverage of ABCA1 variations and their potential association with HDL-C levels in the general population. However, two extensive studies have addressed this question. The promoter and coding region variants of the ABCA1 gene were screened for in the lowest and highest 1% HDL-C level samples from the general population of over 9000 Danes and one common non-synonymous SNP was associated with extremes of the HDL-C distribution (33). In another study in two population cohorts from Canada and US, the coding regions of three genes, associated with Mendelian forms of low HDL-C (ABCA1, APOA1 and LCAT) were sequenced. Interestingly one out of six individuals with HDL-C levels below the fifth percentile had a rare functional mutation either in ABCA1 or APOA1(34). Although not conclusive, these two studies support the hypothesis that indeed variants of the genes, earlier associated with rare monogenic or Mendelian trait also contribute to the HDL-C level in the general population. Finally, for some established examples of distinct Mendelian mutations, common at the population level, the penetrance is so variable that the genotype–phenotype correlation is almost as complicated as in truly complex traits. First, the disease proved to be more heterogeneous than expected, the research of rare families has exposed a total of six genes and has significantly contributed to our understanding of iron homeostasis. Secondly, although one mutation explains most of the cases, a C282Y mutation in the HFE gene, many puzzling questions prevail concerning the impact of the mutation, found in a homozygote form in 64–100% of cases (35). In the European population, the allele frequency of the C282Y variant ranges from 2.6% to 28%, increasing from south-east to north-west. Yet, the prevalence of hereditary hemochromatosis in most populations is <1%. Three prospective studies have followed C282Y homozygotes over time (36–38). Interestingly, these studies did not consistently identify increasing transferrin saturation and serum ferritin levels over time and did not demonstrate a clear link to the overt clinical manifestation of hereditary hemochromatosis. Thus, although in clinical manifest hemochromatosis patients the association to the HFE gene is evident, on the individual level, in asymptomatic subjects, a heterozygote C282Y mutation has a relatively low predictive value of a clinically manifest disease. What the final triggering factors are is still unclear, but environmental components, such as excessive alcohol consumption surely explain a portion of the phenotypical heterogeneity (39). The role of other hemochromatosis genes (e.g. HAMP and HJV) affecting the penetrance of the HFE homozygotes has been demonstrated in some cases, providing an interesting scenario of a potential interplay of multiple genes affecting iron metabolisms and finally the penetrance of the C282Y mutation. Such lessons emerging from studies of many monogenic diseases are highly useful for our thinking of the factors affecting clinical phenotype of common diseases.

RARE GENOMIC REARRANGEMENTS GUIDING TO PATHWAYS BEHIND COMPLEX TRAITS

A number of genes causing Mendelian disorders have been identified based on the gross chromosomal alterations such as translocations that have guided to the mutated gene. Their usefulness in mapping complex traits has not been fully utilized in a systematic genome-wide manner, but some successes exist. Chromosomal translocations have guided to genes associated with speech and language disorders (40,41). The emerging understanding of the complexity of the genome landscape, its submicroscopic changes and polymorphisms pave the way for further insight into the association between genomic rearrangements and complex phenotypes. Thus, not only classic chromosomal translocations but also other smaller genome rearrangements, such as copy number changes, are reasonable candidates for disease susceptibility variants. The association of a copy number change in the CCL3L1 gene and susceptibility to HIV might represent the tip of the iceberg, where polymorphic copy number changes have a link to a phenotype (42). Well established examples of Mendelian disorders caused by genomic
rearrangements include Charcot–Marie–Tooth (CMT), Di George Syndrome, Prader–Willi and Angelman Syndromes. The CMT locus is especially interesting, rich in low copy number repeats, where different rearrangements result in different phenotypes (43). The hypothesis has been put forward that perhaps, low copy number repeats are analogous to the changes introduced by a replication error at a nucleotide base: both are endogenous molecular mechanisms that introduce variation into our genome (43). Thus, it is logical that both types of variations represent potential susceptibility variations to complex phenotypes.

