Perlegen Sciences, Inc.

For years we have recognized that using genetics to target the right medicines to the right patients at the right time could improve healthcare; however, technical and financial constraints have prevented researchers from conducting the comprehensive, whole-genome association studies required to make such advances. Now, Perlegen and its collaborators are making personalized medicine and high density whole-genome association studies a reality. Perlegen quickly and cost-effectively analyzes millions of unique genetic variations in thousands of patients to identify the genetic differences associated with efficacy and side effects. This enables the pharmaceutical drug response to be predicted, and the therapeutic window of a treatment to be expanded. Perlegen licenses compounds that may benefit from this approach, and collaborates with pharmaceutical companies to rescue and better position medicines in their portfolios. Furthermore, the company collaborates with researchers in the public sector to discover the genetic variations associated with common diseases, paving the way for new therapeutics and diagnostics.

Perlegen was founded in late 2000 with the mission of identifying patterns of genetic variation between individuals and using that information to improve patient care and uncover the genetic underpinnings of diseases. With initial funding of US$100 million and technology from Affymetrix, the DNA microchip leader of which Perlegen is an offshoot, the company resequenced over 20 individual human genomes and identified over 1.6 million genetic variations known as single nucleotide polymorphisms (SNPs). This effort was technically possible because of the publicly available first reference sequence created by the Human Genome Project. Perlegen was then able to read multiple whole genomes at a fraction of the cost and time of the original effort and compare them, in order to discover genetic variations that help make each individual human unique.

Perlegen brought unprecedented scale and density to the study of genomics. Allele frequency data for Perlegen's estimated 1.6 million SNPs show that the SNP set is extremely useful for association studies; a large fraction of these SNPs have sizable minor allele frequencies in geographically diverse populations [1]. Perlegen has also identified a subset of approximately 300,000 SNPs, which contain more than 80% of the genetic information of the larger set, suitable for genotyping in individual samples [1,2].

Applying this knowledge, Perlegen quickly developed collaborations with an extensive list of leading research organizations, including six of the world's ten largest pharmaceutical companies, eight separate institutes within the National Institutes of Health (NIH), numerous universities and foundations, and one of the world's largest consumer products companies. The company is now conducting comprehensive genetic analyses of several diseases, including: Alzheimer's disease, Parkinson's disease, autism, breast cancer, and metabolic syndrome, in addition to analyzing nicotine addiction. A complete list of Perlegen's publicly announced collaborations is listed on the Perlegen website [101].

In 2004, Perlegen performed over 2 billion genotyping assays, including an unprecedented large-scale genotyping study to estimate the allele frequencies of 1,586,383 SNPs in three diverse populations, generating more than 112 million individual genotypes [1]. Genotypes for over 150,000 of these SNPs had been previously determined by the International HapMap Project, allowing a direct comparison of Perlegen's results with those from HapMap centers. The concordance between Perlegen's genotype data and that obtained from the multiple different platforms used by the HapMap ap centers was 99.54%, demonstrating the excellent genotyping quality of the Perlegen platform. The results of this study are publicly available [102].

Based on the results of this and other successful genotyping studies, Perlegen was awarded a grant by the National Human Genome Research Institute (NHGRI) in September 2004, to support the International HapMap Project in genotyping SNPs in the public domain for the 270 samples
included in the study. By September 2005, Perlegen will have genotyped the vast majority of all common SNPs in the public domain.

To date, Perlegen has raised US$206 million in private funding, including a US$74 million deal closed in March of 2005, to help the company build its own pipeline of late-stage drugs focused on metabolic, cardiovascular and central nervous system diseases. Perlegen is now actively in-licensing late-stage drugs that can be improved through pharmacogenomics, so that it can apply its technology to bring them to market. In April 2005, the company announced a licensing agreement with Mitsubishi Pharma Company for the late-stage development and commercialization of MCC-555 (a peroxisome proliferator-activated receptor [PPAR] agonist) for the treatment of diabetes and other metabolic disorders.

Perlegen headquarters are in Mountain View, California, and the company also has offices in McLean, Virginia in the USA, as well as in Tokyo, Japan.

Why 'whole genome' is becoming the new standard in genomics
There are three primary reasons why Perlegen focuses on taking a high-density, whole-genome approach when studying the genetic associations with drug response and disease:

- Most human traits are likely to be polygenic. The majority of common diseases are not caused by a single variation or even a single ‘disease gene’ in the genome. Most common diseases are probably associated with more than 20 different genetic loci spread across the whole genome.

- Any single genetic variation may be predictive for only a small portion of a disease or drug response. For example, SNPs that are predictive of only a 2% increase in the likelihood of a disease are important, but not sufficient to predict disease susceptibility on their own. Therefore, it is important to be able to identify many of the SNPs across the genome that may contribute to the phenotype. Results from the research Perlegen has conducted with its pharmaceutical partners suggest that out of the 7-10 million common SNPs, a set of as few as 20-40 different SNPs can be all that is needed to make a prediction of whether a patient is more or less susceptible to an adverse event when taking a drug.

