

Molecular genetic and cytogenetic diagnostics

Editor: H.-G. Klein

**Next generation sequencing in genetic diagnostics<sup>1)</sup>****Saskia Biskup\***Praxis für Humangenetik und CeGaT GmbH, Tübingen,  
Germany**Abstract**

The introduction of next-generation sequencing technologies in human genetic diagnostics is a challenge to many of its aspects. It is mainly positive, even revolutionary, which will be discussed as well as its critical aspects. It used to take several months or years to complete genetic testing. This already belongs to the past. As sequencing technologies are progressing at an extremely high speed, Germany recently introduced the new Gen Diagnostics Law in February 2010. Four common diseases with genetic contribution (dementia, Parkinson's syndrome, epilepsy, and hereditary eye diseases) will be used to exemplify the latest development of human genetic testing.

**Keywords:** diagnostic panels; next-generation sequencing; SOLiD 4 technology.

**Introduction**

Genetic or molecular genetic diagnostics refers to the study of individual genes that, when changed, most probably lead to the manifestation of a disease or are the cause of a disease. A genetic cause must be considered especially when the afflicted persons are of young age. The sum total of hereditary material in humans is stated to be approximately 3 billion DNA base pairs and codes for 20 to 30,000 genes. A change in the hereditary material may be transmitted from generation to generation or it may arise "de novo". A distinction is made between changes in germ cells, which are hereditary, and somatic changes, e.g., tumor cells. In this paper, the term "mutation" (change in the hereditary material) is replaced with

"variation", in order to do better justice to the different types of variations. There are pathogenic variants that are certain to cause disease, but there also are variants whose significance is not explicitly clear. These include the "probably pathogenic", the "probably non-pathogenic", the "probably benign", the "certainly benign" and finally an ever increasing group of variants of uncertain significance (VUS).

The first large catalog of genes that, when changed, are associated with a disease was compiled into a database named OMIM (Online Inheritance In Man) by the American human geneticist Viktor McKusick (1921–2008). As a consequence, he is seen as the founder of medical genetics. The data base grows larger every day and is administered by the Johns Hopkins University in Baltimore, USA. So far variations in 10% of all human genes could be associated with known phenotypes. That is a comparatively large number and yet does not explain the majority of genetic variability in man.

Genetic diagnostics is useful if it brings about consequences from a therapeutic or prophylactic perspective or eases the burden on the individual concerned and the family. The right not to know exists at any time and this should be emphasized during genetic counseling before as well as after testing. What follows are examples from the day-to-day practice that illustrate why human genetics rightly sees technological progress as a quantum leap for its own discipline, for the patient and for the health care system.

**From exomes to diagnostic panels in human genetics**

By means of the sequencing machines available today, high-throughput sequencing permits the simultaneous sequencing of 100 billion basepairs within one week, i.e., the haploid human genome with an average coverage of 30. Because of higher accuracy and higher coverage and for reasons of better understanding of variation it has advantages to focus on the coding regions of the human genome. This is called exome-sequencing; the exome is the totality of all coding regions in the genome. Sequencing an exome is useful if (i) there are strong indications for a genetic disease, (ii) all known genes associated with the disease at issue have been ruled out and (iii) this approach is applied and evaluated by experts in the particular field within the scope of a research project.

In contrast, a diagnostic panel is the targeted simultaneous screening of a list of known genes that have already been described as the cause of a specific disease. Up to now genetic

<sup>1)</sup>Original German online version at: <http://www.reference-global.com/toc/labm/34/6>.

The German article was translated by Compuscript Ltd. and authorized by the authors.

\*Correspondence: Dr. med. Dr. rer. nat. Saskia Biskup, Praxis für Humangenetik und CeGaT GmbH, Paul-Ehrlich-Str. 17, 72076 Tübingen, Germany  
Tel.: +49 7071-5654400  
Fax: +49 7071-5654422  
E-Mail: [saskia.biskup@humangenetik-tuebingen.de](mailto:saskia.biskup@humangenetik-tuebingen.de)

diagnostics in such cases was not only very time-consuming, but was frequently not performed because of high costs.

A diagnostic panel clearly differs from the purely scientific and explorative approach of exome-sequencing. A diagnostic panel is ordered by the physician. Only the genes whose connection with the disease has been established are investigated. Finally, a finding is made and forwarded to the physician. This finding contains detailed information about detected variants that are validated through Sanger sequencing and interpreted in relation to the disease at issue.

