ckj

OXFORD

Leading European Nephrology

Clinical Kidney Journal, 2017, vol. 10, no. 5, 581–585

doi: 10.1093/ckj/sfx075
Advance Access Publication Date: 18 July 2017
Editorial Comment

EDITORIAL COMMENT

Potential and pitfalls in the genetic diagnosis of kidney diseases

Anne Kesselheim[1], Emma Ashton[2,3] and Detlef Bockenhauer[1,2]

[1]Centre for Nephrology, University College London, London, UK, [2]Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK and [3]North East Thames Regional Genetics Service, Molecular Genetics, London, UK

Correspondence and offprint requests to: Detlef Bockenhauer; E-mail: d.bockenhauer@ucl.ac.uk

Abstract

Next-generation sequencing has dramatically decreased the cost of gene sequencing, facilitating the simultaneous analysis of multiple genes at the same time; obtaining a genetic result for an individual patient has become much easier. The article by Ars and Torra in this issue of the *Clinical Kidney Journal* provides examples of the ever-increasing ability to understand a given patient's disease on the molecular level, so that in some cases not only the causative variants in a disease gene are identified, but also potential modifiers in other genes. Yet, with increased sequencing, a large number of variants are discovered that are difficult to interpret. These so-called 'variants of uncertain significance' raise important questions: when and how can pathogenicity be clearly attributed? This is of critical importance, as there are potentially serious consequences attached: decisions about various forms of treatment and even about life and death, such as termination of pregnancy, may hinge on the answer to these questions. Geneticists, thus, need to use the utmost care in the interpretation of identified variants and clinicians must be aware of this problem. We here discuss the potential of genetics to facilitate personalized treatment, but also the pitfalls and how to deal with them.

Key words: exome, genome, inherited renal disease, mutation analysis

Introduction

Massively parallel sequencing, also called 'next-generation sequencing' (NGS), is revolutionizing the way we perform genetic testing [1]. Whereas previously, genetic testing typically assessed one gene at a time, which had to be laboriously amplified in pieces of 500–800 base pairs, prior to submitting each piece to traditional (Sanger) sequencing, we now can assess as much genetic information as desired in one single process. This can range from a panel of genes specific for the patient's condition, to all coding regions of genes (whole exome) to the entire genetic information (whole genome). Panels are especially helpful in diseases with genetic heterogeneity, such as Bartter syndrome or steroid-resistant nephrotic syndrome [2, 3]. Whole exome (WES) or genome sequencing (WGS) can be especially helpful in cases where a genetic cause is suspected, but a specific clinical diagnosis cannot be established. There are numerous examples for this also in patients with kidney diseases, including the discovery of previously unrecognized disease genes and disorders through WES and WGS [4–7]. The majority of these discoveries have been achieved with WES, but, increasingly, there are now also reports of WGS discovering non-coding disease-causing mutations. For example, our own group recently described a promoter mutation underlying a previously unrecognized disorder consisting of hyperinsulinism and polycystic kidney disease [8]. In addition, WGS can sometimes overcome technical difficulties posed by the

Received: 8 June 2017. **Editorial decision:** 13 June 2017

presence of highly homologous pseudogenes, for instance in the genetic diagnosis of autosomal dominant polycystic kidney disease (ADPKD) [9]. These examples highlight the advantages of such comprehensive genomic analysis, helping to identify difficult-to-diagnose diseases more efficiently and lowering the risk of misdiagnosis by identifying the primary underlying genetic problem. Yet, it may go even beyond this. Sequencing of multiple genes may also help to explain the often perplexing variability of genetic diseases, even within families, where affected members all carry the same mutation. By identifying not only the causative mutation in the primary disease gene, but also potentially modifying variants in other related genes (oligogenic inheritance) the variability may be at least partially explained. Modification of the phenotype by additional variants in other 'secondary' disease genes has become a leading hypothesis for the understanding of the phenotypic variability in ciliopathies, for instance [10].

