

ARTICLE OPEN



A cost-effective sequencing method for genetic studies combining high-depth whole exome and low-depth whole genome

Claude Bh er 1,2,3, Robert Eveleigh 2,4, Katerina Trajanoska 1,2,3, Janick St-Cyr 2, Antoine Paccard 2, Praveen Nadukkalam Ravindran 2,3, Elizabeth Caron 2, Nimara Bader Asbah 1,2, Peyton McClelland 1,2,3, Clare Wei 1,2,3, Iris Baumgartner 5,6, Marc Schindewolf 5,6, Yvonne D oring 5,6,7, Danielle Perley 2,4, Fran ois Lefebvre 2,4, Pierre Lepage 2, Mathieu Bourgey 1,8, Guillaume Bourque 1,2,4, Jiannis Ragoussis 1,2, Vincent Mooser 1,2,3 and Daniel Taliun 1,2,3✉

Whole genome sequencing (WGS) at high-depth (30X) allows the accurate discovery of variants in the coding and non-coding DNA regions and helps elucidate the genetic underpinnings of human health and diseases. Yet, due to the prohibitive cost of high-depth WGS, most large-scale genetic association studies use genotyping arrays or high-depth whole exome sequencing (WES). Here we propose a cost-effective method which we call “Whole Exome Genome Sequencing” (WEGS), that combines low-depth WGS and high-depth WES with up to 8 samples pooled and sequenced simultaneously (multiplexed). We experimentally assess the performance of WEGS with four different depth of coverage and sample multiplexing configurations. We show that the optimal WEGS configurations are 1.7–2.0 times cheaper than standard WES (no-plexing), 1.8–2.1 times cheaper than high-depth WGS, reach similar recall and precision rates in detecting coding variants as WES, and capture more population-specific variants in the rest of the genome that are difficult to recover when using genotype imputation methods. We apply WEGS to 862 patients with peripheral artery disease and show that it directly assesses more known disease-associated variants than a typical genotyping array and thousands of non-imputable variants per disease-associated locus.

npj Genomic Medicine (2024)9:8; <https://doi.org/10.1038/s41525-024-00390-3>

INTRODUCTION

Accurate assessment of DNA sequence variation enables insights into the genetic basis of diseases and other traits. Whole genome sequencing (WGS) at high-depth of coverage (30X and above) using next generation sequencing technologies is the current gold standard method for the accurate discovery of single nucleotide variants (SNVs) and short insertions/deletions (InDels) genome-wide^{1,2}. Sequencing offers several advantages over array-based genotyping, notably that variant positions are not fixed, which allows the discovery of novel population-specific variants. Yet, despite the decreasing costs of high-depth WGS, sequencing a large number of samples remains expensive. So far, the use of whole exome sequencing (WES) has dominated large-scale sequencing studies such as gnomAD³ and UK Biobank⁴, but WES is limited to coding regions. As a result, there is still a need for more cost-effective solutions to capture both coding and non-coding variation.

The array-based genotyping coupled with genotype imputation at untyped genomic positions from public haplotype reference panels^{2,5,6} is a popular, cost-effective strategy for increasing statistical power and genomic coverage in current genome-wide association studies (GWAS)⁷. The largest TOPMed haplotype reference panel allows for the imputation of variants down to minor allele frequencies (MAF) of ~0.002–0.003% (imputation quality $r^2 > 0.3$) in individuals of European and African ancestries⁵. However, rare variant imputation with TOPMed still has much

lower accuracy than common variant imputation, especially in non-European or non-African ancestry groups⁶. At the same time, the advantage of local sequencing-based imputation reference panels was demonstrated for multiple populations, such as the Estonian⁸, Finnish⁹ and Sardinian¹⁰.

Several cost-effective sequencing-and-imputation strategies have been described to improve genomic coverage while allowing better assessment of population-specific variants. Those include (a) WGS in a subset of study participants (at a depth ranging from 5X to 30X) to create a customized reference panel⁷ for imputation of the remaining participants who were genotyped using genotyping arrays and (b) ultra-low depth WGS (depth of coverage (DP) down to 0.1X–0.5X) or (c) low-depth (1X–4X) WGS in all study participants followed by imputation using public reference panels^{11–14}. While ultra-low depth WGS can be performed at the same cost as array-based genotyping¹¹, it has also been suggested that ultra-low depth and low-depth sequencing plus imputation are good alternative technologies to imputed genotyping arrays by doubling the number of true association signals discovered¹⁴ and improving the accuracy of polygenic risk prediction models^{12,13}. The latter models have also benefited from the inclusion of rare coding variants in their prediction algorithms^{15–17}. However, recent work suggested that array-based imputation strategies may miss approximately half of the rare coding variants with MAF < 0.05% detected by WES². Although cheaper than WGS, WES is still a more expensive option

¹Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montr al, Qu bec, Canada. ²Victor Phillip Dahdaleh Institute of Genomic Medicine at McGill University, Montr al, Qu bec, Canada. ³Canada Excellence Research Chair in Genomic Medicine, McGill University, Montr al, Qu bec, Canada. ⁴Canadian Centre for Computational Genomics, McGill University, Montr al, Qu bec, Canada. ⁵Division of Angiology, Swiss Cardiovascular Center, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland. ⁶Department for BioMedical Research (DBMR), University of Bern, Bern, Switzerland. ⁷Institute for Cardiovascular Prevention (IPEK), Ludwig-Maximilians University Munich, Pettenkoferstr 9, 80336 Munich, Germany. ⁸Mila - Quebec AI Institute, Montr al, Qu bec, Canada. ✉email: daniel.taliun@mcgill.ca

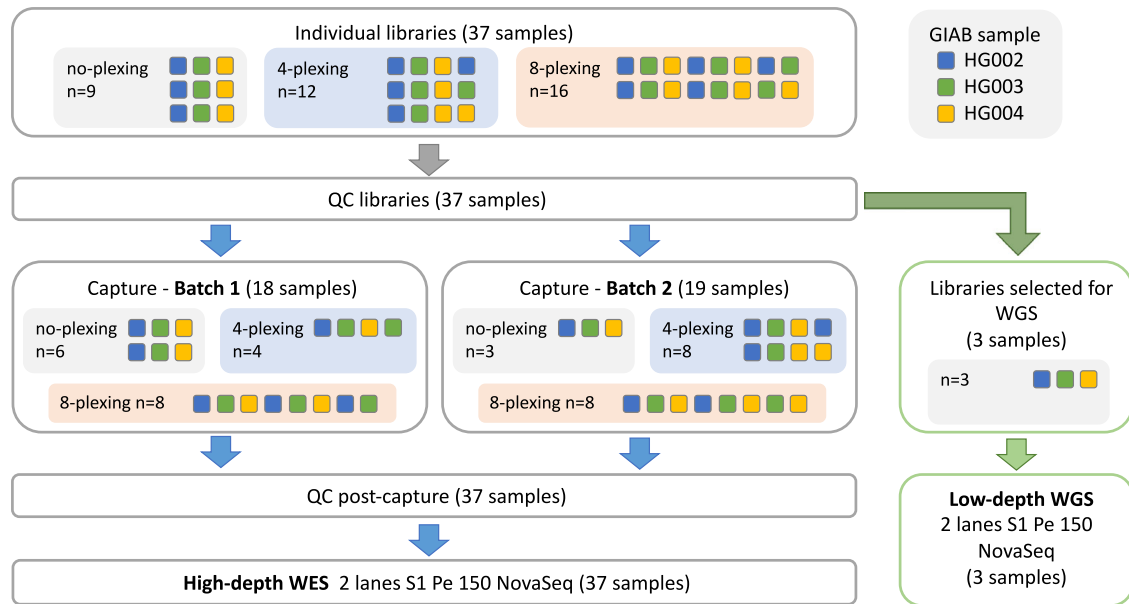


Fig. 1 WEGS experimental design overview. DNA samples from a GIAB family trio (HG002, HG003, HG004) were used to perform WES experiments without and with multiplexing of 4 and 8 samples (no-plexing, 4-plexing and 8-plexing, correspondingly). For each sample in the family trio, we performed library preparation and sequencing to a target coverage of 100X in triplicate for the no-plexing and 4-plexing WES experiments, and in duplicate for the 8-plexing experiment, for a total of 37 samples. Sequencing library QC was performed before and after exome capture. After libraries QC, three individual libraries – one from each sample in the family trio – were selected to perform low-depth WGS on two lanes. Sequencing was performed using the Illumina NovaSeq S1 platform.

than imputation-based strategies, and it ignores the majority of non-coding regions of the genome. Assessment of genetic variation in non-coding regions, which contains the vast majority of genetic variants² and a majority (84%) of GWAS association signals¹⁸, is critical for many genetic analyses, notably understanding regulatory genetic variation.

Here, we propose a cost-effective sequencing method, which we call Whole Exome Genome Sequencing (WEGS), that combines low-depth WGS (2–5X) and high-depth WGS (100X) with up to 8 samples pooled and sequenced simultaneously (multiplexed) to reduce reagents costs¹⁹. We experimentally demonstrate that WEGS, while being 1.7–2.0 times cheaper than standard high-depth WGS (100X) due to multiplexing and 1.8–2.1 times cheaper than 30X WGS, maintains similar precision and recall rates in the discovery of rare coding variants and allows assessment of population-specific variants in the rest of the genome. We demonstrate the scalability and utility of WEGS by applying it to 862 patients with peripheral artery disease (PAD).