Using a diagnostic panel on a newest generation high-throughput sequencer makes sense only if several or very large genes can be considered for the disease. Some examples are hereditary tumor diseases with the known breast cancer-causing genes *BRCA1* and *BRCA2*, cardiomyopathies, cardiac dysrhythmias, familial hypercholesterolemias, mental retardation, epilepsies, hereditary ophthalmic diseases and neurodegenerative diseases. Collaboration with clinical experts is indispensable when compiling the gene list of a panel and also later when interpreting the data. Since gene lists can be from two to more than one hundred genes in length, advance clinical delimitation is important. After the patient has given his written consent his DNA is enriched and sequenced with all genes contained in the gene list. The clinical expression of the disease then determines the sequence of the genes to be evaluated.

Before a diagnostic panel can be offered commercially for diagnostic purposes it must first be validated. Within the framework of such validation patients whose variations are already known are “post”-sequenced on the new panel. This requires that 100% of the known variations are found by means of the panel. Furthermore, the enrichment of the genomic regions must be efficient, specific and reproducible. Only then can a diagnostic panel be used for a clinical objective.

But is it then actually useful? We shall closer examine this question with the help of three examples. Detlef Boehm (CeGaT GmbH), a pioneer in establishing new methods in human genetic diagnostics [2] and the Practice for Human Genetics in Tübingen together with clinical partners and with Applied Biosystems/Life Technologies have put special emphasis on hereditary ophthalmic diseases, epilepsies and neurodegenerative diseases when developing diagnostic panels. The aim here is to clarify the genetic cause in affected families and thereby to (i) secure a clinical diagnosis, (ii) be able to offer a targeted examination of other family members, (iii) make possible an early therapeutic intervention, (iv) provide a prognostic assessment of the course of the disease and (v) provide the basis for new therapeutic methods in the long-term.

### Diagnostic panel and hereditary ophthalmic diseases

Worldwide hereditary ophthalmic diseases affect several million people. The disease usually starts during adolescence and initially often appears as night blindness. Thereafter the loss

of sight as a rule progresses slowly and can take decades. The loss of sight may occur in isolation or in connection with other symptoms. There is no treatment in most cases. As of now a total of more than 180 genes have been described that, when defective, can cause hereditary ophthalmic disease [1]. The description of this list of genes, which has grown so markedly in just the last few years, has provided the first substantial clues for understanding the pathogenesis of the disease and thus has contributed considerably to understanding the disease itself. The gene or its product is not directly sufficient for developing any new medication, but it is the central starting point or point of action of a therapy.

At present approx. 50% of all familial ophthalmic diseases are clarified genetically. Diagnosing frequently takes several years and, due to the investigation of many large genes, is very expensive. Such considerations as well as the need to clarify the other 50% have resulted in the transfer of hereditary ophthalmic disease diagnostics to the high-throughput sequencing arena. The Retina-All-Panel was developed in collaboration with the ophthalmic geneticists Prof. Wolfgang Berger and Dr. John Neidhardt (University of Zürich, Institute for Medical Genetics) and Prof. Bernd Wissinger and Dr. Susanne Kohl (University of Tübingen, Molecular Genetic Laboratory of the Eye Clinic). It contains all presently known genes with a connection to hereditary ophthalmic disease. Sequencing this large panel of genes takes about one to two weeks. Detected variants are issued in a list, classified according to their importance (pathogenic, benign or VUS) and then verified via the Sanger method. Finally, the results are evaluated, summarized and interpreted. Altogether the diagnosis at this time requires no more than two to three months.

A simultaneous study of more than 180 genes does not make sense from a clinical point of view. The list of genes can be divided into smaller groups for the ophthalmic geneticist and clinician. This allows for the simultaneous study of 26 genes in a case of autosomal dominant pigmentary retinopathy, of 28 genes in the case of autosomal recessive pigmentary retinopathy, of 10 genes in the case of Usher’s syndrome, 11 genes in the case of congenital stationary blindness, 14, 9 or 5 genes, respectively in cases of Bardet-Biedl, Joubert or Refsum syndrome. A current and comprehensive article on this subject provides more information on the subdivision of ophthalmic diseases and a description of the associated genes [1].