In this way, NGS may be able to facilitate personalized treatment by enabling more detailed genetic counselling and prognoses. The hope is that it will also inform clinical management, by identifying pathways contributing to the clinical phenotype, which may be amenable to treatment. In some cases, genetic testing may actually correct the clinical diagnosis with clear management implications. Ars and Torra report on a man with isolated proteinuria, subsequently diagnosed with Fabry disease [11]. In another report, patients referred with a clinical diagnosis of Bartter syndrome were found to actually suffer from congenital chloride diarrhoea [12]. In our own experience, patients referred for genetic testing with a clinical diagnosis of 'idiopathic hypokalaemia' or 'Dent disease' were genetically diagnosed with distal renal tubular acidosis, enabling appropriate treatment with alkali supplementation (E. Ashton *et al.*, manuscript submitted for publication). Yet, as always, new technologies provide not only promise, but also problems. With NGS, it is mainly the large number of variants identified and the uncertainty of their interpretation. As the promises of NGS are elegantly discussed in the article by Ars and Torra [11], in this editorial we will focus on the potential problems.

The more you sequence, the more you find

Each human genome contains ~4–5 million variants from the reference genome [13]! Five million! Imagine the scale of the problem a geneticist faces when attributing relevance to them! While WGS is currently still the exception, the ever-decreasing costs are likely to establish it as a (if not 'the') routine test in the near future. But even if we perform 'only' WES, concentrating on the 2–3% of the genome that contains genes, we still find thousands of variants in each individual. In a study from the 1000 Genomes Consortium, each genome contained 149–182 protein-truncating variants (such as nonsense mutations), 10 000–12 000 peptide-sequence-altering variants (missense mutations) and ~500 000 variants in known regulatory regions (i.e. non-coding regions, such as promoters, as well as 5′ and 3′ untranslated regions) [13]. On average 24–30 recognized disease-causing mutations are found in each of us [13]! So, how can we make sense of this large number of variants?

Databases

Key to the first interpretation of identified variants is the comparison with databases that compile identified variants. These include the gnomad browser (www.broadinstitute.org), Clinvar (https://www.ncbi.nlm.nih.gov/clinvar/) HGMD (http://www.hgmd.cf.ac.uk/ac/index.php) and Varsome (https://varsome.com/). Increasingly, laboratories also maintain their own databases, which is important to compile variants identified in the population served, but which may also be related to the artefacts arising due to either the sequencing technology or the subsequent assembly of the sequencing products [14]. By running identified variants against these databases, a large number (>99% for whole genomes) can be excluded, mainly based on the frequency of a given variant in the population. Variants found in >1% of the population are usually deemed benign, i.e. of no direct clinical relevance. On the other hand, variants listed in databases such as Clinvar, which contain confirmed mutations, would immediately be flagged as potentially disease-causing in the patient.

However, these databases are not perfect and there are several potential pitfalls:

- The presence of specific variants obviously varies with ethnicity. Of critical importance is, therefore, to compare the identified variants in a given patient against an ethnically matched dataset. The frequency of a given variant in a database may thus attest more to the predominant ethnicity within the database than its potential pathogenicity.
- Mutations published as pathogenic may in fact be benign variants. This is not unusual. If a researcher identifies a new disease gene in a cohort of patients with a given clinical phenotype, then some of these patients may have bona fide pathogenic mutations, whereas others have benign polymorphisms. If information about the frequency of these variants is not available at the time of publications, these benign polymorphisms may be erroneously reported as pathogenic. There are several examples of such 'mutations', reported even in the most prestigious journals, which were subsequently deemed benign, based on the frequency of these variants in the relevant population [15–18].
- Conversely, polymorphisms can sometimes be pathogenic. A striking example is the Arg229Gln variant in NPHS2 (c.686G>A; rs61747728), which is found in 1–2% of the Caucasian population, but can be disease-causing when present 'in trans' with specific other mutations [19]. Moreover, polymorphisms may still have a modifier effect on an underlying disorder not directly connected to this gene. For instance, variants in *NOS3* and *HBEGF* have been proposed as modifiers in ADPKD and CFHR5 nephropathy, respectively [20, 21].
- As discussed above, we all harbour numerous pathogenic mutations. Various reports suggest that each of us carries at least 100 loss-of-function mutations [22]. The majority of the mutations are heterozygous and affect recessive disease genes, but increasingly, individuals are found with recessive mutations, considered pathogenic and penetrant, yet who have no evidence of being affected, a phenomenon labelled as 'genetic resilience' [23]. Careful correlation of identified variants with the clinical phenotype is, thus, of critical importance.

