https://pubpeer.com/publications/D780DD5C4D665BBC415F45C707D0C4

Because the journal doesn't have a comment system, I wanted to add a comment in this system.

First, here a comment on the earlier preprint:

https://www.biorxiv.org/content/10.1101/2020.04.29.068452v2#disqus_thread

So, I will try not to over-emphasize the question of *directly measured* genotypes versus Minimac4 imputed genotypes, since I asked that before.

In fact, I think that is acknowledged in the introduction: "*a genotyping array will probe a number of variants which is one to two orders of magnitude fewer [than 0.5x lcWGS], albeit with higher average accuracy (The 1000 Genomes Project Consortium 2015)*." This was also in the 1st version of the preprint, so I apologize for not acknowledging that in my earlier comment.

I also have a comment on another recent paper (other authors describing lcWGS versus genotyping arrays):

https://www.cell.com/ajhg/fulltext/S0002-9297(21)00096-3#comments-heading

The questions posted in that comment are a little different (although some overlap the preprint version of this paper). If anybody cannot access the AJHG comment, I also have a draft posted here.

The Martin et al. 2021 paper mentions Gencove worked relatively better at lower coverage, and BEAGLE worked better at the 4-6x WGS (with that being the highest imputed accuracy reported in Table S4 or Table S5).

So, I think that is consistent with that I see in this paper (as I understand it), while also being consistent with what I have countered in my own experiences.

In fact, I believe Supplemental Figure 12 is new, relative to the preprints (even though I think some of the imputation methods are a bit hard to tell apart in that figure).

However, I believe these would be the questions that may have not been asked before:

a) Since the preprint post, I have added at least 1 extra example of a lcWGS imputation having noticeable problems at 0.1x-0.5x (indirectly mentioned in the AJHG comment). So, I believe that I have experiences that may not completely match what is described in this paper.

   While it may be hard to say, I think knowing **why some results / applications with lcWGS were unacceptable** might be helpful in justifying the other situations where it may be OK. So, if that is possible to do, then I think that would help.

For example, I don't have the raw lcWGS data, but I would estimate I had <0.3x lcWGS for my cat samples (based upon the higher coverage WGS that I do have). The coverage for >95% or 99% samples will be lower than the average coverage. Do any of the Sample A samples have **~0.2x** coverage (such as the far-left point in Figure 1)?

If so, would you say that you could expect additional problems with such a sample?

b) My impression from the Martin et al. 2021 paper was that my ~0.5 lcWGS results (for myself and for my cat) **might have been improved with higher coverage lcWGS might have worked better**.

Do you think that is a fair assessment?

c) In the Martin et al. 2021 paper, the GSA had the lowest performance among all of the arrays. For example, in that paper, I would expect the Omni 2.5 array to outperform 0.5x to 2x lcWGS (with either Gencove or BEAGLE).

You do specifically mention the GSA array in the abstract, but I do not believe the title is universally true (even if you look at SNP chip genotypes that I believe include imputations). Is there anything that I might have misunderstood?

For example, the Gencove 0.5x imputation concordance in the Martin et al. paper (**0.876** in Table S4) is **lower** than reported for Sample A in Table 2 (a mean of 0.900 or 0.923). Is there anything that might show cause performance to be lower in new samples not used in this paper?

For my own sample, the level of concordance varied depending upon the variant caller (as well as whether the zygosity of the variant had to match). The F-statistic from precisionFDA is lower than the concordance measure that I defined. However, **either way**, I would say some results were OK, but others were not acceptable (such as for medical applications).

Thank you for communication of your research!

Best Wishes,

Charles