nature portfolio

Peer Review File

Integration of proteomic aging clocks in a phase 2a clinical trial supports simultaneous geroprotective assessment

Corresponding Author: Dr Alex Zhavoronkov

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Biotechnology.

Parts of this Peer Review File have been redacted as indicated as we could not obtain permission to publish the reports from one of the peer reviewers.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Author Rebuttal letter:

Dear editors of Nature Biotechnology,

Prior to transferring to your publication, we received comments from the referees assigned by Nature. Since then, we have worked to improve the clarity of our message, additional experiments to strengthen our claim, and more in-depth discussion that contextualizes our findings. Please, see below our point-by-point response to the received comments.

Referee #1:

1. The high baseline prediction error of the aging clocks (RMSE 5.4–16.2 years), especially for the Galkin clock, is attributed to the burden of IPF; however, the possibility that these clocks—particularly the fibrosis-tailored Galkin clock—incorporate predictive proteins (e.g., COL1A1, MMP7) that are themselves established markers of IPF pathology has not been adequately excluded. This raises the concern that the observed “biological age reversal” may partly reflect disease-state improvement rather than genuine aging reversal. Supplementary analyses are required to strengthen the specificity of this conclusion, such as demonstrating correlations between the clocks’ core predictive proteins and established IPF-severity markers, or evaluating clock performance in a non-fibrotic yet aging-related disease cohort (e.g., COPD).

3. All subjects were late-stage IPF patients, a population characterized by severe disease-associated proteomic perturbations and high mortality risk, which may bias aging-clock readings. Current evidence is inadequate to confirm replicability of similar anti-aging effects in healthy populations. A clearer distinction between “disease-modifying effects” and “systemic anti-aging effects” is warranted.

We have extended the sections dedicated to the differences between the observed anti-aging and anti-fibrotic proteomic changes. More specifically, we show that the dFVC and dBioAge are only weakly correlated meaning that overall health improvement seen in patients does not fully explain the anti-aging effect. Moreover, we have inspected the sets of genes affecting the aging clock predictions, known fibrotic markers, and genes affected by rentosertib. In Figure 6A we display that the rentosertib-modulated proteins are much more widely represented among the aging clock features than fibrotic

markers.

Sure, aging clocks are usually constructed based on many weakly associated proteins, while some fibrotic markers such as PDGFB are extremely important in this process, and mere number of proteins does not suffice to prove the distinct aging-directed mode of action. So, we additionally explored if the proteins modulated in each arm expose any aging specific regulation patterns. Using StringDB, we identified clusters of proteins modulating metabolic (glutathione, cholesterol, others) in the 30BID but not 60QD group (Figure 6B). Note that 60QD is the arm that shows the greatest FVC improvement, yet is not picked up by mortality-trained aging clocks.
To further the point that the aging process is affected, we have checked if rentosertib-modulated proteins are detected as aging-associated proteins elsewhere. As out dataset of reference, we have chosen UK Biobank, as the most massive dataset using the same Olink platform as in the trial. We then show that although weak, there is a statistically significant correlation between rentosertib-induced and normak aging protein trajectories for the 30BID group.

We also sought out datasets that would allow us to contextualize rentosertib among other antifibrotic drugs or lung diseases, however, such datasets are quite rare and come with limitations that render any analytical efforts a moot point. E.g. some teams apply mass spectrometry to study pirfenidone's effects in lungs (PMID:30072508), or study nintedanib effects but in gastric cancer cultures (PMC10893871), or use an Olink platform with a much smaller set of measured proteins (PMC9238057). Even the studies showcased by Olink themselves (https://olink.com/publication/common-physiologic-and-proteomic-biomarkers-in-pulmonary-and-coronary-artery-disease) mostly feature datasets obtained with a narrowly targeted biomarker panel, rather than a full-scale Olink 3072 screening.

2. Only 42 IPF patients were enrolled, with a follow-up of merely 12 weeks. Such short-term outcomes are insufficient to substantiate a robust claim of “anti-aging drug” status. The observed plateau or partial reversal of biological-age reduction after the maximum effect at week 4 (by week 12) lacks direct experimental support for the proposed hypotheses (e.g., compensatory mechanisms, cellular heterogeneity, drug-induced stress). Additional experimental data are needed to clarify the mechanism underlying this phenomenon.
We agree that experimental validation would strengthen our hypotheses, and we have revised the discussion to more clearly frame the findings as directions for future investigation.

At this time, the trial design and legal regulation constrain our ability to conduct additional experiments on this cohort. Nonetheless, our 42 participants across 12 weeks compares favorably to published human geroprotector studies. E.g. in a recent review of 22 trials featuring a well-known geroprotector rapamycin (https://www.thelancet.com/journals/lanhl/article/PIIS2666-7568(23)00258-1), the median N of participants is 38 and the median duration is 82 days (~12 weeks), with 10 trials having <20 participants. Note that this review excluded all published trials with <5 participants, a setting which is also gaining popularity in geroscience due to the difficulties associated with organizing human trials (e.g. PMC10085584).

Moreover, our study features a premium proteomics platform, Olink Explore 3072, which is highly atypical for either geroprotector or clinical trials among which most lack proteomics or utilize panels measuring fewer than 400 proteins.
The consistency of biological age reduction across six methodologically independent clocks, coupled with coordinated shifts in senescence markers, suggests an effect worthy of further investigation. We have revised our conclusions to appropriately frame rentosertib as a candidate geroprotector in need of extra validation, rather than claiming definitive anti-aging status based on our preliminary findings.

4. Although GSEA indicates a decline in senescence markers and links to IGF-axis regulation, direct cellular or tissue-level validation (e.g., lung-biopsy analysis, single-cell transcriptomics, immunohistochemical confirmation) is lacking. The possibility that these molecular changes arise partly from disease alleviation rather than direct geroprotective action cannot be dismissed. The causal chain requires more cautious handling and experimental support.
We agree that tissue-level validation would strengthen our findings, and we have revised the manuscript to update our conclusions. Unfortunately, the trial design precludes retrospective collection of lung biopsies or additional tissue samples from

participants.
Regarding the concern that molecular changes reflect disease alleviation rather than geroprotection, we addressed this with additional experiments, such as comparing the set of rentosertib-modulated proteins versus established fibrotic markers (Fig6A), comparison to aging trajectories in the UKB cohort, measuring the dBioAge variance attributed to dFVC, and PPI cluster analysis. These analytics suggest effects extending beyond anti-fibrotic activity. However, we acknowledge that making a definitive claim about aging-related effects requires dedicated studies in non-IPF populations, which we now explicitly state as a study limitation. We have revised the discussion to frame our IGF-axis and senescence findings as hypothesis-generating rather than established mechanisms.

5. When assessing proteomic aging clocks in aging populations, substantial variability may arise from differences in sample batches and detection methods. Selecting baseline data from other cohorts could introduce considerable bias. It is recommended to collect baseline data from different age groups within the same cohort to establish a more reliable foundation for the aging-protein clock.

The trial's setting, however, does not preclude us from defining adequate baselines. This RCT with repeated proteomic measures enables us to define two baselines and choose one depending on the context. In our analysis, we can both compare to the placebo group or run paired sample tests between timepoints, which is arguably a more reliable baseline than any other. The trial itself was designed to minimize variability due to demographic, clinical, or other confounders and the recruited cohort can be considered homogenous enough to render unbiased between-group or within-individual comparisons.
We appreciate the concern regarding baseline data from external age-matched cohorts. At this time, we have no ability to collect additional baseline data from healthy people or other age groups in China. Due to financial and regulatory constraints, such efforts may take years to organize, and thus, we have chosen to go ahead with the data generated in the trial.

However, our analysis employs two methodologically robust baseline approaches: between-group comparisons versus placebo and within-individual longitudinal assessments. While additional validation in broader populations would be valuable, the RCT framework provides robust internal controls that arguably offer greater reliability than external age-stratified reference data, which would introduce batch effects, population heterogeneity, and technical variability across collection sites and timepoints. The trial's strict inclusion criteria minimize demographic, clinical, and other confounders, yileding a rather homogenous population.

6. The current LMEM includes treatment × time interactions and random intercepts but does not adjust for age, sex, or disease severity. These variables should be included as covariates or examined via stratified analyses; additionally, clarify whether IPF itself accelerates biological aging and whether this acceleration correlates with disease progression.
Unfortunately, the data deposited at the CNCB biobank offers no extra annotations apart from FVC, NPX, and age of the patients. Due to strict data management regulations, sex, disease severity, intermediary FVC measurements or other variables are not avaiable for this trial.

Per your suggestion we added the "chronological age" variable in LMEM. However, adding additional terms for age does not affect the significance of time-dependent protein trajectories, as can be validated in the updated Table S4. In total, only 14 proteins were significantly affected by a participant's chronological age, among which MXRA8 is the only one affected by rentosertib in either age-corrected or uncorrected LMEM.

