Genome Biology

EDITORIAL

The exome factor

Hannah Stower

Abstract

Exome sequencing is rapidly expanding both as a technique and in its biological applications.

The decision to launch a special issue of *Genome Biology* coincided with my first American Society of Human Genetics meeting in Washington, DC [1]; exciting work using exome capture sequencing had been presented on melanoma, somatic mutations in induced pluripotent stem cells and numerous single gene disorders. The technique strongly enriches a sequencing sample for exons by using DNA capture probes targeted only to the portion of the genome constituting the exome. In addition, at this time I met and discussed exome sequencing with one of the leaders in the field, Jay Shendure, and we are delighted that he agreed to be the Guest Editor for this issue. There was no doubt for me that exome sequencing was flourishing, and so it is in this September issue that we present articles with that special 'exome factor'.

Choosing to publish a special issue on a field that is, relatively speaking, in its infancy necessitates many discussions about how best to make use of the technology and analyze the data, and these discussions have led us to present some guidelines and new approaches to exome capture sequencing. In a 'bake-off' between different exome capture techniques, Janna Saarela and colleagues [2] perform a systematic comparison of the two solution-based capture kits commercially available from Agilent and Roche-Nimblegen. The results show that researchers should be confident with the data garnered by either approach, and we hope that the clarity of this study will make it a useful reference, particularly to those just starting out. To provide further coverage and discussion of best methodological practice, we commissioned Shamil Sunyaev and colleagues [3] to review the computational and statistical approaches necessary for prioritization of variants from exome sequencing data. With typically 15,000 to 20,000 variants discovered per exome, whittling these variants down to those that are likely to be causing disease is a significant challenge. An abundance of tools has been developed for this task, and the Review therefore sets out to consider the relative merits of a number of the leading examples of these. Progressing to new methods, Katherine Smith *et al.* [4] present a novel method to further simplify variant prioritization – demonstrating that genetic linkage mapping can be applied to single nucleotide polymorphism (SNP) genotypes extracted from exome data, removing the need for array-based genotyping.

It is the lack of ascertainment bias in SNP genotyping from exome sequencing data that motivates Josh Akey and colleagues [5] to promote its use in the fields of human evolution, population structure and demography. In their Review, he discusses how the power to sequence to high depth the exomes of multiple individuals from several populations will provide us with insights into the mode and timing of population expansions. Indeed, in an article from Gabor Marth and the 1000 Genomes Project Exon Sequencing Pilot Subgroup [6], data are presented from the exomes of 700 individuals from seven populations. A project of this scale allows a unique insight into rare variants and, as might be expected but has never previously been shown, rare exomic variants show increased population specificity and are enriched for functional polymorphisms. For researchers who have previously been unable to find rare variants in their data, Tao Wang and colleagues [7] have developed a new technique that displays an excellent combination of sensitivity and specificity in identifying such variants.

But why should human population geneticists have all the fun? In an article from Eduard Akhunov and colleagues [8], 3.5 Mb of exome regions are captured from two species of wheat, a cultivated variety and a wild variety, and SNPs and copy number variations are compared, to garner information about the evolutionary history of these two varieties. The challenges that had to be overcome by these researchers are many; wheat species have undergone whole genome duplications leading to extensive polyploidy and the wheat genome assembly is significantly underdeveloped and poorly annotated. The polyploidy in particular proves

hannah.stower@genomebiology.com
Special Issues Editor, *Genome Biology*
Genome Biology, BioMed Central, Gray's Inn Road, London WC1X 8HB, UK

challenging, as each probe could potentially hybridize to three homeologous sequences. Given the success of the authors' approach here, such pipelines may, in the future, be applied to similar problems in polyploid cancer genomes in addition to other plant species. Another non-human exome capture article comes from Eric Vallender [9], who has captured the exomes of two non-human primates on human exome arrays in a proof of principle of this approach. Furthermore, he shows the need for better annotation of non-human primate genomes.