RESULTS

Sample multiplexing lowers depth of coverage due to duplicate reads

Sample multiplexing allows multiple samples to be pooled and sequenced simultaneously, resulting in lower per-sample sequencing costs¹⁹. However, multiplexing may also increase the number of false positive variant calls²⁰. To assess sequencing quality, we first compared the DP and variant calling when using WES without and with multiplexing of 4 and 8 samples (no-plexing, 4-plexing and 8-plexing, correspondingly). For this, we generated 37 exome sequences at 100X WES and different levels of multiplexing using DNA from Ashkenazi trio samples (Fig. 1, Supplementary Figure 1 and Methods).

We observed a strong negative correlation (Pearson's $r = -0.69$, P value = 2.31×10^{-6}) between the average DP in targeted exome regions and the number of multiplexed samples (Fig. 2). The

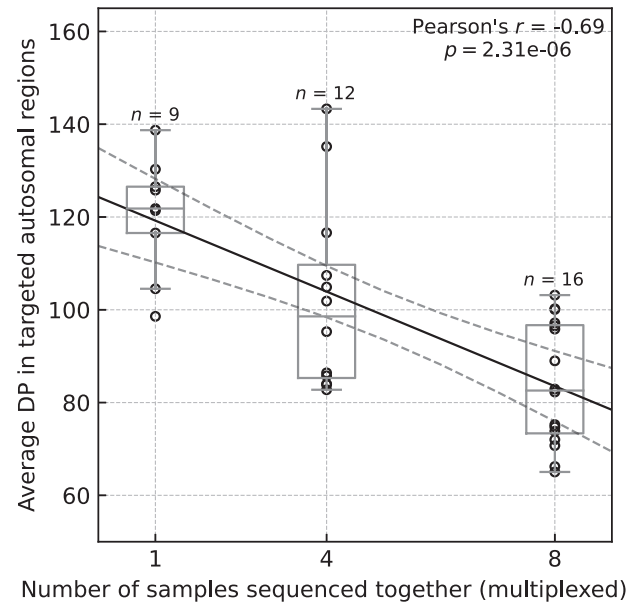


Fig. 2 Average depths of coverage across all targeted regions in autosomal chromosomes in WES experiments without and with multiplexing. The average depth of coverage (DP) was computed across target regions in Agilent V7 capture using paired mapped reads and counting only base-pairs with minimal Phred-scaled mapping and base qualities of 20. The solid black line corresponds to the linear regression line, and the dashed black lines correspond to a 95% confidence interval. The box bounds the IQR, and Tukey-style whiskers extend to a maximum of $1.5 \times$ IQR beyond the box. The horizontal line within the box indicates the median value. Open circles are data points corresponding to the average DP across individual exome.

median values of average DP across individual exomes dropped from 121.8 in no-plexing experiments to 98.6 and 82.6 in 4-plexing and 8-plexing, respectively. The average DP ratio between no-plexing and 4- and 8-plexing experiments was similar in all targeted regions across the exome - showing no evidence that differences in average DP were non-uniform or affected only a subset of targeted regions (Supplementary Fig. 2). When stratifying by library preparation batch, we observed statistically significant differences (P value = 0.048) in average DP between two batches only in experiments without multiplexing (Supplementary Fig. 3A). Nevertheless, these differences did not influence the overall trend - the strong negative correlations between the number of multiplexed samples and DP remained in both library preparation batches (Supplementary Figure 3B–C).

To better understand the cause of lower average DP in multiplex sequencing, we assessed the total number of paired reads, the number of reads flagged as PCR or optical duplicates, the number of unmapped reads, and the average base qualities in reads. There was no correlation (Pearson's $r = -0.08$, P value = 0.631) between the total number of paired reads and the number of samples pooled together for sequencing (Supplementary Figure 4A). However, there was a strong positive correlation (Pearson's $r = 0.92$, P value = 2.13×10^{-15}) between the percent of reads flagged as PCR or optical duplicates and degrees of multiplexing (Supplementary Figure 4B). Compared to the multiplexing-free sequencing experiments, the 4-plexing and 8-plexing experiments showed a 1.7-fold (18.4% vs 31.2%) and 2.3-fold (18.4% vs 43.0%) increase in the median percent of duplicated reads, respectively. The data also suggested a weak, non-statistically significant correlation (Pearson's $r = 0.32$, P value = 0.06) between the percent of unmapped reads and degrees of multiplexing (Supplementary Figure 4C). Also, the percent of unmapped reads did not exceed 0.11 percent of the total number of paired reads and, thus, did not contribute much to the differences in average DP. There was a moderate correlation (Pearson's $r = -0.52$, P -value = 1.09×10^{-3}) between the average base qualities and the degree of multiplexing (Supplementary Fig. 4C). However, when stratified by the library preparation batch and in contrast to the other metrics mentioned above, the first batch did not show the same correlation pattern (Supplementary Figs. 5–8), suggesting that other factors may affect the base qualities. We conclude that the main contributor to the lower average DP in sample multiplexing experiments compared to the experiments without sample multiplexing is the percent of reads flagged as PCR or optical duplicates.

UMI does not recover losses in the depth of coverage

UMI - a unique barcode appended to each DNA fragment before the PCR - helps to distinguish the truly duplicated fragments originating from the same molecule from the very similar fragments originating from a different molecule^{21,22}. In addition, UMI-aware software tools for duplicate read removal help to identify and remove sequencing errors by grouping reads with the same UMI and creating a consensus read²³. We applied the duplex UMI method in our sequencing experiments and evaluated the utility of LocatIt and UmiAwareMarkDuplicatesWithMateCigar (GATK + UMI) UMI-aware read deduplication tools with multiplexing. LocatIt reduced the average DP in experiments without and with multiplexing, compared to the UMI agnostic deduplication approach, while UMI + GATK increased the average DP (Supplementary Figure 9). We explain the different effects on average DP by the difference in strategies between these two tools. For example, in 8-plexing experiments, the GATK + UMI reduced the percent of duplicated reads on average by 0.4 (SE = 0.01), while LocatIt reduced it by 1.56 (SE = 0.03) (Supplementary Table 2). However, LocatIt, on average, marked an additional 4.38% of reads as QC failed, which included reads with

low base qualities in their UMIs and single consensus read pairs without complementary pairs. This additional filtering in LocatIt resulted in lower average DP, fewer unmapped reads, and higher average base qualities. In summary, the UMI-aware read deduplication showed that the vast majority of the duplicated reads in multiplexing experiments are truly PCR/optical duplicates. UMI-aware deduplication didn't help recover the loss in average DP in multiplexing experiments back to the levels of no-plexing experiments.

Sample multiplexing decreases variant recall rates

We observed moderate-to-strong negative correlations between the number of samples sequenced together and the recall rates for SNVs (Pearson's $r = -0.60$, P value = 7.79×10^{-5}) and InDels (Pearson's $r = -0.48$, P value = 2.85×10^{-3}) (Supplementary Figure 10A, C). The average recall rates dropped from 0.983 (SE = 0.0004) and 0.939 (SE = 0.003) in no-plexing experiments to 0.980 (SE = 0.0004) and 0.926 (SE = 0.003) in 8-plexing experiments for SNVs and InDels, respectively (Supplementary Table 3). In many instances, the recall rates were lower in the second library preparation batch, and some of these differences were statistically significant (Supplementary Figure 11A, D). Despite these differences, the statistically significant negative correlations between variant recall rates and the number of multiplexed samples were present in both library preparation batches (Supplementary Figure 11B, C, E, F).

We also observed a drop in precision for both variant types with the increased number of multiplexed samples, but unlike recall, the negative correlations were weaker and not statistically significant (Supplementary Figure 10B, D, Supplementary Table 3). The precision rates were similar between the library preparation batches (Supplementary Figure 12A, D), and they also did not show statistically significant correlations with the number of multiplexed samples when stratified by batch (Supplementary Figure 12B, C, E, F). Only in the first batch we saw a weak positive correlation (Pearson's $r = 0.27$, P value = 0.26) between precision and the number of multiplexed samples (Supplementary Figure 12B).

We looked into the number of true positive (TP), false positive (FP), and false negative (FN) variant calls to explain the statistically significant decrease in recall rates. We found the strongest correlation in the degree of multiplexing and the number of FN calls (Pearson's $r = 0.60$, P value = 8.46×10^{-5} in SNVs and Pearson's $r = 0.44$, P value = 6.34×10^{-3} in InDels), representing the true variants that are not detected (Supplementary Figure 13). For example, the average number of undetected true SNVs increased from 384 (SE = 8) in single-sample sequencing experiments to 446 (SE = 9) in 8-plexing experiments (Supplementary Table 3). On average, 65 (SE = 6) true SNVs missed in 8-plexing experiments were correctly identified across all no-plexing experiments for the corresponding sample, and 61 (SE = 6) of those had a higher depth of coverage in no-plexing experiments than in 8-plexing experiments (Supplementary Table 4). We conclude that the main driver for the decrease in recall rates is the drop in average DP in multiplexing experiments, which leads to the increased number of missed true variants.

