



**HAL**  
open science

## Novel genomic techniques in the analysis of monogenic disorders

Gregor Kuhlenbäumer, Julia Hullmann, Silke Appenzeller

► **To cite this version:**

Gregor Kuhlenbäumer, Julia Hullmann, Silke Appenzeller. Novel genomic techniques in the analysis of monogenic disorders. *Human Mutation*, 2011, 32 (2), pp.144. 10.1002/humu.21400 . hal-00612008

**HAL Id: hal-00612008**

**<https://hal.science/hal-00612008>**

Submitted on 28 Jul 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Novel genomic techniques in the analysis of monogenic disorders

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0423.R1
Wiley - Manuscript type:	Review
Date Submitted by the Author:	12-Oct-2010
Complete List of Authors:	Kuhlenbäumer, Gregor; Christian-Albrechts-University, Institute of Experimental Medicine Hullmann, Julia; Christian-Albrechts-University, Institute of Experimental Medicine Appenzeller, Silke; Christian-Albrechts-University, Institute of Experimental Medicine
Key Words:	monogenic, disease gene identification, exome sequencing, SNP array

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5  
6 **Novel genomic techniques open new avenues in the analysis of monogenic disorders**  
7  
8  
9

10 G. Kuhlenbäumer, MD, PhD; Julia Hullmann, PhD; Silke Appenzeller, PhD  
11

12  
13  
14 Institute of Experimental Medicine, Christian-Albrechts University Kiel, Germany  
15  
16  
17  
18  
19

20 Key words: monogenic, exome sequencing, disease gene identification  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

32 Address for correspondence:  
33

34 Prof. G. Kuhlenbäumer, MD, PhD  
35

36 Institute of Experimental Medicine  
37

38 Dept. of Molecular Neurobiology  
39

40 Christian-Albrechts University Kiel  
41

42 Arnold-Heller Str. 3  
43

44 D-24105 Kiel  
45

46 Phone: 0049/431/5978806  
47

48 Fax: 0049/431/5978812  
49

50 e-mail: g.kuhlenbaeumer@neurologie.uni-kiel.de  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 **Abstract:** The molecular genetic cause of over 3000 monogenic disorders is currently  
3  
4 unknown. This review discusses how novel genomic techniques like Next-Generation-DNA-  
5  
6 Sequencing (NGS) and genotyping arrays open new avenues in the elucidation of genetic  
7  
8 defects causing monogenic disorders. They will not only speed up disease gene identification  
9  
10 but will enable us to systematically tackle previously intractable monogenic disorders. These  
11  
12 are mainly disorders not amenable to classic linkage analysis e.g. due to insufficient family  
13  
14 size. Most monogenic diseases are caused by exonic mutations or splice-site-mutations  
15  
16 changing the amino acid sequence of the affected gene. These mutations can be identified by  
17  
18 sequencing of all exons in the human genome (exome sequencing) rendering whole genome  
19  
20 sequencing unnecessary in most cases. Genotyping-arrays containing  $10^5 - 2 \times 10^6$  single-  
21  
22 nucleotide-polymorphisms (SNP) and non-polymorphic markers allow highly accurate  
23  
24 mapping of genomic deletions and duplications not detectable by exome sequencing which  
25  
26 are the second most common cause of monogenic disorders. However, several hundred rare,  
27  
28 previously unknown sequence variants affecting the amino acid sequence of the encoded  
29  
30 protein are found in the exome of every human individual. Therefore the main problem will be  
31  
32 the differentiation between the many rare benign variants detected by novel genomic  
33  
34 techniques and disease causing mutations.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Introduction

During the last years enormous technical progress has been achieved in the areas of DNA-sequencing and genotyping-array technology. Resequencing of the whole human genome can now be achieved within weeks instead of months to years and genotyping-array technology allows the detection of genomic duplications and deletions on a whole genome scale within days and with unprecedented accuracy. These technologies are leading to new strategies in the analysis of monogenic diseases allowing the identification of causative mutations in disorders which are not amenable to linkage-based positional cloning [Biesecker, 2010; Check Hayden, 2009]. In this review we will focus on two technologies: Genotyping arrays and exome sequencing, and the strategies which can be applied to filter the many detected, previously unknown SNPs for the causative mutation. Exome sequencing is defined as the selective resequencing of all exons in a genome. Since most mutations causing monogenic disorders are exonic this approach is especially promising for the analysis of monogenic disorders. A number of studies has already proven that exome sequencing alone is able to identify genes involved in monogenic disorders [Lalonde, et al., 2010; Ng, et al., 2010a; Ng, et al., 2010b; Ng, et al., 2009].

### **Why do we still need to study monogenic disorders?**

Why do we need to study rare monogenic disorders in the age of complex genetics elucidating the genetic basis of more important common complex disorders [Peltonen, et al., 2006]?

1. Compared to infrequent complex disorders many monogenic disorders are not this rare.

Monogenic variants of common complex disorders can be very frequent, especially in

1  
2 selected ethnic groups. The Leucine-Rich-Repeat-Kinase-2 (*LRRK2*, MIM609007) G2019S  
3  
4 mutation is for example responsible for a large proportion of all Parkinson's disease (PD)  
5  
6 cases in North African Arabic- and Ashkenazi Jewish patients [Lesage, et al., 2006; Ozelius,  
7  
8 et al., 2006].  
9

10 2. Rare monogenic variants of common diseases contribute eminently to our understanding of  
11  
12 complex disorders, e.g. in Alzheimer's disease (AD). The current view of AD pathogenesis,  
13  
14 the so called amyloid hypothesis, is largely based on the discovery of mutations in three  
15  
16 genes (*APP*, MIM104760; *PS1*, MIM104311 and *PS2*, MIM600759) causing monogenic AD  
17  
18 [Ertekin-Taner, 2007].  
19

20 3. Monogenic disorders can be modeled in animal- and cell-systems mainly because they  
21  
22 have a strong phenotypic effect. Animal models may be used as pathogenetic as well as  
23  
24 therapeutic disease models. In contrast most genetic variants known to be associated with  
25  
26 complex disease to date are difficult to analyze in model systems because 1.) they have only  
27  
28 marginal phenotypic effects and 2.) the really causative variant cannot be discerned from  
29  
30 variants in close linkage disequilibrium.  
31

32 4. Mutations causing monogenic diseases provide information about the function of the  
33  
34 affected gene.  
35

Deleted: ¶  
Identification of mutations causing  
monogenic disorders¶

36  
37  
38 Please note, that for simplicity all phrases and words in the following paragraphs are meant to  
39  
40 refer to monogenic disorders only and not to their complex counterparts unless stated  
41  
42 otherwise.  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

The classic strategy of disease gene identification: positional cloning based on linkage

analysis

To understand the new opportunities created by novel genomic techniques one has to comprehend the limitations of classical positional cloning [Botstein and Risch, 2003]. Positional cloning starts with the ascertainment of large families inheriting the disease under investigation in a Mendelian fashion. The disease gene is mapped by linkage analysis [Ott, 1991]. Linkage analysis allows the delineation of a genomic candidate region containing the disease gene. All genes in the candidate region are analyzed for mutations. Causality of genetic variants in a gene is assumed to be shown if the variants meet the following requirements.