Research in the field of hereditary ophthalmic diseases also gains substantially from the Retina-All-Panel. With the use of the Retina-All-Panel many more patients carrying a pathogenic variant in one of the candidate genes are being identified. Hence, the pool of individuals affected by a certain variant is growing worldwide and the clinical observation of the course of the disease in these patients allows better classification and prognosis assessment in additional patients with the same variant. In the future, it would also be useful from a therapeutic perspective to group patients based on their genetic background when testing new medications, in order to better interpret positive effects as well as side effects.

Next generation sequencing also has an additional scientific aspect. From a purely technological standpoint it makes no difference whether two, one hundred or several thousand patient genes are enriched and sequenced. With a clear diagnostic objective this would be senseless, since data are generated whose required evaluation and validation is difficult and very time-consuming. The situation is different with a family where no genetic cause can be found in the known genes. This presents a possibility that genes as yet not associated with the disease could be studied for the first time and named as the new cause for the disease. In the case of ophthalmic diseases candidate genes are genes that have an important function in the eye but that have not yet been studied in the patients. They are candidates for the cause of ophthalmic disease. This is where the enormous potential of next generation sequencing, that would further advance the knowledge about the causes of hereditary ophthalmic diseases, is to be found.

### Diagnostic panel and epilepsy

Epilepsies affect 1% to 3% of the population in the course of a lifetime. The various expressions of the disease are differentiated depending on age and the form of progression. A genetic cause is probable if a symptomatic cause from brain damage, from a tumor, an infection or a metabolic disturbance can be ruled out. Large families with frequently occurring epileptic disease have contributed to the identification of genes and crucially also to the clarification of the pathogenesis of the disease. Identified genes include above all voltage-dependent ion channels and the receptors of neurotransmitters. Conceptually a change in the neural transmission of nerve cell to nerve cell, caused by defective ion channels or neurotransmitter receptors, fits the cause of a convulsive disorder. Specific therapies, e.g., the targeted attack of a medication on a defective sodium channel (valproic acid, carbamazepine, oxcarbazepine and phenytoin) or on a defective receptor (GABA receptor, phenobarbital), provide efficient treatment of a patient. The main goal therefore is the molecular genetic discovery of the cause in as many cases as possible, in order to create a specific individual treatment of the disease. Numerous genes that have been described as the cause for the various forms of familial epilepsy also are possible candidate genes for non-familial cases. Since ion channel genes in particular represent especially large genes, the development of a diagnostics panel for high throughput screening seemed the obvious choice. An epilepsy panel for clinical use was developed by Dr. Johannes Lemke, University of Bern together with the groups around Prof. Holger Lerche and Prof. Ingeborg Krägeloh-Mann (Neurologic and Pediatric University Clinic, Tübingen) in collaboration with CeGaT GmbH. At present the list of genes for purely diagnostic objectives consists of 55 genes and has been clinically subdivided into generalized/myoclonic epilepsies including febrile seizures and absences (a total of 24 genes), epileptic encephalopathies (a total of 8 genes) and syndromal diseases with epilepsy (a total of 23 genes). The panel contains another 450 candidate genes that are being studied for research purposes.

### Diagnostics panel and Parkinson's disease and dementia

Parkinson's disease together with Alzheimer's disease is one of the most frequently occurring neurodegenerative diseases worldwide. Both diseases most often occur sporadically and as a rule manifest themselves in individuals above 65 years of age. With a steadily rising life expectancy Parkinson's and Alzheimer's represent one of the greatest medical and socio-economic challenges of the future. In most cases the cause of the death of nerve cells remains a mystery. At the present time it is impossible to predict whether and when an individual will be affected by the breakdown of nerve cells. Once symptoms occur, however, the majority of affected nerve cells have already died. Hence, current treatment concepts have little or no effect, since the time of intervention is years too late. A molecular genetic examination does not immediately offer itself. Why would an individual want to know whether he or she has a predisposition for a neurodegenerative disease as long as no therapies are available? Genetic causes for both Parkinson's and Alzheimer's have been described for slightly more than 10 years. These genetic studies have made important contributions to understanding the breakdown of nerve cells. Intensive research is being done on the gene products in order to speed up new and innovative therapy concepts. The identification of mutated genes has for the first time made it possible to describe biomarkers, in this case "genetic biomarkers", that can predict the occurrence of the disease at a future point in time with a high degree of probability. This allows us to define a group of individuals who could get access to therapies decades before any manifestation of disease. Even though at present such therapies are not yet available, it seems safe to say that the changed gene products will most probably represent the points of action for the therapies of the future.

