

# Parlez-vous VUS?

Gregory M. Cooper

HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA

Human genome sequencing is routine and will soon be a staple in research and clinical genetics. However, the promise of sequencing is often just that, with genome data routinely failing to reveal useful insights about disease in general or a person's health in particular. Nowhere is this chasm between promise and progress more evident than in the designation, "variant of uncertain significance" (VUS). Although it serves an important role, careful consideration of VUS reveals it to be a nebulous description of genomic information and its relationship to disease, symptomatic of our inability to make even crude quantitative assertions about the disease risks conferred by many genetic variants. In this perspective, I discuss the challenge of "variant interpretation" and the value of comparative and functional genomic information in meeting that challenge. Although already essential, genomic annotations will become even more important as our analytical focus widens beyond coding exons. Combined with more genotype and phenotype data, they will help facilitate more quantitative and insightful assessments of the contributions of genetic variants to disease.

Genomic sequence information can provide general knowledge about genetic contributions to phenotype (e.g., Ng et al. 2010; Cirulli et al. 2015) and benefits to the health and welfare of individual patients and families (e.g., Worthey et al. 2011; Yang et al. 2013). However, while productivity gains in genomics increasingly provide the infrastructure and raw materials needed, major analytical challenges reduce the potential benefits of genomic sequencing. In many projects and for many patients, sequencing fails to yield compelling links between variation and disease. A poignant manifestation of these limitations is the now routine discovery of variants that harbor one or more properties of disease-causal variants, but which lack enough data to support a firm conclusion about phenotypic relevance (or lack thereof). Such variants are trapped in the interpretive void between "benign" (i.e., definitively not relevant to disease) and "pathogenic" (i.e., definitively relevant to disease), and termed to be "variants of uncertain significance" (VUS) (Richards et al. 2015). Defining an appropriate decision-making and communication framework for such variants is essential: Medical and personal consequences to specific individuals and improvements to our general understanding of the human genotype-phenotype map depend on their dissemination. However, a deeper examination of the purpose and consequence of VUS as a means to communicate information is warranted: What, precisely, is meant by this phrase?

The simplest place to start is the nominal meaning of the words, beginning with "variant." There is, fortunately and not to be taken for granted, a consensus on what this term means: a sequence difference between any two homologous chromosomes, denoted via a reference assembly location and the possible sequence states, "reference" and "alternate." While pedantic to spell out, this terminology and definition are essential. They circumvent the semantic disputes arising from ambiguous frequency or disease effect requirements that plague terms like "polymorphism" or "mutation" (Condit et al. 2002). Genome-driven variant descriptions also bypass the constrictions and complications imposed by protein-, transcript-, or locus-specific frameworks (<http://www.hgvs.org/mutnomen/>). Although there remains a legacy of "pre-genome" variant definition schemes that must occa-

sionally be wrestled with, such basic genetic observations can now be described in a manner that is universal, scalable, applicable nearly anywhere within human genomes, and readily linked to oceans of genomic data.

## 'Uncertainty' and 'significance'

It is the terminal two-thirds of the phrase wherein trouble manifests, starting with "uncertain." This term is an ouroboros given the many plausible meanings of "uncertain" in this context (Biesecker et al. 2014). Further, use of "uncertain" for some variants implies that non-VUS variants are classified with "certainty." However, virtually nothing in the natural world is certain, with biological systems being among the least predictable collections of phenomena. Even robust genetic diagnoses often have uncertain implications for individual patients given the limited clinical or prognostic information for many diseases. There are also many variants with large but incompletely penetrant effects (King et al. 2003; Spurdle et al. 2012), including some variants thought to be fully penetrant but which are in reality less so, an overestimation bias from phenotype-driven ascertainment of individuals and families (Antoniou et al. 2003). Such variants would all be classified as "likely pathogenic" or "pathogenic" in standard nomenclature, and yet arguably have "uncertain" implications for any given person owing to our ignorance as to the genetic or environmental risk factors that modulate their effects. Conversely, to the extent that a "polygenic dust" (International Schizophrenia Consortium et al. 2009) model holds true for many common diseases, there exists a plethora of common variants that are bona fide functionally relevant risk factors for disease (Musunuru et al. 2010) yet would be classified as "benign" or "likely benign" in standard clinical nomenclature.

