

The Role of New Sequencing Technology in Identifying Rare Mutations in New Susceptibility Genes for Cancer

Melissa C. Southey

Published online: 13 July 2013
© Springer Science + Business Media New York 2013

Abstract Massively parallel sequencing (MPS) has transformed our capacity to analyze the genome. Technology now facilitates the production of hundreds of gigabases of sequencing data from single instrument runs and is flexible to study design, allowing analyses of full genomes through a range of targeted sequencing strategies involving one to thousands of samples. The search for new cancer susceptibility genes is no longer limited by sequencing technology; theoretically, MPS can be advantageous to studies searching for genetic variation responsible for cancer predisposition across the risk spectrum. Genetically uncharacterized rare syndromes are now being unraveled at a much increased rate (including rare cancer-related syndromes), yet complex diseases such as common cancers have proven to be more challenging. MPS has revealed the complexity of the human genome, and our current study designs and bioinformatic and computational approaches need to be refined to realize the full potential of MPS for it contribute to the identification of new cancer susceptibility genes.

Keywords Massively parallel sequencing · Common cancer · Mendelian disease · Bioinformatics · Familial cancer

Introduction

Sequencing continues to be a dynamic and evolving technology. The so-called next-generation sequencing technologies have become affordable to a large number of

laboratories with increasingly diverse applications. Although cost still limits inspiration and aspiration in some settings, commercial competition has rapidly made significant advances such as whole human exome sequencing, via high-throughput exome-enrichment kits [1, 2], affordable for studies of considerable scope [3, 4, 5, 6, 7]. Constant increases in the amounts of sequencing data that can be generated from single instrument runs now place whole genome sequencing close to parity with whole exome sequencing from several perspectives [8]. Successful incorporation of barcoding and multiplexing has also impacted the cost-effectiveness of new sequencing technology and has allowed a broader number of research initiatives to use this technology in their study designs. Researchers aiming to identify rare mutations in new susceptibility genes for cancer have embraced this technology. Certainly, the application of this technology has revealed the human genome to be much more complex and variable than previous understood [9]—but has (or will) the application of this technology lead to new discoveries?

Genetically uncharacterized rare syndromes are now being unraveled at a much increased rate (including rare cancer-related syndromes [10]). These studies have been attractive to early exome sequencing endeavors as they can be successfully performed on a very small number of well-selected samples and have demanded less from bioinformatic variant filtering pipelines, which are still in their infancy [11–13]. Although intrinsically valuable in their own right, studies of Mendelian disorders have also acted as pilot studies or foundation studies for the application of this new technology to the study of the missing heritability of more complex diseases, such as common cancers. Now the technology is considered routine in a large number of research-active environments, can it play a role in identifying rare mutations in new susceptibility genes for cancer?

M. C. Southey (✉)
Genetic Epidemiology Laboratory, Department of Pathology,
The University of Melbourne, Melbourne, VIC 3010, Australia
e-mail: msouthey@unimelb.edu.au

What Is the New Technology?

Massively parallel sequencing (MPS) includes many high-throughput approaches to sequencing that share the feature of sequencing massive amounts of DNA (or RNA) in parallel. Although the detail of the chemistry used differs between platforms, all platforms work on the basis of sequencing spatially separated, clonally amplified templates (referred to as a library). MPS is also referred to as “next-generation sequencing,” but here I use the term “massively parallel sequencing” (MPS), allowing for several more and anticipated “next generations” of sequencers. At the time of writing, at least five sequencing platforms with various protocols for library preparation, sequencing chemistry, fragment read length, run times, and amount of sequencing data that can be produced in an instrument run are regarded as having established themselves in the marketplace.

The step from the (previous) gold standard Sanger sequencing (based on the chain termination method described by Sanger et al. [14]) to MPS (sequencing by synthesis) has not been another small step in the slow incremental improvements to sequencing technology—it is something like a technological revolution. Genetic research of diverse endeavors has been, or will be, substantially changed by this new technology, and activities are already under way to introduce this technology into the diagnostic/clinical arena [15–17].