- Candidate region studies may not identify clinically relevant markers. Candidate region studies rely on, and are only as good as, our present knowledge of biology. Although they have been useful in discovering single genes, they may miss out on critical regions of the genome because of the limited knowledge of biological mechanisms, or because any single gene effect is likely to contribute only a small portion to a disease. In order to identify sufficient SNPs for use in a clinical setting as a diagnostic test, or to identify novel pathways for new therapeutics, it is critical to look for variations across the genome. Whole-genome association studies allow us to take a hypothesis-free approach in identifying many of the regions of the genome that are associated with the phenotype of interest, and worthy of further study.

Improving performance of clinical trials and marketed therapeutics
Many potential drugs never gain regulatory approval, despite demonstrating efficacy and safety in a large proportion of the participants of late-stage clinical trials. Insufficient efficacy, or adverse events in a small percentage of trial participants, can limit the development of a drug. The same is true for marketed drugs. There is now an increasing number of major drugs recalled entirely for safety reasons. If not recalled, such medicines are often saddled with overly restrictive labelling. In such instances, hundreds of millions of dollars in development expenditures are lost, and thousands of patients that could have benefited from the drug are deprived of access to it.

As illustrated in Figure 1, Perlegen uses pharmacogenomics to identify and screen out those patients most likely to experience adverse drug reactions (ADRs) or poor drug efficacy. By predicting patient drug response, Perlegen helps collaborators bring drugs from development to the market more effectively, improve efficacy relative to competing drugs in their class, prevent unnecessary recalls, and ensure an improved therapeutic benefit for patients.

Pharmacogenomics collaborations
Perlegen's collaborators collect DNA samples from patients who are either involved in clinical trials or taking a drug already on the market. For each patient they also collect various phenotypic measures indicating how well or poorly the patient has responded to the medication. Together, Perlegen and collaborators review the distribution of responses from patients and then design a study. Typical studies begin with...
approximately 400 cases (for example, patients who suffered from an ADR or low efficacy) and 400 controls (for example, patients who responded well to the drug).

Particularly relevant to this effort is Perlegen’s ability to genotype a broad range of SNPs, including those discovered at Perlegen, SNPs from public databases such as the Single Nucleotide Polymorphism database (dbSNP), or from other proprietary SNP databases. Perlegen has developed a unique sample preparation method used in its laboratory that allows it to assay almost any SNP. SNPs can be chosen to provide effective coverage of common genetic variations across the whole genome, and assays are customized to the needs of the study.

With the study design developed and SNPs selected, Perlegen designs custom chips for the study, and forwards the chip designs to Affymetrix for production. Perlegen then performs a whole-genome association study and subsequent replication study in its laboratory, evaluating from hundreds of thousands to more than four million SNPs, for association with adverse events and/or lack of efficacy. In addition to performing the genotyping using its proprietary processes and equipment, Perlegen brings its team of experienced analysts, and a proprietary analysis pipeline that has been established over the past 5 years, to analyze this massive amount of data and draw appropriate conclusions regarding the genetic factors that influence differential drug response. The resultant set of associated genetic variants, typically ranging from 20–40 SNPs from across the genome, provides the knowledge that enables the pharmaceutical company collaborating with Perlegen to predict which patients may experience an ADR or insufficient efficacy in response to the drug. Perlegen is able to estimate the likelihood of an ADR for a given patient using statistical models, which take into account which, and how many, of the predisposing alleles that patient has.
This whole process can be completed rapidly. The typical time taken from when the samples are delivered to Perlegen to when the results and full analysis are completed and delivered to the collaborator is 6 months. The accuracy, cost effectiveness and speed of Perlegen's whole-genome association studies have made them not just feasible – the company believes they are becoming the new standard.

Outlook for pharmacogenomics
The use of a genetic diagnostic based on the markers identified in whole-genome association studies, such as in the example described above, will be of tremendous value for both patients and pharmaceutical companies. The relatively low investment required and significant improvement in patient care that can be achieved allows Perlegen and its collaborators to get drugs to market that might otherwise fail, increase the market share of existing drugs by better positioning them against competing drugs in their class, and prevent unnecessary recalls.

It is now standard practice for leading pharmaceutical companies to collect DNA samples for pharmacogenomic research during clinical trials. Similarly, ongoing studies of ADRs or differential efficacy in marketed drugs will increasingly include DNA collection, in order to assess the genetic aspects associated with drug response.

Perlegen has unique advantages in this area of pharmacogenomics as it can execute very high-density, whole-genome association studies. In order for a diagnostic to be most predictive, it requires the identification of most of the genetic loci that contribute significantly to the drug response. If there are, for example, 30 loci that are predictive of the drug response, each only contributing a small proportion to the total, then it is important that most of them are identified to have an impact in the clinic. Identifying only one or two markers may be interesting, but not very commercially applicable as a diagnostic for a polygenic based response. Perlegen is helping to make the vision of personalized medicine a reality.

Perlegen, Perlegen Sciences, and the Company Logo are trademarks of Perlegen Sciences, Inc.

Bibliography