UMI improves variant calling insufficiently

We investigated how the recall and precision rates changed in SNV calling after applying the UMI-aware duplicate read removal. We wanted to test if a more accurate read deduplication could partially compensate for the loss of variant recall rates in multiplexing experiments. As previously, we considered two UMI-aware deduplication tools: LocatIt and UmiAwareMarkDuplicatesWithMateCigar (GATK + UMI).

We observed small but statistically significant drops in the recall rates when using LocatIt in all samples at all levels of plexing

(Supplementary Figure 14A). For example, on average, the paired difference in the same sample in the 8-plexing experiment between two recall rates, one measured after LocatIt and another measured after the UMI agnostic approach, was only -0.0008 ($SE = 0.0001$) (Supplementary Table 5). The paired differences between the recall rates were consistently negative in all samples in the 8-plexing experiment, and this relationship was statistically significant (P value = 2×10^{-3}) (Supplementary Figure 14A). However, there was no consistent and statistically significant change in the precision rates: precision slightly increased in some samples but dropped in others (Supplementary Figure 14B). For instance, while, on average, a paired difference in the 8-plexing experiment between precision rates increased by 0.0002 ($SE = 0.0002$) (Supplementary Table 5), the paired differences between precision rates were negative in 5 out of 16 samples and did not support this average increase (P value = 0.85) (Supplementary Figure 14B). The statistically significant decreases and increases were also in the total number of called SNVs and the number of missed true SNVs (i.e. FN calls), respectively (Supplementary Table 5). In samples in the 8-plexing experiments, the average paired difference between the total numbers of called SNVs was -20 ($SE = 4$) and between the numbers of missed true SNVs was 17 ($SE = 2$). Although the average paired difference between the numbers of FP calls was -2 ($SE = 3$) and suggested a decrease in the numbers of FP calls when using LocatIt, this relationship was not statistically significant (P -value ≥ 0.05). The reduced number of called SNVs is consistent with our previous observation of reduced average DP when using LocatIt due to additional read filtering.

When using GATK + UMI, we observed slight but statistically significant improvements in the SNV recall rates for samples in multiplexing experiments (Supplementary Figure 14C). In samples in the 8-plexing experiments, the average paired difference between recall rates was 0.0003 ($SE < 0.0001$) (Supplementary Table 5), and the increase in recall rates was observed in the majority of samples and supported the statistical significance of the relationship (P value = 3.1×10^{-5}) (Supplementary Figure 14C). At the same time, there was also a slight statistically significant drop in the precision rates at all levels of plexing (Supplementary Figure 14D). In the same samples in the 8-plexing experiments, the average paired difference between precision rates was -0.0014 ($SE = 0.0001$) (Supplementary Table 5), and the decrease was consistent across all samples leading to the statistically significant relationship (P -value = 1.5×10^{-5}) (Supplementary Table 5). The observed increase in the number of called SNVs (e.g. $M = 39$ [$SE = 2$] in 8-plexing) and the number of FP calls (e.g. $M = 33$ [$SE = 2$] in 8-plexing) with a much smaller decrease in the number of FN calls (e.g. $M = -6$ [$SE = 1$] in 8-plexing) (Supplementary Table 5) can explain the increase in recall and decrease in precision rates. The increase in the number of called SNVs is consistent with our previous observation of increased DP when using GATK + UMI.

In summary, while UMI-aware read deduplication can improve SNV recall or precision rates depending on the approach, this improvement appears minimal in the present experiment. It does not allow to recover these rates back to levels similar to no-plexing experiments.

WEGS significantly improves variant calling in multiplexed samples

To compensate for the losses in variant recall rates when performing multiplexed WES, we introduced reads from low-depth WGS before variant calling. We called this approach WEGS. We evaluated four combinations in comparison to no-plexing WES: (1) 4-plexing WES and WGS at 2X average DP (WEGS_{4P,2X}), (2) 4-plexing WES and WGS at 5X average DP (WEGS_{4P,5X}), (3) 8-plexing WES and WGS at 2X average DP (WEGS_{8P,2X}), and (4) 8-plexing WES and WGS at 5X average DP (WEGS_{8P,5X}). In each

combination, we looked at the paired difference in the same sample between two recall rates, one measured after adding reads from WGS and another before.

Additional reads from 2X and 5X WGS improved variant recall rates in all multiplexing experiments, and the differences were statistically significant (P value < 0.05) (Supplementary Figures 15 and 16). For instance, the average paired difference in SNV recall rates in WEGS_{8P,2X} was 0.0031 ($SE = 0.0002$) (Supplementary Table 6). This paired difference in recall rates was positive across all samples and, thus, supported the statistical significance of the observed increase in recall rates (P value = 1.5×10^{-5}) (Supplementary Figure 15A). The total number of discovered SNVs increased on average by 76 ($SE = 6$), of which 70 ($SE = 5$) were true positives, explaining the improved recall rates (Supplementary Table 6). Similarly, there were statistically significant improvements in InDel recall rates (Supplementary Figures 15C and 18C). As expected, adding reads from 5X WGS improved the recall rates the most. The average paired difference in SNV recall rates in WEGS_{8P,5X} was 0.0044 ($SE = 0.0003$) compared to 0.0031 ($SE = 0.0002$) in WEGS_{8P,2X} (Supplementary Table 6).

The change in variant precision rates after adding reads from low-depth WGS differed for SNVs and InDels. We observed slight drops in SNV precision rates in all combinations of multiplexing levels in WES and read depths in WGS. However, the declines were not systematic, i.e. they were present only in part of the samples, in contrast to increases in SNV recall rates which were, on average, much higher and present in all samples (Supplementary Figures 15B and 16B). For example, the lowest average paired difference in SNV precision rates among all WES and WGS combinations was -0.0003 ($SE = 0.0001$) in WEGS_{4P,2X} (Supplementary Table 6). It was the only combination where this paired difference in SNV precision rates reached statistical significance (P value = 0.026) (Supplementary Figure 15B). Thus, adding reads from low-depth WGS increased the number of called SNVs by a few dozen, but at the same time, some of these additionally called SNVs were FP, which slightly changed the SNV precision rate in either direction.

Differently from SNVs, all combinations of multiplexing levels in WES and read depths in WGS showed statistically significant improvements in InDel precision rates (P value < 0.05). In WEGS_{8P,2X}, the average paired difference in InDel precision rates was 0.0055 ($SE = 0.0009$) (Supplementary Table 6), and only 1 out of 16 pairs had a negative paired difference between InDel precision rates after and before adding WGS reads (Supplementary Figure 15). In contrast to SNVs, additional reads from 2X WGS raised the average number of called InDels by 10 ($SE = 2$) and, at the same time, decreased the average number of FPs by 6 ($SE = 1$) in 8-plexing WES.

WEGS enhances WES with millions of variants genome-wide

We compared the variant recall rates in standard no-plexing WES to those in multiplexing WES combined with low-depth WGS (Fig. 3A, C). The average SNV and InDel recall rates exceeded the corresponding rates in no-plexing WES for most WEGS configurations, except for WEGS_{8P,2X}. Both WEGS_{4P,2X} and WEGS_{4P,5X} resulted in higher average SNV recall rates than no-plexing WES: 0.9842 ($SE = 0.0002$, P value = 6.4×10^{-3}) and 0.9852 ($SE = 0.0001$, P value = 7.1×10^{-5}) against 0.9830 ($SE = 0.0004$), respectively (Fig. 3A). Among 8-plexing experiments, only WEGS_{8P,5X} resulted in higher average SNV recall rates than no-plexing WES: 0.9847 ($SE = 0.0001$, P -value = 5.6×10^{-4}). Similarly, only WEGS_{4P,2X}, WEGS_{4P,5X}, and WEGS_{8P,5X} statistically significantly increased average InDel recall rates compared to no-plex WES (Fig. 3C). The average InDel recall rate showed a statistically significant increase from 0.9390 ($SE = 0.0029$) in no-plex WES to 0.9493 ($SE = 0.0029$, P value = 0.01), 0.9552 ($SE = 0.0019$, P value = 2.8×10^{-4}), and 0.9490 ($SE = 0.0020$,