To date, exome capture followed by MPS has been a much more feasible option compared with whole genome MPS for most applications in cancer genetics. This is a result of (1) the availability of commercially produced exome-enrichment products, (2) most of the currently identified disease-causing mutations in Mendelian disorders being in coding (or flanking intronic) regions of genes, thus making the approach arguably sound, (3) the availability of strategies to interpret exomic and proximal splice junction sequencing variation in terms of possible disease association [18–23], and (4) cost (exome MPS had been, until 2012, approximately ten times more expensive than genome MPS despite the cost of the front-end enrichment process/products).

Very recently there has been less compelling justification for undertaking exome capture followed by MPS rather than whole genome MPS for cancer gene discovery projects. The key factors to change in a relatively short period of time are as follows: (1) the cost differential, which is now more moderate (around threefold to fourfold); (2) computational capacities and analytical pipelines have developed to make possible the analysis of significantly larger data files; (3) there is increasing interest in interpreting and ability to interpret noncoding regions of the genome, and (4) some of the technical limitations

experienced with exome capture followed by MPS (associated with the capture process) are not encountered.

MPS platforms are not limited to the sequencing of genomes and exomes (DNA), transcriptomes, and methylomes, and other sequencing targeted strategies are emerging as innovative approaches to address current research questions.

The Search for Cancer Predisposition Genes

Over the last few decades a small collection of cancer predisposition genes have been identified by applying, at least by today’s standards, rather blunt tools. Many of these early findings were assisted by clinical observations of phenotype that facilitated the selection of rare families most likely to have dominantly inherited disease that were most informative for linkage studies. These families included (1) families selected by having thousands of polyps in the colon and a predisposition to the development of colon cancer (familial adenomatous polyposis) that led to the identification of the adenomatous polyposis coli gene (*APC*) [24–26], (2) families with multiple cases of breast and ovarian cancer that led to the identification of *BRCA1* [27, 28] and (3) families with both male and female breast cancer that made possible the identification of *BRCA2* [29, 30]. Mutations in these genes are typically rare and associated with a high risk of cancer (Fig. 1). Much more diverse clinical phenotypes associated with carrying mutations in these genes are now apparent, and there is evidence that some mutations are associated with more moderate risk [31] and/or attenuated disease [32].

As linkage studies failed to identify further reliable signals, the search for further cancer predisposition genes moved into a phase of candidate gene screening. This screening was supported by advances in Sanger sequencing that allowed a move away from manual chain-termination radioactive sequencing in slab gels to fluorescent capillary-based sequencing that allowed automated base-calling. Candidate genes were predominantly selected on the basis of having some relationship with the previously identified cancer predisposition genes (same pathway and/or function), and led to the identification of another small number of cancer susceptibility genes. Illustrative examples of this include the screening for *MUTYH* mutations (particularly in *APC* mutation negative families with a high number of somatic G:C>T:A mutations in their tumors) owing to its role in base excision repair [33, 34] and the screening for *PALB2* mutations in *BRCA1* and *BRCA2* mutation negative affected women from multiple-case breast cancer families because the *PALB2* gene product was known to interact with *BRCA2*, and biallelic mutations in *PALB2*, similar to biallelic *BRCA2*