Moreover, while careful comparison with databases can dramatically reduce the number of potentially disease-causing variants, there typically remains a large number of variants for which there is insufficient information available, the so-called 'variants of uncertain significance' (VUS). It is these VUS that create the most headaches in interpretation and a variety of strategies is employed to assess potential pathogenicity. Guidelines have been proposed to increase the accuracy of such assessments, which rely mainly on the following strategies [18, 24].

Predictions

An increasing number of mostly web-based tools, such as Mutation taster, SIFT or Polyphen2, are available for the

prediction of pathogenicity of identified variants. Different algorithms are used for these predictions ranging from evolutionary conservation of the affected amino acid to Hidden Markov Models assessing relationships between protein domains to determine the likelihood of a variant affecting protein function [25–27]. However, when tested independently, these tools have variable accuracy, partially dependent on the underlying gene and nature of mutation [28, 29]. Most importantly, these tools only provide a likelihood score of pathogenicity, yet no certainty. This is important to remember, when genetic analysis is needed to make important clinical management decisions.

Another important deficiency of these tools is that by design they primarily predict loss-of-function, yet some, typically dominant diseases are caused by gain-of-function. Consider the example reported by Ars and Torra of a 9-year-old girl with nephrotic syndrome, in whom a heterozygous variant in *TRPC6* was identified [11]. They speculate on pathogenicity, as the prediction algorithms all indicated a high likelihood for this. *TRPC6* encodes a calcium channel important for podocyte function and dominant mutations in this gene cause familial focal segmental glomerulosclerosis (FSGS) type 2 [30]. Yet, these are typically gain-of-function mutations, which increase calcium currents through the channel, whereas the effect of heterozygous loss-of-function mutations is controversial [31]. Indeed, suppression of TRPC6-mediated currents is considered a potential treatment strategy for podocyte injury [32]. Thus, the predictions of pathogenicity by the various tools can be misleading in such a case. This demonstrates the fundamental importance of an understanding of the underlying disease mechanism when considering the potential pathogenicity of an identified variant in a given gene.

Segregation

The same example also highlights another strategy to assess pathogenicity of a given variant, namely, whether it segregates with the disease in the family. In the case of the 9-year-old girl with FSGS, three other family members were found to carry the same variant, yet had no apparent clinical problem [11]. While one could speculate about incomplete penetrance, the absence of segregation in this family strongly argues against pathogenicity.

Conversely, incomplete penetrance can be a major problem in some disorders, such as atypical haemolytic uremic syndrome (aHUS). Heterozygous mutations in various complement-related genes have been identified to cause susceptibility to aHUS, but additional trigger(s) appear necessary [33]. Thus, it is not unusual when assessing families with aHUS to find unaffected members who carry the putative underlying variant, and using segregation of the variant with the disease for proof of pathogenicity is therefore of limited value in such disorders.

Functional studies

Functional studies are typically time and work intensive and it is of course impossible to perform these for thousands of variants in each individual. Yet, if a single candidate variant is identified as a potential cause, this can be considered. For instance, if a splice variant of unknown pathogenicity is identified, analysis of mRNA can help assess the impact of this variant on the mRNA level [34]. Moreover, variants in ion channels can easily be assayed for their functional impact by expression of the mutant channels in experimental cells, as we have done, for instance, for variants identified in patients with EAST syndrome [35–37]. This is especially important when the type of functional impact (loss- versus gain-of-function) is important, as discussed above for TRPC6. Alternatively, identified variants can be expressed in animal models, such as mice, zebrafish or *Drosophila*, to better assess their functional relevance [38, 39]. Yet, most diagnostic genetic labs are not set up to perform functional studies and this is usually performed by research laboratories.

Ethical issues

When a large amount of genetic information is analysed, variants identified may be relevant for other diseases rather than the symptoms or diagnosis that prompted the genetic investigation. These so-called incidental findings can throw up serious ethical issues. The best-known examples are variants in cancer predisposition genes, such as *BRCA1* or *VHL* [40].

The topic becomes particularly sensitive in paediatric medicine, but is far too complex to be discussed in this editorial. The interested reader is instead referred to other detailed reviews and consensus statements on this topic [41–47].

Conclusions

NGS is revolutionizing genetic diagnosis by greatly facilitating analysis of known disease genes and discovery of previously unrecognized ones. It holds the promise of enabling personalized medicine. However, the large amount of variants identified also creates serious problems with respect to over- or misinterpretation. Only with the careful collection and annotation of more and more sequencing data of people from all ethnic backgrounds and with various diseases will we be able to better understand the significance of specific variants and put them into the context of the individual patient.

