

Multi-omic underpinnings of epigenetic aging and human longevity

Lucas A. Mavromatis^{1#}, ScB; Daniel B. Rosoff^{1,2#}, AB, ScB; Andrew S. Bell¹, BA; Jeesun Jung¹, PhD; Josephin Wagner¹, MD, MSc; Falk W. Lohoff^{1*}, MD.

¹Section on Clinical Genomics and Experimental Therapeutics, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD, USA

²NIH-Oxford-Cambridge Scholars Program; University of Oxford, UK

***Corresponding Author:**

Falk W. Lohoff, M.D.

Chief, Section on Clinical Genomics and Experimental Therapeutics (CGET)

Lasker Clinical Research Scholar

National Institute on Alcohol Abuse and Alcoholism (NIAAA)

National Institutes of Health

10 Center Drive (10CRC/2-2352)

Bethesda, MD 20892-1540

Office: 301-827-1542

falk.lohoff@nih.gov

#These authors contributed equally

Supplementary Note 1 (STROBE-MR Checklist)

1. TITLE and ABSTRACT

Indicate Mendelian randomization as the study's design in the title and/or the abstract.

Abstract

INTRODUCTION

2. Background

Explain the scientific background and rationale for the reported study. Is causality between exposure and outcome plausible? Justify why MR is a helpful method to address the study question.

Addressed in the Introduction and Methods

3. Objectives

State specific objectives clearly, including pre-specified causal hypotheses (if any).

Addressed in the Introduction.

METHODS

4. Study design and data sources

Present key elements of study design early in the paper. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following:

- a) Describe the study design and the underlying population from which it was drawn. Describe also the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, if available.*
- b) Give the eligibility criteria, and the sources and methods of selection of participants.*
- c) Explain how the analyzed sample size was arrived at.*
- d) Describe measurement, quality and selection of genetic variants.*
- e) For each exposure, outcome and other relevant variables, describe methods of assessment and, in the case of diseases, the diagnostic criteria used.*
- f) Provide details of ethics committee approval and participant informed consent, if relevant.*

Addressed in the Methods.

5. Assumptions

Explicitly state assumptions for the main analysis (e.g. relevance, exclusion, independence, homogeneity) as well assumptions for any additional or sensitivity analysis.

Addressed in the Methods.

6. Statistical methods: main analysis

Describe statistical methods and statistics used.

a) Describe how quantitative variables were handled in the analyses (i.e., scale, units, model).

b) Describe the process for identifying genetic variants and weights to be included in the analyses (i.e, independence and model). Consider a flow diagram.

c) Describe the MR estimator, e.g. two-stage least squares, Wald ratio, and related statistics.

Detail the included covariates and, in case of two-sample MR, whether the same covariate set was used for adjustment in the two samples.

d) Explain how missing data were addressed.

e) If applicable, say how multiple testing was dealt with.

Addressed in the Methods.

7. Assessment of assumptions

Describe any methods used to assess the assumptions or justify their validity.

Addressed in the Methods.

8. Sensitivity analyses

Describe any sensitivity analyses or additional analyses performed.

Addressed in the Methods.

9. Software and pre-registration

a) Name statistical software and package(s), including version and settings used.

Addressed in the Methods.

b) State whether the study protocol and details were pre-registered (as well as when and where).

Addressed in the Methods.

RESULTS

10. Descriptive data

a) Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow-diagram.

b) Report summary statistics for phenotypic exposure(s), outcome(s) and other relevant variables (e.g. means, standard deviations, proportions).

c) If the data sources include meta-analyses of previous studies, provide the number of studies, their reported ancestry, if available, and assessments of heterogeneity across these studies. Consider using a supplementary table for each data source.

d) For two-sample Mendelian randomization:

i. Provide information on the similarity of the genetic variant-exposure associations between the exposure and outcome samples.

ii. Provide information on extent of sample overlap between the exposure and outcome data sources.

Addressed in the Methods, Results and Supplementary Data.

11. Main results

a) Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale (e.g. comparing 25th and 75th percentile of allele count or genetic risk score, if individual-level data available).

b) Report causal effect estimate between exposure and outcome, and the measures of uncertainty from the MR analysis. Use an intuitive scale, such as odds ratio, or relative risk, per standard deviation difference.

c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time-period.

d) Consider any plots to visualize results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure).

Addressed in the Results and Supplementary Data.

12. Assessment of assumptions

a) Assess the validity of the assumptions.

b) Report any additional statistics (e.g., assessments of heterogeneity, such as I^2 , Q statistic).

Addressed in the Results, Supplementary Data and Discussion.

13. Sensitivity and additional analyses

a) Use sensitivity analyses to assess the robustness of the main results to violations of the assumptions.

b) Report results from other sensitivity analyses (e.g., replication study with different dataset, analyses of subgroups, validation of instrument(s), simulations, etc.).

c) Report any assessment of direction of causality (e.g., bidirectional MR).

d) When relevant, report and compare with estimates from non-MR analyses.

e) Consider any additional plots to visualize results (e.g., leave-one-out analyses).

Addressed in the Results and Supplementary Data.

DISCUSSION

14. Key results

Summarize key results with reference to study objectives.

Addressed in the Discussion.

15. Limitations

Discuss limitations of the study, taking into account the validity of the MR assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias, and any efforts to address them.

Addressed in the Discussion.

16. Interpretation

a) Give a cautious overall interpretation of results considering objectives and limitations.

Compare with results from other relevant studies.

b) Discuss underlying biological mechanisms that could be modelled by using the genetic variants to assess the relationship between the exposure and the outcome.

c) Discuss whether the results have clinical or policy relevance, and whether interventions could have the same size effect.

Addressed in the Discussion.

17. Generalizability

Discuss the generalizability of the study results (a) to other populations (i.e. external validity),

(b) across other exposure periods/timings, and (c) across other levels of exposure.

Addressed in the Discussion.

OTHER INFORMATION

18. Funding

Give the source of funding and the role of the funders for the present study and, if applicable, for the original study or studies on which the present article is based.

Addressed in the Acknowledgements.

19. Data and data sharing

Present data used to perform all analyses or report where and how the data can be accessed.

State whether statistical code is publicly accessible and if so, where.