I now address "significance," a term also fraught with many distinct connotations including not only pathogenicity or clinical relevance but also deleteriousness, molecular consequence, penetrance, population-level disease risk, and other properties that could be considered "significant" to researchers, clinicians, or patients. Further, even when confined to a single, relatively well-defined property, any given level or flavor of significance can have distinct implications when viewed by different

**Corresponding author:** [gcooper@hudsonalpha.org](mailto:gcooper@hudsonalpha.org)

Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.190116.115>. Freely available online through the *Genome Research* Open Access option.

© 2015 Cooper This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

people or in different contexts. Your VUS may be my diagnosis, depending on the manner in which we use the information and the weights that we place on the consequences of false positive and false negative conclusions. “Significance,” like “actionability” (Berg et al. 2013) and related properties is, in no small measure, in the eye of the beholder.

Thus, one can argue that variants across the entire interpretive spectrum can all be comfortably described as of “uncertain significance.” However, as any biologist who has casually used “significant” within earshot of a statistician can attest, this phrase tends to suggest that a specific quantitative assertion ought to lurk somewhere in the background. In particular, significance is a measure of the extent to which random forces are a plausible explanation for the observation at hand—in our case, whether the variant is unrelated to the phenotype. And herein lies a goal for clinical and research genomicists: to replace “VUS,” and for that matter “benign,” “likely benign,” “likely pathogenic,” and “pathogenic” with quantitative assertions about the relevance of variants to phenotypes in any given individual or population. Crucially, as effect size and the extent of error around estimates of effect size result in distinct types of predictive uncertainty—many variants that meet robust definitions of “pathogenic” confer intermediate disease risks, and VUSs may turn out to be either completely irrelevant or fully penetrant—plausible distributions of risk estimates are as important as the point estimates themselves. Thus, VUSs are variants that confer plausible disease risks with some unknown, although likely multimodal, distribution in  $[0,1]$ . Risk-elevating GWAS alleles, on the other hand, strongly cluster near zero but, unlike VUSs, comfortably exclude both zero and one.

## Toward more general and quantitative variant classification

Although dissection of the terminology that we use to describe variant contributions to disease renders some insight into the challenges we face as genome-equipped geneticists, it also makes clear that the challenge lay not in terminology per se. Annotating a variant as a “VUS” is arguably no worse than saying it confers disease risk somewhere between 0 and 1, including both extremes. Both descriptions are efforts to say *something* yet respect the principle that it is better to draw a crude and imprecise, but not incorrect, conclusion than to be elegantly and precisely wrong. Indeed, it is hardly novel to suggest that quantitation is a valuable goal in this context (Plon et al. 2008). Given interpretive and data limitations, perhaps “VUS,” like democracy, is simply the worst choice except for all the other possible options.

Considering the above, improvements in the future will depend on better quantification at every step of the genotype–phenotype evaluation process, along with an overarching model in which to unify those steps. The most crucial need is more genotype/phenotype data, which we are fortunately in a position to expect. Where an agronomist sees the challenges of feeding billions of people (McCouch et al. 2013), we see many carriers of every nonlethal de novo mutation (Kong et al. 2012) and homozygotes or compound heterozygotes for loss-of-function alleles in thousands of genes (MacArthur et al. 2012). Of course, nontrivial obstacles must be overcome, particularly to craft data sharing and research protocols that truly engage, inform, protect, and respect research participants (Greely 2007; Erlich et al. 2014), but the availability of millions of genomes attached to at least some phenotypic information is realistic in the coming years.