Funding

This work was supported by a grant from the JF Moorhead Trust, by the European Union, FP7 [grant agreement 2012-305608 'European Consortium for High-Throughput Research in Rare Kidney Diseases (EURenOmics)'] and by Kids Kidney Research.

Conflict of interest statement

None declared.

References

1. Bockenhauer D, Medlar AJ, Ashton E *et al*. Genetic testing in renal disease. *Pediatr Nephrol* 2012; 27: 873–883
2. Sadowski CE, Lovric S, Ashraf S *et al*. A single-gene cause in 29.5% of cases of steroid-resistant nephrotic syndrome. *J Am Soc Nephrol* 2015; 26: 1279–1289
3. Ashton E, Legrand A, Benoit V *et al*. Simultaneous sequencing of 37 genes identifies likely causative mutations in 70% of children with renal tubulopathies. 2017. Manuscript submitted for publication

4. Nicolaou N, Renkema KY, Bongers EM *et al*. Genetic, environmental, and epigenetic factors involved in CAKUT. *Nat Rev Nephrol* 2015; 11: 720–731
5. Ashraf S, Gee HY, Woerner S *et al*. ADCK4 mutations promote steroid-resistant nephrotic syndrome through CoQ10 biosynthesis disruption. *J Clin Invest* 2013; 123: 5179–5189
6. Lin F, Bian F, Zou J *et al*. Whole exome sequencing reveals novel COL4A3 and COL4A4 mutations and resolves diagnosis in Chinese families with kidney disease. *BMC Nephrol* 2014; 15: 175
7. Braun DA, Schueler M, Halbritter J *et al*. Whole exome sequencing identifies causative mutations in the majority of consanguineous or familial cases with childhood-onset increased renal echogenicity. *Kidney Int* 2016; 89: 468–475
8. Cabezas OR, Flanagan SE, Stanescu H *et al*. Polycystic kidney disease with hyperinsulinemic hypoglycemia caused by a promoter mutation in phosphomannomutase 2. *J Am Soc Nephrol* 2017; doi: 10.1681/ASN.2016121312
9. Mallawaarachchi AC, Hort Y, Cowley MJ *et al*. Whole-genome sequencing overcomes pseudogene homology to diagnose autosomal dominant polycystic kidney disease. *Eur J Hum Genet* 2016; 24: 1584–1590
10. Lu JT, Campeau PM, Lee BH. Genotype-phenotype correlation–promiscuity in the era of next-generation sequencing. *N Engl J Med* 2014; 371: 593–596
11. Ars E, Torra R. Rare diseases, rare presentations: recognizing atypical inherited kidney disease phenotypes in the age of genomics. *Clinical Kidney Journal* 2017; 10: 586–593
12. Choi M, Scholl UI, Ji W *et al*. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 2009; 106: 19096–19101
13. Genomes Project C, Auton A, Brooks LD *et al*. A global reference for human genetic variation. *Nature* 2015; 526: 68–74
14. Lassmann T, Hayashizaki Y, Daub CO. TagDust–a program to eliminate artefacts from next generation sequencing data. *Bioinformatics* 2009; 25: 2839–2840
15. Bell CJ, Dinwiddie DL, Miller NA *et al*. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med* 2011; 3: 65ra64
16. Xue Y, Chen Y, Ayub Q *et al*. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am J Hum Genet* 2012; 91: 1022–1032
17. Norton N, Robertson PD, Rieder MJ *et al*. Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. *Circ Cardiovasc Genet* 2012; 5: 167–174
18. MacArthur DG, Manolio TA, Dimmock DP *et al*. Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014; 508: 469–476
19. Tory K, Menyhard DK, Woerner S *et al*. Mutation-dependent recessive inheritance of NPHS2-associated steroid-resistant nephrotic syndrome. *Nat Genet* 2014; 46: 299–304
20. Papagregoriou G, Erguler K, Dweep H *et al*. A miR-1207-5p binding site polymorphism abolishes regulation of HBEGF and is associated with disease severity in CFHR5 nephropathy. *PLoS One* 2012; 7: e31021
21. Mallett A, Sandford R. NOS3 as a potential modifier of ADPKD phenotypic variability: progress towards an answer. *Nephrology* 2014; 19: 733–734
22. MacArthur DG, Balasubramanian S, Frankish A *et al*. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012; 335: 823–828