7. Trends across the aging clocks are not entirely consistent. For some clocks, predicted age increases after week 4, raising the question of whether continued medication might induce adverse bodily effects.
We acknowledge the heterogeneity in aging clock trajectories (Fig3B). This variability

represents an important finding worthy of sharing with the community.
To our knowledge, this manuscript features the first systematic comparison across such an array of proteomic clocks, with all clocks released in 2024-2025 (or yet in review). The observed discordance, particularly between mortality- versus age-trained clocks, parallels findings from epigenetic clock studies where similar inconsistencies have been reported, including negative correlations between certain clock pairs (PMC11404624, also see https://doi.org/10.1111/acel.70103 for an example of DNAm clocks used in trials).

Regarding the post-week-4 increases in some clocks, we present several mechanistic hypotheses in the discussion for further validation. While not fully understood, this temporal pattern has precedent in anti-aging therapeutics. The TRIIM trial (https://doi.org/10.1111/acel.13028) similarly observed epigenetic age acceleration after a prolonged intervention. We have revised the discussion to explicitly acknowledge the possibility of adverse effects from continued treatment and emphasize the need for intermittent dosing studies to optimize the therapeutic window.

8. The most critical concern is that rentosertib has already been reported to inhibit pulmonary fibrosis. This paper neither proposes any novel drugs nor undertakes an in-depth exploration of the specific pathways and mechanisms of innovative therapeutics. The cohort employed appears to be a subset extracted from PMID: 40461817 (even the database numbers are identical). Furthermore, the entire manuscript relies on computational bioinformatic inference, and its major conclusions await experimental verification.
We respectfully disagree that this work lacks novelty. Yes, the cohort is that reported in the original trial report, but this study addresses fundamentally different questions.

This is the first demonstration that a drug rationally designed to target aging hallmarks shows geroprotective effects in humans when evaluated with aging clocks. It's also the first systematic comparison of six independent proteomic aging clocks in an interventional setting, providing resilience from spurious clock-specific signals. To make this comparison possible, we have developed and publicly released a unified interface for all the models we used (https://github.com/Insilico-org/proteoclock), which will hopefully be used by other researchers and will make it easier to include proteomic aging analytics into any study for all. The release of the full dataset alongside production-grade codebase is exceedingly rare for pharmaceutical trials where proteomics typically remains proprietary due to the perceived commercial risks and exclusivity concerns.
While rentosertib's senomorphic activity was shown in vitro, this manuscript provides the first evidence of its geroprotective potential in humans. While we fully understand the limitations of our exploration, we have done our best to confirm if the in vitro phenomenon translates to the human organism, as based on convergent signals from aging clocks, senescence markers, and pathway analyses. The IGF-axis and metabolic findings represent novel mechanistic hypotheses for TNIK inhibition. We acknowledge the computational nature of our work and have revised the manuscript accordingly.

Referee #2:

At its core, however, this is an exploratory biomarker study derived from a small clinical cohort of an investigational drug trial not designed to evaluate efficacy. Phase 2a studies primarily establish dosing and safety, yet here only ~10 patients per arm across three dosing groups plus placebo were included for proteomic analysis—an underpowered sample to assess therapeutic effects. Furthermore, it is unclear how patients were selected for serum analysis when the original intention-to-treat population included ~18 per group. Given that discontinuations due to adverse events were reported in the trial, any exclusion of these patients could introduce bias and confound interpretation of proteomic changes.
The proteomic substudy included all 42 participants from the phase-2a trial who (i) provided informed consent for omics research, (ii) completed the trial without severe adverse events requiring discontinuation, and (iii) had complete proteomic data at all timepoints. This selection was pre-specified in the trial protocol to ensure data quality and interpretability. We acknowledge the limited statistical power and have revised our claims accordingly, framing findings as hypothesis-generating. The original trial report (PMID: 40461817) provides complete details on the intention-to-treat population and discontinuations. At this point, we have no ability to extend the data selection available at the CNCB biobank as legally we do not control the data or biospecimen. However, we

aim to continue with follow-up studies in vivo to gain deeper mechanistic understanding upon sharing the current extent of our findings.
Another major concern is the poor correlation between the published aging clocks and chronological age in this IPF cohort (Figure 3), which undermines confidence in using these clocks to infer biological age reversal.
A major part of this project was to give the reader a fair comparison of findings from different point of views. To our knowledge, no other study features such a wide array of proteomic clocks applied to the same cohort. Similar cross-clock comparisons have previously been done for epigenetic clocks (e.g. PMC11404624), in which some clock pairs yield predictions with negative correlation coefficients. In this context, proteomic clocks offer a better alternative to older DNAm-based models, with most clock pairs showing high correlatedness of their predictions, except for pairs in which one model is designed to predict mortality and the other one — chronological age. Yet even in such comparisons, the correlation between clock predictions remains positive.

The poor baseline correlation between mortality-trained clocks and chronological age was expected given the severe disease burden—IPF patients have 3-4 year median survival, so mortality clocks appropriately estimate much higher biological age than chronological age. Meanwhile, chronological clocks showed much better correlation with real age.
Finally, we deliberately chose to include as many proteomic clocks as was possible, regardless of their performance, to give the reader a fair and well-rounded perspective of the current models. While some clocks present an image much more consistent with our expectations (i.e. Argentieri's 2024 clock), we have deemed it dishonest to present only such findings and ignore the rest.

Moreover, many clocks showed regression toward baseline between weeks 4 and 12 (Figure 4), a pattern not aligned with reported clinical benefit in the 60 mg QD group at week 12 in the original trial report.
Regarding the week 4-12 regression, this pattern has precedent in aging interventions (e.g. thymic rejuvenation trial; https://doi.org/10.1111/acel.13028) and we've added mechanistic hypotheses to the discussion which we aim to validate in further research projects.

The effects also does not appear to be dose-dependent, with the largest changes sometimes observed in the 30 mg QD group instead.
We accept that "regimen-dependent" is a much more fitting term to describe the observed effects. Yet, in the statistical analysis based on all clocks' measurements (Fig4D), we show that the clock consensus marks 30QD as the arm with the least reliable effect on the pace of aging, if any. In following explorations of proteomic trajectories (Fig5A), we show that this arm displays negligible effect on the participants proteome.

This allows us to conclude that 60 mg daily dose yields greater anti-aging signal, in addition to better clinical improvement, as per the original trial report.

The differences between the activities induced by 30BID and 60QD back that perhaps the administration regimen rather than total daily dose may be the key factor that determines the aging clock response. As we demonstrate in the added analytics (i.e. String DB PPI clusters), 30BID participants show unique metabolic modulation (Fig6B) in addition to the induction of anti-fibrotic pathways. While speculative at this point, we have included our considerations into the Discussion on how the lower plasma Cmax and steadier drug levels seen in 30BID may contribute to more beneficial induction of anti-aging pathways.

Thus, the lack of strict dose-dependency may reflect pharmacokinetic differences rather than absence of effect.

More broadly, the study does not resolve whether age-associated biomarkers are causal mediators of aging or simply correlates. At present, there is little evidence that serum proteomic clocks capture causal biology rather than epiphenomena.
Prior studies in IPF have identified serum proteomic signatures predictive of progression and transplant-free survival (PMID: 37847691), yet these validated disease-related proteins show minimal overlap with the markers highlighted in Figures 5 and 7 of the present work.
We agree this is a fundamental limitation of all aging clock studies. We cannot definitively prove causality, but the convergent signals across six clocks, senescence

signatures, and pathway analyses strengthen our confidence beyond what a single biomarker assay provides.
Regarding overlap with validated IPF progression markers, we've added analysis comparing rentosertib-modulated proteins to UK Biobank aging trajectories, showing statistically significant correlation, and measured the overlap with the suggested fibrotic markers (Fig 6A). The small size of the intersection set does not necessarily mean low disease relevance (i.e. PDGFB is widely studied profibrotic biomarker). Additionally, our PPI clustering of proteins modulated by rentosertib shows the largest clusters to be associated with ECM remodeling and immunity in both 30BID and 60QD groups, confirming the correctness of the chosen methodology by identifying the mediators of anti-fibrotic activity.

Finally, the interpretation of rentosertib as a senomorphic agent rests heavily on the SenMayo gene set, which was derived from RNA-seq of bone tissue biopsies (PMID: 35974106). Applying this dataset to infer senescence from serum proteomics is not validated and is not widely accepted as a gold standard.
We acknowledge that applying bone-derived RNA signatures to serum proteomics has limitations.
However, cross-tissue application of senescence markers is common in aging research given the systemic nature of senescence. Moreover, we used SenMayo as it was the signature used in the preceding report of rentosertib senomorphic activity (https://doi.org/10.14336/ad.2024.1492), thus making SenMayo GSEA a necessity to confirm in vitro findings.

Importantly, we used several independent signatures and observed convergent patterns. Unlike SenMayo, the CellAge signatures obtained from HAGR cover senescence development and prevention in various settings.