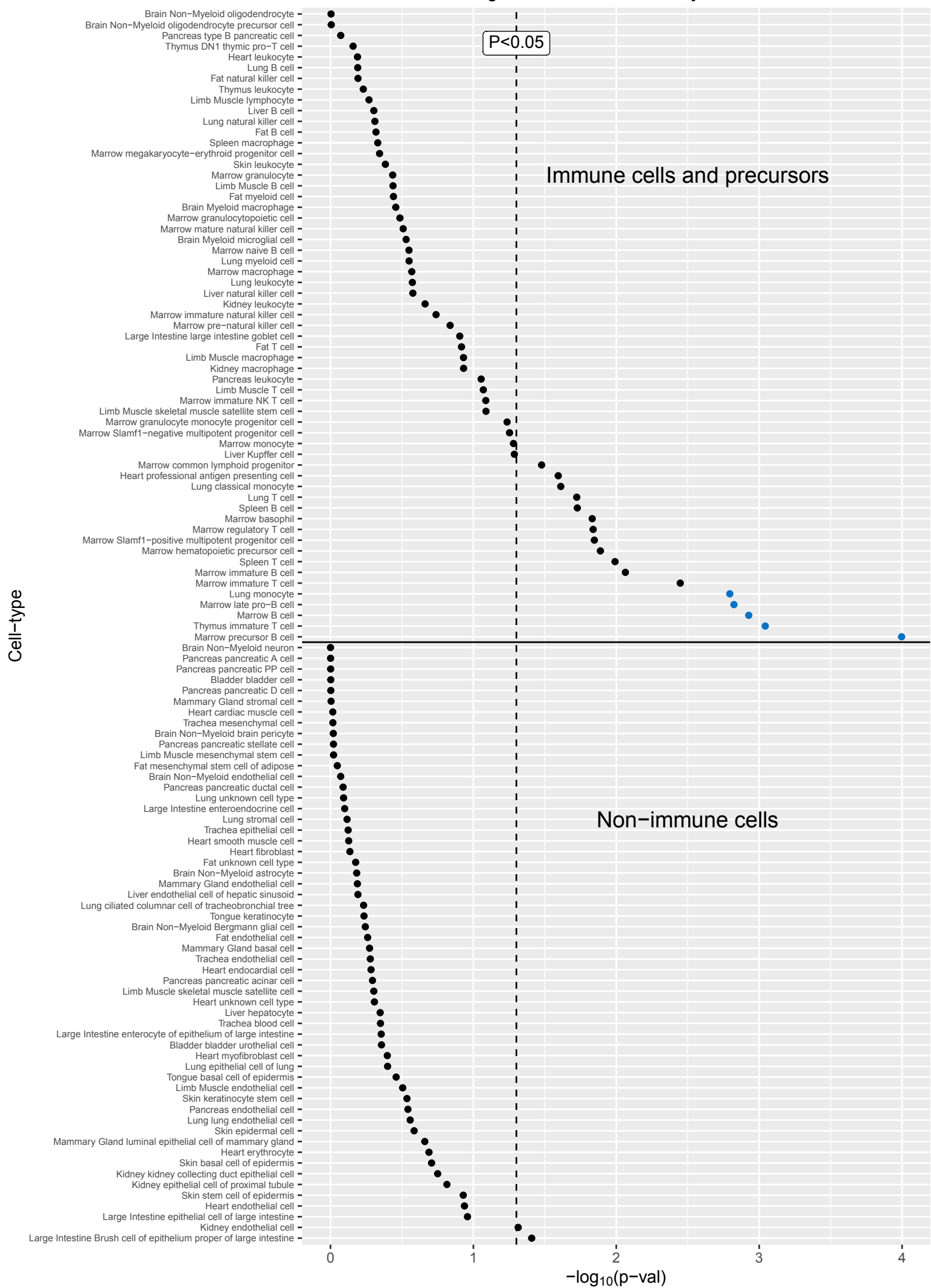
Addressed in the Methods.

20. Conflicts of Interest

All authors should declare all potential conflicts of interest.

Addressed in the Competing interests.

● = Significant at False Discovery Rate of 0.05



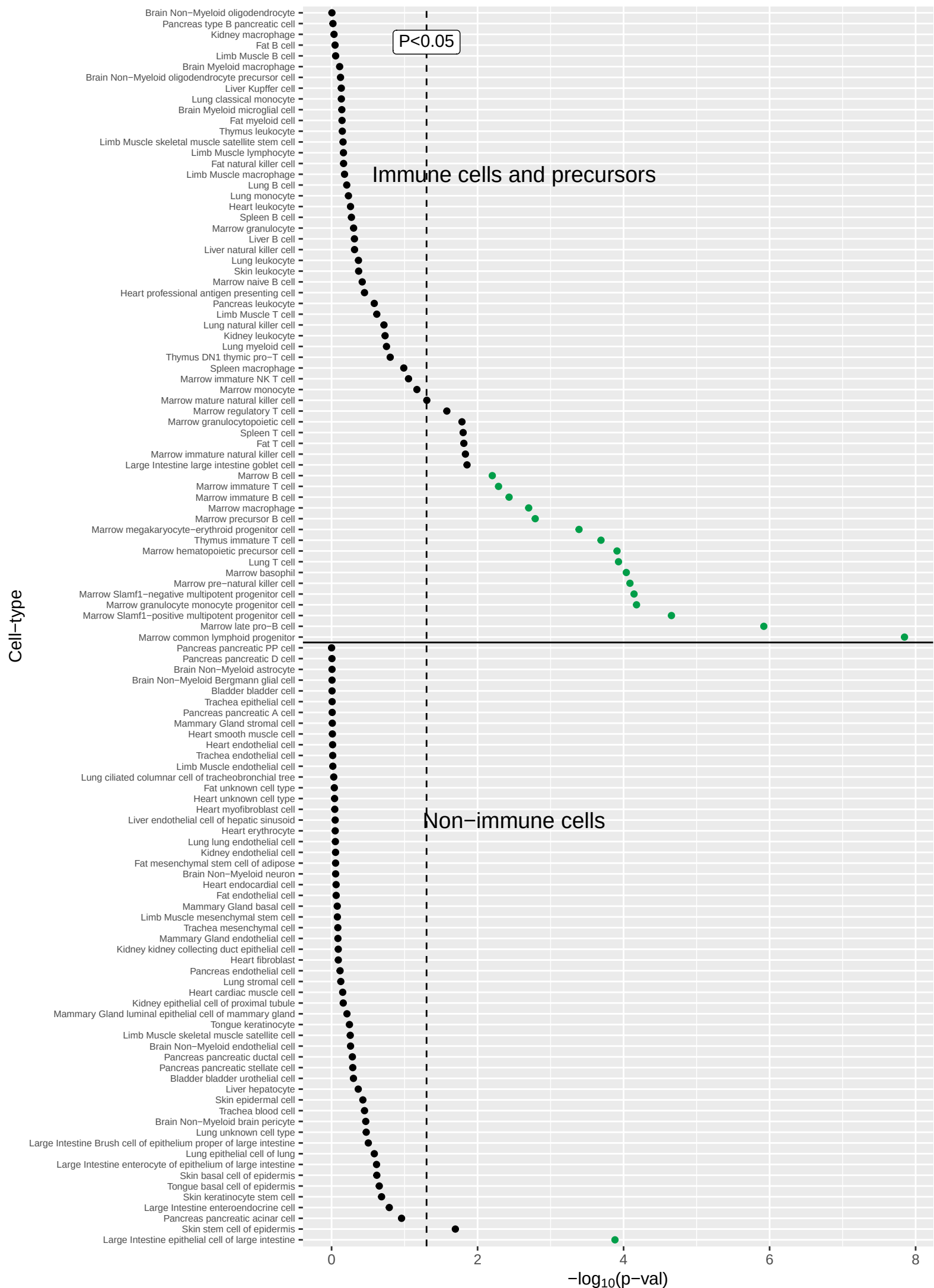
Supplementary Fig. 1: CELLECT-MAGMA cellular associations with IEAA (immune cells and precursors, non-immune cells)

Results from CELLECT-MAGMA cell-type enrichment analysis of IEAA. Cell types are split into immune cells and immune cell precursors (top half of figure) and non-immune cells (bottom half of figure). Circles represent negative, log-transformed, unadjusted P values. Statistical analyses were conducted using one-sided t-tests. Blue circles represent cell types significantly enriched in IEAA-associated SNPs at a FDR of 0.05.

