deCODE genetics, Inc.

Hakon Hakonarson[†1], *Jeffrey R Gulcher* & *Kari Stefansson*[2]

deCODE genetics, Inc., Sturlugata 8, IS-101 Reykjavik, Iceland
Tel: +354 570 1900; Fax: +354 570 1903;
[1]*E-mail: hakonh@decode.is*
[2]*E-mail: kstefans@decode.is*

deCODE has compiled the world's most comprehensive collection of population data on genealogy, genotypes and phenotypes. This combination of resources provides an effective system for identifying key genes linked to common diseases and to the regulation of drug response. deCODE has successfully mapped genes in over 25 common complex diseases and isolated genes in 8 of these. The methodology used to map these genes is based upon the company's genealogical database of the Icelandic population, which enables deCODE scientists to efficiently conduct population- and genome-wide linkage studies to identify key genetic factors involved in phenotypes ranging from diseases to drug response. In its growing pharmacogenomics program, deCODE has combined this linkage approach with high-throughput expression profiling to develop accurate tests that can predict individual responsiveness to virtually any drug of interest. deCODE is applying its unrivaled discovery capabilities to bring to market new drugs, DNA-based diagnostic products and pharmacogenomic tests. The company believes that such tests will play a crucial role in delivering personalized medicine – contributing to the development of more effective means of diagnosing and treating disease by matching each patient with the most suitable drug.

Introduction

Founded in 1996 and headquartered in Reykjavik, Iceland, deCODE genetics has established itself as the world's leading gene discovery organization. Since mapping its first gene in 1997, the company has located key genetic factors involved in more than 25 common diseases and has isolated genes in 8 of these. deCODE is using its discoveries to develop new drugs, disease diagnostics and pharmacogenomic tests that can accurately predict individual responsiveness to specific drugs (**Figure 1**).

deCODE's population genetics approach provides a singularly powerful means for identifying drug and diagnostic targets that are firmly rooted in the biology of disease. Unlike genomics strategies based solely on studies of the function or expression of selected genes, deCODE scientists are able to use the company's unrivaled genealogical resources and genotyping know-how to efficiently scan the entire genome and home in on the key genetic variations that contribute to virtually any phenotype. This approach has enabled the company to begin to develop more effective therapeutic and diagnostic regimes from the bottom up, taking aim directly at the pathology of disease rather than on the signs and symptoms.

In early 2002, deCODE acquired its US-based pharmaceuticals and biostructures groups. Along with its Encode pharmacogenomics and clinical trials contract research organization (CRO), the company now has in place integrated capabilities ranging from gene discovery to drug development to clinical trials, as well as a range of corporate partnerships which are leveraging these assets to create valuable products and services. Last year, deCODE and Roche entered a new phase in their 4 year partnership and deCODE is now conducting drug discovery work in 4 of the 12 diseases covered by their 1998 agreement. In September 2002, deCODE signed a major alliance with Merck to speed the development of new drugs against obesity.

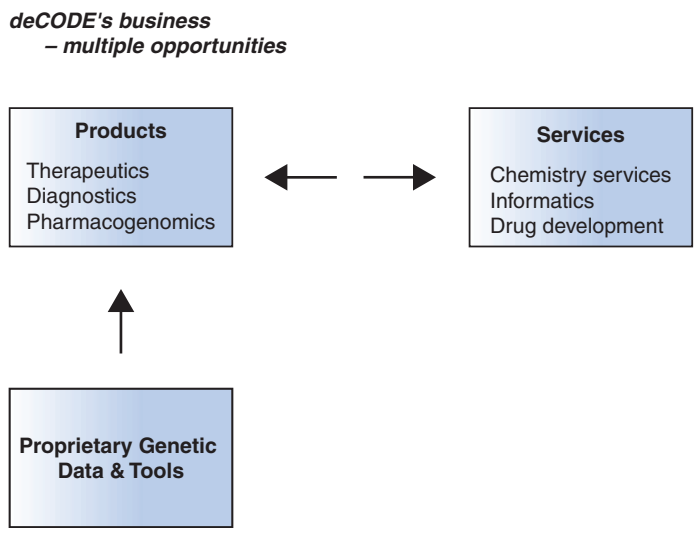
In diagnostics, deCODE has a broad alliance with Roche Diagnostics focused on utilizing many of the same targets deCODE identifies in its drug discovery programs for the development of molecular tests for disease predisposition and clinical applications. With Affymetrix, deCODE has achieved very exciting results in the development of expression-based pharmacogenomics assays for several leading prescription drugs in major therapeutic areas. In January 2002, deCODE formed an alliance with Pharmacia to identify genetic markers that could indicate likely progression from early to more advanced heart disease and thus to identify patients likely to benefit from Pharmacia's developmental cardi-