To reduce the reliance on SenMayo to infer senomorphic properties, the current manuscript version also features KEGG Medicus and Reactome GSEA, an imperfect yet statistically significant correlation with UK Biobank aging trajectories, and StringDB PPI clustering.

We would also argue that the main findings in our research come from the six aging clocks, while GSEA serves a supporting role, and that aging clocks are a more widely accepted method of reporting anti-aging effects in a variety of settings, ranging from iPSC reprogramming to multi-species evolutionary studies (https://doi.org/10.1126/sciadv.adm7273).

In summary, while the study raises interesting hypotheses, it provides limited biological insight and does not convincingly demonstrate that rentosertib—or TNIK pathway inhibition more generally—represents a viable strategy for extending human longevity.
We agree our study does not definitively establish rentosertib as a longevity intervention and we have revised the manuscript to present it in a more measured tone. Yet it is the first demonstration of TNIK inhibition's geroprotective potential in humans.

We acknowledge that the viability of the strategy requires validation in dedicated aging trials, non-IPF populations, and clear mechanistic demonstrations in vivo. Nonetheless, this manuscript serves as an important milestone in pioneering the adoption of aging clocks in clinical trials.

The novelty of comparing a wide range of proteomic clocks within one cohort should not be disregarded either, as the detected correlations, performance differences, and signal inconsistencies will undoubtedly serve other research teams who are unsure which novel tools to use in their projects.

We thank the referees for their critique and hope that the edits we have made to address their concerns have positively improved the quality of our presentation.

Kind regards,
Alex Zhavoronkov

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript entitled “Proteomic Clock Analysis of Rentosertib Clinical Trial Data Shows Attenuation of Aging Signatures” presents an ancillary analysis of a Phase 2a trial evaluating Rentosertib, a novel AI-designed TNIK inhibitor, in patients with Idiopathic Pulmonary Fibrosis (IPF). By deploying six diverse proteomic aging clocks, the authors demonstrate a reduction in biological age across treatment arms, with the most robust and consistent signals observed in the 30 mg BID cohort at Week 4. Through comprehensive pathway enrichment and overlap analyses, the study posits that Rentosertib exerts senomorphic effects, modulating IGF signaling and ECM remodeling, which are mechanistically distinct from its primary anti-fibrotic efficacy.

The robustness of the study’s conclusions is limited by the small sample size (N=11 per group, N=9 for one group). Therefore, the manuscript must transparently address these statistical constraints. The conclusions should be carefully framed to reflect that this is an exploratory, signal-seeking analysis rather than a definitive confirmation of efficacy, particularly regarding the lack of significance in certain clock or group.

The results indicate that regardless of the dosage regimen, the therapeutic efficacy was most pronounced at Week 4, followed by an attenuation of effect by Week 12. The investigators may consider evaluating efficacy at Week 8 to determine if the therapeutic benefit remains constant throughout the trial duration. Furthermore, examining the proteomic data at Week 12 for markers of cellular stress or exhaustion would help to better substantiate the molecular mechanisms underlying this observed attenuation.

The central conclusion regarding the dissociation of Rentosertib’s geroprotective effects from its anti-fibrotic activity requires stronger justification. Given that IPF is characterized by accelerated aging, the authors should critically discuss why the 60 mg QD group, which exhibited the most significant clinical improvement in FVC, paradoxically displayed less consistent reductions in aging clocks compared to the 30 mg BID group.

A dedicated demographic table for the 42-patient proteomic substudy is missing. Given the authors' emphasis on metabolic and IGF-1 signaling pathways (Pages 5 and 16) as key mechanisms, Body Mass Index (BMI) is a critical variable. Please provide age, sex, and specifically BMI data. Additionally, did the authors analyze whether baseline BMI correlates with the magnitude of biological age reduction (dBioAge)? This is crucial to verify if the fixed dosing (30/60 mg) was sufficient for patients with higher body weight.

Adverse Events (AEs) can induce systemic inflammation and alter the plasma proteome, potentially skewing aging clock predictions. Please report the incidence of AEs in this sub-cohort. A sensitivity analysis excluding patients with moderate-to-severe AEs is recommended to ensure the observed "rejuvenation" (and subsequent attenuation) is driven by the drug's senomorphic activity rather than AE-induced physiological stress.

The study relies heavily on algorithmic predictions. To validate that the calculated "Biological Age Reduction" reflects true biological changes, orthogonal wet-lab validation is essential. The authors should verify key findings—such as the reduction in SASP factors or IGFBPs (Figure 9)—using standard assays (e.g., ELISA).

The manuscript infers TNIK inhibition efficacy primarily through downstream phenotypic changes (aging clocks) and pathway enrichment (Figure 1 schematic). However, direct evidence of TNIK inhibition in patient samples is lacking in this manuscript. While TNIK is an intracellular kinase, are there any proximal phosphorylation substrates available in the Olink panel or assessable via ancillary assays to confirm target engagement?

The discussion (Page 14, Lines 496-508) speculates that the 30mg BID regimen performs better than 60mg QD due to "steadier drug levels" and avoiding "surges" that trigger off-target effects. This hypothesis needs data support. The authors should present the plasma concentration (PK) data for the subjects in this study.

While LMEM were used, the data clearly exhibits non-linear dynamics (peaks at Week 4). I suggest applying non-linear models to better characterize the temporal dynamics. This would help clarify if the "attenuation" at Week 12 represents a true reversal of anti-aging effects or stabilization at a new homeostatic equilibrium.

To enhance the interpretability of Figure 1, the authors should revise the legend to clearly demarcate general signaling pathways from the specific pharmacodynamic effects of Rentosertib observed in this trial.

A supplementary table detailing the list of features in Figure 6A should be provided for reference.

It is important to note that proteomic clocks alone cannot authentically reflect the organism's overall health status. They represent a single, albeit powerful, layer of biological information. To more accurately approximate clinical readouts and enhance the translational relevance of the findings, the analysis should be strengthened by correlating or integrating the proteomic clock data with other phenotypic clocks (e.g., clinical chemistry, epigenetic, or transcriptomic clocks) when possible. This multi-modal approach would provide a more holistic view of the drug's impact on biological aging.

Strengthening the mechanistic insight is paramount. The current analysis, while comprehensive in its correlative and descriptive power regarding pathway changes, would be significantly elevated by incorporating targeted experimental

validation or deeper molecular analyses to move beyond association and towards causality. For instance, can the proposed senomorphic mechanism be linked more directly to the observed clock changes? Without this, the study risks remaining primarily descriptive.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Assessment of responses to previous reviewer 2 comments (reviewer 2)
Underpowered sample size and potential selection bias from excluding patients with adverse events
The authors clarify that the proteomic study included 42 participants who completed the trial without severe adverse events and had complete proteomic data. They refer to sample sizes in other geroscience trials and regulatory hurdles preventing access to more data.
While these are fair points made by the authors, they do not resolve the fundamental limitation. More concerning is the selection bias issue, which remains inadequately addressed. If patients who showed adverse effects to the drug were systematically excluded, and this correlates with biological response or baseline health status, this could bias results toward showing benefit. The statement that excluded patients had "severe adverse events" could suggest that these events might reflect biological responses relevant to the outcomes.
Poor correlation between aging clocks and chronological age, and regression toward baseline between weeks 4 and 12
The authors argue that mortality-trained clocks appropriately estimate higher biological age in severely ill patients, while chronological clocks showed better correlations. They cite the TRIIM trial as precedent for regression patterns.
The defense of mortality-trained clocks is reasonable, but the deeper issue is that if clocks perform variably in this population, how confident can we be that detected changes reflect modulation of aging biology rather than disease-specific processes? The TRIIM trial citation was very different (and also not without criticism by the community), but importantly it was performed on healthy men. The week 4-12 regression pattern remains unexplained.. If the drug genuinely modified aging biology, wouldn't one expect sustained or progressive effects with continued dosing. Alternative explanations that the week 4 signal partly reflects noise, or that differential dropout over time introduces bias are not considered.
Dose-dependency
The authors reframe the effects as "regimen-dependent" rather than dose dependent and propose pharmacokinetic explanations.
The pharmacokinetic explanation is plausible but entirely post-hoc. If dosing frequency were critical for geroprotective effects, this should have been hypothesized a priori based on known TNIK biology or pharmacology. The data pattern where the regimen with best clinical efficacy (60 mg QD) does not show the most consistent aging clock effects could equally reflect noise in an underpowered study.
No evidence that clocks capture causal aging biology
The authors added analyses comparing rentosertib-modulated proteins to UK Biobank aging trajectories and examining overlap with aging clock features versus fibrotic markers
The new analysis in Figure 6A has a circularity problem in our opinion. Aging clocks are trained on proteins that change with age, so finding that rentosertib affects proteins used by aging clocks is somewhat tautological and does not establish that the mechanism is aging-specific rather than a reflection of improved health (see our own comments below in major concern 1). The comparison is complicated by the fact that UKBB participants are generally healthy while IPF patients have overall perturbed proteomes.
SenMayo gene set derived is not validated for serum proteomics and based on different tissue
The authors added CellAge signatures and reply that cross-tissue application of senescence markers is common in aging/senescence research.
The inclusion of multiple independent signatures strengthens the pathway analyses. This is a pragmatic approach given available tools.
Additional major concerns not mentioned in previous review
Inability to distinguish aging-specific effects from disease modification i.e. improved health through treatment
This is the central problem with the study that has not been adequately addressed, and I don't know how this be done other than tuning down the language and claims significantly. Proteomic aging clocks are trained on proteins that differ between younger and older individuals in population cohorts, but many of these same proteins also differ between sick and healthy people of the same chronological age. IPF is a severe systemic disease characterized by chronic inflammation, metabolic dysfunction, stress-response proteins, and other pathways aging clocks associate with older biological age.
Successful treatment of IPF reducing fibrosis, systemic inflammation, improving metabolic function would be expected to shift proteomic profiles in directions that aging clocks interpret as "rejuvenation," even if no fundamental aging biology has been modified. The findings are therefore entirely consistent with pure disease modification without any aging-specific effect.
The authors attempt to address this through several arguments, none of which are conclusive to me. Their main argument that the correlation between dFVC and dBioAge is low. Tis may simply reflect that FVC is an incomplete proxy for overall health/disease status. A patient could have stable FVC but reduced systemic inflammation or improved metabolic function, both of which would affect aging clocks. The overlap analysis showing more intersection with aging clock features than fibrotic markers has the circularity problem noted above.
Statistical concerns
With 6 clocks × 3 treatment arms × 3 post-baseline timepoints = 54 primary comparisons, finding 21 significant results at $q < 0.10$ requires careful interpretation. While the authors apply FDR correction, the clocks are not independent in that they share many input proteins and are trained on similar outcomes. The effective number of independent tests is unclear, and