Black circles represent non-significant associations. Black circles to the right of the dotted line represent nominally significant associations (i.e., cells with P value < 0.05 that fail to survive correction for multiple comparisons). Exact P values are contained in Supplementary Data 30.

CELLECT: CELL-type Expression-specific integration for Complex Traits; MAGMA: Multi-marker Analysis of GenoMic Annotation; IEAA: intrinsic epigenetic age acceleration; SNP: single nucleotide polymorphism; FDR: false discovery rate.

● = Significant at False Discovery Rate of 0.05

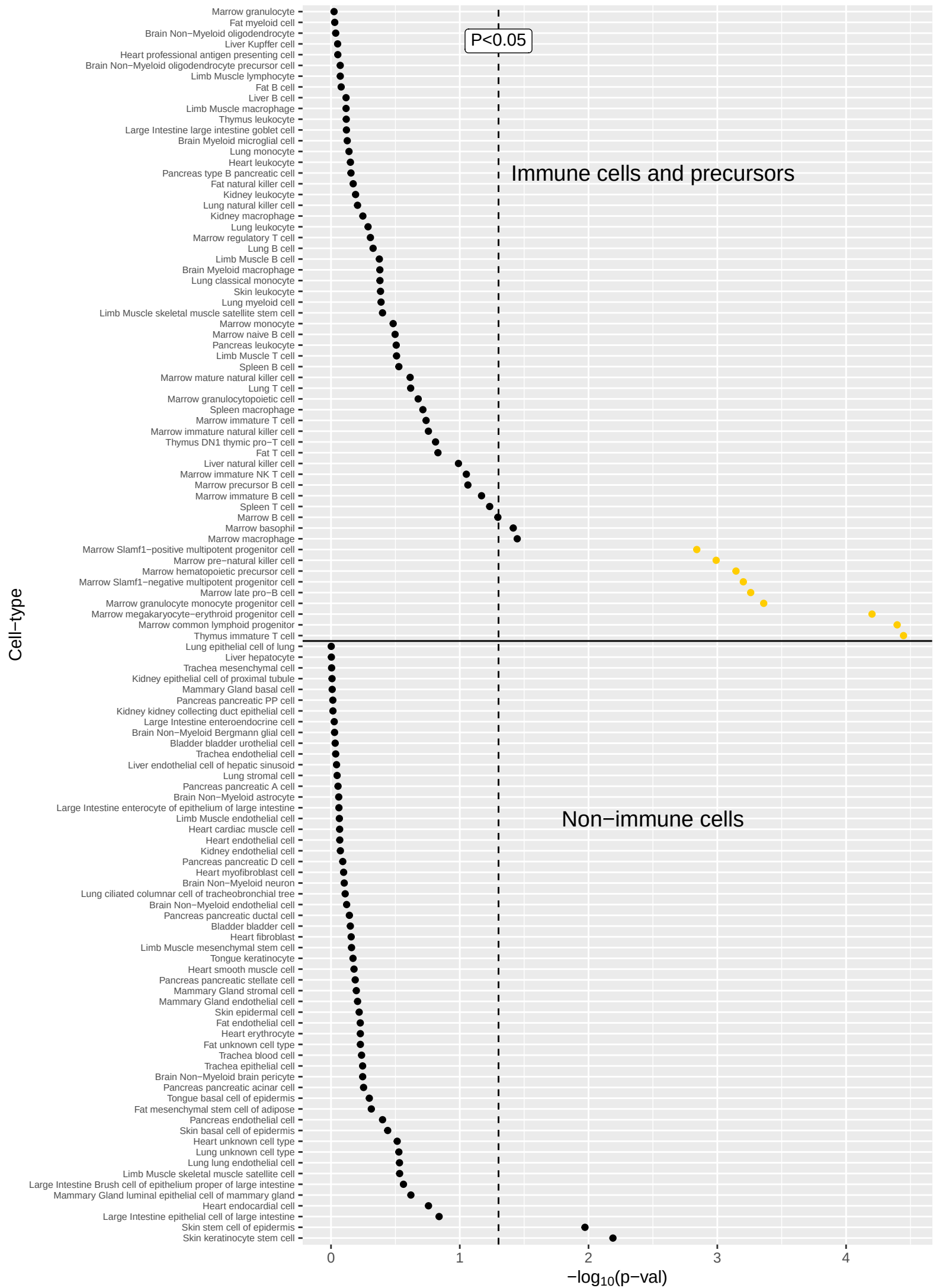


Supplementary Fig. 2: CELLECT-MAGMA cellular associations with HannumAge (immune cells and precursors, non-immune cells)

Results from CELLECT-MAGMA cell-type enrichment analysis of HannumAge. Cell types are split into immune cells and immune cell precursors (top half of figure) and non-immune cells (bottom half of figure). Circles represent negative, log-transformed, unadjusted P values. Statistical analyses were conducted using one-sided t-tests. Green circles represent cell types significantly enriched in HannumAge-associated SNPs at a FDR of 0.05. Black circles represent non-significant associations. Black circles to the right of the dotted line represent nominally significant associations (i.e., cells with P value < 0.05 that fail to survive correction for multiple comparisons). Exact P values are contained in Supplementary Data 31.

CELLECT: CELL-type Expression-specific integration for CompleT Traits; MAGMA: Multi-marker Analysis of GenoMic Annotation; SNP: single nucleotide polymorphism; FDR: false discovery rate.