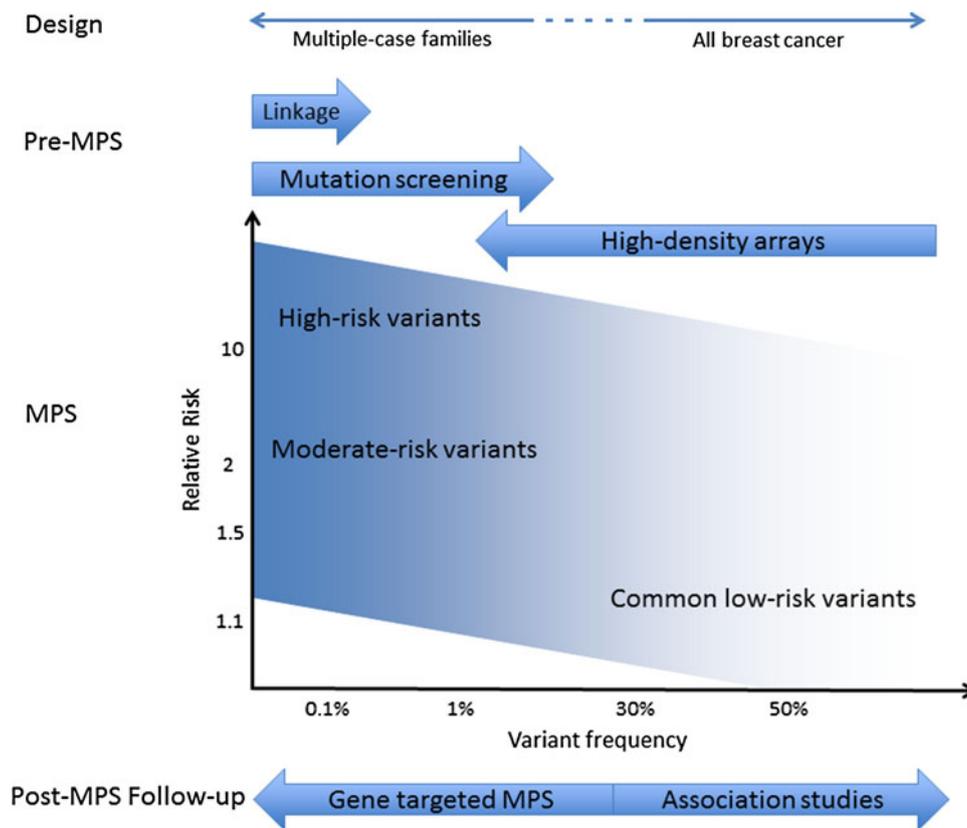


Fig. 1 The figure adapts a popular approach to illustrating the spectrum of genetic variants associated with different magnitudes of cancer risk [69]. The strength of massively parallel sequencing (MPS) to characterize genetic variants associated with cancer risk spectrum and the niches of pre-MPS and post-MPS methodology are indicated. For cancer-predisposing genetic variant discovery research, the resources of highly selected multiple-case families have been successfully used to identify high-risk genetic variants using linkage

and later candidate gene mutation screening projects. Studies of all breast cancer, typically case-control studies, have been successfully used to identify common genetic variants associated with very small increases in risk via high-density SNP array genotyping. MPS has the potential to identify genetic variation associated with a broad spectrum of cancer risk (depending on the study design) and usually requires follow-up validation initiatives that can also be MPS-based

mutations, caused Fanconi anemia [35–37]. Mutations in these genes are also generally very rare and are thought to be associated with, on average, more moderate cancer risk [35] and in some instances more moderate phenotype (e.g. biallelic *MUTYH* mutation associated colon cancer) [33, 38]. Further developments in high-throughput, cost-effective mutation screening methods that generally applied Sanger sequencing in a validation phase made possible considerable candidate gene sequencing and resequencing that further defined mutation frequency and associated cancer risks [39–45] (see Fig. 1). The frequency of some specific mutations in a number of these genes has made possible mutation-specific risk estimation, which provides evidence for much higher cancer risk associated with some mutations (*ATM* c.7271 T>G [46–48]; *PALB2* c.3113G> [49], *PALB2* c.1592delT [50]) and evidence for alternative modes of genetic inheritance of risk (e.g., colorectal cancer risk associated with monoallelic mutations in *MUTYH* [43].

Cancer, “Missing Heritability,” and MPS

A considerable proportion of the heritable risk for common cancers remains unexplained. Current estimates of what proportion of familial risk is explained by what we currently know about genetic cancer risk are in the order of 30–48 % for common cancers. These estimates include contributions from rare mutations in genes that convey high to moderate risk of cancer (discussed earlier) and the large number of common genetic variants (individually associated with very small increased risk) identified via genome-wide association studies. For breast cancer, the current estimate of the proportion of familial risk explained by common genetic variants is approximately 28 %, and another approximately 20 % is explained by rare mutations in breast cancer susceptibility genes such as *BRCA1*, *BRCA2*, and *PALB2* [51]. For prostate cancer, 70 identified common genetic variants account for approximately 30 % of the familial risk [52].

The nature of the missing heritability of common cancer is currently a matter of conjecture and has received much attention. A proportion of this missing heritability is likely to be explained by rare mutations in a large number of different genes [53–55].

Can the application of MPS technology uncover more of the missing heritability of common cancer? In the context of identifying rare mutations in genes that are associated with high to moderate risk of cancer, the challenge is no longer a technical one but rather an issue of study design and data interpretation.