The key to the changed genes were families in which dementia or Parkinson's disease occurred with great frequency. In the case of Parkinson's disease 16 gene locations for familial autosomal recessive and dominant forms have so far been described in the hereditary [3, 4]. Since only a small part of approx. 5% of familial cases can currently be genetically clarified, the expectation is that the list of genes causing Parkinson's disease will grow. The greater part of the knowledge we possess today about the pathogenesis of Parkinson's disease derives from those genes that have been described in connection with the disease. The gene therefore is the first clue concerning the location of the malfunction within the diseased cell. Like pieces of a puzzle other genes will add to the complex picture of neurodegenerative diseases and – it is hoped – allow a further crucial step in our understanding of the disease in the near future.

The current Parkinson's dementia panel is studying 16 genes in the case of Parkinson's disease, 19 genes in the case of dementia. The panel was created together with Prof. Thomas Gasser (Neurology and Hertie-Institute for Clinical Brain Research, University of Tübingen). The list of genes will grow rapidly, not least due to the possibility of examining families affected by the disease for changes in the total

genome through a method free of hypotheses and within the scope of a research objective.

As with a purely diagnostic objective and within the framework of the genetic diagnostics law the individual seeking advice should also be informed concerning the research objective. Most particularly this includes information on the possible handling of incidental findings with relevance for all family members, the possible destruction of the probe after the examination is completed, the anonymization and use of the probe for further studies, and the right not to know at any point in time.

### Outlook and open questions

From the perspective of human genetics the quick and cost efficient sequencing of several thousand human genes within a few days is revolutionary. Also revolutionary is the prospect of personalized medicine in which each single human and each single tumor can be sequenced. It is hoped that with this knowledge diseases might someday be treated individually, i.e., much more targeted than today. Each human is unique, each tumor is unique, each disease with its individual genetic background is unique. Medications therefore have different effects in different people. A changed gene is an essential key for understanding a disease. Even if individualized therapy, e.g., with neurodegenerative diseases, lies in the distant future, the foundation for the therapies of the future is being laid now. For some time tumor genetics has made targeted therapy partially possible, other diseases have followed and more will follow. With all the euphoria that genetics has been experiencing for some time it must be remembered that the results of high-throughput sequencing bring up questions that cannot be answered at this time. This includes the identification of as yet unknown variants in the genome of individuals, variants of uncertain significance (VUS). Add to this that, in spite of high-throughput sequencing, in many cases the cause of a disease cannot be found, be it that no genetic cause exists or that the cause lies in the non-examined regions of the genome, or that it is the synergy of several changed genes with the environment that results in the disease, or that changed gene products (RNA or proteins) are the actual cause of the disease.

High-throughput sequencing or “deep sequencing” is a screening method. Hereditary material is illuminated and it can happen that variations are found that are characterized as incidental findings. Dealing with incidental findings, particularly if they have serious consequences for the patient, is a considerable challenge for the physician and the individual seeking counseling. As with the findings of variants of uncertain significance the patient must be informed about this issue in advance. The result of a genetic examination should be conveyed within the framework of a consultation. The law on gene diagnostics, in force since February 2010, established guidelines for the performance of genetic diagnostics that can also be directly applied to high-throughput diagnostics.

It is hoped that high-throughput diagnostics will arrive in many laboratories and as a consequence will not only contribute to markedly higher clarification quotas of genetically

caused diseases, but also will establish genetic diagnostics as a quick, efficient, cost-effective and useful method in the minds of the individuals seeking counseling, of affected individuals, of physicians and scientists. In this regard the diagnostic panels are a step in the right direction.

