Qeios

Peer Review

Review of: "SteeraMed: A Biomedical World Model for N-of-1 Intervention Reasoning Across Chronic Diseases and Aging"

Guifeng Zhang[1]

1. RWTH Aachen University, Germany

This preprint proposes SteeraMed Core, a pipeline that maps individual blood methylation profiles onto PPI network modules and ranks candidate compounds by a target-versus-non-target alignment score. Each patient receives an auditable four-layer evidence chain. I found the overall aim of structuring N-of-1 intervention reasoning into transparent, auditable evidence genuinely interesting. I also appreciated the author's openness about limitations; the claims-versus-evidence table (Table 2) is a practice I would like to see more often. I offer a few suggestions that I hope may help strengthen the work.

1. Independent validation and the choice of null model. Modules and features are selected from the same cohorts used to evaluate known-drug recovery, and some parameters are calibrated with reference to the positive controls. It would therefore be very helpful to see the pipeline applied, with parameters fixed in advance, to at least one independent cohort per phenotype. I also wonder whether the uniform random baseline may be somewhat generous. Compounds with many STITCH targets (and well-studied drugs in general) may be more likely to rank highly regardless of patient state. Reporting a target-count-matched baseline, or the recall obtained by simply ranking compounds by target count, would make the fold-enrichment results more convincing.
2. Demonstrating the added value of individualisation. Since modules and features are defined at the group level, it would be informative to compare the per-patient rankings with a single group-level ranking applied to all patients. If individual rankings clearly outperform it, that would directly support the N-of-1 rationale. Given the moderate bootstrap stability of individual rankings, a brief analysis of between-patient versus within-patient variability would also be valuable. A sensitivity

 doi.org/10.32388/NLRNBM

analysis adjusting for estimated blood cell-type composition could further help readers judge how much of the signal is cell-intrinsic, particularly for the RA cohort.

3. Cohort and positive-control details. Could the author kindly double-check the description of GSE51032? My understanding is that it contains pre-diagnostic samples from the EPIC-Italy cohort, which would affect how the breast cancer results are interpreted. A short justification for some positive-control choices would also be helpful, for example, allopurinol and imatinib for RA, and niacin versus the predefined NAD+-related controls for aging.

Finally, a careful consistency check across the manuscript may be worthwhile. For example, RA Recall-10 is reported with different values in different sections (51.7%, 35.3%, 24.0%, 7.9%). The panels of Figure 3 and the content of Figure 8 do not seem to match their legends and the text. Some cross-references and citation numbers also appear to be shifted.

I hope these comments are useful, and I look forward to seeing how this framework develops.

Declarations

Potential competing interests: No potential competing interests to declare.