

Short Note

Not peer-reviewed version

What Is the Real Diagnostic Benefit of Whole-Genome Sequencing?

[Florian Battke](#)*, Martin Schulze, Björn Schulte

Posted Date: 3 December 2024

doi: 10.20944/preprints202412.0023.v1

Keywords: Genetic Diagnostics; Whole-exome sequencing; WES; Whole-genome sequencing; WGS; Diagnostic Yield



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Short Note

What Is the Real Diagnostic Benefit of Whole-Genome Sequencing?

Florian Battke ^{1,*}, Martin Schulze ² and Björn Schulte ²

¹ CeGaT GmbH, Tübingen, Germany

² Center for Human Genetics, Tübingen, Germany

* Correspondence: Florian.Battke@cegat.de

From the beginning, genetic diagnostics as a field experienced rapid progress. After single genes, first small and then larger gene panels could be sequenced, later the whole set of coding genes could be investigated with whole-exome sequencing (WES). With each increase in the scope of analysis, more cases could be solved, and patients' waiting time for a diagnosis was reduced. Depending on the disease in question, solution rates reached around 30-35% on average [1–3].

As sequencing prices dropped further, the community started looking to the obvious next step, whole-genome sequencing (WGS), to finally close the remaining gaps in diagnostic yield. Despite larger and faster sequencing machines, WGS still requires substantially larger effort for sequencing, data storage, data processing and interpretation. These large additional efforts have been expected to result in another large step up for solution rates. At the same time, large diagnostic institutions such as ours have continually optimized WES to cover more and more disease-causing variants.

The expected benefits of WGS and the prestige associated with using the latest technology create a strong pressure to transition from WES to WGS in diagnostics. Meanwhile, public health systems strictly regulate reimbursement and cost-efficiency remains highly important. The additional effort for WGS must be justified by additional benefit. Therefore, multiple groups have sought to show the benefits of WGS in terms of increased diagnostic yield and shortened patient journey. In such studies, cohorts of patients analyzed with both methods are used to determine which cases remained unsolved by WES but could be solved by WGS.

Despite these efforts, so far the large benefits of WGS for diagnostics have not materialized: Publications show large additional diagnostic yields, but on closer inspection these can't be attributed to the new technology. There seem to be three main reasons for the observed yield increases: First, where WES was performed long before WGS, scientific knowledge added in the meantime would have led to the same solutions if the WES data were reanalyzed using current databases. Second, data from low-quality WES is sometimes compared with state-of-the-art WGS. And third, WGS variants are considered as disease-causing even though they are classified as VUS (variants of uncertain significance) and don't contribute to a definite case solution. We consider a case as solved only if a significant finding could be reported, in other words, a likely pathogenic or pathogenic variant (ACMG class 4 or 5) was found.

In the first publication we analyzed, the authors find a yield increase of 14.5 percentage points (%pt) [4]. Our reanalysis of their data shows only a yield increase of at most 1.4 %pt [5]. The second publication did not contain enough data for reanalysis but the headline figure of a 9 percent yield increase [6] corresponds to only 1.8 percentage points [7], and some of the variant types mentioned (eg., exonic CNVs) are unlikely to be missed by WES. Interestingly, the pressure of expectation appears so strong as to override what the data actually shows: When confronted with this discrepancy, the authors point to a third publication [8]. Even though they agree that their own data shows a yield increase of 1.8 %pt and the figure in the new publication is 8.2 %pt, they conclude that "our estimation seems to be similar to those of other GS studies", basically saying that everybody knows that WGS must have a large yield increase.

Here we take a closer look on the third publication [9]. Studying a cohort of 744 cases, the authors find 61 case-solving variants only detectable by WGS, a yield increase of 8.2% over WES. As the data was made available, let's have a closer look.

We perform our re-analysis in two steps. First, assessing each variant in the paper's supplement, we confirm the authors' decision that the variant solves the case. Second, we look at all solution variants and confirm that they are not detectable with WES. Wherever we are unsure, be it because of lacking detail in the supplementary data or because interpretations may differ, we decided to agree with the authors' original assessment.

Of 61 solution variants, we find that two lack sufficient description. One is a LINE insertion that is not exactly placed and may or may not be hitting a coding exon, and one is a tandem duplication/deletion affecting *MED13L* but with coordinates far outside of that gene. We will treat these two as case-solving variants that are only detectable by WGS. Of the 59 we can evaluate, 16 are not really case solving: They are classified as VUS by the authors themselves. This leaves 43 case-solving variants that fall into the following categories:

- 14 are SNVs or indels in coding sequence. Stating that WES is incapable of detecting variants in coding exons should give anyone pause. These are all, of course, detectable by WES, including some that are in splice sites 1-2 base pairs outside of the coding exon, which are obviously also covered by WES.
- 9 are deep intronic SNVs or indels. These are too far outside the coding region to be detectable by standard WES. However, modern customized exome kits can include such targets. In our customized exome, all these targets are covered by enrichment targets.
- 5 are tandem repeat expansions. This class of variant is hard to call from short-read sequencing data in general, but given sufficient coverage, there is no reason to believe it works better in WGS data than WES data. Three of the 5 locations are covered in off-the-shelf exomes, all five are covered in our customized exome enrichment.
- 7 deletion/duplication variants affecting at least one complete coding exon (CNV). These can all be called from WES data, so no WGS is needed for their detection.
- 8 larger or more complex structural variants (SV), including one large UTR deletion, one deletion affecting only half of an exon, 5 inversions and one complex variant. Of these difficult variants, three would be at least partially visible in WES data but the data would not show their full complexity. A partial call would be enough to warrant closer inspection, however.

