



Rare Disease Genes—Lessons and Challenges

Leena Peltonen and Annukka Uusitalo

Genome Res. 1997 7: 765-767

Access the most recent version at doi:[10.1101/gr.7.8.765](https://doi.org/10.1101/gr.7.8.765)

References This article cites 17 articles, 4 of which can be accessed free at:
<http://genome.cshlp.org/content/7/8/765.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads 'CRISPR and RNAi Genetic Screening. Your new superpower.' In the center, there is a white-bordered box containing the words 'LEARN MORE'. On the right, there is a photograph of a woman wearing a red superhero mask and a red cape over a white top. To the right of the photo is the Cellecta logo, which consists of a cluster of green dots of varying sizes, with the word 'CELLECTA' in white capital letters below it.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Rare Disease Genes—Lessons and Challenges

Leena Peltonen¹ and Annukka Uusitalo

Department of Human Molecular Genetics, National Public Health Institute and Institute of Biomedicine, University of Helsinki, FIN-00300 Helsinki, Finland

The fiercest competition in the field of human genetics takes place in the area of genetic diseases that are common at the population level—this is primarily attributable to the potential commercial utilization of emerging data. However, dissection of the molecular background of many extremely rare diseases has proven to be highly useful for the detailed characterization of cellular dysfunction and the identification of novel metabolic pathways, broadening our understanding of biological processes in general. This information has largely been obtained from research carried out in populations with exceptional enrichment of a given disease (Fig. 1).

Although the global prevalence of rare diseases is insignificant [e.g., worldwide there are <200 known cases of patients with infantile-type neuronal ceroid lipofuscinosis (INCL) in comparison to a global population prevalence of 1:65 for the recessive cystic fibrosis mutation or 1:8000 for dominant Marfan syndrome], this rarity does not necessarily reflect the impact of these diseases on biological research.

From Population Sample to Gene Identification: Special Statistical Strategies

Rare diseases are characteristically enriched in populations that have been isolated for religious, linguistic or geographical reasons; good examples of this are the Ashkenazi Jews or the Finns. From the viewpoint of population genetics, rare diseases have initially led to the identification of many population bottlenecks and have produced quite precise data on the time period at which the mutation was actually introduced into the population, thus providing initial clues to the genetic history of human populations. Often a common ancestor or inbreeding can be verified in

the history of affected individuals, which suggests that there is one major ancestor mutation. This situation allows the adoption of unique strategies to search for a causative locus: similarity searches for the disease locus using DNA samples from just a few affected individuals and searches for shared alleles or homozygosity of genotypes. Using this approach, an initial genome scan with 400 markers can be carried out in just a few days as compared to the weeks or months required for genotyping all family members in traditional genome scans using linkage analysis in diseases that have a more diverse mutational back-

ground. The strategy of using only a few DNA samples has been highly successful in the locus identification of many rare diseases (Nikali et al. 1995; van Soest et al. 1996, this issue).

The phenomenon of shared alleles in affected individuals or significant deviation in the allelic frequencies of polymorphic markers in disease chromosomes compared to control chromosomes is called linkage disequilibrium. This is sometimes observed with markers over amazingly wide intervals, as exemplified in Table 1, which summarizes the data on Finnish disease alleles. The existence of linkage disequilibrium over

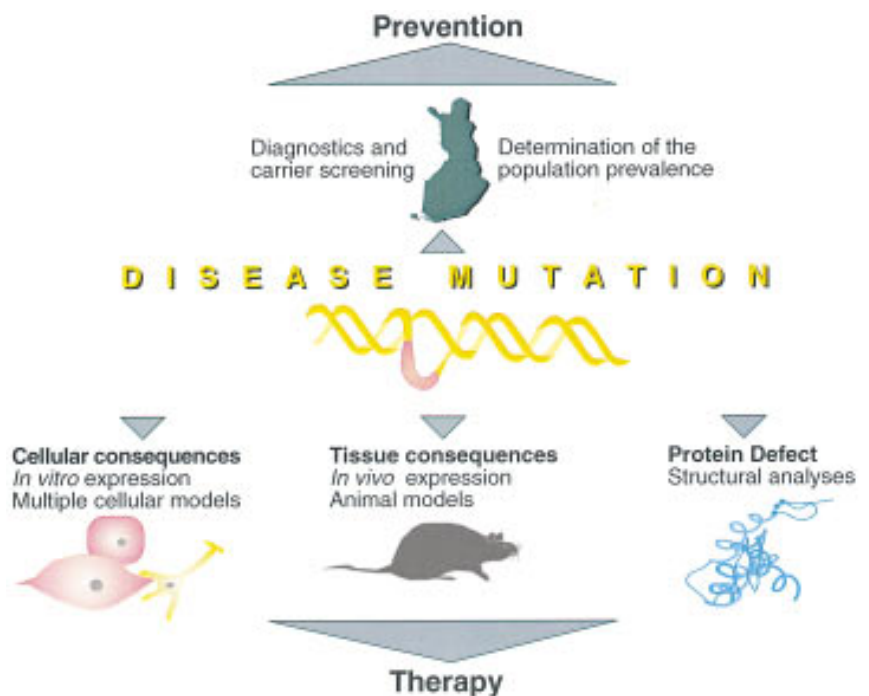


Figure 1 The identification of a disease mutation is considered beneficial, as it provides tools for assessing the pathogenic mechanisms of the disease and ultimately designing prevention and therapy. Research into the molecular background of many rare disorders has also provided the scientific community with various new strategies for the locus identification of a disease gene, as well as given new insight into metabolic pathways and biological processes.