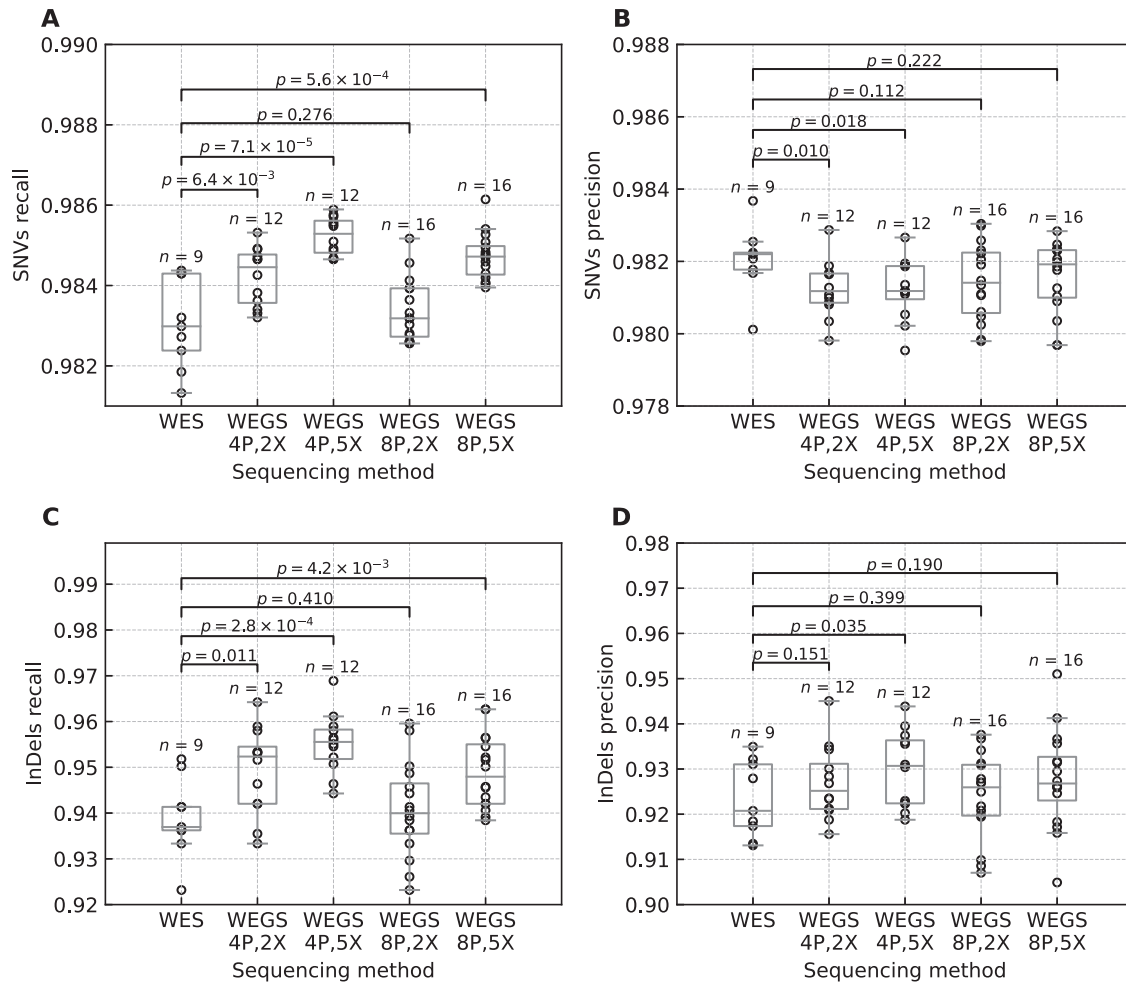


Fig. 3 Variant recall and precision rates in no-plexing WES and WEGS. The figure represents variant calls inside the target regions in Agilent V7 capture and the GIAB high-confidence regions. The box bounds the IQR, and Tukey-style whiskers extend to $1.5 \times$ IQR beyond the box. The horizontal line within the box indicates the median value. Open circles are data points corresponding to the individual WES and WEGS. The p-values above each sequencing method pair correspond to the one-tailed Wilcoxon rank-sum test. **A** Recall rates of the called SNVs. **B** Precision rates of the called SNVs. **C** Recall rates of the called InDels **D** Precision rates of the called InDels. Supplementary Table 6 shows average values and standard errors.

P value = 4.2×10^{-3}) in WEGS_{4P,2X}, WEGS_{4P,5X}, and WEGS_{8P,5X}, respectively. When stratified by the library preparation batch, the average variant recall rates across WEGS remained higher than those in no-plexing WES, except for SNV recall rates in WEGS_{8P,2X} (Supplementary Figures 17E, F, 18E, F). The batch effect in SNV recall rates in WES, described above, also affected WEGS (Supplementary Figure 18D). Despite this, the WEGS_{4P,5X} and WEGS_{8P,5X} had statistically significantly higher SNV recall rates compared to no-plexing WES in both batches, and the increase in WEGS_{4P,2X} was close to statistical significance (Supplementary Figure 17E, F). There were no statistically significant differences in InDel recall rates between the two batches within the no-plexing WES and each WEGS configuration (Supplementary Figure 18D). But only for WEGS_{4P,5X} the increase in InDel recall rates compared to no-plexing WES was statistically significant in both batches. WEGS_{4P,2X} showed a statistically significant increase only in the first batch. WEGS_{8P,2X} showed a statistically significant increase only in the second batch, and the increase in the first batch was close to a statistical significance (P -value = 0.092).

The variant calling precision rates in no-plexing WES compared to WEGS differed depending on the variant type. The average SNV precision rates in every WEGS configuration were slightly lower than in WES, while average InDel precision rates were higher than in WES (Fig. 3B, D, Supplementary Table 7). Only drops in average

SNV precision rates in WEGS_{4P,2X} and WEGS_{4P,5X}, and an increase in the average InDel precision rate in WEGS_{4P,5X} were statistically significant (Supplementary Table 7). Furthermore, when stratified by the library preparation batch, the decreases in average SNV precision rates in WEGS compared to no-plexing WES were statistically significant only in the second batch (Supplementary Figure 17A–C, Supplementary Table 8). In contrast, WEGS_{8P,2X} and WEGS_{8P,5X} demonstrated an increase in average SNV precision rates compared to no-plexing WES in the first batch. We explain this by the initially lower precision rates in multiplexing WES experiments in the second batch as described above. When stratified the average InDel recall rates by the library preparation batches, the average InDel precision rates in WEGS remained higher than in no-plexing WES for all configurations except WEGS_{8P,2X} (Supplementary Figure 18A–C, Supplementary Table 8). However, none of the increases remained statistically significant.

We also compared variant recall and precision rates in WES and WEGS to the 30X WGS, which we generated by downsampling reads from 300X WGS data (see Methods). Average variant recall and precision rates inside regions targeted by WES were higher in 30X WGS compared to WES and WEGS. For SNVs, these differences were below 0.7%, while for InDels, the maximal difference reached 6% (Supplementary Table 9). WEGS_{4P,2X}, WEGS_{4P,5X}, and WEGS_{8P,5X} were closer to 30X WGS than WES in targeted regions. 30X WGS

had no rivals when comparing genome-wide recall and precision rates. On average, it found 1.7–2.5 times more SNVs and 2.5–3.8 times more InDels genome-wide than WEGS (Supplementary Table 10). The average genome-wide SNV and InDel precision rates in WGS were up to 18% and 70% higher than in WEGS, respectively. As expected, WEGS_{4P,5X} and WEGS_{8P,5X} were the closest to the 30X WEGS.

In summary, these results confirm that our WEGS approach eliminates the negative impact of sample multiplexing in WES on variant recall rates in coding regions and brings variant recall rates to the levels of a standard no-plexing WES or higher. Furthermore, these results suggest that WEGS_{4P,2X}, WEGS_{4P,5X} and WEGS_{8P,5X} are the closest alternatives to no-plexing WES, as these sequencing strategies demonstrated statistically significant increases in SNV and InDel recall rates and, at the same time, showed increases in InDel precision rates in targeted regions. WEGS has a clear advantage over WES by allowing the assessment of additional ~2 M SNVs and InDels per individual genome-wide.

WEGS correctly assesses variants which genotype imputation misses

Next, we wanted to understand what other benefits low-depth WGS data could bring to multiplexed WES besides removing the negative effects of sample multiplexing. We compared WEGS to array-based genotyping followed by genotype imputation. For each of our three samples, HG002, HG003, and HG004, we emulated the genotyping array data covering 654,013 genetic positions and performed genotype imputation using the TOPMed reference panel consisting of 97,256 diverse genomes. We compared these imputation results to WEGS_{4P,2X} and WEGS_{8P,5X}, the closest alternatives to no-plexing WES in targeted regions.

First, we investigated regions targeted by WEGS. SNVs imputed from emulated genotyping array data showed high precision rates (> 99%) for all three samples, but imputation missed between 824 to 1,028 SNVs per sample (among them, between 482 to 576 were non-synonymous) compared to WEGS (Supplementary Table 11). For example, in sample HG002, WEGS_{8P,5X} correctly identified 22,390 SNVs on average, and the TOPMed reference panel imputed only 21,458 SNVs, which is 938 SNVs less. The difference in the number of correctly identified InDels was even larger: imputation missed around 60% of true InDels (40% recall), while WEGS only missed around 5% (95% recall).

Second, we investigated the number of imputed and sequenced variants genome-wide (Supplementary Table 12). In contrast to the WEGS targeted regions, the genotyping array-based imputation approach outperformed WEGS by the number of correctly identified SNVs: imputation missed 4–5% (95–96% recall), WEGS_{4P,2X} missed 54–65% (35–46% recall), and WEGS_{8P,5X} missed 36–50% (50–64% recall) of true SNVs. The differences in correctly identified InDels were much lower: imputation missed around 61% (39% recall), WEGS_{4P,2X} missed 69–78% (22–31% recall), and WEGS_{8P,5X} missed 53–67% (33–47% recall) of true InDels.