A fundamental part of the leap from the methodology applied in past candidate gene mutation scanning projects (see earlier) to MPS platforms is the capability to search for cancer susceptibility genes in a so-called agnostic manner. To successfully achieve an agnostic approach to find additional cancer susceptibility genes, several aspects of the study design need to be carefully considered, although most require the incorporation of several assumptions about the underlying genetic architecture [56••].

Early Experiences

It could be well argued that the most successful application of MPS is the discovery of rare variants in new susceptibility genes for cancer is the identification of *HOXB13* as a prostate cancer susceptibility gene [57••]. This is a fusion of old data and the new technology as the investigators applied targeted capture of genes within a region of 17q21-22 that demonstrated linkage in a proportion of multiple-case families [58]. Targeted capture followed by MPS in selected individuals from families selected for linkage to this region revealed a rare (found in four of 94 DNA screened) but recurrent mutation in *HOXB13* (G84E) that has subsequently been extensively validated as associated with prostate cancer risk [59–61]. Similar large collections of linkage data are available for other cancer studies, and several similar initiatives are under way.

There are some recent reports from breast cancer researchers who have applied exome capture followed by MPS to multiple, affected members of multiple-case breast cancer families. This design offers researchers some capacity to use the family design in data filtering pipelines to both manage sequencing artifacts and annotate variant sharing between relatives [3, 56••, 62•].

Some demonstration of the intrafamily exome-sequencing approach to identify breast cancer susceptibility genes has been realized. Thompson et al. [5•] have reported rare variants in *FANCC* and *BLM*, which are responsible for the autosomal recessive disorders Fanconi anemia and Bloom syndrome, respectively, and Park et al. [63] have reported rare variants in *XRCC2*. These genes

have been identified as strong candidates from early initiatives, with strong reliance on prior gene function knowledge, rather than application of an agnostic gene assessment algorithm. Larger-scale validation of these genes as breast cancer susceptibility genes has provided further insight into their possible role in cancer predisposition [63, 64], yet have also illustrated the challenges when conducting these studies for common cancers, with variable phenotypes, mutations that are likely to be rare (and in a number of different genes), and penetrance far from complete [3, 4•, 5•, 56••, 63].

Integration with Other MPS Datasets

Other data sets generated via the application of MPS, such as cancer genomes and transcriptomes, can provide additional information that can be useful in the search for new cancer susceptibility genes. A study of familial pancreatic cancer that applied exome capture followed by MPS of the cancer genome identified a *PALB2* germline mutation and found a further three mutation carriers when *PALB2* was sequenced in a further 96 highly selected pancreatic cancer families [65].

Transcriptome sequencing could offer a complement or indeed a substitute for exome sequencing [66, 67]. There are indeed some promising aspects to this approach as it facilitates the analysis of variants within the coding region, bypasses the need for exome enrichment, and offers some cost advantages. Naturally, there are also some limitations, notably tissue specificity, which needs to be considered in the study design and application [68].

“Next-Generation” Collaboration

Coordinated international collaborations are beginning to emerge that have the potential to advance the discovery of additional cancer susceptibility genes by increasing the likelihood of identifying functionally relevant genetic variants in the same genes in multiple families by combining MPS data [7] (COMPLEXO; <http://www.path.unimelb.edu.au/research/labs/southey/Complexo.dwt>). Pooling of information about common genetic variation in case–control studies has proven to be a successful way to measure the role of common variants associated with cancer predisposition via the consortium model such as the Breast Cancer Association Consortium (<http://ccge.medschl.cam.ac.uk/consortia/bcac/index.html>), the Consortium of Investigators of Modifiers of BRCA1/2 (<http://ccge.medschl.cam.ac.uk/consortia/cimba/index.html>), the Ovarian Cancer Association Consortium (<http://ccge.medschl.cam.ac.uk/consortia/ocac/aims/aims.html>), and the Collaborative Oncology Gene-

Environment Study (<http://www.cogseu.org>). However, combining data from MPS studies has additional challenges that demand refinement of our bioinformatics analysis pipelines to both handle the volume of data available and conform to data interpretation and filtering methods. International collaboration also offers the opportunity to gather the magnitude of resources necessary to validate potential cancer susceptibility genes identified via combined MPS studies. However, even with international coordination, the rarity of mutations in some of the yet to be identified cancer susceptibility genes may mean that empirical demonstration of an association with cancer risk is not possible, and that other, potentially functionally based, assays may need to be relied on.