Insight/Outlook

Table 1. Interval Showing the Linkage Disequilibrium in the Disease Alleles of Finnish Disease Heritage

Disease	Gene location	Interval (cM)
Autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED)	21q22.3	3
Cartilage-hair hypoplasia (CHH)	9p21–p13	3
Choroideremia	Xq21	9
Congenital chloride diarrhoea (CCD)	7q31	13
Congenital nephrosis (CNF)	19q12–q13.1	3
Cornea plana congenita	12q21	2.5
Diastrophic dysplasia (DTD)	5q31–q34	2
Familial amyloidosis, Finnish type (FAF)	9q33	3.5
Infantile neuronal ceroid-lipofuscinosis (INCL)	1p32	2.5
Infantile onset spinocerebellar ataxia (IOSCA)	10q23.3–q24.1	5
Northern epilepsy (Kainuu epilepsy)	8ptel.	10
Progressive myoclonus epilepsy (PME)	21q22	5
Retinoschisis	Xp22.2–p22.1	10
Salladisease	6q14–q15	10
Usher syndrome, type III (USH3)	3q21–q25	7.5
Variant form of late infantile NCL (vLINCL)	13q21.1–q32	11

For original references, see Peltonen et al. (1995).

such broad intervals suggests that genome-wide association-based analyses—perhaps not well-justified in mixed, heterogeneous populations—could also be successful for locus searches in complex polygenic diseases in genetic isolates (Lander and Schork 1994).

The linkage disequilibrium-based strategy has also been utilized successfully to restrict the critical DNA region after initial assignment of the disease locus. This is perhaps best exemplified in the isolation of the gene defective in a severe cartilage disorder called diastrophic dysplasia. Utilizing the rules of bacterial genetics (Luria and Delbrück 1943) and a simplified population genetic model based on the assumption of a small number of ancestors in the Finnish population, the gene was originally predicted to be located 64 kb from the polymorphic marker showing the strongest linkage disequilibrium. When the gene, which encodes a sulfate transporter protein, was isolated, it was 70 kb from this marker (Hästbacka et al. 1994). Again, applications of this method for complex diseases and their predisposing genes are obvious. Characteristically the DNA region identified to be associated with a common trait is very wide because of statistical problems combined

with the incomplete penetrance of hypothetical disease gene(s). Monitoring of linkage disequilibrium in population isolates ought to be a highly powerful tool for the more precise localization of predisposing genes. However, the situation in complex diseases is likely to be more problematic than in the case of rare disease alleles because of the expected high population prevalence and multiplicity of predisposing mutations—even in an isolate.

Shortcuts to Gene Functions and Clues to Essential Biological Pathways

Well-characterized tissue symptoms in patients with rare diseases provide immediate understanding of the function and tissue expression of the identified disease gene. On many occasions, totally new metabolic pathways have either been identified or associated for the first time with a specific cell type or developmental stage. Again, some Finnish diseases exemplify this: The identification of the mutated gene for diastrophic dysplasia provided evidence of the exceptional sensitivity of cartilage cells to the relative lack of sulfate ions as compared to other cell types (Hästbacka et

al. 1994). Similarly, the finding that the causative gene for the lethal infantile brain disorder INCL encoded a palmitoyl protein thioesterase demonstrated for the first time that proper removal of palmitoyl residues from lipid-modified proteins is an absolute requirement for the normal development and maturation of neocortical neurons (Vesa et al. 1995). INCL is characterized by the rapid death of cortical neurons, whereas neurons in lower parts of the CNS remain intact. Although details of both diastrophic dysplasia and INCL remain to be characterized further, these findings have opened new avenues for research of the metabolism of both sulfate and palmitoyl residues in specific cell types and tissues.

A lysosomal enzyme deficiency, aspartylglucosaminuria (AGU), resulting in progressive mental retardation, serves as a good example of a very rare disease that might initially be regarded as rather uninteresting. However, years of intensive research into the pathogenesis of AGU exposed molecular details of more general relevance. At first, the need for elucidation of the disease mechanism stimulated research into the defective enzyme, aspartylglucosaminidase (AGA), which had so far been hampered by failure to purify the protein to sufficient homogeneity because of its low quantity in tissues. Following identification of the gene and the major disease-causing mutation, *in vitro* expression studies of normal and mutated AGA polypeptides clarified the details of its biosynthesis and intracellular processing and provided clues to some general features of the cell biology of lysosomal enzymes (Ikonen et al. 1991a,b, 1993; Mononen et al. 1993; Riikonen et al. 1994, 1996). Subsequent crystallization data on the AGA enzyme revealed a novel catalytic mechanism based on the amino-terminal nucleophile and indicated that AGA is the first eukaryotic member of a new enzyme family of amidohydrolases (Oinonen et al. 1995; Tikkanen et al. 1996a,b). Structural analyses also provided the means for characterizing the lysosomal targeting process of AGA in detail and increased our comprehension of the targeting of lysosomal enzymes in general (R. Tikkanen, M. Peltonen, C. Oinonen, J. Rouvinen, and L. Peltonen, in prep.).