Third, we looked at how many variants missed or wrongly imputed outside non-protein coding regions can be recovered by WEGS. We grouped TOPMed-imputed variants outside WEGS-targeted non-protein-coding regions into three categories: (1) the number of imputed alleles matches the number of true alleles (i.e. imputation is correct); (2) the number of imputed alleles is less than the number of true alleles; (3) the number of imputed alleles is higher than the number of true alleles. For each of these groups, we looked at the median fold change in alternate AF between the ASJ population and TOPMed. The median fold-change in AF was higher (i.e. AF in ASJ was higher than in TOPMed) when imputation was systematically missing alleles (group 2) and lower when imputation was wrongly imputing extra allele(s) (group 3) (Supplementary Table 13). This result is in line with previous

studies^{24,25}, which showed that the imputation accuracy depends on the genetic similarity between the study individual and the reference panel. WEGS_{4P,2X} correctly identified true alleles in 38–46% of variants in group 2 and 89–92% of variants in group 3, while WEGS_{8P,5X} correctly identified true alleles in 55–67% of variants in group 2 and 91–94% in group 3.

Finally, to improve the variant recall in non-coding regions in WEGS, we evaluated the applicability of the GLIMPSE method²⁶, developed to impute missing variants from low-depth WGS data. After applying GLIMPSE to WEGS_{4P,2X} and WEGS_{8P,5X} with local reference haplotypes from the 1000 Genomes Project and Human Genome Diversity Project (see Methods), genome-wide SNV recall rates and precision increased drastically. In imputed WEGS_{4P,2X}, the average genome-wide SNV recall rate and precision increased from ~35–46% to ~69–81% and from ~80–82% to ~94–95%, respectively (Supplementary Tables 12 and 14). In imputed WEGS_{8P,5X}, the average genome-wide SNV recall rate and precision increased from ~50–65% to ~79–89% and from ~87–90% to ~95–96%, respectively. The genome-wide recall rate and precision also increased for InDels. When considering the GLIMPSE-imputed variants only, i.e. without merging them with variants called in sequencing data only, the SNVs precision rate was very high and greater than 99% across all sequencing experiments (Supplementary Table 15). The SNV recall rate in sequence-based imputation was still lower than in genotyping array-based imputation. One of the possible explanations is that the state-of-the-art TOPMed reference panel contains >20 times more haplotypes than our local reference panel. To confirm this, we run the genotyping array-based imputation using our local reference panel and the Minimac4 tool²⁷. The Minimac4-imputed recall rates for SNVs were only slightly higher than the GLIMPSE-imputed WEGS_{8P,5X} and much higher than the GLIMPSE-imputed WEGS_{8P,2X} (Supplementary Table 16). However, the precision rates of GLIMPSE-imputed SNVs were always much higher than those of the Minimac4-imputed SNVs. When considering imputed variants only, the recall rates for imputed InDels were similar in the sequence-based imputation and genotyping array-based imputation using TOPMed, but were lower compared to genotyping array-based imputation using the local reference panel. The InDels recall rates became closer to Minimac4-imputation results when combining imputed and called InDels together, but at the expense of precision.

In summary, these results showed that WEGS outperforms the genotyping array and imputation approach in terms of the number of identified variants, especially InDels, inside protein-coding regions. Outside protein-coding regions, WEGS allows one to discover genetic variants missed by genotyping array-based imputation due to their population specificity. Sequencing-based imputation methods can be applied to WEGS to recover variants missed due to lower depth of coverage outside protein-coding regions. WEGS_{8P,5X} has a clear advantage over WEGS_{4P,2X} outside the protein-coding region due to the higher depth of coverage in the WGS experiment.

WEGS is substantially cheaper than high-depth WES and WGS

We compared costs for WEGS scenarios relative to genotyping arrays, low-depth WGS, 30X WGS and no-plexing 100X WES. Per sample cost estimates for the genotyping array included DNA QC and genotyping using Affymetrix Axiom UKBB array. Sequencing costs per sample were based on current pricing and a scenario of 1,000 samples sequenced on the Illumina NovaSeq 6000, S4 platform. We note that sequencing costs can vary depending on multiple factors, including reagents pricing, flow cell volume and sequencing platform, while genotyping array prices are less affected by sample size.

Our estimates show that the combinations of WEGS_{4P,2X} and WEGS_{8P,5X} are half the price compared to standard 100X WES (no-

Table 1. Relative genotyping and sequencing costs per sample given current pricing.

	Cost relative to array	Cost relative to 2X WGS	Cost relative to 5X WGS	Cost relative to 30X WGS	Cost relative to 100X WES
Axiom array	1.00	0.52	0.33	0.08	0.09
2X WGS	1.91	1.00	0.63	0.15	0.17
5X WGS	3.04	1.59	1.00	0.24	0.26
30X WGS	12.48	6.52	4.10	1.00	1.08
100X WES	11.50	6.01	3.78	0.92	1.00
100X WES 4plex	5.05	2.64	1.66	0.40	0.44
100X WES 8plex	3.95	2.07	1.30	0.32	0.34
100X WES 4plex + 2X WGS	5.80	3.03	1.91	0.47	0.50
100X WES 4plex + 5X WGS	6.93	3.63	2.28	0.56	0.60
100X WES 8plex + 2X WGS	4.71	2.46	1.55	0.38	0.41
100X WES 8plex + 5X WGS	5.84	3.05	1.92	0.47	0.51

Per sample cost estimates for the genotyping array include DNA QC and genotyping using Affymetrix Axiom UKBB array. Sequencing scenarios are based on 1000 samples sequenced on Illumina NovaSeq 6000, S4 platform. These cost estimates include sample preparation steps from DNA QC (QC, gDNA, high throughput) to Illumina library preparation and capture for Agilent SureSelect XT HS2 V7, sequencing library QC, and Illumina sequencing. Bold values represent the combinations which are considered possible alternatives to standard WES and 30X WGS

plexing) and ~47% of the price of 30X WGS (Table 1). The combination of 5X WGS with 4-plexing WES is slightly more expensive but still 56% of the cost of 30X WGS and 60% of the cost of no-plexing 100X WES. As such, the WEGS scenarios representing the most economical strategies relative to WGS and WES are again the combinations of 2X WGS with 4-plexing WES and 5X WGS with 8-plexing WES. Yet, as shown above, while WEGS_{4P,2X} and WEGS_{8P,5X} show comparable precision and recall in targeted coding regions relative to standard WES, the latter combination is more effective at capturing noncoding variation. As such, we conclude that the most cost-effective WEGS strategy to capture both coding and non-coding variants is 5X WGS with 8-plexing high-depth WES.

WEGS applied to the study of peripheral artery disease

We applied WEGS_{8P,4X} to 862 patients diagnosed with PAD. Based on the genetic ancestry analyses (see Methods), 780 (90.5%) PAD patients were inferred as Europeans, 60 (7.9%) as Admixed Americans, 7 (0.8%) as Africans, 4 (0.6%) as Asians and 3 (0.3%) as middle Eastern (Supplementary Figure 19). The GIAB control sample included in each of the 10 plates showed similar precision and recall for SNVs and InDels as in the benchmark experiment (Supplementary Figure 20). After variant level filtering (see Methods), we identified 44,747,114 genetic variants (33,505,105 SNVs and 11,242,009 InDels) in PAD samples (Table 2). A total of 12,893,703 of these variants were novel (not described in dbSNP v109.3), from which 63.8% were singletons (carried by one individual). Inside the coding regions, we observed 35.4% synonymous (11,053 per individual), 59.0% non-synonymous (11,636 per individual), 1.1% stop/essential splice (490 per individual), 2.1% frameshift (298 per individual), and 2.3% (371 per individual) inframe genetic variants.

We evaluated the WEGS ability to capture known loci associated with PAD identified by large-scale GWAS²⁸. All lead variants mapping to these loci were present in the PAD WEGS data (Supplementary Table 17). The majority of the lead variants are intergenic, with an average read depth of 13.7. Only 6 out of the 19 lead variants are directly typed onto the Global Screening Array (GSA) 24.v3; demonstrating the WEGS potentials to assess disease-causing variants beyond the genotyping arrays. In addition, we observed that WEGS captured, on average, 4056 (SE = 295) genetic variants within the known PAD loci that are not present in the TOPMed imputation reference panel and, thus, could not be

imputed (Supplementary Table 18). Although the majority of these loci are intergenic, WEGS was able to identify additional missense variants within these regions.

DISCUSSION

In this work, we propose and evaluate a new sequencing method which we call WEGS, designed to be more economical than WES and WGS. We considered WEGS based on WES (100X) with sample multiplexing, i.e. pooling and sequencing up to 8 samples simultaneously, combined with the low-depth WGS (2-5X). First, we evaluated the effect of sample multiplexing in WES. We demonstrated that an increased number of PCR/optical read duplicates in multiplexing WES experiments leads to the loss of depth of coverage and, consecutively, to a higher number of missed true variants. Second, we showed that although the UMI-aware read deduplication helps improve variant calling recall or precision rates, the improvements are minimal and do not compensate for the losses due to multiplexing. Third, we demonstrated that combining reads from low-depth WGS and reads from multiplexing WES brings variant calling recall and precision rates in protein-coding regions to the levels of no-plexing WES or above. Specifically, based on our experiments, we recommend using combinations of 2X WGS with 4-plexing WES and 5X WGS with 8-plexing WES as an alternative to standard WES.