- 1.) The variants segregate with the disease phenotype in the families.
- 2.) The variants are unknown, meaning that they are neither found in databases (e.g. dbSNP, ) nor in controls.
- 3.) The variants lead to a (non-conservative) amino acid exchange, a protein truncation or a splicing aberration.
- 4.) Different variants in the same gene segregating with the disease are found in different affected families.

In the best case these data are backed by:

- 5.) Functional data supporting causality [Collins, 1995].
- 6.) Finding de-novo mutations in some families.

Experience has shown that this approach is highly reliable. The main advantage of linkage analysis compared to genome or exome sequencing is the identification of a small candidate region increasing the a priori likelihood that an identified variant in this region is causal. In most cases only one variant fulfilling the above mentioned criteria is found. The main

drawback of linkage analysis is the requirement for large families. To achieve statistical significance a multigenerational pedigree with 6-12 affected individuals is required for autosomal dominant disorders. Autosomal recessive disorders in consanguineous families can often be solved in small families using homozygosity mapping but non-consanguineous families with suspected causative compound-heterozygote mutations are not amenable to homozygosity mapping because the affected children do not inherit identical copies of the candidate region from both parents. In addition phenocopies and significantly reduced disease penetrance strongly impede linkage-based disease gene identification.

Deleted: s

Deleted: ¶

Deleted: Novel strategies: Direct disease gene identification by Next-Generation-DNA-Sequencing and genotyping-arrays¶

Formatted: Underline

### Novel genomic techniques: Overview of Next-Generation-DNA-sequencing and Genotyping-array technology

**Next-Generation-Sequencing (NGS)** bears some fundamental differences compared to conventional Sanger-sequencing. Sanger-sequencing of candidate genes is directed to a specific target, mostly by sequencing PCR-amplified-DNA. Sanger-sequencing generates long “reads”, meaning that one reaction deciphers 400-800 bases allowing easy sequence assembly. In principle, one sequence suffices to identify homozygous and heterozygous sequence variants because every reaction contains and determines both possible alleles. Sanger-sequencing always follows the same technical principle but NGS uses a number of different technical approaches. The following characteristics of NGS are somewhat generalized and not all of them apply to all methods. Technical details can be found in other reviews [Dufva, 2009; Mardis, 2008; Tucker, et al., 2009]. NGS as such is based on sequencing clonally amplified single molecules of DNA. NGS needs between 10–50 reads of the same base to reliably identify a heterozygous sequence variant, because every read shows only one of the possible two alleles in a stochastic distribution. NGS generates comparatively

Deleted: not directed to a target sequence. All DNA in the reaction mixture is sequenced. NGS generates comparatively short reads making assembly more difficult.



1  
2 short reads making assembly more difficult. The main advantage of NGS is the massive  
3  
4 amount of generated sequence data. Depending on the method used between ~400 and ~3000  
5  
6 megabases can be generated in one “run” of the instrument [Mardis, 2008]. For comparison,  
7  
8 the human genome has a size of ~3000 megabases. While NGS is much cheaper than Sanger-  
9  
10 sequencing on a per base basis, it is still very expensive on a genome scale. Because the  
11  
12 overwhelming majority of all single base mutations causing monogenic disorders are located  
13  
14 in exons or in splice sites in the immediate vicinity of exons, it should be sufficient to  
15  
16 sequence all exons, the so called “exome” to detect most mutations causing monogenic  
17  
18 diseases. The exome has a size of only ~30 megabases. Exome sequencing is therefore  
19  
20 currently much cheaper than whole genome sequencing. The starting material used for  
21  
22 mutation analysis is usually whole genomic DNA. Therefore exonic DNA has to be enriched  
23  
24 from genomic DNA to perform exome sequencing. A number of enrichment-methods is  
25  
26 currently available. Most of them use hybridization of fragmented genomic DNA to  
27  
28 immobilized exonic target DNA [Mamanova, et al., 2010; Summerer, 2009; Turner, et al.,  
29  
30 2009]. None of these methods is perfect and able to extract pure exonic DNA which  
31  
32 homogeneously covers the whole target. Therefore not all sequences produced by exome  
33  
34 sequencing will be “on-target”, meaning within the targeted exons.

Deleted:

Deleted: {Summerer, 2009 #31 }

35  
36  
37 **Genotyping-arrays** also called SNP-chips are devices allowing the simultaneous  
38  
39 determination of the genotypes of  $10^4$  to  $2 \times 10^6$  SNPs [Bier, et al., 2008]. Genotyping-array  
40  
41 technology has been available for years and is technically mature. Genotyping-arrays are now  
42  
43 the method of choice for linkage analysis because they are faster and allow to construct a  
44  
45 much denser linkage map than conventional Short-Tandem-Repeat marker based linkage  
46  
47 analysis. Most current genotyping-arrays also contain probes of non-polymorphic DNA  
48  
49 allowing the detection of small genomic duplications and deletions in the range of some  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Deleted: Determination of the SNP-genotypes is usually based on allele-specific hybridization of genomic DNA followed by a detection step involving fluorophores.

1  
2 kilobases (=Copy Number Variants, CNV). Analysis of the signal intensities is used to  
3 determine CNVs [Komura, et al., 2006; Peiffer, et al., 2006]. Normal diploid DNA contains  
4 two copies of each DNA fragment while genomic duplications contain three copies and  
5 deletions only one copy. Hence, probes within heterozygous duplications have a signal  
6 intensity of about 1.5 and probes within heterozygous deletions of 0.5, while homozygous  
7 duplications show a signal intensity of 2.0 and homozygous deletions of 0 compared to wild-  
8 type segments with two copies.

Deleted: .

Deleted: Hybridization is a quantitative process and therefore these probes can be used to detect genomic duplications and deletions (= Copy Number Variants, CNV).

Deleted: The high density of probes allows the reliable detection of small duplications and deletions in the range of some kilobases.

