WEScover: whole exome sequencing vs. gene panel testing

William Jefferson Alvarez[1], In-Hee Lee[1], Carles Hernandez-Ferrer[1], Jose Negron La Rosa[1], Kenneth D. Mandl[1] and Sek Won Kong[1, 2, *]

[1] Computational Health Informatics Program, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA, 02115, USA
[2] Department of Pediatrics, Harvard Medical School, 10 Shattuck Street, Boston, MA 02115, USA
[*] To whom correspondence should be addressed.

Abstract

Motivation: Whole exome sequencing (WES) is widely adopted in clinical and research settings; however, the potential for false negatives due to incomplete breadth and depth of coverage for some exons has been reported. In some cases, targeted gene panel testing could be a reliable option to ascertain true negatives for phenotype-associated genomic variants. We developed a tool for quickly gauging whether all genes of interest are comprehensively covered by WES or whether targeted gene panel testing should instead be considered to minimize false negatives in candidate genes.
Results: *WEScover* is a novel web application providing an interface for discovering breadth and depth of coverage across population scale WES datasets, searching either by phenotype, by targeted gene panels and by genes. Moreover, the application shows metrics from the Genome Aggregation Database to provide gene-centric view on breadth of coverage.
Availability: *WEScover* is available at http://gNOME.tchlab.org/WEScover/. The source code is available at https://github.com/bch-gnome/WEScover.
Contact: sekwon.kong@childrens.harvard.edu
Keywords: bioinformatics

Introduction

As the cost of whole exome sequencing (WES) drops, it is beginning to replace broad and/or targeted gene panel testing. (Stavropoulos, et al., 2016; Wang, et al., 2014). WES for example, is superior in measurement of the ever-growing number of driver and passenger mutations in diverse genes across different cancer types as well as increasing awareness of polygenic contribution to most genetic disorders. However, WES may not capture all exons in clinically implicated genes in the human genome (Kong, et al., 2018; Meienberg, et al., 2015). Whole genome sequencing (WGS) faces a similar challenge for some genes including highly polymorphic ones. Moreover, population scale aggregation of WES and WGS clearly shows limited breadth of coverage for some clinically implicated genes (Kong, et al., 2018; Wang, et al., 2017). Therefore, gene panel testing, whether for a single gene or for hundreds of candidate genes, is still a clinically useful measure where false negatives due to suboptimal coverage are likely. However, it is difficult to predict whether the exons known to harbor disease-associated variants would be covered with sufficient per-site depth of coverage to reliably call variants or not.

Implementation

To help biomedical investigators to select the reliable genetic test – i.e., WES vs. targeted gene panel testing, we developed the WEScover web application that highlights global gene level coverage and inter-individual variation in breadth of coverage for genes along with corresponding genetic tests listed in the National Institutes of Health Genetic Testing Registry (GTR) (Rubinstein, et al., 2013). A total of 5,309 putative disease-associated genes are listed across 54,612 genetic tests for both clinical and research usage including 37,746 CLIA-certified ones in GTR. For each entry in the Consensus Coding Sequence (CCDS) (Pruitt, et al., 2009), we calculated breadth of coverage at >10x, >20x and >30x (the percentage of sites where per-site depth of coverage is higher than 10x, 20x, and 30x, respectively) across the exomes from the 1000 Genomes Project (1KGP) phase 3 (Genomes Project, et al., 2015) (N=2,504, alignment files remapped to GRCh38 human reference genome). Additionally, we took the average value among the entire exomes (N=123,136) from the Genome Aggregation Database project (gnomAD) (Lek, et al., 2016)) as a global estimate from large-scale data (the continent-level data is not currently available from gnomAD project). Using the relationship between phenotypes, genetic test names from GTR, and genes, we created a database and a query interface as a R Shiny application (Chang, et al., 2017).