From a purely biological standpoint, the study of rare diseases has potential

similar to research with knockout mice, which have been used traditionally for characterizing the tissue consequences of a gene defect. The phenotype of these rare diseases is extremely well described, and despite the complexity and individual variability of the genetic background, skillful clinicians have typically confirmed the unifying clinical features between patients. Consequently, once the mutation has been identified, the function of a defective gene can be dissected immediately at a tissue- or cell lineage-specific level. Moreover, well-monitored follow-ups of these patients over many years produce detailed information on the disease process—and all information is directly applicable to humans! When the monitoring of individual variations in multiple genes becomes more feasible, for example, by array technology, even higher accuracy will be possible in the perception of the tissue-specific consequences of single genes mutated in human diseases.

The examples presented here for the enormous potential that comes from the molecular characterization of rare diseases were chosen from one population. Similar information has emerged from numerous other rare disorders, the classical example being the low-density lipoprotein (LDL)-receptor mutation in familial hypercholesterolemia that revolutionized our concepts of lipid metabolism (Brown and Goldstein 1986). Within the next several years all human genes will be cloned. This development will provide scientists with more precise tools to tackle the basic molecular defects in all of the >4000 rare human diseases (McKusick 1994), the causes of which still remain unknown. The vast amount of information emerging from this endeavor will have immense significance for the basic concepts of the multiple metabolic routes in human cells and tissues.

REFERENCES

- Brown, M.S. and J.L. Goldstein. 1986. *Science* 232: 34–47.
- Hästbacka, J., A. de la Chapelle, M.M. Mahtani, G. Clines, M.P. Reeve-Daly, M. Daly, B.A. Hamilton, K. Kusumi, B. Trivedi, A. Weaver et al. 1994. *Cell* 78: 1073–1087.
- Ikonen, E., M. Baumann, K. Grön, A.-C. Syvänen, N. Enomaa, R. Halila, P. Aula, and L. Peltonen. 1991a. *EMBO J.* 10: 51–58.
- Ikonen, E., N. Enomaa, I. Ulmanen, and L. Peltonen. 1991b. *Genomics* 11: 206–211.
- Ikonen, E., I. Julkunen, O.-K. Töllersrud, N. Kalkkinen, and L. Peltonen. 1993. *EMBO J.* 12: 295–302.
- Lander, E.S. and N.J. Schork. 1994. *Science* 265: 2037–2048.
- Luria, S.E. and M. Delbrück. 1943. *Genetics* 28: 491–511.
- McKusick, V.A. 1994. *Mendelian inheritance in man. Catalogs of human genes and genetic disorders*, 11th ed. Johns Hopkins University Press, Baltimore, MD.
- Mononen, I., K.J. Fisher, V. Kaartinen, and N.N. Aronson, Jr. 1993. *FASEB J.* 7: 1247–1256.
- Nikali, K., A. Suomalainen, J. Terwilliger, T. Koskinen, J. Weissenbach, and L. Peltonen. 1995. *Am. J. Hum. Genet.* 56: 1088–1095.
- Oinonen, C., R. Tikkanen, J. Rouvinen, and L. Peltonen. 1995. *Nature Struct. Biol.* 2: 1102–1108.
- Peltonen, L., P. Pekkarinen, and J. Aaltonen. 1995. *Biol. Chem.* 376: 697–704.
- Riikonen, A., E. Ikonen, R. Sormunen, V.-P. Lehto, L. Peltonen, and A. Jalanko. 1994. *DNA Cell Biol.* 13: 257–264.
- Riikonen, A., J. Rouvinen, R. Tikkanen, I. Julkunen, L. Peltonen, and A. Jalanko. 1996. *J. Biol. Chem.* 271: 21340–21344.
- van Soest, S., S. te Nijenhuis, L.I. van den Born, E.M. Bleeker-Wagemakers, E. Sharp, L.A. Sandkuijl, A. Westerweld, and A.A.B. Bergen. 1996. *Cytogenet. Cell Genet.* 73: 81–85.
- van Soest, S., J. Swart, N. Tijmes, L.A. Sandkuijl, J. Rommers, and A.A.B. Bergen. 1997. *Genome Res.* (this issue).
- Tikkanen, R., A. Riikonen, C. Oinonen, J. Rouvinen, and L. Peltonen. 1996a. *EMBO J.* 15: 2954–2960.
- Tikkanen, R., J. Rouvinen, A. Törrönen, N. Kalkkinen, and L. Peltonen. 1996b. *Proteins* 24: 253–258.
- Vesa, J., E. Hellsten, L.A. Verkruyse, L.A. Camp, J. Rapola, P. Santavuori, S.L. Hoffmann, and L. Peltonen. 1995. *Nature* 376: 584–587.