Deleted: ¶

### New strategies in the analysis of monogenic disorders

18  
19  
20  
21  
22 In the following paragraphs we will describe the novel strategies in the analysis of monogenic  
23 disorders made possible by the technologies introduced in the previous paragraphs, starting  
24 with the genetic characteristics of suitable monogenic disorders which were previously  
25 intractable by classical positional cloning. Sections on CNV-analysis and NGS describe the  
26 application of these techniques to monogenic disorders in more detail and give examples  
27 from the most recent literature. Three further paragraphs are devoted to the question how the  
28 causative mutant can be filtered out from the background of thousands of innocuous variants  
29 present in every exome. The last sections deal with possible problems, diagnostic applications  
30 and finally give a summary and an outlook.

Formatted: Font: Not Bold

### **Genetic characteristics of Mendelian disorders suited for analysis by novel genomic techniques**

41  
42  
43  
44  
45  
46  
47  
48 Positional linkage based disease gene identification is a proven reliable strategy and will  
49 remain the gold-standard if suitable families are available. In this context, NGS may be used  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 to sequence whole candidate regions determined by linkage analysis. However, in many cases  
3  
4 large families are not available or one might even find single affected individuals only. These  
5  
6 situations could be simply due to ascertainment problems, but the more common issue is that  
7  
8 the disease under investigation confers a severe reproductive disadvantage. Either statistical  
9  
10 power is too small for linkage analysis in the families or, in the case of single patients only,  
11  
12 linkage analysis is principally not applicable. A number of scenarios is conceivable (Fig. 1).  
13  
14 In all scenarios it is assumed that all patients share the same clinical phenotype but not  
15  
16 necessarily the same genetic cause of the disease:

17  
18 1.) Only a number of small families and some single individuals are available (Fig. 1.1). The  
19  
20 disease might be autosomal dominant or autosomal recessive. A number of examples  
21  
22 demonstrating the feasibility of disease gene identification in this setting are already available  
23  
24 and described in more detail in the following paragraphs. Theoretically linkage-analysis  
25  
26 could be achieved using several hundred small families, e.g. affected sibling pairs. This  
27  
28 approach has been used for complex disorders. However, such large numbers of families with  
29  
30 a monogenic disorder are hardly ever available and in most disorders high locus  
31  
32 heterogeneity would severely hamper this approach. Recessively inherited diseases with  
33  
34 unrelated parents due to compound heterozygote mutations are not amenable to  
35  
36 homozygosity mapping and will be excellent targets for disease gene identification by exome  
37  
38 sequencing.

39  
40 2.) The disease is seemingly sporadic because it is found in multiple single individuals with  
41  
42 unaffected parents (Fig. 1.2). This may be the case for diseases leading to death or severe  
43  
44 handicap in childhood as well as disorders directly impairing fertility. That these diseases  
45  
46 might in many instances be in fact monogenic and caused by de-novo mutations has been  
47  
48 shown for some disorders by candidate gene sequencing and for others by exome sequencing

49  
50 [Hoischen, et al., 2010; Ng, et al., 2010a; Saitsu, et al., 2008]. Alternatively, the disease

Deleted: .

1  
2 might be autosomal recessive with only one affected child per family. This could e.g. be the  
3  
4 case if the disease leads in to stillbirths most cases.

**Deleted:** The strategy and the results of the latter will be described in more detail in the following paragraphs.

5  
6 3.) Multiple small families with presumed X-linked inheritance are available (Fig. 1.3). In  
7  
8 this scenario, targeted NGS of all genes on the X-chromosome will be a solution [Tarpey, et  
9  
10 al., 2009].

11 4.) It can also be envisaged to apply exome sequencing in distantly related individuals of  
12  
13 large families who share a common founder many generations ago. These individuals share  
14  
15 only a small proportion of their genome allowing to reduce the number of putatively disease  
16  
17 causing mutations. Linkage based cloning will still be the method of choice for large families  
18  
19 with monogenic disorders. However, sometimes significant linkage can not be established  
20  
21 despite sufficient family size. This might be due to misclassification of patients, problems  
22  
23 with low disease penetrance, clinically identical phenocopies and further to date unknown  
24  
25 causes. In addition a number of rather common disorders exists for which many families with  
26  
27 presumed monogenic inheritance have been collected and for some of these diseases even  
28  
29 linkage has been established in a few families but no causative mutations could be identified.  
30  
31 Two prototypical examples are Restless-legs syndrome and Essential tremor [Deng, et al.,  
32  
33 2007; Winkelmann and Müller-Myhsok, 2008]. It will be very interesting to see whether  
34  
35 exome sequencing can contribute to the elucidation of the genes underlying these disorders.  
36  
37  
38  
39

#### 40 **Identification of genomic duplications and deletions using genotyping-arrays**

41  
42  
43 Genomic rearrangements, especially larger genomic duplications or deletions, are the second  
44  
45 most common cause of monogenic diseases. Very small genomic duplications or deletions  
46  
47 can be detected by NGS or Sanger-sequencing. Genotyping-array analysis for CNVs will  
48  
49 often be performed before exome sequencing because it costs only between 300\$ and 700\$  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 per individual. Duplications or deletions which are absent in databases and large control data  
3 sets are suspicious. One study for example screened 3699 controls negative for the candidate  
4 CNV, others even more [Helbig, et al., 2009; Shinawi, et al., 2009]. Duplications and  
5  
6 CNV, others even more [Helbig, et al., 2009; Shinawi, et al., 2009]. Duplications and  
7  
8 deletions detected in genotyping-arrays can be verified using methods like quantitative PCR,  
9  
10 Fluorescence In Situ Hybridization (FISH) and long range PCR [Helbig, et al., 2009;  
11  
12 Shinawi, et al., 2009]. Unknown rearrangements segregating with the disease in families and  
13  
14 present in multiple genetically independent affected individuals are excellent, most likely  
15  
16 causal candidate mutations. The same is true for de-novo duplications or deletions in patients  
17  
18 since de-novo mutations are extremely rare. In many cases, a monogenic disease is caused by  
19  
20 CNVs in a subset of patients while others bear a point mutation in a gene contained within  
21  
22 the duplication or deletion. Two good examples are Charcot-Marie-Tooth disease type 1A  
23  
24 (CMT1A), caused in most patients by a duplication on chromosome 17p11.2 harboring the  
25  
26 *PMP22* gene (MIM601097) but in some cases also by point mutations in the same gene  
27  
28 [Kuhlenbäumer, et al., 2002] and early infantile epileptic encephalopathy caused by CNVs as  
29  
30 well as point mutations involving the *STXBPI* gene (MIM602926) [Saitsu, et al., 2008].  
31  
32  
33