Conclusion

MPS has revealed the complexity of the human genome and has great potential to reveal more of the missing heritability of diseases such as common cancer. However, our current study designs and bioinformatic and computational approaches need to be refined to realize the full potential of MPS for it to contribute further to the identification of new cancer susceptibility genes.

Acknowledgments The author is a Senoir Research Fellow of the National Health and Medical Research Council (Australia) and a Group Leader of the Victorian Breast Cancer Research Consortium.

Disclosure M.C. Southey declares no conflicts of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- Asan, Xu Y, Jiang H, Tyler-Smith C, et al. Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biol.* 2011;12(9):R95.
- Clark MJ, Chen R, Lam HY, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol.* 2011; 29(10):908–14.
- Snape K, Ruark E, Tarpey P, et al. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. *Breast Cancer Res Treat.* 2012;134(1):429–33.
- Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in XRCC2 increase the risk of breast cancer. *Am J Hum Genet.* 2012;90(4):734–9. *An early example of the application of MPS in the setting of multiple-case breast cancer families designed to identify new breast cancer susceptibility genes.*
- Thompson ER, Doyle MA, Ryland GL, et al. Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. *PLoS Genet.* 2012;8(9):e1002894. *An early example of the application of MPS in the setting of multiple-case breast cancer families designed to identify new breast cancer susceptibility genes.*
- DeRycke MS, Gunawardena SR, Middha S, et al. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. *Cancer Epidemiol Biomarkers Prev.* 2013 (in press).
- Southey MC, Park DJ, Nguyen-Dumont T, et al. COMPLEXO: identifying the missing heritability of breast cancer via next generation collaboration. *Breast Cancer Res.* 2013 (in press). *This letter describes an international collaboration that could expedite the identification of new breast cancer susceptibility genes that could provide a useful working model for research into other complex diseases.*
- Ku CS, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet.* 2011;129(4):351–70.
- Genomes Project Consortium, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467(7319):1061–73.
- Bogliolo M, Schuster B, Stoepker C, et al. Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. *Am J Hum Genet.* 2013;92(5):800–6.
- Comino-Méndez I, Gracia-Aznárez FJ, Schiavi F, et al. Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat Genet.* 2011;43(7):663–7.
- Gilissen C, Hoischen A, Brunner HG, et al. Unlocking Mendelian disease using exome sequencing. *Genome Biol.* 2011;12(9):228.
- Rabbani B, Mahdieh N, Hosomichi K, et al. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet.* 2012;57(10):621–32.
- Sanger F, Air GM, Barrell BG, et al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature.* 1977;265(5596):687–95.
- Glusman G. Clinical applications of sequencing take center stage. *Genome Biol.* 2013;14(3):303.
- Sikkema-Raddatz B, Johansson LF, de Boer EN, et al. Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. *Hum Mutat.* 2013;34(7):1035–42.
- Rattenberry E, Vialard L, Yeung A, et al. A comprehensive next generation sequencing based genetic testing strategy to improve diagnosis of inherited pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab.* 2013. doi:10.1210/jc.2013-1319.
- Southey MC, Tesoriero A, Young MA, et al. A specific GFP expression assay, penetrance estimate, and histological assessment for a putative splice site mutation in BRCA1. *Hum Mutat.* 2003;22(1):86–91.
- Goldgar DE, Easton DF, Deffenbaugh AM, et al. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Genet.* 2004;75(4):535–44.
- Tesoriero AA, Wong EM, Jenkins MA, et al. Molecular characterization and cancer risk associated with BRCA1 and BRCA2 splice site variants identified in multiple-case breast cancer families. *Hum Mutat.* 2005;26(5):495.
- Easton DF, Deffenbaugh AM, Pruss D, et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. *Am J Hum Genet.* 2007;81(5):873–83.
- Tavtigian SV, Byrnes GB, Goldgar DE, et al. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat.* 2008;29(11):1342–54.
- Tavtigian SV, Greenblatt MS, Lesueur F, et al. In silico analysis of missense substitutions using sequence-alignment based methods. *Hum Mutat.* 2008;29(11):1327–36.
- Miyoshi Y, Ando H, Nagase H, et al. Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients. *Proc Natl Acad Sci USA.* 1992;89(10):4452–6.