● = Significant at False Discovery Rate of 0.05

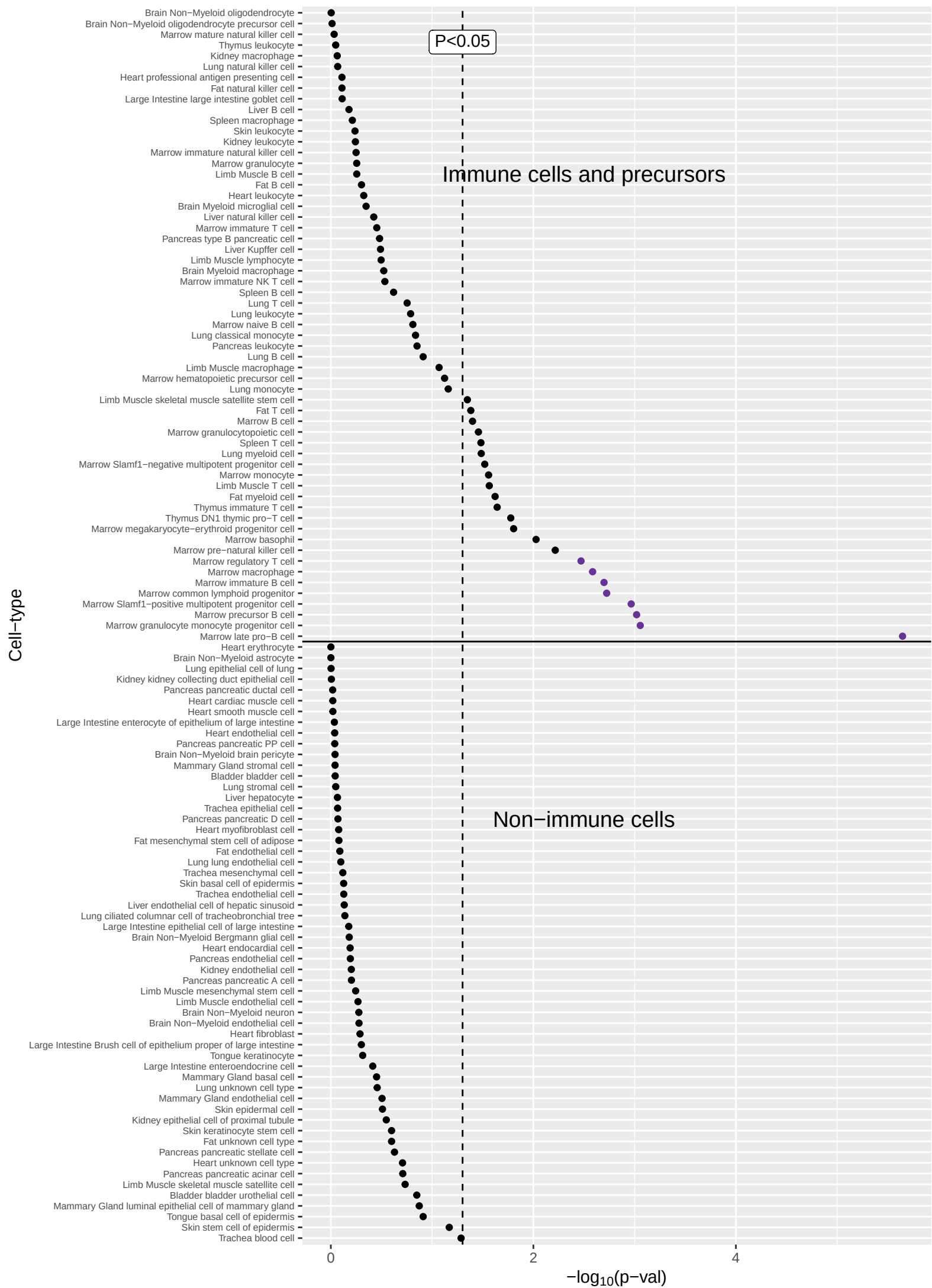


Supplementary Fig. 3: CELLECT-MAGMA cellular associations with GrimAge (immune cells and precursors, non-immune cells)

Results from CELLECT-MAGMA cell-type enrichment analysis of GrimAge. Cell types are split into immune cells and immune cell precursors (top half of figure) and non-immune cells (bottom half of figure). Circles represent negative, log-transformed, unadjusted P values. Statistical analyses were conducted using one-sided t-tests. Yellow circles represent cell types significantly enriched in GrimAge-associated SNPs at a FDR of 0.05. Black circles represent non-significant associations. Black circles to the right of the dotted line represent nominally significant associations (i.e., cells with P value < 0.05 that fail to survive correction for multiple comparisons). Exact P values are contained in Supplementary Data 32.

CELLECT: CELL-type Expression-specific integration for Complex Traits; MAGMA: Multi-marker Analysis of GenoMic Annotation; SNP: single nucleotide polymorphism; FDR: false discovery rate.

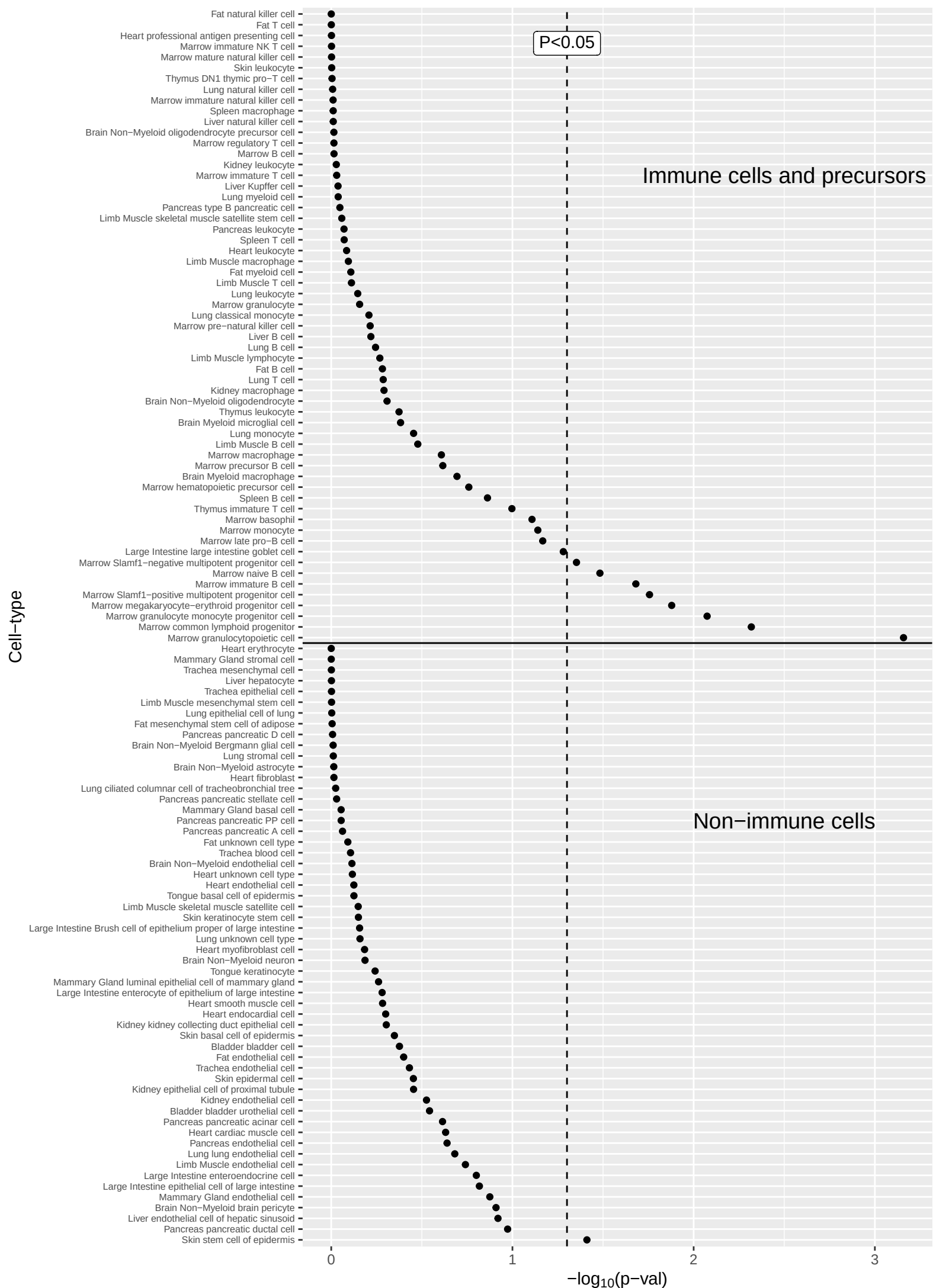
● = Significant at False Discovery Rate of 0.05