the correlation structure among clocks (Fig 3B) suggests substantial redundancy. The authors should provide some estimate of what would be expected under the null hypothesis given the correlation structure.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Biotechnology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)
Summary and Importance

This manuscript analyzes proteomic aging clocks using data from a Phase 2a, double-blind, placebo-controlled trial of rentosertib, a novel TNIK inhibitor being initially developed for IPF. There is compelling preclinical evidence for TNIK's geroprotective effects, and this study provides the first human evidence of consistent reductions in biological age across six proteomic clocks with rentosertib treatment. The largest effects are observed at week 4 in the 30 mg BID arm. Downstream analyses are consistent with mechanisms related to senescence suppression and modulation of IGF signaling, inflammation, and matrix remodeling.

The work is important: It provides an "in-the-wild" evaluation of aging biomarkers in a human clinical trial. This includes head-to-head comparisons of six recent proteomic clocks. Proteomics is a well-established clinical biomarker modality, and while the clocks show some heterogeneity they also show a clear overall age attenuation with rentosertib treatment. The findings are also intriguing with respect to rentosertib's potential as a geroprotector. The authors transparently address clock inconsistencies across doses and timepoints. More broadly, the paper addresses a key geroscience bottleneck: aging biomarkers must be stress-tested in interventional trials to become clinically actionable. This study is a meaningful step in that direction.

My additional comments are minor.

Evaluation of Authors' Response to Reviewer #2 (as requested)

In my view, the authors provide a thoughtful response and adequately address most of Reviewer #2's concerns. They clarify the exploratory nature of the analyses, limited power, and the pre-specified criteria used to define the proteomics subset. They also explain why mortality-trained clocks can decouple from chronological age in a diseased IPF cohort. On the week 4 improvements, and partial attenuation by week 12, they cite precedent (e.g., TRIIM) and add hypotheses while acknowledging that validation is needed. They also add analyses aimed at distinguishing disease improvement from aging clock changes. Overall, these additions are helpful and strengthen the manuscript.

On the "causality" point, I think requiring causality from an aging clock sets too high a bar. Biomarkers can be clinically useful without being causal, but they should be reproducible, technically robust, and predictive of meaningful outcomes across settings. The manuscript could be framed more explicitly around that standard.

Other Comments

Senescence pathway language: The enrichment analyses are interesting, but some statements are too definitive (eg "confirm senomorphic activity in humans"). Suggest replacing "confirm" with "consistent with".

Plasma vs serum inconsistency: Abstract says "plasma proteomics" but samples are described as "serum samples".

Expand discussion on clock discordance and implications for clinical trial design.

Add a short paragraph in Discussion on requirements for clinical utility of aging clocks (predictive validity / reproducibility / impact on decision making).

(Remarks on code availability)

Author Rebuttal letter:

Response to Reviewers
We thank the editor and the reviewers for the opportunity to respond and re-write the manuscript according to their comments and feedback. We have expanded the data we present and reframe the paper from a report on clinical age reversal to the first implementation of a clinical trial design template where a novel therapeutic discovered with a deep focus on aging biology is evaluated in the context of a real clinical study.

We now have re-framed the central achievement of our work as a "blueprint" demonstrating how proteomic aging clocks can be embedded in disease-focused clinical trials to generate aging biology hypotheses, detect or interpret clock signals, and identify pharmacological candidates warranting dedicated geroprotective studies. All claims regarding geroprotective activity of rentosertib have thus been toned down throughout the title, abstract, and main text in accordance with the review comments. We also now highlight the key limitations of the aging clock methodology and how complementary methods can compensate for them.

Of note, to address the concerns expressed by multiple reviewers about potential contributions of unaccounted for variables, such as BMI and others, we obtained and are now permitted to report a richer set of participant metadata. We now disclose BMI, sex, and higher-grade adverse events. We also performed additional analyses to confirm insignificant BMI ~ BioAge correlations, insensitivity of our key findings to adverse events, and present a more detailed exploration of proteomic trajectories, correcting for BMI and sex in all models. A comprehensive demographic summary is now provided in Table S1.
We have also significantly restructured the text itself to improve its readability and maintain the focus on the essential. In total, the reference list has been compacted from 128 to 89 items, and the main text's volume has been reduced from 6.3K words to 5.7K while supporting all the new results. All existing figures and tables have been updated and some new ones have been added.
Please find below the point-by-point response to all critical comments we received:
Insilico Medicine Restricted (Level 3)

Response to Reviewer 1
Comment 1.1
The robustness of the study's conclusions is limited by the small sample size (N=11 per group, N=9 for one group). Therefore, the manuscript must transparently address these statistical constraints. The conclusions should be carefully framed to reflect that this is an exploratory, signal-seeking analysis rather than a definitive confirmation of efficacy, particularly regarding the lack of significance in certain clock or group.

We fully agree. The revised manuscript now consistently frames the analysis as exploratory and hypothesis-generating. We have added a permutation test (Figure S1) demonstrating that the 21 significant comparisons across all clock-timepoint-regimen combinations far exceed the null expectation (mean under permutation = 0.15), providing a benchmark for what chance-level performance would look like in this dataset. In the sections where we discuss limitation, the statistical power constraints and their implications for interpretation are explicitly stated: “Modest sample size, relatively short duration, and dominance of computational approaches weaken the conclusions pertaining to the geroprotective potential of rentosertib.”

Comment 1.2
The results indicate that regardless of the dosage regimen, the therapeutic efficacy was most pronounced at Week 4, followed by an attenuation of effect by Week 12. The investigators may consider evaluating efficacy at Week 8 to determine if the therapeutic benefit remains constant throughout the trial duration. Furthermore, examining the proteomic data at Week 12 for markers of cellular stress or exhaustion would help to better substantiate the molecular mechanisms underlying this observed attenuation.

Proteomic profiling was performed only at the prespecified timepoints (baseline, week 2, week 4, and week 12) and intermediate timepoints are unavailable. We have addressed the temporal dynamics question by introducing a categorical time analysis that directly compares protein levels between timepoints, enabling identification of the proteins most responsible for the week-4 peak and subsequent partial attenuation.

Regarding cellular stress and exhaustion markers: after applying the minimum gene set size filter (n > 25 proteins shared with Olink), only one relevant Reactome pathway, CELLULAR_ RESPONSE_TO_CHEMICAL_STRESS, passed the threshold. Results for this pathway were inconclusive (30 mg BID: NES = +1.66 at week 12 was the only significant one with q-value = 0.019; for 60 mg QD: q-value > 0.28 at both timepoints) and are not reported in the main text. We acknowledge this as a technological limitation of the Olink platform in the discussion: "This technological constraint makes some statistical tests incomplete or impossible, as was the case for our GSEA when enrichment scores could be obtained for only 145 of 1,787 available Reactome pathways due to limited intersection between array and pathway features".