## Genomic annotations in variant assessment

Although more, better, and broadly available human genetic data are obviously necessary, genomic annotations are and will in the future be essential. They are informative at nearly every level of human genetic inference, allowing for refined estimates of variant-level prior probabilities to enrich for and prioritize disease-causal variation; grouping of variants into genetically coherent targets within which to aggregate mutations (e.g., linking exons together into transcripts and regulatory elements to their target genes); and mechanistic understanding of pathophysiology. Put simply, they tie the room together. In fact, as we inevitably shift our focus away from exomes to genomes, annotations will be even more essential to filter, prioritize, group, and classify variants to better discriminate signal from noise in human genetics. Focusing, then, on the future of genomic annotations, there are two main areas of past success that point to future advancements: conservation-based approaches that exploit the relationship between disease and natural selection and functional approaches that exploit the relationship between disease and molecular function. These two broad and complementary categories encompass most of the information content used to evaluate genetic variants.

### Information from evolution

To the extent that many diseases are deleterious (i.e., reduce survival and reproductive success) and result from perturbations of ancient sequence-driven functions, comparative genomic identification of sites that have experienced purifying selection is a conceptually powerful approach to enrich for disease-causal variation. Such information can be rendered into discrete forms—e.g., “conserved” versus “nonconserved”—but also lends itself quite naturally to quantitative assessments that relate to both the evolutionary age of any given genomic site and the strength of selection operating on variants that affect it, which in turn correlate with the probabilities that variants are pathogenic. Evidence now overwhelmingly points to the value of comparative genomics for both large-scale studies and individual clinical interpretation. This utility became apparent within protein-altering variants more than 10 years ago (Ng and Henikoff 2001; Sunyaev et al. 2001), and information from genome-wide comparisons has proven in more recent years to be valuable in both coding and noncoding DNA (Cooper et al. 2010; Weedon et al. 2014; Amendola et al. 2015).

Excitingly, we are far from realizing the full informational potential of comparative genomics, as current methods and approaches will scale readily with more genomes and increase in effectiveness as a result. As divergence rates among neutrally evolving positions establish the null expectation in comparative genomics, this parameter strongly influences the specificity and sensitivity with which natural selection can be inferred from any given analysis. Current neutral depths of mammalian genome alignments provide useful enrichment for sites under selection, especially among positions that are perfectly or nearly perfectly conserved across the entire tree (Cooper et al. 2003; Eddy 2005). However, considerable amounts of imprecision remain, especially at the per-position level best suited to evaluate SNVs and indels, in selective inferences; observing three substitutions at a site where six are expected is consistent with both neutral and selective evolutionary histories, whereas observing 12 where 24 are expected is much more likely to result from natural selection.

Thus, as comparative genomic depths increase, sites that have experienced weaker or more recent selection will become more

visible. This improvement should, in turn, make conservation scores even more effective at discrimination of disease-causal variants. This obviously applies to those variants that are highly penetrant and deleterious, where current evolutionary depths are already helpful, but increasingly deeper evolutionary comparisons should disproportionately increase sensitivity to weaker-effect, yet nevertheless deleterious, disease-causal variants. A comprehensive catalog of primate reference genomes is an additional avenue for better discovery power. Although such alignments will always be more limited in depth than pan-mammalian comparisons, sufficient genomic diversity of primates exists to provide useful specificity (Boffelli et al. 2003), whereas the greater extent of ancestrally shared sequence-driven biology should allow for more comprehensive detection of phenotypically relevant variants (Cooper and Brown 2008; Ponting and Hardison 2011).

The specificity and sensitivity of comparative genomics can therefore be readily improved with more genomes from more species, and fortunately, hundreds of primates and thousands of mammals exist that may yet be sequenced (Genome 10K Community of Scientists 2009). While “more sequencing” is sometimes a consequence of seeing nails when one possesses a hammer, this is one area in which needs and resources align well.