23. Chen R, Shi L, Hakenberg J *et al*. Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat Biotechnol* 2016; 34: 531–538
24. Strande NT, Riggs ER, Buchanan AH *et al*. Evaluating the clinical validity of gene-disease associations: an evidence-based framework developed by the clinical genome resource. *Am J Hum Genet* 2017; 100: 895–906
25. Gough J, Karplus K, Hughey R *et al*. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* 2001; 313: 903–919
26. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 2011; 39: W29–W37
27. Radivojac P, Clark WT, Oron TR *et al*. A large-scale evaluation of computational protein function prediction. *Nat Methods* 2013; 10: 221–227
28. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat* 2011; 32: 358–368
29. Hicks S, Wheeler DA, Plon SE *et al*. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum Mutat* 2011; 32: 661–668
30. Winn MP, Conlon PJ, Lynn KL *et al*. A mutation in the TRPC6 cation channel causes familial focal segmental glomerulosclerosis. *Science* 2005; 308: 1801–1804
31. Riehle M, Buscher AK, Gohlke BO *et al*. TRPC6 G757D loss-of-function mutation associates with FSGS. *J Am Soc Nephrol* 2016; 27: 2771–2783
32. Sonneveld R, Hoenderop JG, Isidori AM *et al*. Sildenafil prevents podocyte injury via PPAR-gamma-mediated TRPC6 inhibition. *J Am Soc Nephrol* 2017; 28: 1491–1505
33. Sansbury FH, Cordell HJ, Bingham C *et al*. Factors determining penetrance in familial atypical haemolytic uraemic syndrome. *J Med Genet* 2014; 51: 756–764
34. Wang K, Zhao X, Chan S *et al*. Evidence for pathogenicity of atypical splice mutations in autosomal dominant polycystic kidney disease. *Clin J Am Soc Nephrol* 2009; 4: 442–449
35. Bockenhauer D, Feather S, Stanescu HC *et al*. Epilepsy, ataxia, sensorineural deafness, tubulopathy, and KCNJ10 mutations. *N Engl J Med* 2009; 360: 1960–1970
36. Freudenthal B, Kulaveerasingam D, Lingappa L *et al*. KCNJ10 mutations disrupt function in patients with EAST syndrome. *Nephron Physiol* 2011; 119: p40–p48
37. Parrock S, Hussain S, Issler N *et al*. KCNJ10 mutations display differential sensitivity to heteromerisation with KCNJ16. *Nephron Physiol* 2013; 123: 7–14
38. Helmstadter M, Simons M. Using drosophila nephrocytes in genetic kidney disease. *Cell Tissue Res*; doi: 10.1007/s00441-017-2606-z
39. Mahmood F, Mozere M, Zdebik AA *et al*. Generation and validation of a zebrafish model of EAST (epilepsy, ataxia, sensorineural deafness and tubulopathy) syndrome. *Dis Model Mech* 2013; 6: 652–660
40. Burke K, Clarke A. The challenge of consent in clinical genome-wide testing. *Arch Dis Child* 2016; 101: 1048–1052
41. Jarvik GP, Amendola LM, Berg JS *et al*. Return of genomic results to research participants: the floor, the ceiling, and the choices in between. *Am J Hum Genet* 2014; 94: 818–826

Clinical Kidney Journal
ckj

42. Committee on Bioethics, Committee on Genetics, The American College of Medical Genetics *et al.*, Ethical and policy issues in genetic testing and screening of children. *Pediatrics* 2013; 131: 620–622
43. Watson M. Incidental findings in clinical genomics: a clarification. *Genet Med* 2013; 15: 664–666
44. Hehir-Kwa JY, Claustres M, Hastings RJ *et al.* Towards a European consensus for reporting incidental findings during clinical NGS testing. 2015
45. Abdul-Karim R, Berkman BE, Wendler D *et al.* Disclosure of incidental findings from next-generation sequencing in pediatric genomic research. *Pediatrics* 2013; 131: 564–571
46. McGuire AL, Caulfield T, Cho MK. Research ethics and the challenge of whole-genome sequencing. *Nat Rev Genet* 2008; 9: 152–156
47. Cho MK. Understanding incidental findings in the context of genetics and genomics. *J Law Med Ethics* 2008; 36: 280–285

Clinical Kidney Journal
ckj