Keywords: complex disease, diagnostic test, disease-specific drugs, genealogical approach, human genetics, pharmacogenomics, target validation

Ashley Publications Ltd
www.ashley-pub.com

 ISSN 1462-2416

Figure 1. deCODE genetics business model.



ovascular drugs. In November 2002, deCODE signed a pharmacogenomic collaboration with Wyeth, under which deCODE is using its expertise in *in vitro* pharmacogenomics to generate gene expression data related to one of Wyeth's candidate drugs for respiratory disease. In December 2002, deCODE signed a pharmacogenomic collaboration with Vertex in relation to one of their developmental drugs indicated for the treatment of skin disorders.

deCODE's population approach to target discovery

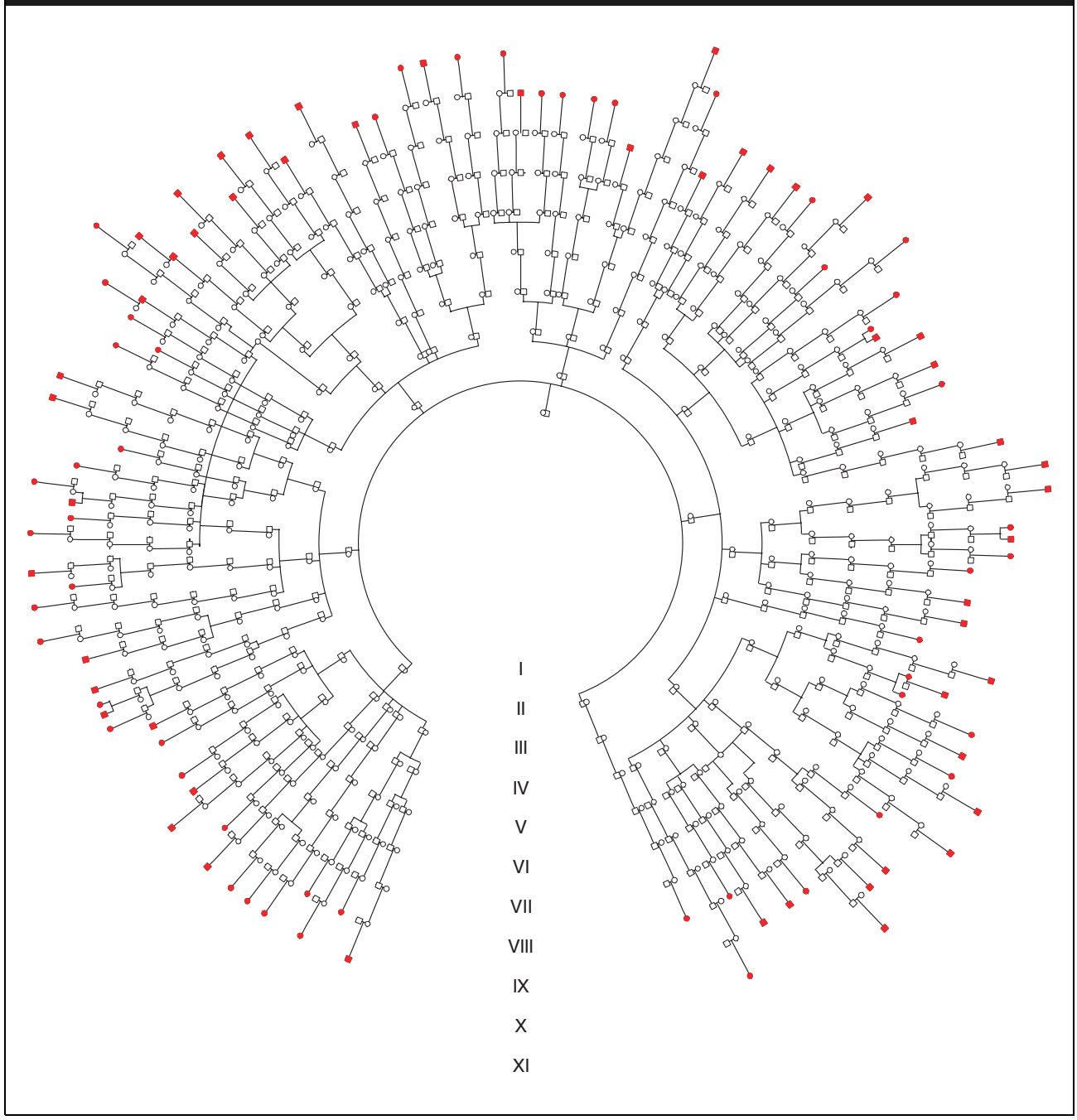
From the very beginning, deCODE has approached gene discovery as an information challenge. In the common diseases, unlike the simple Mendelian disorders, there is a complex relationship between genetic variation, environmental factors and the manifestation of the disease. Teasing out these correlations is a task ideally suited to the modern computer. However, it is critical to have a system of information large and complete enough, as well as the right data-mining tools and software to be able to pick out the signals from the noise. With a foothold in the genetics one has by definition gained entry to the rate-limiting biological pathway of disease - and can thereby analyze gene function, protein interactions, and validated drug targets that can be used to design compounds to disrupt the disease process.