Supplementary Fig. 4: CELLECT-MAGMA cellular associations with PhenoAge (immune cells and precursors, non-immune cells)

Results from CELLECT-MAGMA cell-type enrichment analysis of PhenoAge. Cell types are split into immune cells and immune cell precursors (top half of figure) and non-immune cells (bottom half of figure). Circles represent negative, log-transformed, unadjusted P values. Statistical analyses were conducted using one-sided t-tests. Purple circles represent cell types significantly enriched in PhenoAge-associated SNPs at a FDR of 0.05. Black circles represent non-significant associations. Black circles to the right of the dotted line represent nominally significant associations (i.e., cells with P value < 0.05 that fail to survive correction for multiple comparisons). Exact P values are contained in Supplementary Data 33.

CELLECT: CELL-type Expression-specific integration for Complex Traits; MAGMA: Multi-marker Analysis of GenoMic Annotation; SNP: single nucleotide polymorphism; FDR: false discovery rate.

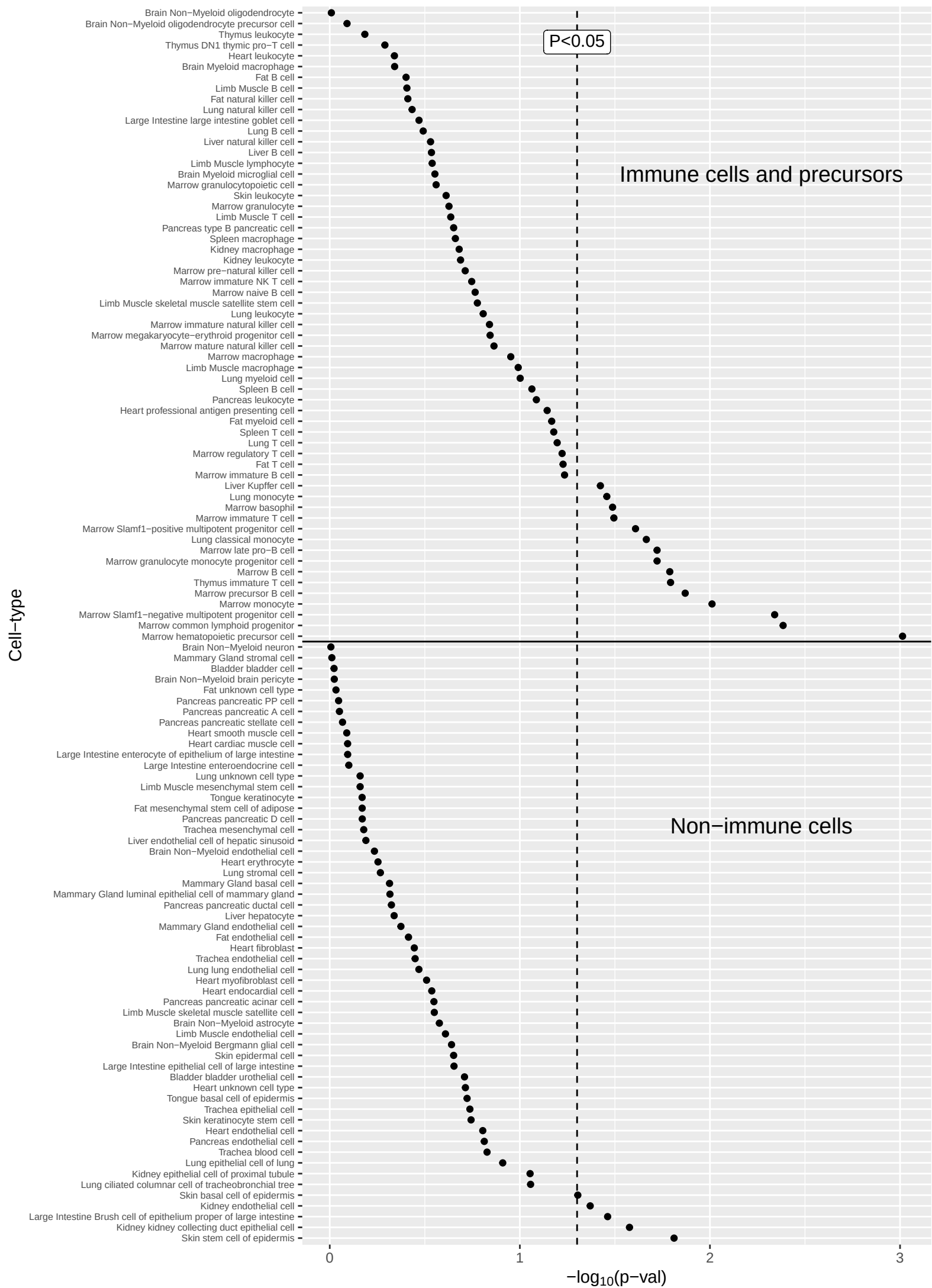


**Supplementary Fig. 5: CELLECT-MAGMA cellular associations with multivariate longevity
(immune cells and precursors, non-immune cells)**

Results from CELLECT-MAGMA cell-type enrichment analysis of multivariate longevity. Cell types are split into immune cells and immune cell precursors (top half of figure) and non-immune cells (bottom half of figure). Circles represent negative, log-transformed, unadjusted P values. Statistical analyses were conducted using one-sided t-tests. Black circles represent non-significant cell-trait associations at a FDR of 0.05. Black circles to the right of the dotted line represent nominally significant associations (i.e., cells with P value < 0.05 that fail to survive correction for multiple comparisons). Exact P values are contained in Supplementary Data 34.