### Information from molecular function

Functional annotations, while tending to correlate with measures of conservation (Margulies et al. 2007), are also essential. Beyond allowing identification of phenotypically relevant elements that may be missed by conservation-based approaches (Blow et al. 2010), functional annotations are essential to facilitate insights into molecular mechanisms and to link together elements in which mutations are more likely to give rise to similar phenotypes (e.g., exons within a gene). Fortunately, and as for comparative genomic data sets, functional genomic annotations are likely to become ever more abundant and rich. Gene and transcript catalogs have steadily improved and expanded over time (Frankish et al. 2015). Additionally, over the past several years, thousands of genome-wide maps of other molecular activities have been generated across many cell types/tissues and individuals, including binding locations of many transcription factors, chromatin modifications, and open chromatin (The ENCODE Project Consortium 2012; Roadmap Epigenomics Consortium et al. 2015); such efforts are likely to not only continue but accelerate in pace.

In addition to baseline maps of molecular activities in human cells and tissues, assessing mutational perturbations to these activities is also valuable and likely to grow in the future. Systematic assessments of mutational effects on transcriptional regulatory element activities, both *in vitro* and *in vivo*, is now straightforward and scalable given that the natural measures of regulatory functions are sequence-based (e.g., Patwardhan et al. 2009, 2012; Kwasniewski et al. 2012). Recent efforts in systematic protein mutagenesis also promise to scale, at least to levels useful for evaluating many medically relevant proteins or protein domains (e.g., Fowler et al. 2010; Starita et al. 2015). The ability to rapidly generate mutations in animal models (Jinek et al. 2012) also promises to greatly accelerate and improve functional assessment of human genetic variants.

Critically, as important to the basic mapping of biochemical activities in our genomes is better understanding of the organization and logic that relate these elements and their underlying sequences to phenotype. Features like transcription-factor binding consensus motifs hold obvious importance to evaluating func-

tional and disease impacts of variants (Weedon et al. 2014). However, higher-level information content is also crucial. Identification of looping or other three-dimensional arrangements that connect regulatory elements to target genes (Dekker et al. 2013), for example, will be essential for genome-powered gene “burden” tests in the future. Further, considering human genome and effective population sizes and concomitant tolerance for molecular “noise” (Lynch and Conery 2003; Khaitovich et al. 2004; Cooper and Brown 2008; Graur et al. 2015), most of the information in functional genomic data sets is irrelevant to disease. In fact, most is not detectably relevant even to molecular phenotypes like gene expression (Reddy et al. 2009). Although the exceptions are obviously important, common sense and the neutral theory (Kimura 1983) provide a strong “likely benign” null model applicable to both genetic variation and the molecular functions they may impact. A major challenge moving forward in this light is to identify the features and rules that differentiate biochemically active but biologically inert elements from those that matter to disease, with healthy skepticism and rigorous standards being essential to ensure a robust genetic and genomic knowledge base (MacArthur et al. 2014).

### Looking forward

Of course, the best path forward, both in terms of understanding molecular functional consequences and developing useful predictive algorithms for identifying variants that matter to phenotype, will consist of integrative measures and analyses that exploit both comparative and functional genomic data. Integration of, for example, gene structure effects, conservation levels at different phylogenetic scales, and regulatory element features is clearly more useful in variant analysis than such features alone (Kircher et al. 2014), with recent evidence pointing to such approaches being effective for clinical interpretation (Amendola et al. 2015; van der Velde et al. 2015). Future explorations of the complementarities and synergies between various types of annotations will lead to further improvements in automated computational assessment of variant effects.

Much as the previous (first) “post-genome” decade has seen major advances in knowledge about the human genotype–phenotype map, it is exciting to imagine the progress one might anticipate in the coming decade. Identifying and quantifying the links that relate evolution and molecular function to disease promises to greatly improve our understanding of human genetic variation in the laboratory and the clinic. In contrast to our current muddled descriptions of genetic variation, we will surely soon speak a better language.