In the Mendelian disorders, the use of small families has proven sufficient for conducting effective linkage studies that can identify disease-causing mutations. In the common diseases, the exponentially greater complexity of the task requires a much larger system of information – ideally, an entire population. Moreover, effective linkage analysis on a population-wide basis requires not only the generation of broad, high-quality phenotypic and genotypic data but also accurate knowledge of the structure of the population. Genealogies provide the avenues by which genetic variation can be traced from generation to generation (**Figure 2**).

In Iceland, deCODE has assembled an unparalleled set of population resources for unraveling the genetics of common diseases and has designed informatics products for the job. Using the population linkage approach, deCODE has mapped genes in more than 25 common complex diseases [101] and isolated genes in 8 of these. The methodology used to map these genes is based on the uniquely complete genealogical information on the Icelandic population gathered together in deCODE's genealogical database. This resource includes living Icelanders and over two-thirds of all Icelanders who have lived to adulthood since the country was settled in 874. As of late 2002, more than 80,000 Icelanders (approximately one-third of the total adult population) had contributed blood for DNA analysis and undergone a detailed medical examination in one or more of the company's 50 disease gene research programs. The company has also built up the world's highest throughput genotyping facility and developed the highest density genetic map yet developed of the human genome [1]. deCODE's research is conducted under rigorous standards of consent, ethical oversight and data protection [2]. The company's research protocols are reviewed by the Icelandic government's National Bioethics Committee and all personal identifiers are encrypted PIN codes created by the country's Data Protection Authority. All participants in deCODE's genetics research are required to provide their signed, informed consent.

In deCODE's research, an encrypted list of all the patients in Iceland with a particular disease trait or other phenotype can be run against a similarly encrypted version of the genealogy database [3]. This system then draws up, in real time, large pedigrees delineating the relationships of all patients, and deCODE statisticians examine the pedigrees to pick out those whose DNA would be most informative for genotypic analysis. A list of encrypted IDs is then sent back