25. Leppert M, Dobbs M, Scambler P, et al. The gene for familial polyposis coli maps to the long arm of chromosome 5. *Science*. 1987;238(4832):1411–3.
26. Nakamura Y, Lathrop M, Leppert M, et al. Localization of the genetic defect in familial adenomatous polyposis within a small region of chromosome 5. *Am J Hum Genet*. 1988;43(5):638–44.
27. Goldgar DE, Fields P, Lewis CM, et al. A large kindred with 17q-linked breast and ovarian cancer: genetic, phenotypic, and genealogical analysis. *J Natl Cancer Inst*. 1994;86(3):200–9.
28. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*. 1994;266(5182):66–71.
29. Wooster R, Neuhausen SL, Mangion J, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science*. 1994;265(5181):2088–90.
30. Stratton MR, Ford D, Neuhasen S, et al. Familial male breast cancer is not linked to the BRCA1 locus on chromosome 17q. *Nat Genet*. 1994;7(1):103–7.
31. Spurdle AB, Whiley PJ, Thompson B, et al. BRCA1 R1699Q variant displaying ambiguous functional abrogation confers intermediate breast and ovarian cancer risk. *J Med Genet*. 2012;49(8):525–32.
32. Nieuwenhuis MH, Vasen HF. Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. *Crit Rev Oncol Hematol*. 2007;61(2):153–61.
33. Cheadle JP, Sampson JR. Exposing the MYtH about base excision repair and human inherited disease. *Hum Mol Genet*. 2003;12 Spec No 2:R159–65.
34. Al-Tassan N, Chmiel NH, Maynard J, et al. Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat Genet*. 2002;30(2):227–32.
35. Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet*. 2007;39(2):165–7.
36. Xia B, Dorsman JC, Ameziane N, et al. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. *Nat Genet*. 2007;39(2):159–61.
37. Reid S, Schindler D, Hanenberg H, et al. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nat Genet*. 2007;39(2):162–4.
38. Morak M, Laner A, Bacher U, et al. MUTYH-associated polyposis—variability of the clinical phenotype in patients with biallelic and monoallelic MUTYH mutations and report on novel mutations. *Clin Genet*. 2010;78(4):353–63.
39. Eng C, Brody LC, Wagner TM, et al. Interpreting epidemiological research: blinded comparison of methods used to estimate the prevalence of inherited mutations in BRCA1. *J Med Genet*. 2001;38(12):824–33.
40. Garritano S, Gemignani F, Palmero EI, et al. Detailed haplotype analysis at the TP53 locus in p.R337H mutation carriers in the population of southern Brazil: evidence for a founder effect. *Hum Mutat*. 2010;31(2):143–50.
41. Tavtigian SV, Oefner PJ, Babikyan D, et al. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am J Hum Genet*. 2009;85(4):427–46.
42. Le Calvez-Kelm F, Lesueur F, Damiola F, et al. Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case–control mutation-screening study. *Breast Cancer Res*. 2011;13(1):R6.
43. Win AK, Cleary SP, Dowty JG, et al. Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer. *Int J Cancer*. 2011;129(9):2256–62.
44. Edwards SM, Kote-Jarai Z, Meitz J, et al. Two percent of men with early-onset prostate cancer harbor germline mutations in the BRCA2 gene. *Am J Hum Genet*. 2003;72(1):1–12.
45. Kote-Jarai Z, Leongamornlert D, Saunders E, et al. BRCA2 is a moderate penetrance gene contributing to young-onset prostate cancer: implications for genetic testing in prostate cancer patients. *Br J Cancer*. 2011;105(8):1230–4.
46. Bernstein JL, Teraoka S, Southey MC, et al. Population-based estimates of breast cancer risks associated with ATM gene variants c.7271T>G and c.1066-6T>G (IVS10-6T>G) from the Breast Cancer Family Registry. *Hum Mutat*. 2006;27(11):1122–8.
47. Chenevix-Trench G, Spurdle AB, Gatei M, et al. Dominant negative ATM mutations in breast cancer families. *J Natl Cancer Inst*. 2002;94(3):205–15.
48. Goldgar DE, Healey S, Dowty JG, et al. Rare variants in the ATM gene and risk of breast cancer. *Breast Cancer Res*. 2011;13(4):R73.