Comment 1.3
The central conclusion regarding the dissociation of Rentosertib's geroprotective effects from its anti-fibrotic activity requires stronger justification. Given that IPF is characterized by accelerated aging, the authors should critically discuss why the 60 mg QD group, which exhibited the most significant clinical improvement in FVC, paradoxically displayed less consistent reductions in aging clocks compared to the 30 mg BID group.

We agree that the evidence for mechanistic dissociation is inferential given the constraints of the dataset. This claim has been substantially toned down throughout the manuscript. The divergence between the 60 mg QD and 30 mg BID groups is now presented as an informative biomarker observation rather than established pharmacological evidence. The observed dissociation is noteworthy and hypothesis-generating, but we no longer frame it as proof of independent anti-aging and anti-fibrotic mechanisms. Our revised Discussion includes a hypothesis that the 2–3-fold difference in serum concentration (Cmax) reported in the original clinical trial report may lead to differential engagement and saturation of target kinases. The Discussion also includes a methodological note acknowledging that aging clocks cannot distinguish between beneficial processes and compensatory responses that may artifactually shift age predictions.

Comment 1.4
A dedicated demographic table for the 42-patient proteomic substudy is missing. Given the authors' emphasis on metabolic and IGF-1 signaling pathways as key mechanisms, Body Mass Index (BMI) is a critical variable. Please provide age, sex, and specifically BMI data. Additionally, did the authors analyze whether baseline BMI correlates with the magnitude of biological age reduction (dBioAge)?

We originally did not have permission to publish participant-level data, but we have now obtained written permission from the organization collecting data in the study to include participant-level metadata and now present a full demographic summary in Table S1 including BMI, sex, age, and AE grade. All mixed-effects models were re-run with BMI and sex included as covariates, although the results were qualitatively unchanged. We additionally tested Spearman correlations between baseline BMI and ΔBioAge across all six clocks and three treated arms (54 tests total). Three reached nominal significance ($p < 0.05$), and none survived multiple testing correction (all q-values > 0.25), confirming that the observed biological age changes were not substantively modulated by BMI within the studied weight range.

Comment 1.5
Adverse Events (AEs) can induce systemic inflammation and alter the plasma proteome, potentially skewing aging clock predictions. Please report the incidence of AEs in this sub-cohort. A sensitivity analysis excluding patients with moderate-to-severe AEs is recommended to ensure the observed "rejuvenation" (and subsequent attenuation) is driven by the drug's senomorphic activity rather than AE-induced physiological stress.

AE data are now reported in Table S1. Six participants experienced adverse events of grade 3 or greater, including 2 in the placebo arm. All were retained in the primary analysis. A sensitivity

analysis (Figure S1) excluding these 6 participants confirmed that the pattern of significant clock comparisons and the direction of proteomic trajectories were preserved, indicating that the

primary findings are not driven by AE-associated proteomic perturbations.

Comment 1.6
The authors should verify key findings—such as the reduction in SASP factors or IGFBPs (Figure 9)—using standard assays (e.g., ELISA).

Additional experimental validation using participants' biomaterials is unfortunately not permitted under the trial protocol. We have made the exploratory nature of the proteomic and IGFBP findings explicit throughout, and we commit to seeking experimental validation in dedicated follow-up studies.

Comment 1.7
While TNIK is an intracellular kinase, are there any proximal phosphorylation substrates available in the Olink panel or assessable via ancillary assays to confirm target engagement?

We would like to thank the reviewer for this valuable suggestion. Target engagement was established using binding affinity and kinase selectivity assays reported previously (Ren et al., Nature Biotechnology 2024). References to the target engagement evidence have been made more prominent in the revised manuscript. Rentosertib is relatively clean in its inhibitory selectivity for TNIK, yet it is not possible or practical to establish direct target engagement from the Olink panel because TNIK itself is not assayed. However, we did identify indirect evidence of target engagement through pathway enrichment and PPI analytics briefly described in the revised manuscript. We also now discuss in more depth the limitations of the proteomic panels.

Comment 1.8
The discussion speculates that the 30mg BID regimen performs better than 60mg QD due to "steadier drug levels." This hypothesis needs data support. The authors should present the plasma concentration (PK) data.

We agree that this hypothesis is post-hoc and have revised the discussion to label it explicitly as such. PK data from the phase-1 and phase-2a trials are available in the published trial reports (Ren et al., Nature Biotechnology 2024; Xu et al., Nature Medicine 2025) and are now included and cited explicitly: "As shown in our previously published rentosertib clinical trial report, 60 mg QD dosing achieves two- to three-fold higher maximum plasma concentrations (Cmax) than either 30 mg regimen." The PK discussion is presented as context for a hypothesis to be tested, not as an established explanation.

Comment 1.9
While LMEM were used, the data clearly exhibits non-linear dynamics (peaks at Week 4). I suggest applying non-linear models to better characterize the temporal dynamics. This would help clarify if the "attenuation" at Week 12 represents a true reversal of anti-aging effects or stabilization at a new homeostatic equilibrium.

We thank the reviewer for this comment guiding further analysis of time-dependent effects. We added an analysis treating time as a categorical variable, which allows for a more granular comparison of individual timepoints. This new section appears in the results (“Stability of rentosertib's effects”, Figure 5) and provides complementary insight into peak and attenuation dynamics. For example, we find that only 5-9% of proteomic shifts were transient, that is, significant at earlier timepoints but not 12, whereas the vast majority are sustained or arise only later. These and other findings suggest to us that the induced anti-aging expression programs are largely durable and may appear attenuated at week 12 due to limitations in clock construction, as we now elaborate on in the Discussion.

Comment 1.10
To enhance the interpretability of Figure 1, the authors should revise the legend to clearly demarcate general signaling pathways from the specific pharmacodynamic effects of Rentosertib observed in this trial.

Figure 1 (the TNIK pathway diagram) has been removed from the revised manuscript. Following the editorial guidance to center the narrative on the methodological contribution, we concluded that a molecular pharmacology figure directing the reader's attention to a specific target would be inconsistent with the revised, more accurate framing.

Comment 1.11
A supplementary table detailing the list of features in Figure 6A should be provided for reference.

Figure 6A and the associated proteomic overlap analysis have been removed from the manuscript following concerns raised by Reviewer 3 regarding circularity (see response to Reviewer 3, Comment 3.3 below). The protein lists are nevertheless available through the data and code repositories cited in the manuscript.

Comment 1.12
To more accurately approximate clinical readouts and enhance the translational relevance of the findings, the analysis should be strengthened by correlating or integrating the proteomic clock data with other phenotypic clocks (e.g., clinical chemistry, epigenetic, or transcriptomic clocks) when possible. This multi-modal approach would provide a more holistic view of the drug's impact on biological aging.
We completely agree with the reviewer on this point suggesting correlation with other omics-based clocks to strengthen the conclusions. Integration with epigenetic or transcriptomic clocks,

however, would require matched multi-omic data not collected in this trial and impossible to collect from trial participants at this point. We note this as a direction for future trial design.

Comment 1.13
The current analysis, while comprehensive in its correlative and descriptive power regarding pathway changes, would be significantly elevated by incorporating targeted experimental validation or deeper molecular analyses to move beyond association and towards causality. For instance, can the proposed senomorphic mechanism be linked more directly to the observed clock changes?
We have included notes on this key limitation of computational omics approaches throughout the manuscript. Validating the hypothesized senomorphic mechanism in humans, however, is not currently feasible, although we cite our previous work on the subject (Tang et al., Aging and Disease 2026).

Response to Reviewer 3
Comment 3.1
Underpowered sample size and potential selection bias from excluding patients with adverse events.

The previous description of the substudy population as excluding participants "with serious adverse effects" was imprecise and has been corrected. The actual inclusion criteria were trial completion, provision of exact age, and written consent for omics research. As now reported in Table S1, 6 of 42 substudy participants experienced adverse events of grade 3 or higher, 2 of whom were in the placebo arm. All were retained in the primary analysis, and a sensitivity analysis excluding these 6 participants confirmed the robustness of key findings (Figure S1). We also highlight the small samples size as a limitation in the Discussion.

Comment 3.2
How confident can we be that detected changes reflect modulation of aging biology rather than disease-specific processes?

We agree that computational evidence alone cannot resolve this question definitively, and we no longer claim that it does. The revised manuscript reframes the study as a methodological evaluation: a demonstration of what proteomic clocks and aging pathway analytics can and cannot reveal when applied to a disease-focused trial. The UK Biobank comparison is retained as a reference for normal aging directions, with explicit acknowledgment of the limitations

introduced by comparing a healthy aging cohort to an IPF population: “However, we acknowledge the limitations of such computational and indirect evidence and intend to seek direct experimental validation.” The statements suggesting established geroprotective activity in humans has been removed.

Comment 3.3
The week 4-12 regression pattern remains unexplained.