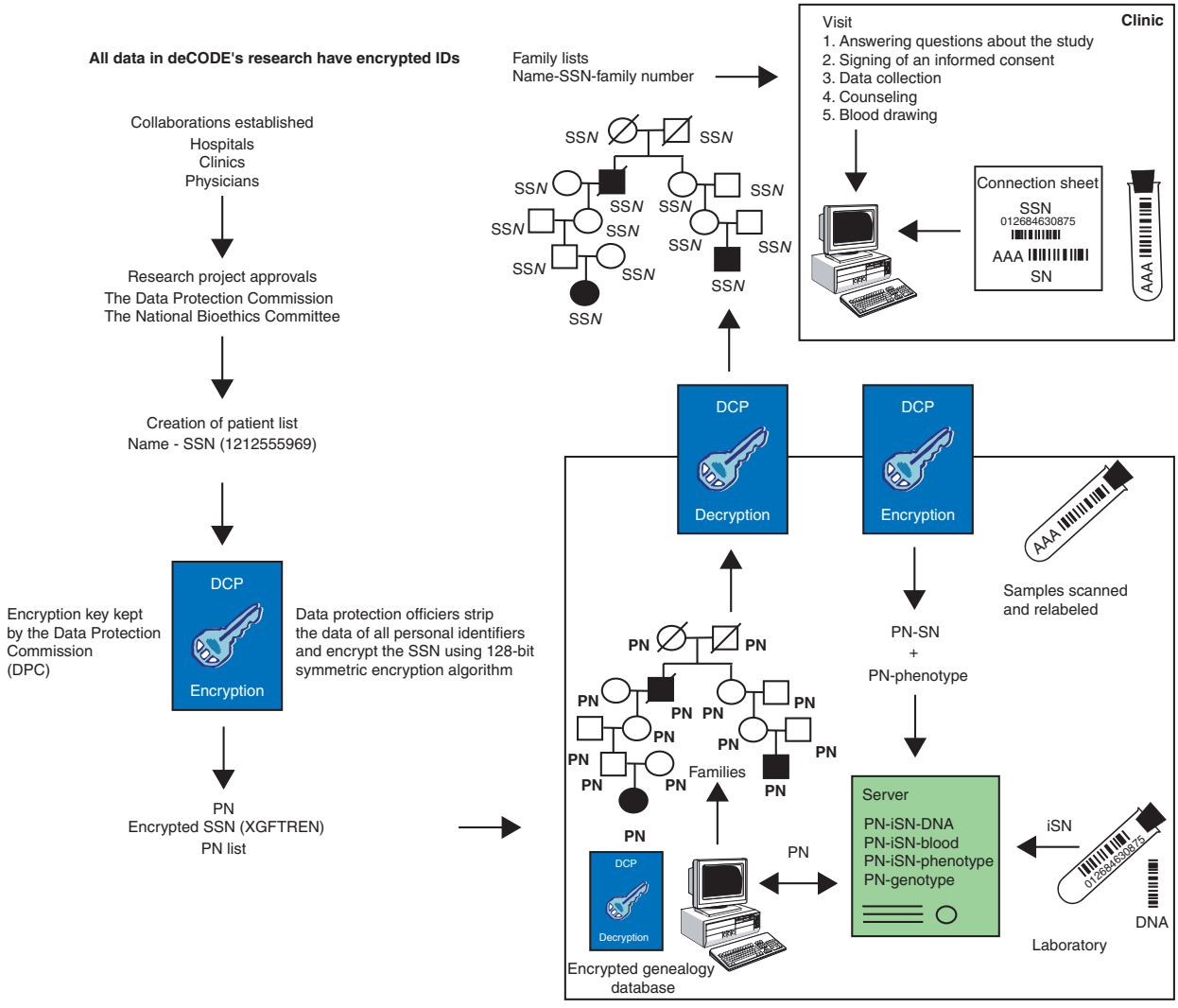
Figure 2. Eleven generation pedigree of asthma patients, showing 107 living patients descended from a founder couple in the late 17th century.



to the Data Protection Authority, which decrypts it and sends the names to collaborating doctors in the national health system. The doctors contact the patients and explain the nature of deCODE's research in a particular project. Those who wish to participate then sign an informed consent form, undergo a detailed medical examination and contribute a blood sample. To date, over 90% of patients asked to participate do so; the reconsent rate for those asked to participate in additional disease projects is over 99% (**Figure 3**).

By utilizing the genealogies, deCODE scientists are able to conduct population- and genome-wide genotypic scans in a virtually hypothesis free manner [4]. DNA samples from all participating patients are genotyped with the same framework set of over 1000 microsatellite

Figure 3. A schematic overview of the encryption systems used to anonymize all biological samples and medical data at deCODE.