Mouse geneticists, too, profit from exome capture techniques in this issue: Laura Reinholdt and colleagues [10] have developed and validated a set of mouse exome capture tools that will be of significant value to research in this beloved of model organisms, valuable for so much of our disease research. In another study in this issue, Karen Steel *et al.* [11] have applied commercially available mouse exome tools to a model for otitis media, finding a mutation in the *Islet1* gene that is likely to predispose carriers to middle ear infections.

The understanding of the genetics and biology of Mendelian disease is surely the area in which exome sequencing has seen the largest growth in the past couple of years. In particular, genes associated with nonsyndromic mental retardation were pinpointed using this approach [12]. Some of the authors of this work, including Christian Gilissen, have reviewed the progress in our understanding of Mendelian disease in this issue [13]. Because of the ease of variant prioritization, highly penetrant recessive diseases are the most widely studied by exome sequencing. However, that is not to say that these are the diseases exclusively studied by these techniques. Heterotaxy is a cardiovascular disorder characterized by incomplete penetrance and heterogeneity and thought to be related to defective left-right patterning. By performing exome sequencing on a complex heterotaxy phenotype proband from a consanguineous family, Stephanie Ware and colleagues [14] have identified a recessive mutation in *SHROOM3*, a gene not previously associated with left-right patterning. Furthermore, while the identification of novel disease genes is exciting, there is other knowledge to be garnered from exome sequencing data, such as an improved understanding of modes of inheritance and of phenotypic spectra associated with known disease genes. An article from Tadafumi Kato and colleagues [15] identifies variants associated with autosomal progressive external ophthalmoplegia (PEO), an inherited mitochondrial disease characterized by slowly progressive paralysis of the extraocular muscles. The authors identify a novel mutation in *RRM2B*, a gene whose heterozygous rare variant has also been shown to cause an autosomal dominant form of PEO. Thus, mutations in a single gene are the root cause of both forms of the disease.

Finally, in an elegant use of targeted capture, Karen Avraham and colleagues [16] identify multiple genes unique to different populations that are responsible for heritable hearing loss. The authors use an approach in which capture probes are designed to hybridize to genes associated with hearing loss in mouse, in addition to those responsible for human deafness. Such capture probes were applied to families from several Israeli and Palestinian genetic populations, identifying causative alleles in both Arab and Jewish probands. Again, some help from the mouse has significantly progressed our understanding of human disease [16].

After reading these articles, you may find yourself wondering what the future holds for exome sequencing. In a question and answer piece [17] answered by Jim Mullikin, Les Biesecker and Kevin Shianna, Mullikin and Biesecker argue that exome capture will endure as a useful tool in the clinic, while Shianna instead believes that it has an inevitable expiry date owing to the rapidly declining cost of whole genome sequencing. In his editorial, Jay Shendure [18] discusses the 'high yield' of exome sequencing and how it is this benefit that may ensure its longevity. It seems to me that the greatest legacy of exome sequencing may in fact be the application of lessons learned in a future in which whole genome sequencing is king.

I am proud to present to you the collection of articles contained within this issue and I would like to thank all of those involved, notably the Guest Editor Jay Shendure, who has contributed invaluably to this exciting project, and *Genome Biology*'s Editor Clare Garvey, in addition to the rest of the *Genome Biology* Editorial and Editorial Production teams. I hope you find the issue as timely and as interesting as we do.

Published: 14 September 2011

References

1. American Society of Human Genetics 60th meeting [http://www.ashg.org/2010meeting/]
2. Sulonen A-M, Ellonen P, Almusa H, Lepistö M, Eldfors S, Hannula S, Miettinen T, Tyynismaa H, Salo P, Heckman C, Joensuu H, Raivio T, Suomalainen A, Saarela J: **Comparison of solution-based exome capture methods for next generation sequencing.** *Genome Biology* 2011, **12**:r94
3. Stitziel NO, Kiezun A, Sunyaev S: **Computational and statistical approaches to analyzing variants identified by exome sequencing.** *Genome Biology* 2011, **12**:227
4. Smith KR, Bromhead CJ, Hildebrand MS, Shearer AE, Lockhart PJ, Najmabadi H, Leventer RJ, McGillivray G, Amor DJ, Smith RJ, Bahlo M: **Reducing the exome search space for Mendelian diseases using genetic linkage analysis of exome genotypes.** *Genome Biology,* 2011, **12**:r85
5. Tennessen JA, O'Connor TD, Bamshad MJ, Akey JM: **The promise and limitations of population exomics for human evolution studies.** *Genome Biology* 2011, **12**:127
6. Gabor T Marth, Fuli Yu, Amit R Indap, Kiran Garimella, Simon Gravel, Wen Fung Leong, Chris Tyler-Smith, Matthew Bainbridge, Thomas Blackwell, Xiangqun Zheng-Bradley, Yuan Chen, Danny Challis, Laura Clarke, Edward V Ball, Kristian Cibulskis, David N Cooper, Bob Fulton, Chris Hartl, Dan Koboldt, Donna Muzny, Richard Smith, Carrie Sougnez, Chip Stewart, Alistair Ward, Jin Yu, Yali Xue, David Altshuler, Carlos D Bustamante, Andrew G Clark, Mark Daly, Mark