### Method/Technology

Next-generation sequencing stands for high-throughput sequencing and allows sequencing of up to 100 billion bases within days. A complete human genome can be sequenced with an average coverage of 30 (i.e., each base is read 30 times). Various technology platforms for high-throughput sequencing are commercially available; the main suppliers include Roche, Illumina and Applied Biosystems/Life Technologies. While the Roche platform achieves the comparatively lowest throughput of 400 million bases per run, it offers advantages in regard to the length of the sequenced fragments (approx. 400 bases). Illumina and Applied Biosystems/Life Technologies with up to 100 billion bases achieve far higher throughputs, but with a shorter read length (50 to 100 base pairs). Since all three platforms are based on different sequencing strategies, which we cannot go into here, each of them is also suited for different objectives. The “Sequencing by Ligation” method used by Applied Biosystems/Life Technologies on its SOLiD platform delivers the highest throughput with by far the lowest error rate. This is of great importance for diagnostic applications, since all variants found in high-throughput sequencing are verified conventionally, i.e., via the Sanger method, and the lower the initial error rate the more feasible and safe it will be to introduce high-throughput sequencing into the human genetics practice.

The process of high-throughput sequencing is as follows. A patient’s genomic DNA is obtained from blood or tissue, fragmented randomly by sonication and subsequently equipped with adaptors. The fragments to be sequenced are fished out with complementary RNA baits that are linked to magnetic beads (SureSelect Method by Agilent). The enriched DNA is finally amplified. This is called targeted enrichment. Here as well different technologies are available in the marketplace but cannot be discussed in any detail in this paper. The method used to perform a large number of amplifications in parallel is called EmulsionPCR. This EmulsionPCR is a special kind of PCR that takes place on beads in an aqueous droplets separated by an oil phase. These beads, equipped with several ten thousands of universal PCR-primer molecules, are magnetic and can later on separated by magnetic forces. The enriched DNA is bound via its attached adaptors to the primers on the beads. One DNA fragment is then multiplied by means of PCR in one bead per aqueous droplet at a time. The newly formed fragments bind to additional primers on the same bead. Following breaking the emulsion, those beads, on which a PCR has successfully occurred, are cross-linked by polystyrene beads and separated through centrifugation based on their size and weight of the cross-linked beads when compared to those to which nothing has bound. Then about

700 million beads are deposited on a slide, which in turn is clamped into the high-throughput sequencer.

Sequencing occurs through detection of light signals that are given off by hybridizing octamers (8 nucleotides with one specific fluorescence signals). The octamers contain two nucleotides that code a color space and gives off one of four fluorescent signals immediately after binding, which are photographed by a digital camera with a CCD sensor. The color signals are then separated and a further binding cycle can be initiated until approx. 50 nucleotides can be converted in a series of color signals, whereby one color always corresponds to two nucleotides. The staggered multi-step read of the original sequence ensures that each base is read at least twice and variations can be identified reliably. This results in the special reliability of the SOLiD system by Applied Biosystems/Life Technologies in regard to the interpretation of the data. At the end of the analysis the color signals are converted back into the nucleotide sequence that then enters into the continuing analysis. Five to seven days are required for the sequencing step. The subsequent data analysis consists of the following three individual steps: Step 1: Initially the data generated by the sequencer arrive in the form of pictures consisting of individual color dots, whereby one particular color dot corresponds to two nucleotides. The color dots are examined by the software for their intensity and quality, both necessary for definitely determining the nucleotides. A software supplied by the manufacturer converts the color dots into the color space

sequence coded by the color. The automatic comparison with the reference sequence converted into color space, in this case the human genome, takes place in Step 2. The individual 50 color space base pair long fragments are compared to the reference sequence and the color spaces are displayed by side-by-side (alignment). This two base pair encoding of the SOLiD system allows the quick and easy recognition of miscalls and the clear differentiation between miscalls and actual sequence variations. All true deviations from the reference sequence are transmitted and verified via Sanger sequencing. The last and most difficult step involves the interpretation of the data.

## References

1. Berger W, Kloeckener-Gruissem B, Neidhardt J. The molecular basis of human retinal and vitreoretinal diseases. *Prog Retin Eye Res* 2010;29:335–75.
2. Boehm, D, Herold S, Kuechler A, Liehr T, Laccone F. Rapid detection of subtelomeric deletion/duplication by novel real-time quantitative PCR using SYBR-green dye. *Hum Mutat* 2004;23:368–78.
3. Biskup S, Gerlach M, Kupsch A, Reichmann H, Riederer P, Vieregge P, et al. Genes associated with Parkinson syndrome. *J Neurol* 2008;255(Suppl 5):8–17. Review.
4. Gasser T. Mendelian forms of Parkinson's disease [review]. *Biochim Biophys Acta* 2009;1792:587–96.