#### 34 **Next-Generation-DNA sequencing: Identification of point mutations**

35  
36  
37 If whole exome sequencing is applied to monogenic disorders, it is not trivial to identify the  
38  
39 causative variant among the many identified variants present in the exome. The a priori  
40  
41 likelihood that a variant is causal is much lower than in the classic linkage based approach,  
42  
43 because the whole exome and not only the genes in a small candidate region identified by  
44  
45 linkage analysis are sequenced. Disease causing mutations are in most cases either missense  
46  
47 mutations which change an amino acid, nonsense mutations which introduce a premature stop  
48  
49 codon or small duplications/deletions causing a frameshift altering the amino acid sequence  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 C-terminal of the mutation which often leads to a premature stop codon and protein  
3  
4 truncation. Currently available exome sequencing data suggest that ~5000-10000 of these  
5  
6 three types of mutations are found in the average Caucasian exome [Ng, et al., 2008; Ng, et  
7  
8 al., 2009]. The overwhelming majority of these variants are most likely benign.  
9  
10 Approximately 400 to 700 of these variants are absent in SNP-databases, in the eight fully  
11  
12 sequenced exomes available in the HapMap database and the currently available data of the  
13  
14 1000 Genomes project [Lalonde, et al., 2010; Ng, et al., 2008; Ng, et al., 2010b; Ng, et al.,  
15  
16 2009]. In a small family, many of these variants will segregate with the disease by chance and  
17  
18 many identified variants will not be present in a reasonable number of controls (100-1000),  
19  
20 because only rare variants have been selected for examination beforehand. In addition, it is  
21  
22 costly and time consuming to check all variants for segregation and absence in controls by  
23  
24 Sanger-sequencing or other genotyping-methods. Completion of the thousand genomes  
25  
26 project establishing the complete genome sequence of more than 1000 humans will enable us  
27  
28 to eliminate a large number of variants which are present at low frequencies in the general  
29  
30 population but not found in current SNP-databases (<http://www.1000genomes.org>).  
31  
32 Nonetheless, data will have to be painstakingly filtered to allow reliable disease gene  
33  
34 identification.

### Approaches for filtering the identified variants (Fig. 2)

#### Exploiting the characteristics of Mendelian inheritance

Formatted: Font: Bold, Italic

45 The first step will be to exploit the characteristics of Mendelian inheritance (Fig. 2, “Genetic  
46 filter”). If a number of unrelated families and single individuals with the same clinical  
47  
48 phenotype are available the first and most powerful step is to look for different mutations  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 affecting the same gene in the affected individuals from unrelated families. That this  
3  
4 approach is feasible has recently been shown e.g. for the autosomal-recessive Miller-  
5  
6 syndrome (MIM263750) [Ng, et al., 2010b] and the Fowler-syndrome (MIM610865)  
7  
8 [Lalonde, et al., 2010]. In Miller-syndrome exome sequencing was performed in four affected  
9  
10 individuals belonging to three independent non-consanguineous families. It was unknown if  
11  
12 Miller-syndrome is autosomal recessive or autosomal dominant. Therefore both models were  
13  
14 taken into account. About 400 to 500 candidate variants not present in any SNP-database  
15  
16 were identified in each patient. Thirty-one and 33 of these variants were compatible with  
17  
18 recessive inheritance in each of the two sequenced siblings from family one but both  
19  
20 probands of this family taken together showed only 9 SNPs compatible with recessive  
21  
22 inheritance. Adding the single probands from families two and three allowed to reduce the  
23  
24 number of candidate genes to only one in the recessive and eight in the dominant model. The  
25  
26 authors calculated the Bonferroni corrected p-value that two different variants of the ~500  
27  
28 candidate variants are found in the same gene in each of three independent affected  
29  
30 individuals is  $p \sim 0.000015$  arguing strongly against appearance by chance. Sanger-sequencing  
31  
32 of the implicated *DHODH* gene (MIM126064) in three additional singleton cases revealed  
33  
34 compound heterozygote mutations in every affected individual. In Fowler-syndrome two  
35  
36 patients with this recessively inherited disorder sufficed to pinpoint the *FLVCR2* gene  
37  
38 (MIM610865) [Lalonde, et al., 2010]. A number of other very recent reports suggests that  
39  
40 even single patients may suffice to identify recessive mutations causing monogenic disorders,  
41  
42 especially if enough is known about the pathomechanism to filter candidate mutations [Byun,  
43  
44 et al., 2010; Gilissen, et al., 2010; Rios, et al., 2010]. Naturally, it is easier to reduce the  
45  
46 number of candidate mutations for an autosomal recessive disease than for an autosomal  
47  
48 dominant disease because only genes with both alleles bearing an unknown or a very rare  
49  
50 variant have to be considered. Nevertheless exome sequencing can also be applied  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Deleted: {Lalonde, 2010 #75}

1  
2 successfully to autosomal dominant disorders. Ten unrelated probands with autosomal  
3  
4 dominant Kabuki-syndrome (MIM147920) were analyzed but exome sequencing did not  
5  
6 identify a candidate gene in which all affected individuals showed an unknown variant [Ng,  
7  
8 et al., 2010a]. This problem could be due to locus heterogeneity, to mutations outside the  
9  
10 targeted exome or to the fact that exome sequencing failed to detect all mutations. However,  
11  
12 further analyses allowed to find mutations in the *MLL* gene (MIM159555) in 7 out of the 10  
13  
14 sequenced exomes. Sanger sequencing revealed small frameshift indels not detected by NGS  
15  
16 in 2 of the remaining 3 patients and sequencing of the *MLL* gene in 43 additional, seemingly  
17  
18 sporadic patients detected mutations in 26 of these. This investigation proves that exome  
19  
20 sequencing is able to identify genes causing monogenic disorders even in the presence of  
21  
22 moderate locus heterogeneity. The degree of locus heterogeneity still permitting the  
23  
24 identification of disease genes will depend to a large degree on the number of families  
25  
26 available for exome sequencing. Interestingly, all mutations in the *MLL* gene had arisen de-  
27  
28 novo in the 12 patients for whom parental DNA was available underpinning that seemingly  
29  
30 sporadic disorders might in fact be due to dominant de-novo mutations. It seems not unlikely  
31  
32 that severe, often syndromal disorders conferring a large reproductive disadvantage and  
33  
34 always affecting only one sibling might often be caused by de-novo mutations. It has been  
35  
36 estimated that on average only two de-novo mutations are present per exome. [Nachman and  
37  
38 Crowell, 2000] Exome sequencing of one healthy triplet consisting of the parents and one  
39  
40 offspring did not identify any de-novo mutation in the exome of the offspring [Hedges, et al.,  
41  
42 2009]. Therefore exome sequencing of triplets consisting of an affected child and the healthy  
43  
44 parents will be an ideal way to drastically reduce the number of potentially disease causing  
45  
46 candidate variants. In the best case, only one de-novo variant will be found per-patient.  
47  
48 Sanger-sequencing of the identified gene in further individuals affected by the same disease  
49  
50 might reveal additional de-novo mutations and make causality highly likely. [The](#)