The categorical time analysis now added to this revision addresses this directly (please see also Comment 1.9 above). We identify the proteins whose shift affect age prediction between week 4 and week 12 the most significantly, characterize their functional associations, and discuss the methodological limitations of the clocks themselves. Since these models are trained to track cumulative aging trajectories in population data, they may respond differently to transient pharmacological perturbations than to sustained biological change, potentially explaining the partial attenuation observed at week 12.

Comment 3.4
The pharmacokinetic explanation is plausible but entirely post-hoc. If dosing frequency were critical for geroprotective effects, this should have been hypothesized a priori based on known TNIK biology or pharmacology. The data pattern where the regimen with best clinical efficacy (60 mg QD) does not show the most consistent aging clock effects could equally reflect noise in an underpowered study.

This is a fair assessment, and the revised discussion now incorporates these concerns. We now present the 30 mg BID versus 60 mg QD divergence primarily as a biomarker observation that potentially reveals how proteomic clocks can decouple from clinical endpoints, rather than as evidence for a specific pharmacological mechanism. The possibility that the pattern is spurious is acknowledged (“…establishing the functional differences of the two regimens remains to be established with more direct experimental procedures”), along with further clarifications that these insights are primarily hypothesis-generating without additional validation experiments or clinical data.

Comment 3.5
No evidence that clocks capture causal aging biology.

We fully accept this point, consistent with the editorial decision to set aside the causality requirement raised by Reviewer 1. The revised manuscript more clearly avoids claims that the clocks capture causal aging biology. Instead, we evaluate them as candidate biomarkers for interventional settings and discuss their behavior, consistency, and limitations in that context. This framing is aligned with the editorial guidance.

Comment 3.6
The new analysis in Figure 6A has a circularity problem in our opinion. Aging clocks are trained on proteins that change with age, so finding that rentosertib affects proteins used by aging clocks is somewhat tautological and does not establish that the mechanism is aging-specific rather than a reflection of improved health (see our own comments below in major concern 1). The

comparison is complicated by the fact that UKBB participants are generally healthy while IPF patients have overall perturbed proteomes.

We thank the reviewer for this thoughtful critique. We agree and have removed Figure 6A and the associated overlap analysis from the manuscript. The circularity argument is valid, in that demonstrating that a drug whose discovery was partly based on aging biology modulates proteins embedded in age-trained models does not constitute independent evidence that the mechanism is aging-specific. The UK Biobank trajectory comparison, which does not rely on clock features

and uses a directional criterion (whether treatment-induced changes oppose normal aging slopes), is retained as a complementary and non-circular line of evidence, framed appropriately as correlational.

Comment 3.7
Statistical concerns. The effective number of independent tests is unclear, and the correlation structure among clocks (Fig 3B) suggests substantial redundancy. The authors should provide some estimate of what would be expected under the null hypothesis given the correlation structure.

We appreciate this statistical point and thank the reviewer for suggesting analyses that could substantiate the confidence of our findings. A permutation test with patient-level label shuffling has been added (Figure S1). Under the null, the expected number of significant comparisons across 54 clock-timepoint-arm tests is <1. The 21 significant comparisons observed thus represent a substantial departure from chance. The same permutation analysis was repeated in the sensitivity cohort (excluding participants with AEs of grade 3 or higher) with consistent results.

Response to Reviewer 5
We thank Reviewer 5 for the broadly positive assessment and constructive suggestions for minor revision points. Their framing of the work as an "in-the-wild" evaluation of aging biomarkers in a human clinical trial is precisely the lens through which we have refocused the revised manuscript. We appreciate the specific endorsement of the proteomic-versus-methylation clock comparison as an argument well-suited to this journal, and the clarification regarding the appropriate bar for biomarker utility.
The revised manuscript now includes the reviewer's suggestions for evaluating the clocks on reproducibility, technical robustness, and their ability to produce consistent and interpretable signals in an interventional setting. We now discuss the limitations where clock analyses fall short explicitly, concluding that additional, orthogonal analyses are needed to validate such findings, particularly in situations where data is sparse.

Senescence pathway language: The enrichment analyses are interesting, but some statements are too definitive (eg “confirm senomorphic activity in humans”). Suggest replacing “confirm” with “consistent with”.

We have substantially refocused the manuscript to avoid such strong conclusions about senomorphic activity, now including language as suggested to imply correlations and hypothesis-generating findings.
Plasma vs serum inconsistency: Abstract says “plasma proteomics” but samples are described as “serum samples”.

Thank you for identifying this inconsistency. It has been addressed.
Expand discussion on clock discordance and implications for clinical trial design.

We now include discussion on the mortality-based clocks having significantly less concordance with chronology-based clocks, likely reflective of their encountering difficulty in settings of high disease burden characterized by broad dysregulation across tissues and molecular pathways. We emphasize that integrating results from multiple clocks can identify the most confident correlations and biologically meaningful findings.
Add a short paragraph in Discussion on requirements for clinical utility of aging clocks (predictive validity / reproducibility / impact on decision making).

We now include descriptions in our Discussion of the challenges we encountered implementing these clock analyses in a clinical trial, including informed consent, regulatory, and preclinical validation hurdles to even consider these for clinical use. These challenges inform future design considerations for bringing these analyses into more regular use as clinical trial endpoints.

We believe the revised manuscript represents a substantial improvement over the previously submitted text. The narrative is now tightly focused on technological advances, rather than the

putative geroprotective properties of the drug. Our claims have been calibrated according to the available evidence, and the additional analyses, metadata, and sensitivity checks requested by the reviewers have been incorporated. We look forward to the Editor's and reviewers' assessment of the revision.

Sincerely,
Alex Zhavoronkov, on behalf of all authors
(alex@insilico.com)

Version 2:

Author Rebuttal letter:

Response to Reviewers
We thank the editor and the reviewers for the opportunity to respond and re-write the manuscript according to their comments and feedback. We have expanded the data we present and reframe the paper from a report on clinical age reversal to the first implementation of a clinical trial design template where a novel therapeutic discovered with a deep focus on aging biology is evaluated in the context of a real clinical study.

We now have re-framed the central achievement of our work as a "blueprint" demonstrating how proteomic aging clocks can be embedded in disease-focused clinical trials to generate aging biology hypotheses, detect or interpret clock signals, and identify pharmacological candidates warranting dedicated geroprotective studies. All claims regarding rejuvenation, senomorphic, or geroprotective activity of rentosertib in humans have thus been removed or substantially toned down throughout the title, abstract, and main text in accordance with the review comments. We also now highlight the key limitations of the aging clock methodology and how complementary methods can compensate for them, making this piece valuable as a stress-test of an emerging technology, not previously applied to clinical trial analytics.

Of note, to address the concerns expressed by multiple reviewers about potential contributions of unaccounted for variables, such as BMI and others, we obtained and are now permitted to report a richer set of participant metadata. We now disclose BMI, sex, and higher-grade adverse events. We also performed additional analyses to confirm insignificant BMI ~ BioAge correlations, insensitivity of our key findings to adverse events, and present a more detailed exploration of proteomic trajectories, correcting for BMI and sex in all models. A comprehensive demographic summary is now provided in Table S1.
We have also significantly restructured the text itself to improve its readability and maintain the focus on the essential. In total, the reference list has been compacted from 128 to 85 items, and the main text's volume has been reduced from 6.3K words to 5.6K while supporting all the new results. All existing figures and tables have been updated and some new ones have been added.
Please find below the point-by-point response to all critical comments we received:
Response to Reviewer 1
Comment 1.1
The robustness of the study's conclusions is limited by the small sample size (N=11 per group, N=9 for one group). Therefore, the manuscript must transparently address these statistical constraints. The conclusions should be carefully framed to reflect that this is an exploratory, signal-seeking analysis rather than a definitive confirmation of efficacy, particularly regarding the lack of significance in certain clock or group.

We fully agree. The revised manuscript now consistently frames the analysis as exploratory and hypothesis-generating. We have added a permutation test (Figure S2) demonstrating that the 21 significant comparisons across all clock-timepoint-regimen combinations far exceed the null expectation (mean under permutation = 0.15), providing a benchmark for what chance-level performance would look like in this dataset (lines 151-153). In the sections where we discuss limitation, the statistical power constraints and their implications for interpretation are explicitly stated: "The modest sample size, relatively short duration, dominance of computational approaches, and lack of complementary omics modalities limit our ability to conclusively claim rentosertib's human geroprotector status." (lines 476-478)

Comment 1.2
The results indicate that regardless of the dosage regimen, the therapeutic efficacy was most pronounced at Week 4, followed by an attenuation of effect by Week 12. The investigators may consider evaluating efficacy at Week 8 to determine if the therapeutic benefit remains constant throughout the trial duration. Furthermore, examining the proteomic data at Week 12 for markers of cellular stress or exhaustion would help to better substantiate the molecular mechanisms underlying this observed attenuation.