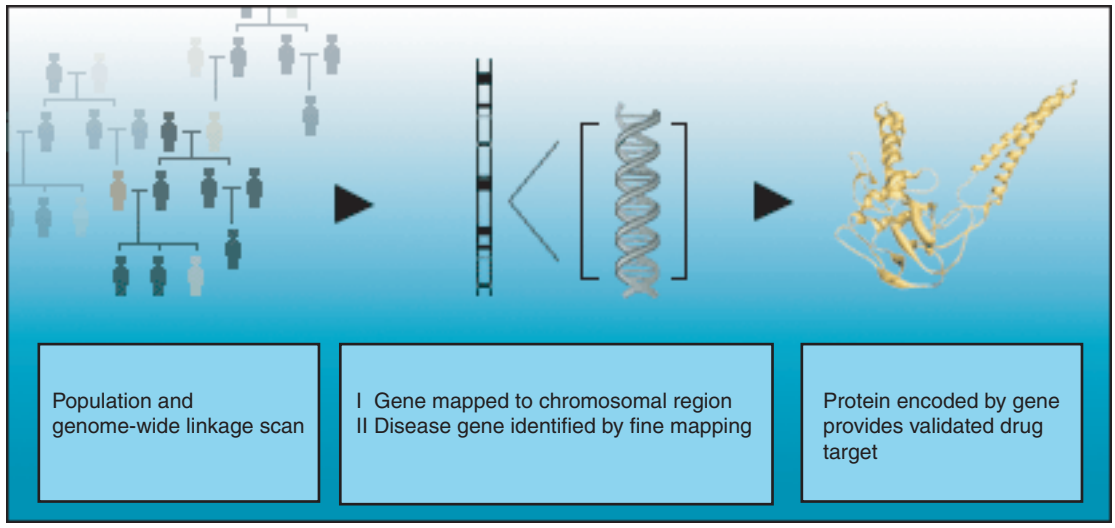


iSN: Internal sample number; PN: Patient number; SN: Sample number; SSN: Social security number.

markers – which can be thought of as fingerprints or signposts along the genome – and all genotypes are auto-edited by deCODE's Allele Calling (DAC) software. The pedigrees are automatically created and over 1000 patients can be run together in the same linkage run, providing great power for mapping disease genes within regions that are shared within and between different families. The Allegro linkage program (for more information e-mail allegro@decode.is) is fully automated and determines if the patients share specific DNA blocks beyond what would be expected by chance alone given their genealogical relationship.

Regions of the genome showing significant linkage with the phenotype are then analyzed with a denser set of microsatellite markers positioned within the deCODE genetic map [1]. The map contains over 6000 microsatellite markers correctly ordered and placed through the analysis of more than 1200 meiotic events. It also provides a framework for the correct positioning of some two million single nucleotide polymorphisms (SNPs) and is now the standard reference map being used by the Human Genome Project for the final assembly of the human genome. After an average marker density of one microsatellite per 20–30 kb is reached, to define all crossover events and linkage disequilibrium blocks in

Figure 4. deCODE's target discovery approach.



the regions of highest interest, the most relevant haplotypes are mapped by genotyping similar numbers of unrelated patients versus control subjects, usually 500–1000. Association studies are then performed to search for haplotypes that are more frequent in patients than in controls. Both the patients and control subjects are genotyped for SNPs in the region and any sequence gaps are filled in using BAC and then reassembled.