### Acknowledgments

This work was supported by the National Institutes of Health (NIH): the National Human Genome Research Institute (NHGRI, grant 1UM1HG007301-02) and the National Cancer Institute (NCI, grant 1RO1CA197139-01). I thank Greg Barsh, Christopher Brown, Richard Myers, and three anonymous reviewers for comments on previous drafts.

### References

- Amendola LM, Dorschner MO, Robertson PD, Salama JS, Hart R, Shirts BH, Murray ML, Tokita MJ, Gallego CJ, Kim DS, et al. 2015. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Res* **25**: 305–315.

- Antoniou A, Pharoah PDP, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg Å, et al. 2003. Average risks of breast and ovarian cancer associated with *BRCA1* or *BRCA2* mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* **72**: 1117–1130.
- Berg JS, Amendola LM, Eng C, Van Allen E, Gray SW, Wagle N, Rehm HL, DeChene ET, Dulik MC, Hisama FM, et al. 2013. Processes and preliminary outputs for identification of actionable genes as incidental findings in genomic sequence data in the Clinical Sequencing Exploratory Research Consortium. *Genet Med* **15**: 860–867.
- Biesecker BB, Klein W, Lewis KL, Fisher TC, Wright MF, Biesecker LG, Han PK. 2014. How do research participants perceive “uncertainty” in genome sequencing? *Genet Med* **16**: 977–980.
- Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet* **42**: 806–810.
- Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, Pachter L, Rubin EM. 2003. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299**: 1391–1394.
- Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, Couthouis J, Lu YF, Wang Q, Krueger BJ, et al. 2015. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347**: 1436–1441.
- Condit CM, Achter PJ, Lauer I, Sefcovic E. 2002. The changing meanings of “mutation:” a contextualized study of public discourse. *Hum Mutat* **19**: 69–75.
- Cooper GM, Brown CD. 2008. Qualifying the relationship between sequence conservation and molecular function. *Genome Res* **18**: 201–205.
- Cooper GM, Brudno M, Green ED, Batzoglou S, Sidow A. 2003. Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Res* **13**: 813–820.
- Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, Shendure J, Nickerson DA. 2010. Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat Methods* **7**: 250–251.
- Dekker J, Marti-Renom MA, Mirny LA. 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* **14**: 390–403.
- Eddy SR. 2005. A model of the statistical power of comparative genome sequence analysis. *PLoS Biol* **3**: e10.
- The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.
- Erlich Y, Williams JB, Glazer D, Yocum K, Farahany N, Olson M, Narayanan A, Stein LD, Witkowski JA, Kain RC. 2014. Redefining genomic privacy: trust and empowerment. *PLoS Biol* **12**: e1001983.
- Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, Fields S. 2010. High-resolution mapping of protein sequence-function relationships. *Nat Methods* **7**: 741–746.
- Frankish A, Uszczyńska B, Ritchie GR, Gonzalez JM, Pervouchine D, Petryszak R, Mudge JM, Fonseca N, Brazma A, Guigo R, et al. 2015. Comparison of GENCODE and RefSeq gene annotation and the impact of reference geneset on variant effect prediction. *BMC Genomics* **16** (Suppl 8): S2.
- Genome 10K Community of Scientists. 2009. Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. *J Hered* **100**: 659–674.
- Graur D, Zheng Y, Azevedo RB. 2015. An evolutionary classification of genomic function. *Genome Biol Evol* **7**: 642–645.
- Greely HT. 2007. The uneasy ethical and legal underpinnings of large-scale genomic biobanks. *Annu Rev Genomics Hum Genet* **8**: 343–364.
- International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**: 748–752.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816–821.
- Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W, Pääbo S. 2004. A neutral model of transcriptome evolution. *PLoS Biol* **2**: E132.
- Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge [Cambridge, England]; New York.