Workflow

The initial query interface allows users to enter phenotype, genetic test name (retrieved from the GTR website), or official gene symbol(s) of interest. For each gene matching the query, the global mean of breadth of coverage along with its maximum and minimum values is shown as a table in an ascending order of global means (Fig. 1A). By default, we used breadth of coverage at 20x – a threshold sufficient to achieve 99% sensitivity for detecting single nucleotide variant (Meynert, et al., 2014). The test statistic and p-value for a one-way analysis of variance, performed to test for differences between means of populations, are also reported in this table. The button at the end of each row opens a panel with further details about the coverage of the gene. The panel first shows a table with the mean of breadth of coverage stratified by continent-level population. The second tab shows a violin plot for breadth of coverage stratified by continent-level population with the mean value from exomes in gnomAD project as a black line (Fig. 1B). A plot for coverage at each genomic position of the selected gene, based on gnomAD coverage data, is shown next to the violin plot (Fig. 1C). Lastly, the panel reports all genetic test involving the gene. Insufficient

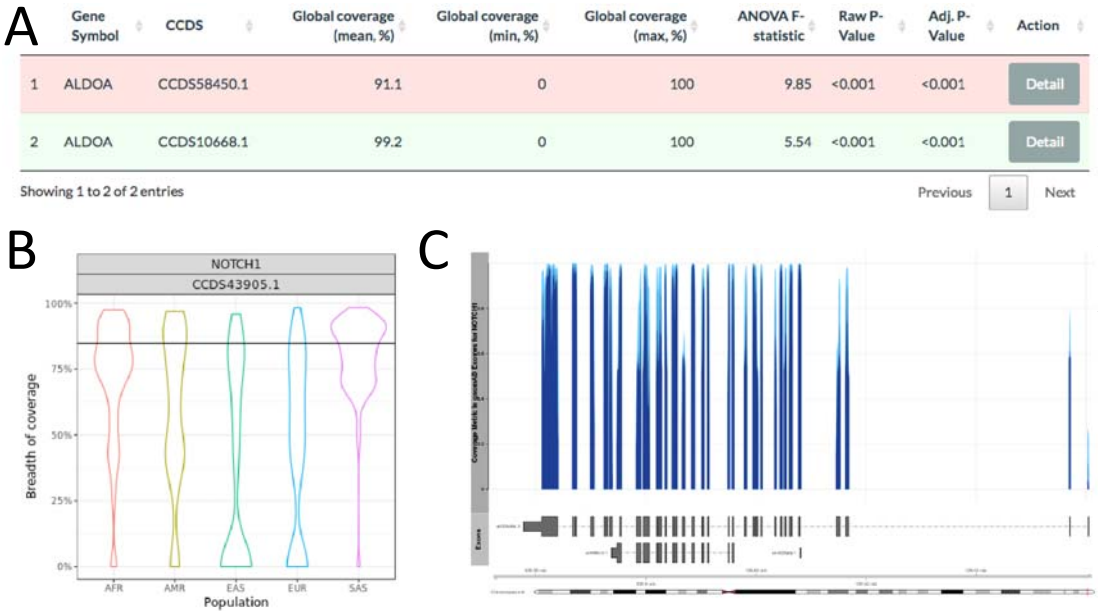


Fig. 1. A) The initial screen for selected genes. Clicking 'Detail' button (red highlighted box) shows window with more information for the selected transcript. B) The violin plot shows the distribution of coverage metrics from 1KGP exomes in each of 5 population. The black horizontal line denotes the global average value from gnomAD exomes. C) The coverage plot shows the transcript model and coverage metric from gnomAD exomes. The upper part of graph shows metric values at 10x (most light blue), 20x, and 30x (most dark blue).

breadth of coverage in both projects, 1KGP and gnomAD, should warn the user that the candidate genes may not be well covered in WES and targeted gene panel tests should be considered to minimize potential false negatives.

Conclusion

WES and WGS could provide more comprehensive evaluation of genomic variants in various conditions; however, users need to be informed regarding possible false negatives due to incomplete breadth and depth of coverage (ideally, from sequencing vendors). In such cases, targeted gene panel tests should be considered as a primary choice over the others. Together with GTR, which provides a transparent and comprehensive list of genetic tests with indications, WEScover can guide users to select optimal genetic tests per phenotype and/or genes of interest. Users can quickly check breadth and depth of coverage for candidate genes and genetic test labs prior to ordering genetic tests in the clinical settings.

Acknowledgements

We appreciate Ms. Diaz-Cuadros for helpful comments for the improvement of the web interface.

Funding

This work has been supported by the Boston Children's Hospital Precision Link initiative. S.W.K. was supported in part by grants from the National Institutes of Health (NIMH R01MH107205 and NHGRI U01HG007530).