CELLECT: CELL-type Expression-specific integration for Complex Traits; MAGMA: Multi-marker Analysis of GenoMic Annotation; SNP: single nucleotide polymorphism; FDR: false discovery rate.

No Cells Significant at False Discovery Rate of 0.05



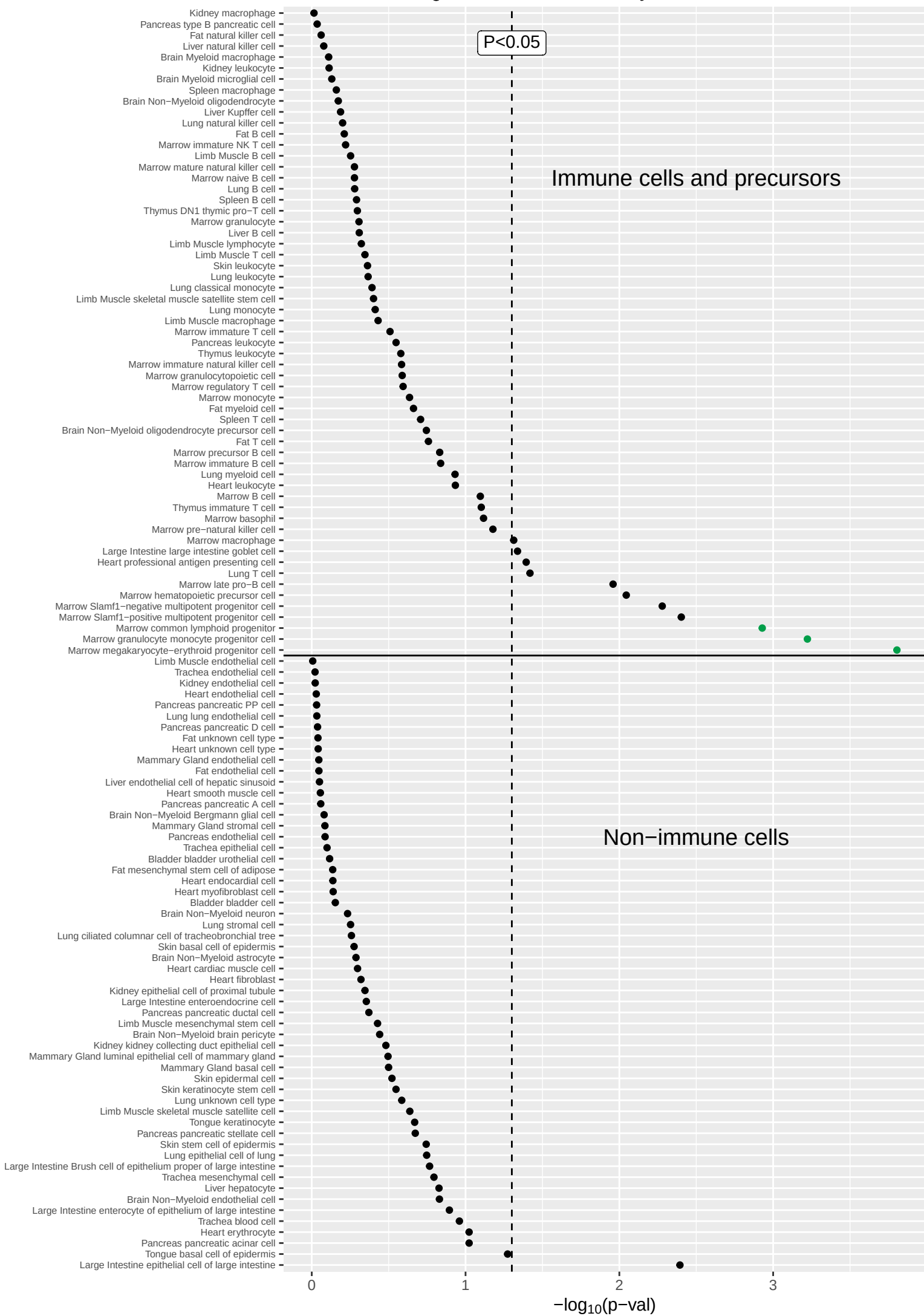
Supplementary Fig. 6: CELLECT-LDSC cellular associations with IEAA (immune cells and precursors, non-immune cells)

Results from CELLECT-LDSC cell-type enrichment analysis of multivariate longevity. Cell types are split into immune cells and immune cell precursors (top half of figure) and non-immune cells (bottom half of figure). Circles represent negative, log-transformed, unadjusted P values. Statistical analyses were conducted using one-sided t-tests. Black circles represent non-significant cell-trait associations at a FDR of 0.05. Black circles to the right of the dotted line represent nominally significant associations (i.e., cells with P value < 0.05 that fail to survive correction for multiple comparisons). Exact P values are contained in Supplementary Data 30.

CELLECT: CELL-type Expression-specific integration for Complex Traits; LDSC: LD Score Regression; IEAA: intrinsic epigenetic age acceleration; SNP: single nucleotide polymorphism; FDR: false discovery rate.