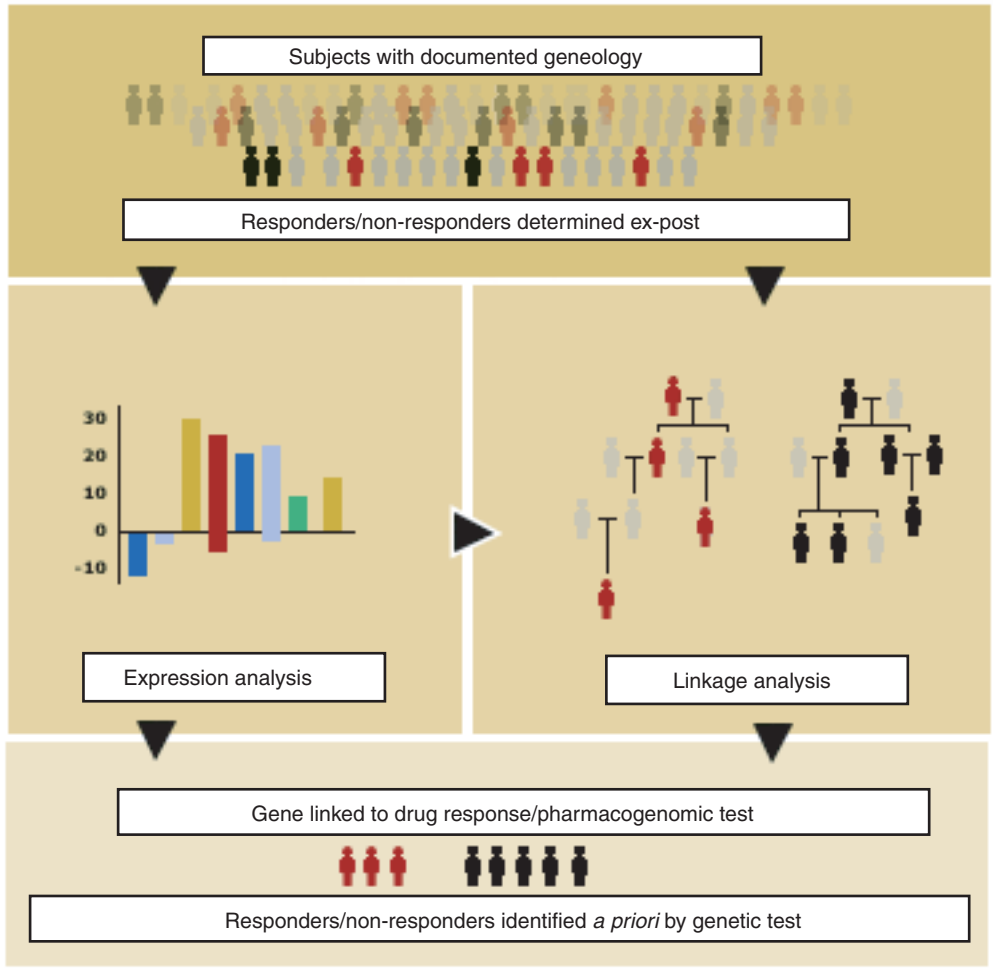
deCODE has narrowed 8 of the 25 genome-wide significant loci its scientists have identified to single genes. In each case, the most significant haplotype found in a particular pointed to a single gene although we find several linkage regions. The relevance of these genes outside of Iceland is then confirmed by genotyping several hundred patients and controls from heterogeneous populations. deCODE, as well as collaborating groups and independent researchers, has validated the company's findings in several diseases using cohorts from the UK, Scandinavia, the US and elsewhere. Once isolated, the genes identified are sequenced for mutations in both the patients and controls, and the biology is examined in animal models and pertinent *in vitro* cell culture assays (**Figure 4**).

One of the most advanced of deCODE's projects is that in schizophrenia. In this work, deCODE has identified a haplotype within the neuregulin 1 gene on chromosome 8p that confers more than twofold increased risk of the disease [5]. The locus had previously been identified by four other groups (in the US, Canada, Finland and Ireland). The seven-marker haplotype identified in Iceland is estimated to account for almost 30% of schizophrenia in Iceland. This exact haplotype also confers a very similar increased risk for schizophrenia in Scottish patients, accounting for an estimated 22% of cases there [6].

The signaling process regulated by neuregulin 1 results in modulation of NMDA receptor expression and activation at the synapse. Knockout mice for the neuregulin 1 gene or its ErbB4 receptor demonstrate hyperlocomotion and decreased prepulse inhibition, behavioral abnormalities which are typical of mice that have received phencyclidine (PCP), the basis of the mouse model for schizophrenia.

In stroke, deCODE has mapped a gene to chromosome 5q that encodes an enzyme, phosphodiesterase 4D (PDE4D) [7]. The gene has several 3–6 marker SNP haplotypes that are significantly more common in patients than controls and which collectively account for over 50% of stroke in Iceland. Some of these haplotypes as well as some unique haplotypes are also present in other populations, and account for a similar proportion of stroke in these populations. Once activated, PDE4D stimulates vascular smooth muscle cell proliferation and migration, a hallmark in the pathogenesis of the atherosclerotic plaque that is at the heart of the pathogenesis of stroke. In these examples and several others, the genealogical approach has proven to be a highly effective means of establishing linkage, enabling the isolation of key genes involved in the pathogenesis of common

Figure 5. A schematic overview of deCODE's pharmacogenomics approach.



diseases and identifying druggable targets that are now the basis for drug discovery work at deCODE.

deCODE's approach to disease susceptibility and drug response marker discovery

Molecular diagnostics promise to transform the practice of medicine by enabling physicians to assess disease susceptibility, permit early detection of disease, determine likely responses to medication and ultimately select the best courses of therapy. Since drug treatment constitutes the mainstay of modern medicine and since most genes contain multiple SNPs, identifying those that are most relevant with respect to disease or drug response is important. Genes and markers identified using deCODE's population genetics approach offer a solid basis for undertaking the development of diagnostics, complemented by the company's expression profiling work for the development of tests predictive of both disease susceptibility and response to drugs (**Figure 5**).