Summarizing, we find that of 61 additional case-solving variants obtained by WGS in a cohort of 744 cases, only 45 are actually solving a case, and only 17 really require WGS for detection (or only 7 if WES is performed with our customized exome). This gives an additional yield from WGS of 2.3 %pt over standard exome and 0.9 %pt over a modern customized exome.

Just as WGS has improved over time, so have WES methods. It is no surprise that current WGS methods outperform WES when the latter is performed with outdated protocols or when the interpretation of the WES data ignored current scientific evidence. State-of-the-art WES assays use up-to-date evidence for variant interpretation and also for target design, for instance when including known disease-causing variants in non-coding regions. Still, it seems plausible that WGS can offer increased yield even over state-of-the-art WES, especially due to its ability to detect complex structural variants. But a likely figure for yield increase would seem to be around 1 %pt, as we have shown in our comments on the three publications. At the same time, WGS misses low-frequency mosaic variants found by WES. In our patient cohort, such mosaic variants contribute around 0.5 %pt of the diagnostic yield. Customized exome designs allow for additional diagnostic advantages over WGS. For example, the inclusion of viral sequences in our ExomXtra® adds important information in prenatal settings and for patients with cancer.

Interestingly, six variants claimed as case-solving, four of the deep intronic variants and two of the SVs, can only be considered solution variants after performing an orthogonal test, whole-transcriptome sequencing (WTS). This method can find causes for genetic disease that are either not visible in WGS or WES, such as monoallelic expression, or that are extremely hard to assess from WGS/WES data, such as previously unknown intergenic variants affecting promotor function and resulting in expression changes, or intronic variants affecting transcript splicing.

High-quality WES - such as our customized exome with its coverage of known non-coding variants and whole-genome CNV backbone - is still the most cost-efficient method to solve most cases of genetic disorders [10]. By including data from both parents, trio analyses can boost diagnostic yield considerably [10–12]. For some unsolved cases, WGS may find a solution at considerable additional expense (for example by detecting a structural variant to complement a single pathogenic variant

found by WES), yet is just as likely to just turn up variants of unknown significance that require further testing, most often with RNA sequencing. We postulate that going directly from WES to RNA sequencing is the most promising route for most patients.

References

1. Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, et al. *Clinical exome sequencing: results from 2819 samples reflecting 1000 families*. Eur J Hum Genet. 2017;25(2):176-182. <https://doi.org/10.1038/ejhg.2016.146>
2. Alotibi RS, Sannan NS, AlEissa M, et al. *The diagnostic yield of CGH and WES in neurodevelopmental disorders*. Front Pediatr. 2023;11:1133789. Published 2023 Mar 1. <https://doi.org/10.3389/fped.2023.1133789>
3. Artech-López, A., Ávila-Fernández, A., Riveiro Álvarez, R. et al. *Five years' experience of the clinical exome sequencing in a Spanish single center*. Sci Rep 12, 19209 (2022). <https://doi.org/10.1038/s41598-022-23786-6>
4. Bertoli-Avella AM, Beetz C, Ameziane N, Rocha ME, Guatibonza P, Pereira C, et al. *Successful application of genome sequencing in a diagnostic setting: 1007 index cases from a clinically heterogeneous cohort*. Eur J Hum Genet. 2020. <https://doi.org/10.1038/s41431-020-00713-9>.
5. Battke, F., Schulze, B., Schulze, M. et al. *The question of WGS's clinical utility remains unanswered*. Eur J Hum Genet 29, 722–723 (2021). <https://doi.org/10.1038/s41431-021-00823-y>
6. Park, J.; Sturm, M.; Seibel-Kelemen, O.; Ossowski, S.; Haack, T.B. *Lessons Learned from Translating Genome Sequencing to Clinical Routine: Understanding the Accuracy of a Diagnostic Pipeline*. Genes 2024, 15, 136. <https://doi.org/10.3390/genes15010136>
7. Battke, F.; Schulze, M.; Schulte, B.; Biskup, S. *Comment on Park et al. Lessons Learned from Translating Genome Sequencing to Clinical Routine: Understanding the Accuracy of a Diagnostic Pipeline*. Genes 2024, 15, 136. Genes 2024, 15, 1322. <https://doi.org/10.3390/genes15101322>
8. Park, J.; Sturm, M.; Haack, T.B. *Reply to Battke et al. Comment on "Park et al. Lessons Learned from Translating Genome Sequencing to Clinical Routine: Understanding the Accuracy of a Diagnostic Pipeline"*. Genes 2024, 15, 136". Genes 2024, 15, 1323. <https://doi.org/10.3390/genes15101323>
9. Wojcik, M.H.; Lemire, G.; Berger, E.; Zaki, M.S.; Wissmann, M.; Win, W.; White, S.M.; Weisburd, B.; Wieczorek, D.; Waddell, L.B.; et al. *Genome Sequencing for Diagnosing Rare Diseases*. N. Engl. J. Med. 2024, 390, 1985–1997.
10. Ewans, L.J., Minoche, A.E., Schofield, D. et al. *Whole exome and genome sequencing in mendelian disorders: a diagnostic and health economic analysis*. Eur J Hum Genet 30, 1121–1131 (2022). <https://doi.org/10.1038/s41431-022-01162-2>
11. Gabriel H, Korin D, Ritthaler M, et al. *Trio exome sequencing is highly relevant in prenatal diagnostics*. Prenat Diagn. 2022;42(7):845-851. doi:10.1002/pd.6081
12. Farwell KD, Shahmirzadi L, El-Khechen D, et al. *Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions*. Genet Med. 2015;17(7):578-586. doi:10.1038/gim.2014.154

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.