1  
2 identification of heterozygous de-novo mutations in the *SETBP1* gene (MIM611060) causing  
3  
4 autosomal dominant Schnitzel-Giedion syndrome (MIM 269150) by exome sequencing  
5  
6 demonstrates the feasibility of this approach [Hoischen, et al., 2010]. If one considers to  
7  
8 identify the gene causing a monogenic disorder in one large family, some degree of filtering  
9  
10 could be achieved by sequencing two or more distantly related individuals which share only a  
11  
12 fraction of their genome. If the two related individuals have a common distant known or  
13  
14 unknown founder the shared part of the genome will be reduced by  $2^n$  per generation. The  
15  
16 variants identified in different genes have to be tested for segregation in the families as an  
17  
18 additional filter. As previously mentioned, the causative variants of rare disorders should  
19  
20 usually be absent in ethnically matched unaffected controls with most published reports using  
21  
22 200-1000 controls. However, using modern genotyping techniques one could easily analyze  
23  
24 5000-10000 controls to differentiate between rare non-causal variants present also in the  
25  
26 unaffected population and causative mutations. If a large number of controls is used for a  
27  
28 disorder with a moderate disease frequency, it becomes mandatory that all controls  
29  
30 individuals are screened negative for the investigated phenotype or that individuals positive  
31  
32 for a variant under question can be re-examined clinically. In addition, recessive disorder  
33  
34 might be caused by mutations found in the unaffected population at a low frequency in the  
35  
36 heterozygous state. A recent study describes that severe brain malformations can be caused  
37  
38 by recessive mutations in the *WDR62* gene (MIM 613583) [Bilguvar, et al., 2010]. One  
39  
40 particular mutation was present in a homozygous or compound heterozygous state in some of  
41  
42 the patients but also in 0.2% of the ethnically matched controls in the heterozygous state.  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Formatted: Font: Italic

Deleted: In principal, the same approach can be applied to X-chromosomal disorders using enrichment methods targeting only the X-chromosomal DNA. Filtering genetic variants to identify the potentially causative one will be facilitated by the much smaller size of the X-chromosomal exome compared to the whole exome resulting in a much smaller number of novel SNPs among which the disease causing one has to be identified.¶

Deleted: . Most

Deleted: ed

Formatted: Font: Italic

*The incorporation of expression and functional data in the filtering process*

Functional, expression and evolutionary data as well as the computational prediction of functional effects of mutations can be incorporated in the filtering process (Fig. 2, “Functional filter”). Candidate genes can be ranked according to their function and expression. An easily understandable and successful example is the selection of ion channels as candidate genes for epilepsies. However, there are also numerous examples in which genes with an entirely unknown relation to the pathogenesis of the investigated disease were identified as disease genes. Expression data tell us which genes containing a mutation are expressed in the tissue affected by the disease. Genes which are not expressed in the target tissue are bad candidate genes. Disease causing mutations are more common in evolutionary conserved genes and in well conserved amino acid residues than in non-conserved ones. There are numerous methods to include the assessment of evolutionary conservation in the filtering process [Mooney, 2005; Ng and Henikoff, 2006; Ng, et al., 2008; Tarpey, et al., 2009]. Computational prediction of functional SNP effects often employs evolutionary conservation as well as physicochemical properties of the affected amino acid in the protein [Mooney, 2005; Ng and Henikoff, 2006]. Details regarding these methods and algorithms can be found in a number of reviews [Mooney, 2005; Ng and Henikoff, 2006]. Well known examples for programs which can be used to identify possibly deleterious SNPs are the programs SIFT (Sorting Intolerant From Tolerant, <http://sift.jcvi.org/>), PolyPhen and the recently developed PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>) and SNAP (Screening for Non-acceptable Polymorphisms, <http://cubic.bioc.columbia.edu/services/SNAP/>) [Adzhubei, et al., 2010; Kumar, et al., 2009; Ramensky, et al., 2002]. Programs such as NNSPLICE can be used to assess mutations possibly affecting splicing [Reese, et al., 1997]. All programs are only partly reliable in

1  
2 predicting functional SNP effects and can therefore not be used to prove the deleterious effect  
3  
4 of a SNP. However, they are good tools to supplement the other strategies and methods  
5  
6 mentioned above and help to prioritize novel variants detected by exome sequencing for  
7  
8 further analysis.  
9

Deleted: proof

10  
11 *Next-Generation-Sequencing of large candidate gene collections and combination of*  
12  
13 *mapping and sequencing approaches*  
14  
15

Formatted: Font: Italic

Formatted: Font: Bold, Italic

16  
17  
18 These “filtering” approaches rely on selecting only the functionally or positionally relevant  
19  
20 parts of the exome for NGS. The approaches are conceptually similar to the traditional  
21  
22 “candidate gene” or “linkage” approaches but NGS has the major advantage that massive  
23  
24 amounts of candidate genes can be sequenced at once. The strategy requires selective  
25  
26 enrichment of the DNA regions of interest, e.g. by using custom enrichment arrays. Due to  
27  
28 the reduction of the amount of DNA which needs to be sequenced, this strategy does not only  
29  
30 reduce the number of candidate variants but also allows to achieve higher sequencing  
31  
32 coverage, i.e. more reads of every sequenced base. Such candidate gene approaches have  
33  
34 already been successfully applied to identify causative mutations in human complex 1-  
35  
36 deficiency and autosomal recessive retinal-renal ciliopathy [Calvo, et al., 2010; Otto, et al.,  
37  
38 2010]. In the latter very high locus heterogeneity necessitated the ability to identify disease  
39  
40 genes in single individuals or families. The selection of ~13.000 ciliopathy “candidate exons”  
41  
42 allowed to reduce the number of exons which needed to be sequenced by ~15-fold compared  
43  
44 to the whole exome [Otto, et al., 2010]. In addition, variants were filtered for those which  
45  
46 mapped to homozygous stretches in the individual genome identified by the use genotyping  
47  
48 arrays which facilitates the detection of truly homozygous but not of compound heterozygous  
49  
50 mutations. Further successful examples of combined mapping/sequencing approaches are the  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 [selection of X-chromosomal genes in forms of X-linked mental retardation and identity by](#)  
3  
4 [descent filtering in autosomal recessive hyperphosphatasia mental retardation syndrome](#)  
5  
6 [\(MIM 239300\) caused by compound heterozygous mutations in the PIGV gene \(MIM](#)  
7  
8 [610274\)](#) [Krawitz, et al., 2010].  
9