- King MC, Marks JH, Mandell JB; The New York Breast Cancer Study Group. 2003. Breast and ovarian cancer risks due to inherited mutations in *BRCA1* and *BRCA2*. *Science* **302**: 643–646.
- Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**: 310–315.
- Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. 2012. Rate of *de novo* mutations and the importance of father’s age to disease risk. *Nature* **488**: 471–475.
- Kwasniewski JC, Mogno I, Myers CA, Corbo JC, Cohen BA. 2012. Complex effects of nucleotide variants in a mammalian *cis*-regulatory element. *Proc Natl Acad Sci* **109**: 19498–19503.
- Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* **302**: 1401–1404.
- MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, et al. 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**: 823–828.
- MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, et al. 2014. Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**: 469–476.
- Margulies EH, Cooper GM, Asimeno G, Thomas DJ, Dewey CN, Siepel A, Birney E, Keefe D, Schwartz AS, Hou M, et al. 2007. Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. *Genome Res* **17**: 760–774.
- McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, Burke JM, Charest D, Cloutier S, Cole G, et al. 2013. Agriculture: feeding the future. *Nature* **499**: 23–24.
- Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, et al. 2010. From noncoding variant to phenotype via *SORT1* at the 1p13 cholesterol locus. *Nature* **466**: 714–719.
- Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. *Genome Res* **11**: 863–874.
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, et al. 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42**: 30–35.
- Patwardhan RP, Lee C, Litvin O, Young DL, Pe’er D, Shendure J. 2009. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol* **27**: 1173–1175.
- Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM, et al. 2012. Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nat Biotechnol* **30**: 265–270.
- Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FBL, Hoogerbrugge N, Spurdle AB, Tavtigian SV, et al. 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* **29**: 1282–1291.
- Ponting CP, Hardison RC. 2011. What fraction of the human genome is functional? *Genome Res* **21**: 1769–1776.
- Reddy TE, Pauli F, Sprouse RO, Neff NE, Newberry KM, Garabedian MJ, Myers RM. 2009. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res* **19**: 2163–2171.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**: 405–423.
- Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenyk M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317–330.
- Spurdle AB, Whaley PJ, Thompson B, Feng B, Healey S, Brown MA, Pettigrew C; kConFab, Van Asperen CJ, Ausems MG, et al. 2012. *BRCA1* R1699Q variant displaying ambiguous functional abrogation confers intermediate breast and ovarian cancer risk. *J Med Genet* **49**: 525–532.
- Starita LM, Young DL, Islam M, Kitzman JO, Gullingsrud J, Hause RJ, Fowler DM, Parvin JD, Shendure J, Fields S. 2015. Massively parallel functional analysis of *BRCA1* RING domain variants. *Genetics* **200**: 413–422.
- Sunyaev S, Ramensky V, Koch I, Lathe W III, Kondrashov AS, Bork P. 2001. Prediction of deleterious human alleles. *Hum Mol Genet* **10**: 591–597.
- van der Velde KJ, Kuiper J, Thompson BA, Plazzer JP, van Valkenhoef G, de Haan M, Jongbloed JD, Wijmenga C, de Koning TJ, Abbott KM, et al. 2015. Evaluation of CADD scores in curated mismatch repair gene variants yields a model for clinical validation and prioritization. *Hum Mutat* **36**: 712–719.
- Weedon MN, Cebola I, Patch AM, Flanagan SE, De Franco E, Caswell R, Rodriguez-Seguí SA, Shaw-Smith C, Cho CH, Lango Allen H, et al. 2014. Recessive mutations in a distal *PTF1A* enhancer cause isolated pancreatic agenesis. *Nat Genet* **46**: 61–64.
- Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, et al. 2011. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med* **13**: 255–262.
- Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, et al. 2013. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* **369**: 1502–1511.