Conflict of Interest: none declared.

References

1. Genomes Project, C., *et al.* A global reference for human genetic variation. *Nature* 2015;526(7571):68-74.
2. Kong, S.W., *et al.* Measuring coverage and accuracy of whole-exome sequencing in clinical context. *Genet Med* 2018.
3. Lek, M., *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 4. 2016;536(7616):285-291.
4. Meienberg, J., *et al.* New insights into the performance of human whole-exome capture platforms. *Nucleic Acids Res* 2015;43(11):e76.
5. Meynert, A.M., *et al.* Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics* 2014;15:247.
6. Pruitt, K.D., *et al.* The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res* 2009;19(7):1316-1323.
7. Rubinstein, W.S., *et al.* The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency. *Nucleic Acids Res* 2013;41(Database issue):D925-935.
8. Stavropoulos, D.J., *et al.* Whole Genome Sequencing Expands Diagnostic Utility and Improves Clinical Management in Pediatric Medicine. *NPJ Genom Med* 2016;1.
9. Wang, J., *et al.* Diagnostic yield of clinical next-generation sequencing panels for epilepsy. *JAMA Neurol* 2014;71(5):650-651.
10. Wang, Q., et al. Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. Sci Rep 2017;7(1):885.

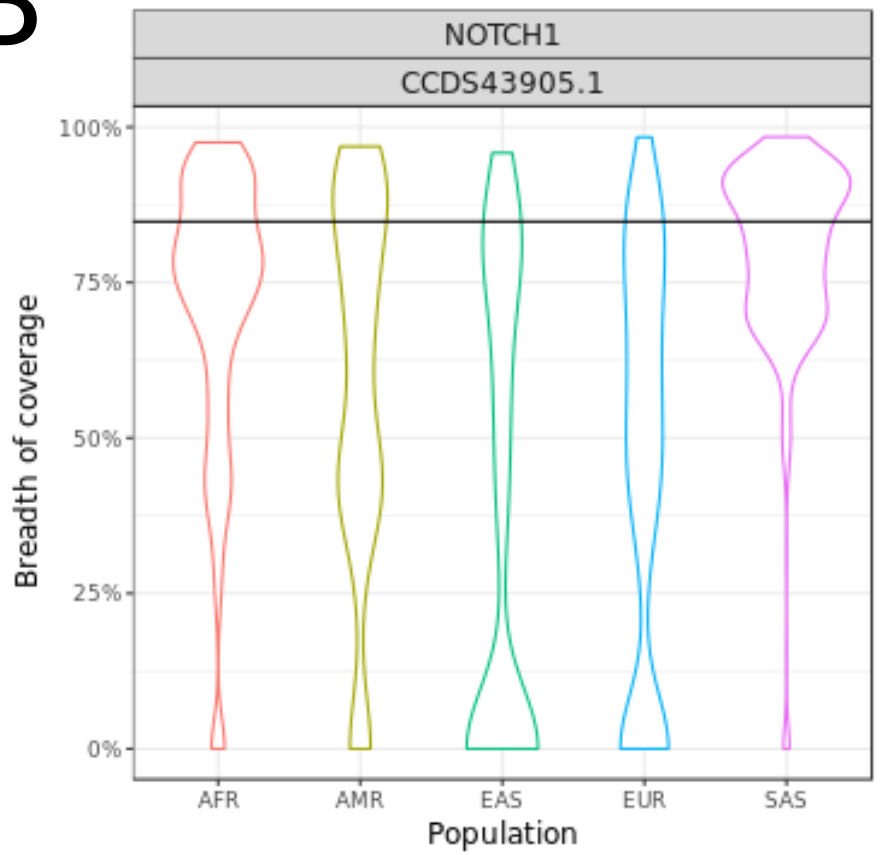
A

	Gene Symbol	CCDS	Global coverage (mean, %)	Global coverage (min, %)	Global coverage (max, %)	ANOVA F-statistic	Raw P-Value	Adj. P-Value	Action
1	ALDOA	CCDS58450.1	91.1	0	100	9.85	<0.001	<0.001	Detail
2	ALDOA	CCDS10668.1	99.2	0	100	5.54	<0.001	<0.001	Detail

Showing 1 to 2 of 2 entries

Previous 1 Next

B



C