### 10 11 **Problems of novel genomic methods in monogenic disease gene identification**

12  
13  
14  
15  
16 One has to discern between technical and genetical/biological problems. Technical problems  
17  
18 are e.g. inhomogeneous coverage of the target region. This problem will lead to a widely  
19  
20 varying number of reads covering different regions of the exome which causes – in  
21  
22 conjunction with unequal sequencing quality – the detection of false positive and false  
23  
24 negative variants. False positive variants can be eliminated by Sanger-resequencing. In  
25  
26 contrast, false negative variants are potentially a big problem because if the filtering strategy  
27  
28 relies on the detection of mutations in the same gene in different affected individuals these  
29  
30 might be missed. However, these problems will be alleviated by technical progress within the  
31  
32 next years. The main genetical problem is locus heterogeneity. Locus heterogeneity is the  
33  
34 presence of multiple distinct disease genes/disease loci for one clinical phenotype. Locus  
35  
36 heterogeneity might complicate the application of the “genetic filters” already described.  
37  
38 Some neurological disorders e.g. CMT (>40 loci), hereditary ataxia and hereditary spastic  
39  
40 paraplegia show a high degree of locus heterogeneity. The main biological problem is  
41  
42 proving causality of a mutation. Ultimately, causality has to be shown using cell- and animal  
43  
44 model systems.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Next-Generation-Sequencing in the diagnosis of monogenic diseases for which the causative genes are known

There are two main arguments against using NGS in genetic diagnostics. The first one is that NGS of the whole exome or genome is still expensive compared to Sanger-sequencing of a few known disease causing genes for every condition. This problem will be solved within the near future and one can predict that NGS will be cheaper than Sanger-sequencing of a number of possibly mutated genes. The second problem is an ethical one. NGS of whole exomes/genomes will not only discover the causative mutation for the investigated disease but also mutations causing other monogenic disorders and genetic risk factors for complex disorders. The first example for the former has been the identification of mutations causing primary ciliary dyskinesia in some patients who were primarily investigated because of Miller-syndrome [Ng, et al., 2010b]. Genetic risk factor determination for complex disorders e.g. *APOE*-genotype, the most important risk factor for complex AD, is a byproduct of exome/genome sequencing. A solution to this problem might be to discard all sequence-data outside the genes required for diagnosis before the analysis of the NGS-data. That genetic diagnostics using NGS are feasible has been shown very recently by the identification of a mutation in the *SH3TC2* gene (MIM608206) in a family with CMT [Lupski, et al., 2010].

### Summary and outlook

Undoubtedly, novel genomic techniques will have a large impact on monogenic disease gene identification. Most importantly, they will enable us to tackle monogenic disorders not amenable to classic linkage-based positional cloning due to insufficient family size and reveal how many seemingly sporadic disorders are in fact monogenic. Duplication/deletion analysis

1  
2 using genotyping arrays is already used in routine genetic diagnostics and NGS will most  
3  
4 likely become the method of choice for diagnostic point mutation identification within in the  
5  
6 next years. The next big step in the investigation of genetic disorders will presumably be the  
7  
8 integration of data obtained by the analysis of noncoding RNAs and epigenetic marks.

Deleted: small

9  
10 Noncoding RNAs are transcripts that do not code for proteins and are involved in gene  
11  
12 regulation. Epigenetic marks like DNA-methylation and Histone-modifications also regulate  
13  
14 gene activity and are to a certain degree passed on from generation to generation. However,  
15  
16 epigenetic modifications are to a much higher degree influenced by the environment than  
17  
18 DNA itself is. Therefore environmental data will have to be taken into account as well. In  
19  
20 addition, a recent publication demonstrates that NGS can also be applied to identify the  
21  
22 mutations underlying linkage peaks for complex traits, at least if the linkage signal is mainly  
23  
24 derived from a few large families in the cohort [Bowden, et al., 2010]. One can easily predict  
25  
26 that the rapidly declining prices of NGS will lead to a surge in it's application to monogenic  
27  
28 and complex disorders.

Deleted: Small

### 31 Acknowledgements

32  
33  
34  
35 GK is a member of the Cluster of Excellence "Inflammation at Interfaces" of the Christian-  
36  
37 Albrechts University Kiel. Work elucidating the genetic causes of monogenic disorders has  
38  
39 been supported by grants to GK of the Deutsche Forschungsgemeinschaft (DFG), the  
40  
41 "Innovative Medizinische Forschungsförderung" of the University of Münster, the  
42  
43 "Intramurale Forschungsförderung" of the University of Kiel, the Heinrich-Hertz-Foundation  
44  
45 and the "Association Francaise contre les Myopathies".  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 Figure legends:  
3

4  
5 Figure 1: Example pedigrees of Mendelian disorders suited for analysis by novel genomic  
6 techniques. Circle – Female, Square – Male, Filled symbol – Affected.  
7  
8

9  
10 Figure 2: Analysis strategy for the analysis of monogenic disorders using novel genomic  
11 techniques. Ovals – Techniques and strategies, Rectangles – Results. The figure does not  
12 encompass filtering by a priori selection of large numbers of candidate genes or by the  
13 combination of mapping and sequencing approaches because these strategies act prior to the  
14 application of genotyping arrays and exome sequencing.  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

## References

- 1  
2  
3  
4  
5  
6  
7  
8 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS,  
9 Sunyaev SR. 2010. A method and server for predicting damaging missense mutations.  
10 Nat Methods 7:248-9.  
11  
12  
13 Bier FF, von Nickisch-Rosenegk M, Ehrentreich-Forster E, Reiss E, Henkel J, Strehlow R,  
14 Andresen D. 2008. DNA microarrays. Adv Biochem Eng Biotechnol 109:433-53.  
15  
16  
17 Biesecker LG. 2010. Exome sequencing makes medical genomics a reality. Nat Genet 42:13-  
18 4.  
19  
20  
21 Bilguvar K, Ozturk AK, Louvi A, Kwan KY, Choi M, Tatli B, Yalnizoglu D, Tuysuz B,  
22 Caglayan AO, Gokben S and others. 2010. Whole-exome sequencing identifies  
23 recessive WDR62 mutations in severe brain malformations. Nature 467:207-10.  
24  
25  
26  
27 Botstein D, Risch N. 2003. Discovering genotypes underlying human phenotypes: past  
28 successes for mendelian disease, future approaches for complex disease. Nat Genet 33  
29 Suppl:228-37.  
30  
31  
32  
33 Bowden DW, An SS, Palmer ND, Brown WM, Norris JM, Haffner SM, Hawkins GA, Guo  
34 X, Rotter JI, Chen YD and others. 2010. Molecular basis of a linkage peak: Exome  
35 sequencing and family-based analysis identifies a rare genetic variant in the ADIPOQ  
36 gene in the IRAS Family Study. Hum Mol Genet e-pub ahead of print.  
37  
38  
39  
40  
41 Byun M, Abhyankar A, Lelarge V, Plancoulaine S, Palanduz A, Telhan L, Boisson B, Picard  
42 C, Dewell S, Zhao C and others. 2010. Whole-exome sequencing-based discovery of  
43 STIM1 deficiency in a child with fatal classic Kaposi sarcoma. J Exp Med.  
44  
45  
46  
47 Calvo SE, Tucker EJ, Compton AG, Kirby DM, Crawford G, Burt NP, Rivas M, Guiducci  
48 C, Bruno DL, Goldberger OA and others. 2010. High-throughput, pooled sequencing  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