49. Southey MC, Teo ZL, Dowty JG, et al. A PALB2 mutation associated with high risk of breast cancer. *Breast Cancer Res*. 2010;12(6):R109. doi:10.1186/bcr2796.
50. Erkkö H, Dowty JG, Nikkilä J, et al. Penetrance analysis of the PALB2 c.1592delT founder mutation. *Clin Cancer Res*. 2008;14(14):4667–71.
51. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. 2013;45(4):353–61, 361e1–2.
52. Eeles RA, Olama AA, Benlloch S, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet*. 2013;45(4):385–91, 391e1–2.
53. Smith P, McGuffog L, Easton DF, et al. A genome wide linkage search for breast cancer susceptibility genes. *Gene Chromosome Canc*. 2006;45(7):646–55.
54. Cicek MS, Cunningham JM, Fridley BL, et al. Colorectal cancer linkage on chromosomes 4q21, 8q13, 12q24, and 15q22. *PLoS One*. 2012;7(5):e38175. doi:10.1371/journal.pone.0038175.
55. Christensen GB, Baffoe-Bonnie AB, George A, et al. Genome-wide linkage analysis of 1,233 prostate cancer pedigrees from the International Consortium for Prostate Cancer Genetics using novel sumLINK and sumLOD analyses. *Prostate*. 2010;70(7):735–44.
56. •• Feng BJ, Tavtigian SV, Southey MC, et al. Design considerations for massively parallel sequencing studies of complex human disease. *PLoS One*. 2011;6(8):e23221. *Design considerations for studies applying MPS are fundamentally important to successful outcomes. This paper explores several aspects of design and associated assumptions.*
57. •• Ewing CM, Ray AM, Lange EM, et al. Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med*. 2012;366(2):141–9. *Identification of HOXB13 as a prostate cancer susceptibility gene via the application of targeted MPS utilizing information from previous linkage studies conducted in multiple-case prostate cancer families.*
58. Lange EM, Robbins CM, Gillanders EM, et al. Fine-mapping the putative chromosome 17q21-22 prostate cancer susceptibility gene to a 10 cM region based on linkage analysis. *Hum Genet*. 2007;121(1):49–55.
59. Chen Z, Greenwood C, Isaacs WB, et al. The G84E mutation of HOXB13 is associated with increased risk for prostate cancer: results from the REDUCE trial. *Carcinogenesis*. 2013;34(6):1260–4.
60. MacInnis RJ, Severi G, Baglietto L, et al. Population-based estimate of prostate cancer risk for carriers of the HOXB13 missense mutation G84E. *PLoS One*. 2013;8(2):e54727.
61. Xu J, Lange EM, Lu L, et al. HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). *Hum Genet*. 2013;132(1):5–14.
62. • Pope BJ, Nguyen-Dumont T, Odefrey F, et al. FAVR (Filtering and Annotation of Variants that are Rare): methods to facilitate the analysis of rare germline genetic variants from massively parallel sequencing datasets. *BMC Bioinformatics*. 2013;25(14):65. *Bioinformatic and analytical pipelines for analysis of MPS data need further development to support scientific questions being asked*

- with MPS data—this report is an early example of a well considered improvement.*
63. Park DJ, Odefrey FA, Hammet F, et al. FAN1 variants identified in multiple-case early-onset breast cancer families via exome sequencing: no evidence for association with risk for breast cancer. *Breast Cancer Res Treat.* 2011;130(3):1043–9.
 64. Hilbers FS, Wijnen JT, Hoogerbrugge N, et al. Rare variants in XRCC2 as breast cancer susceptibility alleles. *J Med Genet.* 2012;49(10):618–20.
 65. Jones S, Hruban RH, Kamiyama M, et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science.* 2009;324(5924):217.
 66. Chepelev I, Wei G, Tang Q, et al. Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. *Nucleic Acids Res.* 2009;37(16):e106.
 67. Cirulli ET, Singh A, Shianna KV, et al. Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. *Genome Biol.* 2010;11(5):R57.
 68. Ku CS, Wu M, Cooper DN, et al. Exome versus transcriptome sequencing in identifying coding region variants. *Expert Rev Mol Diagn.* 2012;12(3):241–51.
 69. Varghese JS, Easton DF. Genome-wide association studies in common cancers—what have we learnt? *Curr Opin Genet Dev.* 2010;20(3):201–9.