When choosing between different WEGS configurations, it is essential to also consider performance outside the protein-coding regions. Specifically, we demonstrated that WEGS allows for the identification of population-specific non-coding genetic variants, which large genotype imputation panels impute less accurately due to differences in allele frequencies between the study population and reference. If there is no available imputation reference panel closely matching the study population, then the 8-plex WES with 5X WGS would be the best option compared to the 4-plex WES with 2X WGS. Also, our cost estimates suggest that WEGS relying on 8-plexing WES and 5X WGS is the most cost-effective configuration and is 2X cheaper than standard no-plex WES and 2.1X cheaper than high-depth WGS. We used this WEGS configuration on 862 samples with PAD to demonstrate the scalability and applicability of the method in a practical setting, assessing almost 3 M variations (24,000 in coding regions) per individual genome on average. Most novel variants were rare and/or present in only one PAD patient. Thus, we expect WEGS to have a major contribution to the discovery of novel rare variants

Table 2. The number of variants discovered in WEGS sequencing data from 862 patients with peripheral artery disease.

	All Individuals (N = 862)			Per Individual
	Total	Singletons	Rare (MAF < 1%)	Average
Depth at non-targeted regions (X)	--	--	--	4.5 (± 0.88)
Depth at targeted regions (X)	--	--	--	114.8 (± 4.47)
Total variants	44,747,114	14,542,812	32,497,956	2,964,080
SNVs	33,505,105	11,291,220	24,576,783	2,587,752
InDels	11,242,009	3,226,457	7,921,173	449,829
Novel variants	12,893,703	8,226,183	12,712,777	26,984
SNVs	9,363,157	5,739,615	9,233,592	18,922
InDels	3,530,546	2,481,974	3,529,310	8062
Coding variants	348,410	181,870	284,434	23,854
Synonymous	123,337	57,968	94,275	11,053
Non-synonymous	205,570	113,475	173,546	11,636
Stop/essential splice	3967	2325	3426	490
Frameshift	7375	4484	6520	298
Inframe	8069	3566	6589	371
Novel coding variants	31,821	28,448	31,713	66
Synonymous	7680	6920	7,675	11
Non-synonymous	19,761	17,884	19,731	31
Stop/essential splice	1071	865	1045	12
Frameshift	2166	1897	2159	8
Inframe	1106	872	1067	7

This table reports the total number of sequenced variants in the overall patient group and the average number of sequenced variants per individual across different functional classifications. Novel variants were defined as variants not present in dbSNP (version 109.3).

implicated in the disease under investigation. We further show that at a lower price than WES, WEGS captures variants in loci known to be associated with PAD, including non-synonymous variants to be investigated in future studies. Lastly, in intergenic regions, WEGS captures a larger number of GWAS lead variants compared to the common genotyping array.

The WEGS data processing pipelines are built on existing open-source software tools and, thus, do not require time and financial investments in tool development. This work demonstrated how the industry-standard GATK toolset²⁹ could be utilized for SNVs and InDels calling and filtering from WEGS data (see Data and Code Availability). Novel genotype imputation methods, such as GLIMPSE²⁶, are available for sequencing data and can be applied to WEGS to further increase the number of identified non-coding variants.

Our study has several limitations. First, benchmarking analyses relied on high-confidence variant calls from a GIAB trio. As benchmarking call sets will become available for regions difficult for variant detection (i.e. outside high-confidence regions), it will be interesting to investigate WEGS performance in these regions. Second, our analysis focused on SNVs and InDels only, as WES and low-depth WGS are known to have limited utility for structural variant calling. Third, our experiments were based on DNA extracted from cell lines and blood. It was shown previously that WGS based on DNA extracted from blood yields better sequencing data metrics compared to saliva and buccal swabs, but this has negligible impact on the accuracy of short variant detection, although some saliva and buccal swab samples can show higher false positive rate³⁰. As such, we expect that WEGS based on DNA extracted from saliva or buccal cells to show similar performance as in our experiment, as long as samples have sufficient DNA concentration and are not contaminated. Fourth, while our precision and recall estimates were broadly consistent across replicates, we acknowledge that they are based on only 3

individual genomes from a single ancestry. Extension of this work could include an investigation of WEGS performance in individuals from other ancestries. Yet, based on our results and recent work assessing the advantages of low-depth WGS¹¹, we expect WEGS to be of particular interest for populations currently underrepresented in public reference panels, enabling the discovery of novel population-specific variants. Fifth, our benchmark experiments using GIAB samples aimed to identify the effects of sample multiplexing and additional WGS reads in WES on raw variant calls. However, in clinical and research settings, we suggest applying recommended automated filtering (e.g. GATK's Variant Quality Score Recalibration [VQSR]³¹) or hard-filtering of variant calls³², which can eliminate many technical artifacts. When we applied GATK's recommended hard filters to GIAB samples, the variant calling precision increased, and for SNVs, it was 99% or above at depths greater than 40X but never reached 100% (Supplementary Figure 20). Thus, similarly to WES³³, additional validation of variants detected using WEGS, especially with lower depths³⁴, may be needed in clinical settings. InDels need more careful interpretation since their precision reaches only 98%, even at the highest depths and after filtering. Lastly, although our results suggest that the precision of the variants after imputation can reach 99% (depending on the imputation reference panel, variant type, sample size, strategy used when merging imputed and called variants, and whether variants were inside targeted regions), the additional quality control of imputed variants is required. For example, the GLIMPSE method, used in this work, provides an IMPUTE information measure (INFO score)³⁵, which is widely used in GWAS to select well-imputed variants (e.g. INFO score > 0.8). We note that the INFO score and other similar imputation accuracy measures are less reliable for rare variants, requiring more sophisticated measures such as those described by Sun et al., 2022³⁶.

We anticipate that WEGS will become a method of choice for studies of the molecular genetic basis of diseases and disease-related traits. Such genetic association studies require many sequenced individuals to reach sufficient statistical power and capacity to detect rare variants. Today, it remains costly to use high-depth WGS; for example, high-depth WGS for 1,000 samples currently costs close to 1 million US dollars, and standard WES can be up to 90% of this figure. As such, a 50% cost reduction when using WEGS will enable high-depth sequencing of up to twice the number of exomes while providing additional information genome-wide. Our cost estimates are based on current pricing, but these relative costs should hold as long as WES reagents costs remain low compared to WGS costs. As such, WEGS should remain competitive until WGS costs become substantially lower than currently. The real impact on association studies will be shown in future studies using WEGS or similar technologies.

METHODS

DNA samples for benchmarking experiments

To benchmark our new method, we used DNA samples derived from cell lines obtained from the US National Institute of Standards and Technology (NIST) RM 8392, a family trio of Ashkenazi Jewish origin including a son (HG002), father (HG003) and mother (HG004), consented by the Personal Genome Project (PGP)³⁷. These DNA samples were developed for the Genome in a Bottle (GIAB) Consortium to generate reference datasets for benchmarking genomic analyses³⁸, and have broad, open consent for all research uses under the terms of the PGP.

Benchmarking experimental study design

To assess the relative performance of different WEGS protocols, we used DNA samples from the Ashkenazi trio to perform a series of WES and low-depth WGS sequencing experiments. For WES, we performed experiments without and with multiplexing of 4 and 8 samples (no-plexing, 4-plexing and 8-plexing, correspondingly). For each sample in the family trio, we performed library preparation and sequencing to a target DP of 100X in triplicate for the 1-plex and 4-plex WES experiments, and in duplicate for the 8-plex experiment, for a total of 37 samples (Fig. 1). For WGS, using pre-capture libraries prepared for WES, we sequenced the trio samples to a target DP of 5X on 2 separate lanes. This allowed us to use a single lane to obtain a target DP of 2.5X. This gave us the possibility to evaluate four WEGS combinations: WEGS_{4P,2X}, WEGS_{4P,5X}, WEGS_{8P,2X}, and WEGS_{8P,5X}, where 4P and 8P denote 4- and 8-plexing, respectively, and 2X and 5X correspond to target DP of WGS.

Sequence data production

WES and WGS sequencing was performed at the McGill Genome Centre in October 2021. Processing included sample quality control (QC) using a QUBIT 1X DSDNA HS ASSAY KT from Life Technologies Inc. to measure DNA concentration quality. An aliquot of 200 ng input in 50 µl total was used to perform DNA fragmentation (shearing) with Covaris LE220 (Covaris Inc.) method to a target of 300 bp fragments. Sample library preparation was carried out using Agilent SureSelect XT HS2. Subsequent captures were performed using Agilent SureSelect XT HS2 V7 capture panel with different plexing strategies: 4-plex (12 samples) and 8-plex (16 samples) (Fig. 1). Unique dual sample indexing barcodes (2x8bp) were added to multiplexed samples during library preparation. Library QC was performed before and after capture in 2 steps: quantification using qPCR (Kapa Biosystems, part #KK4602) and QC using LabChip GX Touch HT Nucleic Acid Analyzer. Exome captures were performed in 2 batches using Agilent SureSelect Human All Exon V7 capture for a total 48.2-Mb

target. Sequencing was performed on 2 lanes of the Illumina NovaSeq platform using S1 flowcells and 150-bp paired-end reads to a target coverage of 100X. Sample pre-capture libraries were used to perform WGS sequencing to a target coverage of 5X in 2 separate lanes on the Illumina NovaSeq platform using S1 flowcells to 150-bp paired-end reads.