● = Significant at False Discovery Rate of 0.05

Cell-type

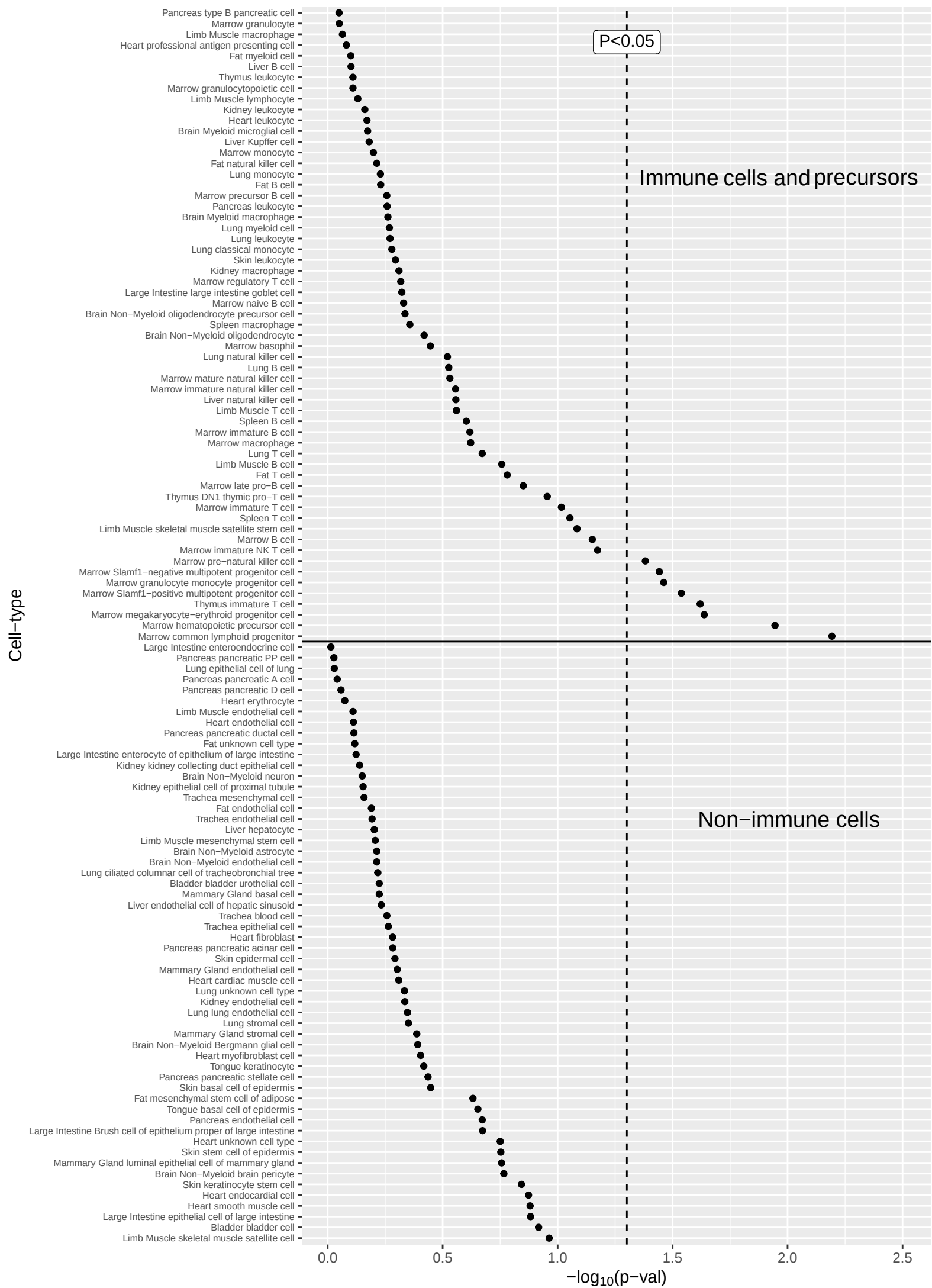


Supplementary Fig. 7: CELLECT-LDSC cellular associations with HannumAge (immune cells and precursors, non-immune cells)

Results from CELLECT-LDSC cell-type enrichment analysis of HannumAge. Cell types are split into immune cells and immune cell precursors (top half of figure) and non-immune cells (bottom half of figure). Circles represent negative, log-transformed, unadjusted P values. Statistical analyses were conducted using one-sided t-tests. Green circles represent cell types significantly enriched in HannumAge-associated SNPs at a FDR of 0.05. Black circles represent non-significant associations. Black circles to the right of the dotted line represent nominally significant associations (i.e., cells with P value < 0.05 that fail to survive correction for multiple comparisons). Exact P values are contained in Supplementary Data 31.

CELLECT: CELL-type Expression-specific integration for Complex Traits; LDSC: LD Score Regression; SNP: single nucleotide polymorphism; FDR: false discovery rate.

No Cells Significant at False Discovery Rate of 0.05

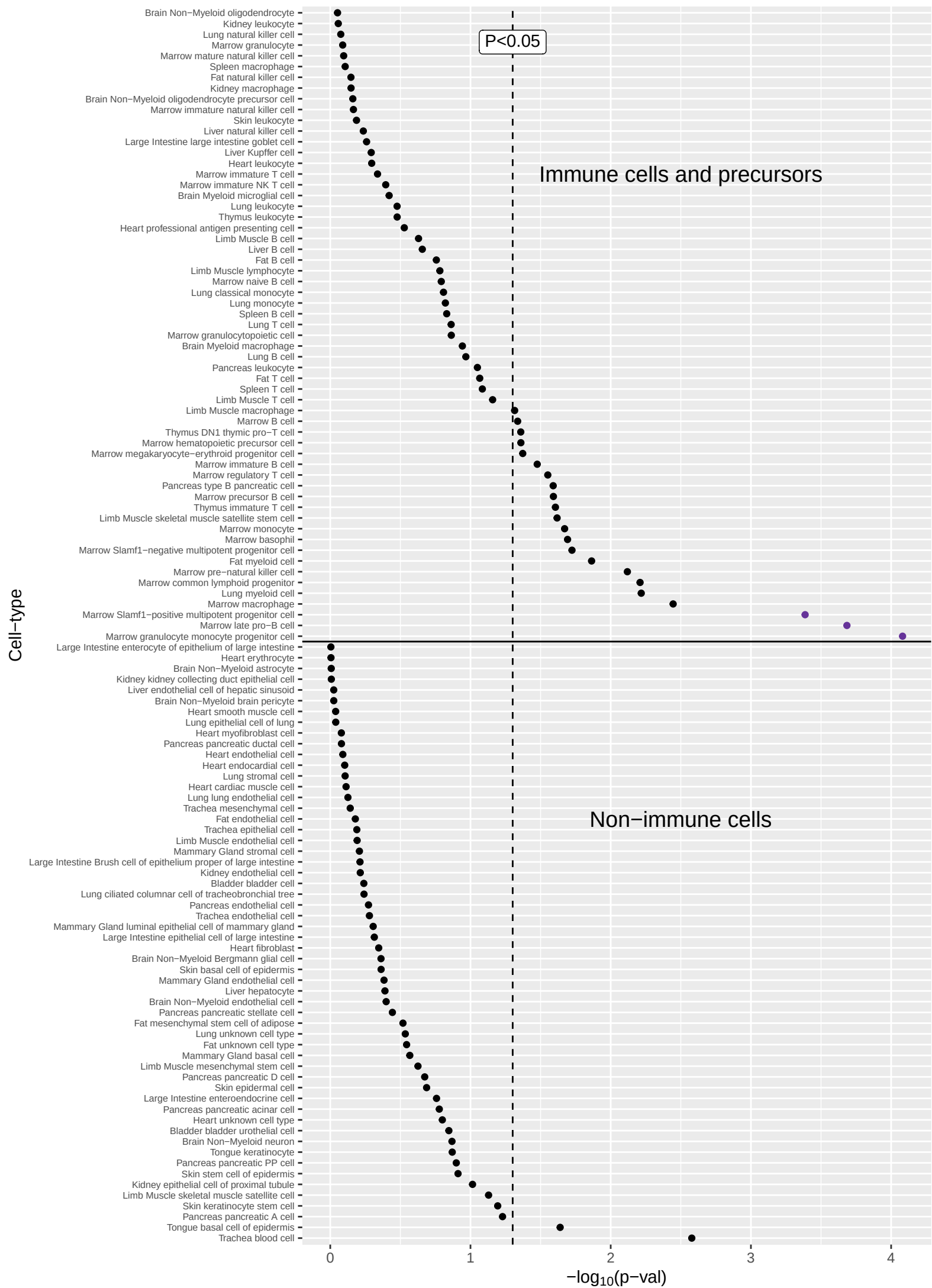


Supplementary Fig. 8: CELLECT-LDSC cellular associations with GrimAge (immune cells and precursors, non-immune cells)