Proteomic profiling was performed only at the prespecified timepoints (baseline, week 2, week 4, and week 12) and intermediate timepoints are unavailable. We have addressed the temporal dynamics question by introducing a categorical time analysis that directly compares protein levels between timepoints, enabling

identification of the proteins most responsible for the week-4 peak and subsequent partial attenuation (line 222, Figure S7).
Regarding cellular stress and exhaustion markers: after applying the minimum gene set size filter (n > 25 proteins shared with Olink), only one relevant Reactome pathway, CELLULAR_ RESPONSE_TO_CHEMICAL_STRESS, passed the threshold. Results for this pathway were inconclusive (30 mg BID: NES = +1.66 at week 12 was the only significant one with q-value = 0.019; for 60 mg QD: q-value > 0.28 at both timepoints) and are not reported in the main text. We acknowledge this as a technological limitation of the Olink platform in the discussion: "This technological constraint makes some statistical tests incomplete or impossible, as was the case for our GSEA when enrichment scores could be obtained for only 145 of 1,787 available Reactome pathways due to limited intersection between array and pathway features" (line 425).

Comment 1.3
The central conclusion regarding the dissociation of Rentosertib's geroprotective effects from its anti-fibrotic activity requires stronger justification. Given that IPF is characterized by accelerated aging, the authors should critically discuss why the 60 mg QD group, which exhibited the most significant clinical improvement in FVC, paradoxically displayed less consistent reductions in aging clocks compared to the 30 mg BID group.

We agree that the evidence for mechanistic dissociation is inferential given the constraints of the dataset. This claim has been substantially toned down throughout the manuscript. The divergence between the 60 mg QD and 30 mg BID groups is now presented as an informative biomarker observation rather than established pharmacological evidence. The observed dissociation is noteworthy and hypothesis-generating, but we no longer frame it as proof of independent anti-aging and anti-fibrotic mechanisms (e.g., lines 272-275). Our revised Discussion includes a hypothesis that the 2–3-fold difference in serum concentration (Cmax) reported in the original clinical trial report may lead to differential engagement and saturation of target kinases (line428). The Discussion also includes a methodological note acknowledging that aging clocks cannot distinguish between beneficial processes and compensatory responses that may artifactually shift age predictions (lines 398-399).

Comment 1.4
A dedicated demographic table for the 42-patient proteomic substudy is missing. Given the authors' emphasis on metabolic and IGF-1 signaling pathways as key mechanisms, Body Mass Index (BMI) is a critical variable. Please provide age, sex, and specifically BMI data. Additionally, did the authors analyze whether baseline BMI correlates with the magnitude of biological age reduction (dBioAge)?

We originally did not have permission to publish participant-level data, but we have now obtained written permission from the organization collecting data in the study to include participant-level metadata and now present a full demographic summary in Table S1 including BMI, sex, age, and AE grade (lines 128, 670-674). All mixed-effects models were re-run with BMI and sex included as covariates, although the results were qualitatively unchanged (line 159, model details at line 712-714). We additionally tested Spearman correlations between baseline BMI and ΔBioAge across all six clocks and three treated arms (54 tests total). Three reached nominal significance ($p < 0.05$), and none survived multiple testing correction (all q-values > 0.25), confirming that the observed biological age changes were not substantively modulated by BMI within the studied weight range.

Comment 1.5
Adverse Events (AEs) can induce systemic inflammation and alter the plasma proteome, potentially skewing aging clock predictions. Please report the incidence of AEs in this sub-cohort. A sensitivity analysis excluding patients with moderate-to-severe AEs is recommended to ensure the observed "rejuvenation" (and subsequent attenuation) is driven by the drug's senomorphic activity rather than AE-induced physiological stress.

AE data are now reported in Table S1 (line 128). Six participants experienced adverse events of grade 3 or greater, including 2 in the placebo arm. All were retained in the primary analysis. A sensitivity analysis (Figure S1) excluding these 6 participants confirmed that the pattern of significant clock comparisons and the direction of proteomic trajectories were preserved, indicating that the primary findings are not driven by AE-associated proteomic perturbations (line 155-158): “Similar findings persisted when the six participants with higher-grade adverse events were removed from the analysis. Among the three regimens, 30 mg BID produced the most consistent signal (9 significant comparisons), followed by 60 mg QD (7) and 30 mg QD (5)”.

Comment 1.6
The authors should verify key findings—such as the reduction in SASP factors or IGFBPs (Figure 9)—using

standard assays (e.g., ELISA).

Additional experimental validation using participants' biomaterials is unfortunately not permitted under the trial protocol. We have made the exploratory nature of the proteomic and IGFBP findings explicit throughout, and we commit to seeking experimental validation in dedicated follow-up studies.

Comment 1.7
While TNIK is an intracellular kinase, are there any proximal phosphorylation substrates available in the Olink panel or assessable via ancillary assays to confirm target engagement?

We would like to thank the reviewer for this valuable suggestion. Target engagement was established using binding affinity and kinase selectivity assays reported previously (Ren et al., Nature Biotechnology 2024). References to the target engagement evidence have been made more prominent in the revised manuscript (line 434): “As shown in our previously published rentosertib clinical trial report, 60 mg QD dosing achieves two- to three-fold higher maximum plasma concentrations…”.
Rentosertib is relatively clean in its inhibitory selectivity for TNIK, yet it is not possible or practical to establish direct target engagement from the Olink panel because TNIK itself is not assayed (line 426): “Importantly, TNIK itself is not part of the Olink 3072 platform used in this work, allowing only for indirect (PPI, GSEA) validations of target engagement.” However, we did identify indirect evidence of target engagement through pathway enrichment and PPI analytics briefly described in the revised manuscript. We also now discuss in more depth the limitations of the proteomic panels (line 396): “Proteomic aging clocks still carry the caveats common among other omics-based clocks” Also see lines 391-396.

Comment 1.8
The discussion speculates that the 30mg BID regimen performs better than 60mg QD due to "steadier drug levels." This hypothesis needs data support. The authors should present the plasma concentration (PK) data.

We agree that this hypothesis is post-hoc and have revised the discussion to label it explicitly as such. PK data from the phase-1 and phase-2a trials are available in the published trial reports (Ren et al., Nature Biotechnology 2024; Xu et al., Nature Medicine 2025) and are now included and cited explicitly: "As shown in our previously published rentosertib clinical trial report, 60 mg QD dosing achieves two- to three-fold higher maximum plasma concentrations (Cmax) than either 30 mg regimen." (line 434) The PK discussion is presented as context for a hypothesis to be tested, not as an established explanation.
Comment 1.9
While LMEM were used, the data clearly exhibits non-linear dynamics (peaks at Week 4). I suggest applying non-linear models to better characterize the temporal dynamics. This would help clarify if the "attenuation" at Week 12 represents a true reversal of anti-aging effects or stabilization at a new homeostatic equilibrium.

We thank the reviewer for this comment guiding further analysis of time-dependent effects. We added an analysis treating time as a categorical variable, which allows for a more granular comparison of individual timepoints. This new section appears in the results (“Stability of rentosertib's effects”, Supplemental Figure S7, line 218) and provides complementary insight into peak and attenuation dynamics. For example, we find that only 5-9% of proteomic shifts were transient, that is, significant at earlier timepoints but not 12, whereas the vast majority are sustained or arise only later. These and other findings suggest to us that the induced anti-aging expression programs are largely durable and may appear attenuated at week 12 due to limitations in clock construction, as we now elaborate on in the Discussion.

Comment 1.10
To enhance the interpretability of Figure 1, the authors should revise the legend to clearly demarcate general signaling pathways from the specific pharmacodynamic effects of Rentosertib observed in this trial.

Figure 1 (the TNIK pathway diagram) has been removed from the revised manuscript. Following the editorial guidance to center the narrative on the methodological contribution, we concluded that a molecular pharmacology figure directing the reader's attention to a specific target would be inconsistent with the revised, more accurate framing. Now, Figure 1 displays the proposed translational pipeline, and contextualizes the reported work within it.

Comment 1.11
A supplementary table detailing the list of features in Figure 6A should be provided for reference.

Figure 6A and the associated proteomic overlap analysis have been removed from the manuscript following concerns raised by Reviewer 3 regarding circularity (see response to Reviewer 3, Comment 3.3 below). The protein lists are nevertheless available through the data and code repositories cited in the manuscript.