Data processing and variant calling

As defined by Genome Analysis Tool Kit³¹ (GATK v4.2.0.0) best practice recommendations, preprocessed reads trimmed by the removal of adapters and low quality bases, were aligned to the decoy version of GRCh37 human genome build (hs37d5) using bwa-mem³⁹ (v0.7.17) (Supplementary Figure 1). Mapped reads were further refined using GATK InDel realignment³¹ (v3.8) to improve the mapping of reads near InDels, marking of duplicated reads using GATK mark duplicates, and improve base quality scores using Base Quality Score Recalibration (BQSR). For WEGS processing, WGS and WES were analyzed by applying the above methods but using different trimming and mark duplication procedures to take advantage of the UMIs present in the WES data. The trimmer and locatIT programs from Agilent's AGeNT tool set (v2.0.5) were used to first identify and remove the adaptor sequences, extract the molecular barcodes (MBC), and then merge duplicated reads by leveraging the MBC information embedded in the aligned BAM file. WGS data were processed using the read trimmer skewer⁴⁰ (v0.2.2), and duplicated reads were assessed using GATK mark duplicates. Variant calling for all the experiments was performed using the GATK's HaplotypeCaller.

Benchmark variant calls and regions

Benchmark (or "high-confidence") variant calls for SNVs and short InDels from GIAB Consortium for each sample in the Ashkenazi family trio were obtained for build GRCh37 (v.4.2.1)⁴¹ at URL: <https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/>. The family structure information was used by the GIAB consortium when constructing high-confidence variant calls through Mendelian inheritance analyses and trio-based phasing⁴². We used Illumina hap.py benchmarking tool (version v.0.3.10) to compare our study variant calls and imputed variants to GIAB "high-confidence" variant calls in previously described "high-confidence regions"^{43,44}. Variant calling recall rate was estimated as the total number of true positive variant calls divided by the total number of variant calls, and precision as the total number of true positive variant calls over the sum of true positive and false positive variant calls. We used imputed best-guess genotypes when estimating recall and precision rates for imputed variants against GIAB "high-confidence" variant calls. Before benchmarking, we did not apply filters on our variant calls (e.g. using variant calling annotations or information on Mendelian inheritance errors from the family structure) or imputed genotypes (e.g. using imputation quality scores) to limit the contribution of other factors when interpreting differences between methods.

High-depth WGS data

To generate 30X WGS data for the Ashkenazi Jewish trio (HG002, HG003, and HG004), we downloaded 300X WGS data from GIAB produced using Illumina HiSeq 2500 in Rapid mode (v1) (PCR-free, pair-end, mean read length 2 x 148 bp). The reads were aligned to the GRCh37 genome build using Novoalign version 3.02.07. Then, we randomly subset 10% of the reads using the samtools⁴⁵ tool to reach 30X coverage on average. For each individual, we generated five 30X WGS datasets using different random seeds. Then, we performed variant calling using GATK v4.2 in the same way as for other experiments.

Tests for statistical significance

We used Wilcoxon rank-sum test to test for statistical significance (1) of differences between two library preparation batches and (2) of variant recall and precision rates between no-plexing WES and WEGS. We used Wilcoxon signed-rank test when comparing the same WES experiments (1) before and after UMI-aware read deduplication and (2) before and after adding WGS reads. We used a one-sided version of the tests depending on the means of two samples, i.e. the alternative hypothesis was that the distribution underlying the sample with a larger mean is stochastically greater than the distribution underlying the sample with a smaller mean. We used the implementation of both tests available in SciPy⁴⁶. To assess the correlation strength between levels of multiplexing and different sequence data metrics, we used the Pearson correlation coefficient and the corresponding *P* values for the two-sided alternative hypothesis that the correlation is non-zero implemented in SciPy⁴⁶. We used *P* value < 0.05 for the statistical significance threshold.

Genotype imputation using genotyping arrays

To mimic genotyping array data for HG002, HG003, and HG004 samples, we subset 654,013 GRCh37 positions on the Infinium Global Screening Array 24 v3 (https://support.illumina.com/array/array_kits/infinium-global-screening-array/downloads.html) from the corresponding GIAB's WGS data. At subset positions from WGS data, each GIAB sample on average carried 150,865 SNVs and 1,309 InDels (150,865 SNVs and 1382 InDels in HG002, 150,383 SNVs and 1228 InDels in HG003, 151,348 SNVs and 1,318 InDels in HG004). The median absolute length of InDels was three base pairs in all samples, and the average absolute length varied between 7 and 14 base pairs (14 in HG002 and HG003, 7 in HG004). 99% of all InDels were shorter than 39 base pairs in all samples, and only 4 InDels spanned more than 100 base pairs. Then, we imputed each sample individually using the multi-ethnic TOPMed reference panel (*N* = 97,256) available at NHLBI TOPMed Imputation Server. In addition to genotype imputation, the server lifted positions from GRCh37 to GRCh38 genome build and performed reference-based statistical phasing. There were 20,880,237 InDels out of 292,058,121 imputed variants in the TOPMed reference panel with an average absolute length of 3 base pairs and a maximal length of 69 base pairs, and 99% of them were shorter than 20 base pairs. The imputed genotypes were on the GRCh38 genome build. To compare them to WEGS, we used the GATK LiftoverVcf tool⁴⁷ to lift imputed positions back to GRCh37. We annotated the data after imputation with alternate allele frequencies (AF) in the Ashkenazi Jewish (ASJ) population from gnomAD v3.1.1³ and overall AF in the BRAVO variant browser, which includes all individuals in the TOPMed reference panel. For both databases, we lifted the GRCh38 positions to GRCh37 using GATK LiftoverVcf. We used only those variants, which passed all quality filters described by gnomAD and TOPMed, correspondingly. When comparing AF distributions in ASJ vs BRAVO, we restricted our analyses to nonmonomorphic genetic variants where at least 1000 ASJ individuals were sequenced.

Genotype imputation using WEGS and local reference panel

We used the GLIMPSE method²⁶ to impute variants from the local reference panel using sequencing reads in WEGS. To build our local reference panel, we used genotypes from the 1000 Genomes Project (1000 G)⁶ and Human Genome Diversity Project (HGDP)⁴⁸ (*N* = 4150) from gnomAD v3³. We kept only variants, which passed all quality filters defined by gnomAD v3, were missing in <1% of individuals, and for which alternate allele count was ≥2. We phased the genotypes using statistical phasing implemented in SHAPEIT⁴⁹ and lifted positions of phased genotypes from

GRCh38 to GRCh37 genome build using the GATK LiftoverVcf tool. There were 5,781,236 InDels out of 59,158,489 variants in the local reference panel with an average absolute length of 4 base pairs and a maximal length of 307 base pairs, and 99% of them were shorter than 32 base pairs. We merged the GLIMPSE-imputed variants with variants directly called from WEGS by GATK. We kept the imputed version when a variant was imputed and called at the same time (i.e. had the same position and alleles).

WEGS application

A total of 862 patients diagnosed with PAD were recruited and provided written consent to use their health-related data and samples for research purposes between April 2, 2017, and September 21, 2021, in the Division of Angiology at the Insel University Hospital of Bern, Switzerland. The PAD study was reviewed and approved by the cantonal ethics committee for research of the Directorate of Health, Social Affairs and Integration of the Canton of Bern Switzerland (Kantonale Ethikkommission Bern) (Project-ID:2021-00055). The WEGS data analyses in the PAD study were also approved by the Research Ethics Office (IRB) (IRB Study Number: A07-M42-21B (21-07025)) of the Faculty of Medicine and Health Sciences at McGill University, Canada. Recruited patients had whole blood samples collected and stored in the Liquid Biobank Bern (LBB). We applied the above WEGS method to each sample using WGS at an average depth close to 5X and WES at 100X. Exomes were captured in 8-plex using the Agilent SureSelect All Exons Human V7 capture. The exome and whole genome libraries were sequenced on MGI T7 sequencers. All sequence reads were mapped to build GRCh38. We followed GATK best practices pipelines for jointly calling SNVs and InDels. We used only those variants, which passed all variant filters after GATK's VQSR and had less than 1% missing genotypes. To control for possible batch effects and assess quality of the sequencing, we included a control sample (HG002) on each of the 10 used plates, resulting in 10 replicates. On average, in the control sample, we obtained 96.85% precision and 97.88% recall for SNVs and 84.90% precision and 89.55% recall for InDels in target regions before applying variant filters (Supplementary Figure 21, Supplementary Table 1). The precision was consistently higher than 99% for SNVs and 90% for InDels at DP > 40X after applying variant-level filtering (Supplementary Figure 22). Genetic ancestry was estimated for PAD patient samples by projecting sample sequenced data to publicly available genotypes from the 1000 Genomes (1KG) and Human Genome Diversity Project (HGDP) [HGDP + 1KG] callset from gnomAD v3.1³. Projected principal component (PC) scores were generated with LASER software⁵⁰ and used to infer genetic ancestry by a random forest model trained on the HGDP + 1KG callset using scikit-learn software⁵¹.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The GIAB family trio datasets, including the WES, WGS, and WEGS, are publicly available for reproducibility purposes at the BioProject ID PRJNA1043666 on the NCBI Sequence Read Archive (SRA), and at the Genetics and Genomics Analysis Platform (GenAP) using the <https://datahub-778-pbbb.p.genap.ca/> link. Access to the generated genetic data for the PAD study is subject to approval and a data-sharing agreement due to participants' consent; requests to access the data should be directed to Marc Schindewolf.