- 1  
2 identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat*  
3  
4 *Genet* 42:851-8.  
5  
6 Check Hayden E. 2009. Genomics shifts focus to rare diseases. *Nature* 461:458.  
7  
8 Collins FS. 1995. Positional cloning moves from perditional to traditional. *Nat Genet* 9:347-  
9  
10 50.  
11  
12 Deng H, Le W, Jankovic J. 2007. Genetics of essential tremor. *Brain* 130:1456-64.  
13  
14 Dufva M. 2009. Introduction to microarray technology. *Methods Mol Biol* 529:1-22.  
15  
16 Ertekin-Taner N. 2007. Genetics of Alzheimer's disease: a centennial review. *Neurol Clin*  
17  
18 25:611-67.  
19  
20 Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Steehouwer M,  
21  
22 van Reeuwijk J, Kant SG and others. 2010. Exome sequencing identifies WDR35  
23  
24 variants involved in Sensenbrenner syndrome. *Am J Hum Genet* 87:418-23.  
25  
26 Hedges D, Burges D, Powell E, Almonte C, Huang J, Young S, Boese B, Schmidt M,  
27  
28 Pericak-Vance MA, Martin E and others. 2009. Exome sequencing of a  
29  
30 multigenerational human pedigree. *PLoS One* 4:e8232.  
31  
32 Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, Muhle H, de Kovel C,  
33  
34 Baker C, von Spiczak S and others. 2009. 15q13.3 microdeletions increase risk of  
35  
36 idiopathic generalized epilepsy. *Nat Genet* 41:160-2.  
37  
38 Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de  
39  
40 Reuver R, Wieskamp N, Mortier G and others. 2010. De novo mutations of SETBP1  
41  
42 cause Schinzel-Giedion syndrome. *Nat Genet* 42:483-5.  
43  
44 Komura D, Shen F, Ishikawa S, Fitch KR, Chen W, Zhang J, Liu G, Ihara S, Nakamura H,  
45  
46 Hurles ME and others. 2006. Genome-wide detection of human copy number  
47  
48 variations using high-density DNA oligonucleotide arrays. *Genome Res* 16:1575-84.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2 Krawitz PM, Schweiger MR, Rodelsperger C, Marcelis C, Kolsch U, Meisel C, Stephani F,  
3  
4 Kinoshita T, Murakami Y, Bauer S and others. 2010. Identity-by-descent filtering of  
5  
6 exome sequence data identifies PIGV mutations in hyperphosphatasia mental  
7  
8 retardation syndrome. *Nat Genet* 42:827-9.  
9
- 10 Kuhlenbäumer G, Young P, Hünermund G, Ringelstein B, Stögbauer F. 2002. Clinical  
11  
12 features and molecular genetics of hereditary peripheral neuropathies. *J Neurol*  
13  
14 249:1629-1650.  
15
- 16 Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous  
17  
18 variants on protein function using the SIFT algorithm. *Nat Protoc* 4:1073-81.  
19
- 20 Lalonde E, Albrecht S, Ha KC, Jacob K, Bolduc N, Polychronakos C, Dechelotte P,  
21  
22 Majewski J, Jabado N. 2010. Unexpected allelic heterogeneity and spectrum of  
23  
24 mutations in Fowler syndrome revealed by next-generation exome sequencing. *Hum*  
25  
26 *Mutat* 31:918-23.  
27
- 28 Lesage S, Durr A, Tazir M, Lohmann E, Leutenegger AL, Janin S, Pollak P, Brice A. 2006.  
29  
30 LRRK2 G2019S as a cause of Parkinson's disease in North African Arabs. *N Engl J*  
31  
32 *Med* 354:422-3.  
33
- 34 Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge  
35  
36 M, Dinh H, Jing C, Wheeler DA and others. 2010. Whole-genome sequencing in a  
37  
38 patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 362:1181-91.  
39
- 40 Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure  
41  
42 J, Turner DJ. 2010. Target-enrichment strategies for next-generation sequencing. *Nat*  
43  
44 *Methods* 7:111-8.  
45
- 46 Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. *Trends*  
47  
48 *Genet* 24:133-41.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2 Mooney S. 2005. Bioinformatics approaches and resources for single nucleotide  
3 polymorphism functional analysis. *Brief Bioinform* 6:44-56.  
4  
5  
6 Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans.  
7  
8 *Genetics* 156:297-304.  
9  
10 Ng PC, Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein  
11 function. *Annu Rev Genomics Hum Genet* 7:61-80.  
12  
13 Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, Li K, Axelrod N, Busam DA,  
14 Strausberg RL, Venter JC. 2008. Genetic variation in an individual human exome.  
15  
16 *PLoS Genet* 4:e1000160.  
17  
18 Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck  
19 AE, Tabor HK, Cooper GM, Mefford HC and others. 2010a. Exome sequencing  
20 identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* e-pub ahead of  
21 print.  
22  
23 Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT,  
24 Jabs EW, Nickerson DA and others. 2010b. Exome sequencing identifies the cause of  
25 a mendelian disorder. *Nat Genet* 42:30-5.  
26  
27 Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M,  
28 Bhattacharjee A, Eichler EE and others. 2009. Targeted capture and massively  
29 parallel sequencing of 12 human exomes. *Nature* 461:272-6.  
30  
31 Ott J. 1991. *Analysis of human genetic Linkage*. Baltimore and London: The John Hopkins  
32 University Press.  
33  
34 Otto EA, Hurd TW, Airik R, Chaki M, Zhou W, Stoetzel C, Patil SB, Levy S, Ghosh AK,  
35 Murga-Zamalloa CA and others. 2010. Candidate exome capture identifies mutation  
36 of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nat Genet* 42:840-50.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2 Ozelius LJ, Senthil G, Saunders-Pullman R, Ohmann E, Deligtisch A, Tagliati M, Hunt AL,  
3  
4 Klein C, Henick B, Hailpern SM and others. 2006. LRRK2 G2019S as a cause of  
5  
6 Parkinson's disease in Ashkenazi Jews. *N Engl J Med* 354:424-5.  
7  
8 Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, Haden K, Li J, Shaw CA,  
9  
10 Belmont J and others. 2006. High-resolution genomic profiling of chromosomal  
11  
12 aberrations using Infinium whole-genome genotyping. *Genome Res* 16:1136-48.  
13  
14 Peltonen L, Perola M, Naukkarinen J, Palotie A. 2006. Lessons from studying monogenic  
15  
16 disease for common disease. *Hum Mol Genet* 15 Spec No 1:R67-74.  
17  
18 Ramensky V, Bork P, Sunyaev S. 2002. Human non-synonymous SNPs: server and survey.  
19  
20 *Nucleic Acids Res* 30:3894-900.  
21  
22 Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie.  
23  
24 *J Comput Biol* 4:311-23.  
25  
26 Rios J, Stein E, Shendure J, Hobbs HH, Cohen JC. 2010. Identification by whole-genome  
27  
28 resequencing of gene defect responsible for severe hypercholesterolemia. *Hum Mol*  
29  
30 *Genet*.  
31  
32 Saitou H, Kato M, Mizuguchi T, Hamada K, Osaka H, Tohyama J, Uruno K, Kumada S,  
33  
34 Nishiyama K, Nishimura A and others. 2008. De novo mutations in the gene encoding  
35  
36 STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. *Nat Genet*  
37  
38 40:782-8.  
39  
40 Shinawi M, Schaaf CP, Bhatt SS, Xia Z, Patel A, Cheung SW, Lanpher B, Nagl S, Herding  
41  
42 HS, Nevinny-Stickel C and others. 2009. A small recurrent deletion within 15q13.3 is  
43  
44 associated with a range of neurodevelopmental phenotypes. *Nat Genet* 41:1269-71.  
45  
46 Summerer D. 2009. Enabling technologies of genomic-scale sequence enrichment for  
47  
48 targeted high-throughput sequencing. *Genomics* 94:363-8.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2 Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, O'Meara S, Latimer C,  
3  
4 Dicks E, Menzies A and others. 2009. A systematic, large-scale resequencing screen  
5  
6 of X-chromosome coding exons in mental retardation. *Nat Genet* 41:535-43.  
7  
8 Tucker T, Marra M, Friedman JM. 2009. Massively parallel sequencing: the next big thing in  
9  
10 genetic medicine. *Am J Hum Genet* 85:142-54.  
11  
12 Turner EH, Ng SB, Nickerson DA, Shendure J. 2009. Methods for genomic partitioning.  
13  
14 *Annu Rev Genomics Hum Genet* 10:263-84.  
15  
16 Winkelmann J, Müller-Myhsok B. 2008. Genetics of restless legs syndrome: a burning urge  
17  
18 to move. *Neurology* 70:664-5.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