Comment 1.12
To more accurately approximate clinical readouts and enhance the translational relevance of the findings, the analysis should be strengthened by correlating or integrating the proteomic clock data with other phenotypic clocks (e.g., clinical chemistry, epigenetic, or transcriptomic clocks) when possible. This multi-modal approach would provide a more holistic view of the drug's impact on biological aging.
We completely agree with the reviewer on this point suggesting correlation with other omics-based clocks to strengthen the conclusions. Integration with epigenetic or transcriptomic clocks, however, would require matched multi-omic data not collected in this trial and impossible to collect from trial participants at this point. We note this as a direction for future trial design (line 401): “Future trials featuring multi-modal clocks may provide more robust conclusions.”
Comment 1.13
The current analysis, while comprehensive in its correlative and descriptive power regarding pathway changes, would be significantly elevated by incorporating targeted experimental validation or deeper molecular analyses to move beyond association and towards causality. For instance, can the proposed senomorphic mechanism be linked more directly to the observed clock changes?
We have included notes on this key limitation of computational omics approaches throughout the manuscript, e.g. lines 476-478: “The modest sample size, relatively short duration, dominance of computational approaches, and lack of complementary omics modalities limit our ability to conclusively claim rentosertib’s human geroprotector status.” Validating the hypothesized senomorphic mechanism in humans, however, is not currently feasible, although we cite our previous work on the subject (Tang et al., Aging and Disease 2026), e.g. on lines 405-406: “Our senescence marker analysis builds on previous in vitro demonstrations of rentosertib’s ability to prevent senescent cell formation in IMR-90 cells.”

Response to Reviewer 3
Comment 3.1
Underpowered sample size and potential selection bias from excluding patients with adverse events.

The previous description of the substudy population as excluding participants "with serious adverse effects" was imprecise and has been corrected. The actual inclusion criteria were trial completion, provision of exact age, and written consent for omics research (lines 669-674). As now reported in Table S1, 6 of 42 substudy participants experienced adverse events of grade 3 or higher, 2 of whom were in the placebo arm. All were retained in the primary analysis, and a sensitivity analysis excluding these 6 participants confirmed the robustness of key findings (Figure S1, lines 155-158): “Similar findings persisted when the six participants with higher-grade adverse events were removed from the analysis. Among the three regimens, 30 mg BID produced the most consistent signal (9 significant comparisons), followed by 60 mg QD (7) and 30 mg QD (5).” We also highlight the small samples size as a limitation in the Discussion (line 476).

Comment 3.2
How confident can we be that detected changes reflect modulation of aging biology rather than disease-specific processes?

We agree that computational evidence alone cannot resolve this question definitively, and we no longer claim that it does. The revised manuscript reframes the study as a methodological evaluation: a demonstration of what proteomic clocks and aging pathway analytics can and cannot reveal when applied to a disease-focused trial. The UK Biobank comparison is retained as a reference for normal aging directions, with explicit acknowledgment of the limitations introduced by comparing a healthy aging cohort to an IPF population (line 274): “However, we acknowledge the limitations of such computational and indirect evidence and intend to seek direct experimental validation.” The statements suggesting established geroprotective activity in humans has been removed.

Comment 3.3
The week 4-12 regression pattern remains unexplained.

The categorical time analysis now added to this revision addresses this directly (please see also Comment 1.9 above, line 222). We identify the proteins whose shift affect age prediction between week 4 and week 12 the most significantly, characterize their functional associations, and discuss the methodological limitations of the clocks themselves. Since these models are trained to track cumulative aging trajectories in population data, they may respond differently to transient pharmacological perturbations than to

sustained biological change, potentially explaining the partial attenuation observed at week 12.

Comment 3.4
The pharmacokinetic explanation is plausible but entirely post-hoc. If dosing frequency were critical for geroprotective effects, this should have been hypothesized a priori based on known TNIK biology or pharmacology. The data pattern where the regimen with best clinical efficacy (60 mg QD) does not show the most consistent aging clock effects could equally reflect noise in an underpowered study.

This is a fair assessment, and the revised discussion now incorporates these concerns. We now present the 30 mg BID versus 60 mg QD divergence primarily as a biomarker observation that potentially reveals how proteomic clocks can decouple from clinical endpoints, rather than as evidence for a specific pharmacological mechanism. The possibility that the pattern is spurious is acknowledged ("Conclusively establishing the functional differences between the two regimens, however, will require more direct experimental procedures.", line 395), along with further clarifications that these insights are primarily hypothesis-generating without additional validation experiments or clinical data (e.g., line 442).

Comment 3.5
No evidence that clocks capture causal aging biology.

We fully accept this point, consistent with the editorial decision to set aside the causality requirement raised by Reviewer 1. The revised manuscript more clearly avoids claims that the clocks capture causal aging biology, e.g. on line 398: "Establishing causality remains a challenge, as the statistical trends they capture may be associated with both compensatory and harmful aging processes" Instead, we evaluate them as candidate biomarkers for interventional settings and discuss their behavior, consistency, and limitations in that context. This framing is aligned with the editorial guidance.

Comment 3.6
The new analysis in Figure 6A has a circularity problem in our opinion. Aging clocks are trained on proteins that change with age, so finding that rentosertib affects proteins used by aging clocks is somewhat tautological and does not establish that the mechanism is aging-specific rather than a reflection of improved health (see our own comments below in major concern 1). The comparison is complicated by the fact that UKBB participants are generally healthy while IPF patients have overall perturbed proteomes.

We thank the reviewer for this thoughtful critique. We agree and have removed Figure 6A and the associated overlap analysis from the manuscript. The circularity argument is valid, in that demonstrating that a drug whose discovery was partly based on aging biology modulates proteins embedded in age-trained models does not constitute independent evidence that the mechanism is aging-specific. The UK Biobank trajectory comparison, which does not rely on clock features and uses a directional criterion (whether treatment-induced changes oppose normal aging slopes), is retained as a complementary and non-circular line of evidence, framed appropriately as correlational.

Comment 3.7
Statistical concerns. The effective number of independent tests is unclear, and the correlation structure among clocks (Fig 3B) suggests substantial redundancy. The authors should provide some estimate of what would be expected under the null hypothesis given the correlation structure.

We appreciate this statistical point and thank the reviewer for suggesting analyses that could substantiate the confidence of our findings. A permutation test with patient-level label shuffling has been added (Figure S2, lines 154-155). Under the null, the expected number of significant comparisons across 54 clock-timepoint-arm tests is <1. The 21 significant comparisons observed thus represent a substantial departure from chance. The same permutation analysis was repeated in the sensitivity cohort (excluding participants with AEs of grade 3 or higher) with consistent results.

Response to Reviewer 5
We thank Reviewer 5 for the broadly positive assessment and constructive suggestions for minor revision points. Their framing of the work as an "in-the-wild" evaluation of aging biomarkers in a human clinical trial is precisely the lens through which we have refocused the revised manuscript. We appreciate the specific endorsement of the proteomic-versus-methylation clock comparison as an argument well-suited to

this journal, and the clarification regarding the appropriate bar for biomarker utility.
The revised manuscript now includes the reviewer's suggestions for evaluating the clocks on reproducibility, technical robustness, and their ability to produce consistent and interpretable signals in an interventional setting. We now discuss the limitations where clock analyses fall short explicitly, concluding that additional, orthogonal analyses are needed to validate such findings, particularly in situations where data is sparse.

Senescence pathway language: The enrichment analyses are interesting, but some statements are too definitive (eg “confirm senomorphic activity in humans”). Suggest replacing “confirm” with “consistent with”.
We have substantially refocused the manuscript to avoid such strong conclusions about senomorphic activity, now including language as suggested to imply correlations and hypothesis-generating findings.
Plasma vs serum inconsistency: Abstract says “plasma proteomics” but samples are described as “serum samples”.

Thank you for identifying this inconsistency. It has been addressed and the manuscript now clearly reads “serum”.
Expand discussion on clock discordance and implications for clinical trial design.

We now include discussion on the mortality-based clocks having significantly less concordance with chronology-based clocks, likely reflective of their encountering difficulty in settings of high disease burden characterized by broad dysregulation across tissues and molecular pathways. We emphasize that integrating results from multiple clocks can identify the most confident correlations and biologically meaningful findings (lines 401-404).
Add a short paragraph in Discussion on requirements for clinical utility of aging clocks (predictive validity / reproducibility / impact on decision making).

We now include descriptions in our Discussion of the challenges we encountered implementing these clock analyses in a clinical trial, including informed consent, regulatory, and preclinical validation hurdles to even consider these for clinical use (lines 451-468). These challenges inform future design considerations for bringing these analyses into more regular use as clinical trial endpoints.

We believe the revised manuscript represents a substantial improvement over the previously submitted text. The narrative is now tightly focused on technological advances, rather than the putative geroprotective properties of the drug. Our claims have been calibrated according to the available evidence, and the additional analyses, metadata, and sensitivity checks requested by the reviewers have been incorporated. We look forward to the Editor’s and reviewers’ assessment of the revision.

Sincerely,
Alex Zhavoronkov, on behalf of all authors
(alex@insilico.com)

Version 3:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this revised manuscript, the authors have incorporated additional analytical insights and moderated the original conclusions accordingly, and the major concerns of the reviewers have been addressed. The manuscript is suitable for publication in Nature Biotechnology.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have addressed my concerns, mainly by backtracking on their claims. From my perspective, the manuscript is technically sound and acknowledges limitations now.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Biotechnology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)
The authors have addressed all my comments

(Remarks on code availability)