DePristo, Paul Flicek, Stacey Gabriel, Elaine Mardis, Aarno Palotie and Richard A Gibbs: **The functional spectrum of low-frequency coding variation.** *Genome Biology* 2011, **12**:r84
7. Niranjan TS, Adamczyk A, Bravo HC, Taub MA, Wheelan SJ, Irizarry R, Wang T: **Effective detection of rare variants in pooled DNA samples using cross-pool tailcurve analysis.** *Genome Biology* 2011, **12**:r93
8. Saintenac C, Jiang D, Akhunov ED: **Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome.** *Genome Biology* 2011, **12**:r88
9. Vallender EJ: **Expanding whole exome resequencing into non-human primates.** *Genome Biology* 2011, **12**:r87
10. Fairfield H, Gilbert GJ, Barter M, Corrigan RR, Curtain M, Ding Y, D'Ascenzo M, Gerhardt DJ, He C, Huang W, Richmond T, Rowe L, Probst F, Bergstrom DE, Murray SA, Bult C, Richardson J, Kile B, Gut I, Hager J, Sigurdsson S, Mauceli E, Di Palma F, Lindblad-Toh K, Cunningham ML, Cox TC, Justice MJ, *et al.*: **Mutation discovery in mice by whole exome sequencing.** *Genome Biology* 2011, **12**: In press
11. Hilton JM, Lewis MA, Grati M, Ingham N, Pearson S, Laskowski RA, Adams DJ, Steel KP: **Exome sequencing identifies a missense mutation in Isl1 associated with low penetrance otitis media in dearisch mice.** *Genome Biology* 2011, **12**:r90
12. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, van Bon BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA: **A de novo paradigm for mental retardation.** *Nature Genetics* 2011, **43:**729-31.
13. Gilissen C, Hoischen A, Brunner H, Veltman J: **Unlocking Mendelian disease using exome sequencing.** *Genome Biology* 2011, **12**:228
14. Tariq M, Belmont J, Lalani S, Smolarek T, Ware SM: **SHROOM3 is a novel candidate for heterotaxy identified by whole exome sequencing.** *Genome Biology* 2011, **12**:r91
15. Takata A, Kato M, Nakamura M, Yoshikawa T, Kanba S, Sano A, Kato T: **Exome sequencing identifies a novel missense variant in RRM2B associated with autosomal recessive progressive external ophthalmoplegia.** *Genome Biology* 2011, **12**:r92
16. Brownstein Z, Friedman LM, Shahin H, Oron-Karni V, Kol N, Rayyan AA, Parzefall T, Lev D, Shalev S, Frydman M, Davidov B, Shohat M, Rahile M, Lieberman S, Levy-Lahad E, Lee M, Shomron N, King M-C, Walsh T, Kanaan M, Avraham KB: **Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in Middle Eastern families.** *Genome Biology* 2011, **12**:r89
17. Biesecker LG, Shianna KV, Mullikin JC: **Exome sequencing: the expert view**. *Genome Biology* 2011, **12**:128
18. Shendure J: **Next-generation human genetics.** *Genome Biology* 2011, **12**:408

doi:10.1186/gb-2011-12-9-407
Cite this article as: Stower H: **The exome factor.** *Genome Biology* 2011, **12:**:407