Figure 1 Example pedigrees

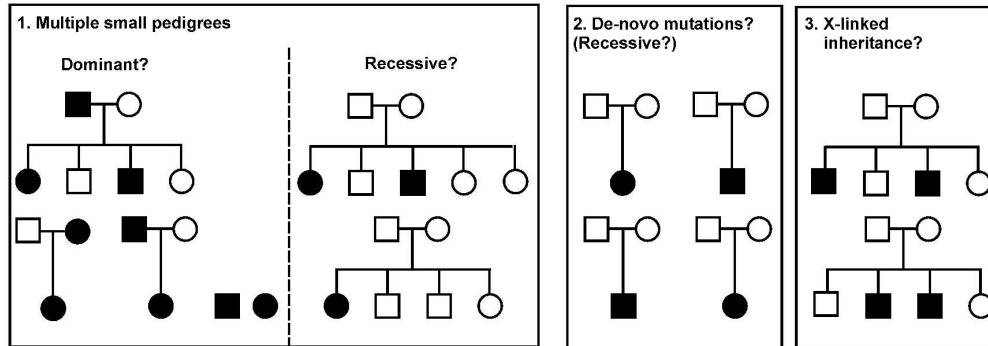


Figure 1: Example pedigrees of Mendelian disorders suited for analysis by novel genomic techniques. Circle – Female, Square – Male, Filled symbol – Affected.

Peer Review

Figure 2 Analysis strategy

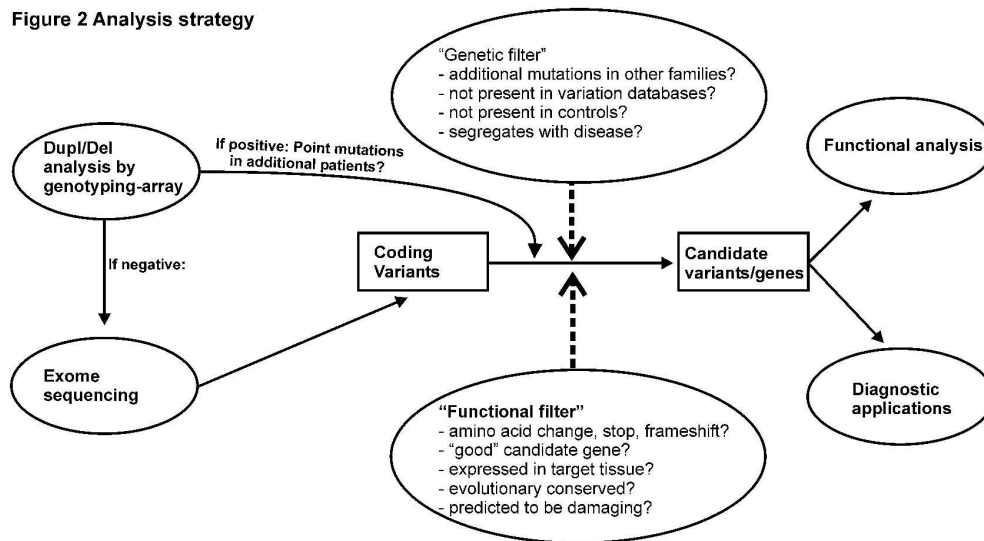


Figure 2: Analysis strategy for the analysis of monogenic disorders using novel genomic techniques. Ovals – Techniques and strategies, Rectangles – Results. The figure does not encompass filtering by a priori selection of large numbers of candidate genes or by the combination of mapping and sequencing approaches because these strategies act prior to the application of genotyping arrays and exome sequencing.