CODE AVAILABILITY

The variant calling and analysis pipelines and source code are publicly available through the following version control repositories: Source code for the tables and figures (https://github.com/CERC-Genomic-Medicine/WEGS_paper); Scalable Nextflow pipelines for variant calling in WGS, WES and WEGS using GATK (https://github.com/CERC-Genomic-Medicine/WEGS_pipelines); A scalable Nextflow pipeline for genotype imputation from low-depth sequencing data using GATK and GLIMPSE (https://github.com/CERC-Genomic-Medicine/glimpse_pipeline); A scalable Nextflow pipeline for statistical genotype phasing using SHAPEIT4 (https://github.com/CERC-Genomic-Medicine/shapeit_pipeline).

Received: 12 May 2023; Accepted: 7 December 2023;

Published online: 07 February 2024

REFERENCES

- Halldorsson, B. V. et al. The sequences of 150,119 genomes in the UK Biobank. *Nature* **607**, 732–740 (2022).
- Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
- Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
- Backman, J. D. et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).
- McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
- 1000 Genomes Project Consortium. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- Quick, C. et al. Sequencing and imputation in GWAS: Cost-effective strategies to increase power and genomic coverage across diverse populations. *Genet. Epidemiol.* **44**, 537–549 Preprint at <https://doi.org/10.1002/gepi.22326> (2020).
- Mitt, M. et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet.* **25**, 869–876 (2017).
- Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
- Pistis, G. et al. Rare variant genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs. *Eur. J. Hum. Genet.* **23**, 975–983 (2015).
- Martin, A. R. et al. Low-coverage sequencing cost-effectively detects known and novel variation in underrepresented populations. *Am. J. Hum. Genet.* **108**, 656–668 (2021).
- Homburger, J. R. et al. Low coverage whole genome sequencing enables accurate assessment of common variants and calculation of genome-wide polygenic scores. *Genome Med.* **11** (2019).
- Li, J. H., Mazur, C. A., Berisa, T. & Pickrell, J. K. Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared to genotyping arrays. *Genome Res.* **31** (2021).
- Gilly, A. et al. Very low-depth whole-genome sequencing in complex trait association studies. *Bioinformatics* **35** (2019).
- Darst, B. F. et al. Combined Effect of a Polygenic Risk Score and Rare Genetic Variants on Prostate Cancer Risk. *Eur. Urol.* **80** (2021).
- Lali, R. et al. Calibrated rare variant genetic risk scores for complex disease prediction using large exome sequence repositories. *Nat. Commun.* **12** (2021).
- Dornbos, P. et al. A combined polygenic score of 21,293 rare and 22 common variants improves diabetes diagnosis based on hemoglobin A1C levels. *Nat. Genet.* **54** (2022).
- Sollis, E. et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985 (2023).
- Wong, K. H., Jin, Y. & Moqtaderi, Z. Multiplex Illumina Sequencing Using DNA Barcoding. *Curr. Protoc. Mol. Biol.* Chapter 7 **Unit 7.11**, (2013).
- Vod ak, D. et al. Sample-Index Misassignment Impacts Tumour Exome Sequencing. *Sci. Rep.* **8**, 5307 (2018).
- Marx, V. How to deduplicate PCR. *Nat. Methods* **14**, 473–476 (2017).
- Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2011).
- Tsagiopoulou, M. et al. UMIc: A Preprocessing Method for UMI Deduplication and Reads Correction. *Front. Genet.* **12**, 660366 (2021).
- Chou, W.-C. et al. A combined reference panel from the 1000 Genomes and UK10K projects improved rare variant imputation in European and Chinese samples. *Sci. Rep.* **6**, 39313 (2016).
- Roshyara, N. R. & Scholz, M. Impact of genetic similarity on imputation accuracy. *BMC Genet.* **16**, 90 (2015).
- Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J. & Delaneau, O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* **53**, 120–126 (2021).
- Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
- Klarin, D. et al. Genome-wide association study of peripheral artery disease in the Million Veteran Program. *Nat. Med.* **25**, 1274–1279 (2019).
- DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
- Trost, B. et al. Impact of DNA source on genetic variant detection from human whole-genome sequencing data. *J. Med. Genet.* **56**, 809–817 (2019).
- Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1–11.10.33 (2013).
- De Summa, S. et al. GATK hard filtering: tunable parameters to improve variant calling for next generation sequencing targeted gene panel data. *BMC Bioinformatics* **18** (2017).
- Koboldt, D. C. Best practices for variant calling in clinical sequencing. *Genome Med.* **12** (2020).
- Zheng, J. et al. A comprehensive assessment of Next-Generation Sequencing variants validation using a secondary technology. *Mol. Genet. Genomic Med.* **7** (2019).
- Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
- Sun, Q. et al. MagicalRsq: Machine-learning-based genotype imputation quality calibration. *Am. J. Hum. Genet.* **109**, 1986–1997 (2022).
- Ball, M. P. et al. A public resource facilitating clinical use of genomes. *PNAS* **109**, 11920–11927 (2012).
- Zook, J. M. et al. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci. Data* **3**, 1–26 (2016).
- Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. <https://doi.org/10.48550/arXiv.1303.3997> (2013).
- Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinforma.* **15**, 182 (2014).
- Olson, N. D. et al. PrecisionFDA Truth Challenge V2: Calling variants from short and long reads in difficult-to-map regions. *Cell Genom.* **2** (2022).
- Wagner, J. et al. Benchmarking challenging small variants with linked and long reads. *Cell Genom.* **2**, 100128 (2022).
- Zook, J. M. et al. An open resource for accurately benchmarking small variant and reference calls. *Nat. Biotechnol.* **37**, 561–566 (2019).
- Zook, J. M. et al. A robust benchmark for detection of germline large deletions and insertions. *Nat. Biotechnol.* **38**, 1347–1355 (2020).
- Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
- Van der Auwera, G. A. & O’Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. (O’Reilly Media, 2020).
- Cavalli-Sforza, L. L. The Human Genome Diversity Project: past, present and future. *Nat. Rev. Genet.* **6**, 333–340 (2005).
- Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L. & Dermizakis, E. T. Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 1–10 (2019).
- Wang, C., Zhan, X., Liang, L., Abecasis, G. R. & Lin, X. Improved ancestry estimation for both genotyping and sequencing data using projection Procrustes analysis and genotype imputation. *Am. J. Hum. Genet.* **96**, 926–937 (2015).
- Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *arXiv [cs.LG]* (2012).

ACKNOWLEDGEMENTS

The authors would like to thank Claire Le Moigne from the McGill Canada Excellence Research Chair Program in Genomic Medicine for administrative support. We also thank the team in Bern who collected the samples and provided administrative support: Gy orgyi Hamvas, Claudine Andres, Rebecca Scheidegger, Marina Rumenovic, Nathalie Minder. The authors thank the Faculty of Medicine and Biology at Lausanne University, Switzerland, and the Swiss Foundation for Vascular Medicine (Schweizerische Stiftung f ur Gef assmedizin) for funding. This work was supported by the McGill Canada Excellence Research Chair Program in Genomic Medicine (VM), CFI awards 41012 & 40104 (JR), CFI 33408, CFI MSI 35444 (JR and GB) and Genome Canada (JR). VM holds a Canada Excellence Research Chair. GB is supported by a Canada Research Chair Tier 1 award and a FRQS Distinguished Research Scholar award. The Canadian Center for Computational Genomics (C3G) is supported by the Canadian Government through Genome Canada. We thank Calcul Quebec and the Digital Research Alliance of Canada for computing resources.

AUTHOR CONTRIBUTIONS

C.B., D.T., J.R. and V.M. conceived the study. C.B. and D.T. lead data analyses. C.B., D.T., K.T. and P.M. performed data analyses. R.E., P.N.R., C.W., and D.P. processed sequencing data. J.R. provided samples and materials. I.B., M.S., Y.D. and V.M. oversaw data acquisition and sequencing of samples. J.S.C., E.C., N.B. and P.L. generated sequencing data. A.P., J.S.C. managed sequencing logistics and M.B., J.F.L. and G.B. managed sequencing data processing. A.P. estimated sequencing costs. D.T., C.B. and K.T. wrote the manuscript. V.M. oversaw the study. All authors read the manuscript.

COMPETING INTERESTS

CB and VM serve as advisors for Medeloop Inc. and hold shares in the company. The remaining authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41525-024-00390-3>.

Correspondence and requests for materials should be addressed to Daniel Taliun.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024