An interesting example of a rare translocation that has guided to a novel, previously unknown pathway, potentially affected in psychiatric disorders emerges from recent data of the involvement of the DISC1 gene in the genetic background of schizophrenia and other psychoses. Initially, the genetic locus on 1q42 was identified as a potential site for mental illness susceptibility gene(s) based on a translocation (1;11) (q42.1; q14.3) in a large Scottish family with multiple family members affected with major mental illness. The observation of the linkage in the Finnish schizophrenia pedigrees on 1q42 targeted the research effort to the 1q42 locus (44,45) and linkage for 1q42 to schizophrenia but also to other forms of psychosis was subsequently reported from multiple populations making this locus one of the most encouraging among the multiple loci reported for psychosis. When the gene interrupted by the translocation at the 1q42 locus was identified, it was found that in fact two genes on opposite strands were directly disrupted by the translocation, and got named as disrupted in schizophrenia 1 and 2 (DISC1 and DISC2) (46). Dissection of the biological role of these genes has greatly strengthened their positions as candidates for a role in the etiology of psychiatric disorders. Protein-protein interaction studies show that DISC1 acts as a ‘scaffold’ for proteins known to be involved in numerous neuronal functions, impacting on neurite outgrowth, neuronal migration, cytoskeletal modulation and signal transduction (47). The DISC2 gene is not known to encode a protein but rather through its RNA acts as a regulator of DISC1. Such examples demonstrate what a wonderful resource for studies of complex traits exists in rare genomic rearrangements that can currently be identified with dramatically increased resolution using array-based technologies (48).

CONCLUSION

Have studies on Mendelian, nearly monogenic diseases educated us about common, more complex traits? During past decades, the scientific community has actually quite successfully used special families and other study samples with restricted number of ancestral chromosomes (like those from isolated populations) for genetic studies of inherited disease phenotypes and characterized the underlying molecular

Figure 2. Overall strategy of the use of exceptional families with Mendelian information of disease trait in studies addressing genetic predisposition of common diseases.
defect. Accurate cataloging of clinical details and following genetic investigations of over 7000 well defined Mendelian disorders have facilitated the characterization of the molecular defect for over 1700 among them (http://www.ncbi.nlm.nih.gov/Omim). Our success to define the molecular background of Mendelian diseases has in many ways formed the basis for most promising initial findings in complex disease entities (13,49). Most so far proposed complex disease genes have been identified based on the ‘Mendelian’ approach, simply because it has been a most powerful strategy. The successes in the identification of genes for both monogenic diseases and rare familial forms of common diseases have opened new avenues for common disease studies via exposing new pathways or molecules proteins previously unsuspected to be associated with a given trait (Fig. 2). Although this ‘Mendelian’ strategy will probably not identify all the DNA variants affecting complex disease phenotypes, it still might be the only surviving or at least the most efficient strategy to expose the critical pathways defective in human disease processes. New hypothesis can be created and novel pathway components monitored in complex diseases on the basis of the findings in Mendelian diseases. We are far from exhausting this wonderful resource of exceptional families and subpopulations on the global scale. Hopefully the examples provided in this review would further stimulate investigators to intensify the efforts to efficiently use the worldwide resources of rare families, they are most relevant for improved characterization of molecular processes resulting in common diseases.

ACKNOWLEDGEMENTS

The study was in part supported by grants from the Academy of Finland (213506), the GenomEUtwin (QLG2-CT-2002-01254) and Eurohead (LSHM-CT-2004-504837) projects, the Sigrid Juselius Foundation, The Nordic Center of Excellence for Disease Genetics and the National Institutes of Health, USA [RO1NS37675 (A.P.), RO1NS43559 (L.P.), RO1HL70150-01A1 (L.P.)].

Conflict of Interest statement. None declared.

REFERENCES