deCODE recently completed a pharmacogenomic study of glucocorticoid (GC) response in asthma. Several studies have been designed to demonstrate association of GC resistance with variations in genes that constitute the GC response pathway. While both structural and functional alterations in the GC receptor units or their response elements are important determinants of GC responsiveness, no clinically relevant prediction has emerged from these studies. In a recent study at deCODE, gene microarray technology was used to examine gene expression profiles of thousands of genes in peripheral blood mononuclear cells (PBMC) obtained from training and test sets of asthmatic patients who were either responders or non-responders to

GCs. The GC resistant patients were also clustered into families and the GC response trait mapped using the population linkage approach. The results were subsequently integrated to determine if differentially expressed genes would match to loci identified by linkage.

The results of this work demonstrated that GC-responders could be separated from non-responders in an independent patient cohort with > 85% accuracy by analyzing the expression profile of only a handful of genes (In Press). Moreover, SNPs in three of the genes that matched to loci were found to carry a relative risk of two or higher with respect to the GC-resistant trait. A comparable prediction was obtained for:

- serotonin-uptake inhibitors in patients with depression
- lipid-lowering drugs in hyperlipidemia
- anti-inflammatory drugs in rheumatoid arthritis
- drugs used to treat high blood pressure and migraine

At deCODE, the translation of this sort of genetic information into novel molecular diagnostic products is well under way. The company expects to be able to bring new diagnostic products to market in the near future, utilizing new technology platforms such as probe-based nucleic acid assays, microarrays, immunoassays and sequencing. In order to identify the most powerful markers, however, it is critical to have access to a population that is well understood genetically, so that meaningful interactions (epistases) between the range of major disease susceptibility genes or genes that regulate drug response can be effectively established. A high signal-to-noise ratio is also critical in order for the signal to come through. Due to its comprehensive genealogical information and relative homogeneity, both with respect to genetic factors and the environment, the Icelandic population provides probably the best available system for pinpointing key disease-linked genes and for examining drug response by expression profiling. The full range of genetic variation that correlates to a given phenotype can then be examined in more heterogeneous populations.

A powerful system for discovery and product development

deCODE has built an infrastructure that allows for efficient identification of key genes and markers linked to diseases and to the regulation of drug response. deCODE has compiled an unmatched record of success in mapping and cloning genes involved in common complex diseases. By applying its discoveries to the development of new drugs and DNA-based disease and pharmacogenomic diagnostic tests, the company is well positioned to play a significant role in the development of better and more personalized medicine. The new genetics promises to transform the practice of medicine by enabling physicians to assess the risks of disease, permit early detection of disease, determine likely responses to medication, choose the best courses of therapy and have at their disposal new therapies that target the disease process itself. deCODE is focused on turning its uniquely powerful discovery capabilities into valuable new products for the market.

Acknowledgement

We are greatly appreciative of Edward Farmer's review of the manuscript.

Bibliography

1. Kong A, Gudbjartsson DF, Sainz J *et al.*: A high-resolution recombination map of the human genome. *Nat. Genetics* 31(3), 241-247 (2002).
2. Gulcher JR, Kristjansson K, Gudbjartsson H, Stefansson K: Protection of privacy by third-party encryption in genetic research in Iceland. *Eur. J. Hum. Genet.* 8, 739-742 (2000).
3. Gulcher, JR Kong, A, Stefansson K: The role of linkage studies for common diseases. *Curr. Opin. Genet. Dev.* 11, 264-267 (2001).
4. Stefansson H, Geirsson RT, Steinthorsdottir V *et al.*: Genetic factors contribute to the risk of developing endometriosis. *Hum. Reprod.* 17(3), 555-559 (2002).
5. Stefansson H, Sigurdsson H, Steinthorsdottir V *et al.*: Neuregulin 1 and susceptibility to schizophrenia. *Am. J. Hum. Genet.* 71, 877-892 (2002).
6. Stefansson H, Sarginson J, Kong A *et al.*: Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. *Am. J. Hum. Genetics* 72, 83-87 (2003).
7. Gretarsdottir S, Sveinbjornsdottir S, Jonsson HH *et al.*: Localization of a susceptibility gene for common forms of stroke to 5q12. *Am. J. Hum. Genet.* 70(3), 593-603 (2002).

Website

101. http://www.decode.com deCODE genetics.