Results from CELLECT-LDSC cell-type enrichment analysis of GrimAge. Cell types are split into immune cells and immune cell precursors (top half of figure) and non-immune cells (bottom half of figure). Circles represent negative, log-transformed, unadjusted P values. Statistical analyses were conducted using one-sided t-tests. Black circles represent non-significant cell-trait associations at a FDR of 0.05. Black circles to the right of the dotted line represent nominally significant associations (i.e., cells with P value < 0.05 that fail to survive correction for multiple comparisons). Exact P values are contained in Supplementary Data 32.

CELLECT: CELL-type Expression-specific integration for Complex Traits; LDSC: LD Score Regression; SNP: single nucleotide polymorphism; FDR: false discovery rate.

● = Significant at False Discovery Rate of 0.05

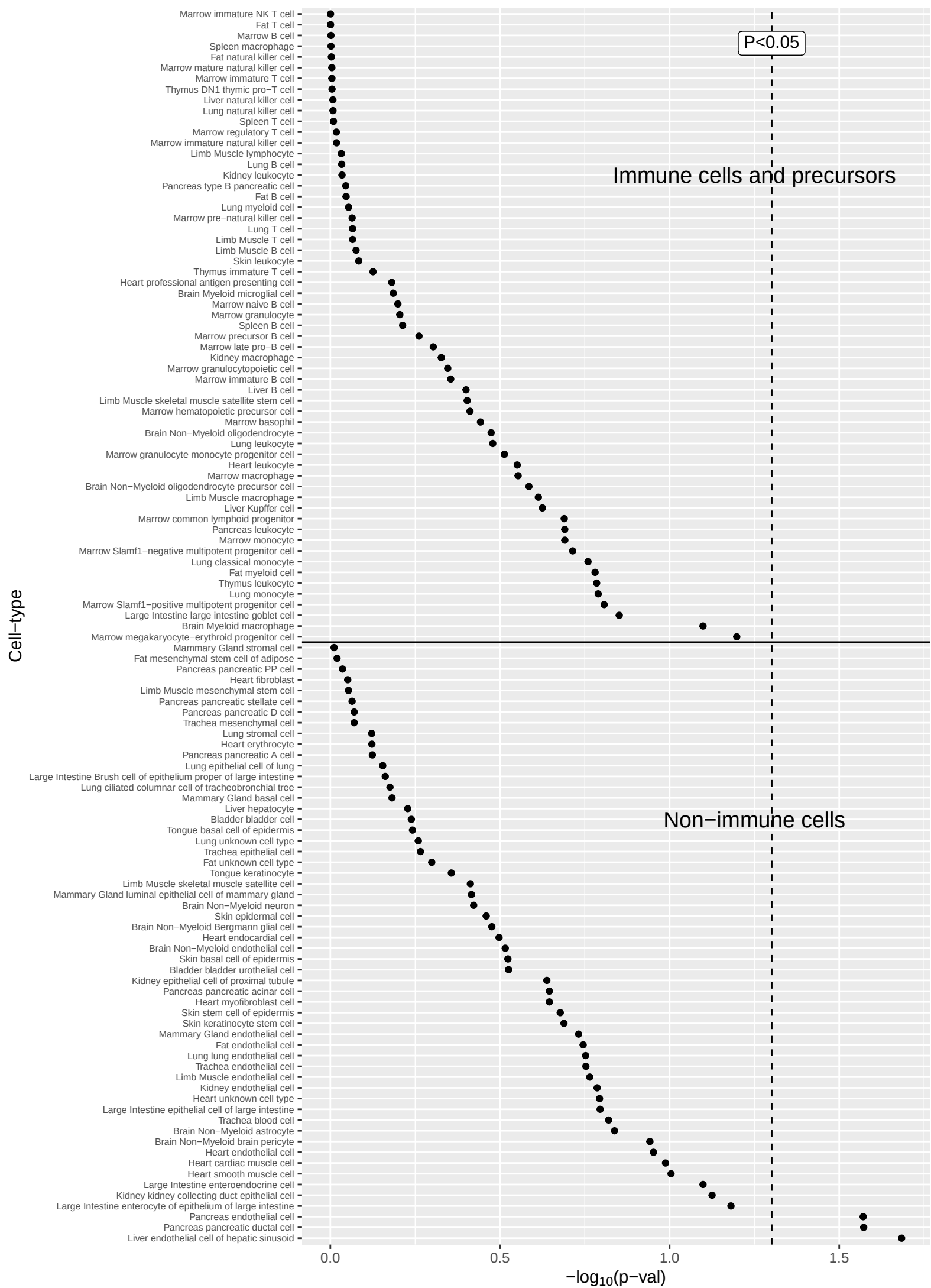


Supplementary Fig. 9: CELLECT-LDSC cellular associations with PhenoAge (immune cells and precursors, non-immune cells)

Results from CELLECT-LDSC cell-type enrichment analysis of PhenoAge. Cell types are split into immune cells and immune cell precursors (top half of figure) and non-immune cells (bottom half of figure). Circles represent negative, log-transformed, unadjusted P values. Statistical analyses were conducted using one-sided t-tests. Purple circles represent cell types significantly enriched in PhenoAge-associated SNPs at a FDR of 0.05. Black circles represent non-significant associations. Black circles to the right of the dotted line represent nominally significant associations (i.e., cells with P value < 0.05 that fail to survive correction for multiple comparisons). Exact P values are contained in Supplementary Data 33.

CELLECT: CELL-type Expression-specific integration for Complex Traits; LDSC: LD Score Regression; SNP: single nucleotide polymorphism; FDR: false discovery rate.

No Cells Significant at False Discovery Rate of 0.05



Supplementary Fig. 10: CELLECT-LDSC cellular associations with multivariate longevity (immune cells and precursors, non-immune cells)

Results from CELLECT-LDSC cell-type enrichment analysis of multivariate longevity. Cell types are split into immune cells and immune cell precursors (top half of figure) and non-immune cells (bottom half of figure). Circles represent negative, log-transformed, unadjusted P values. Statistical analyses were conducted using one-sided t-tests. Black circles represent non-significant cell-trait associations at a FDR of 0.05. Black circles to the right of the dotted line represent nominally significant associations (i.e., cells with P value < 0.05 that fail to survive correction for multiple comparisons). Exact P values are contained in Supplementary Data 34.

CELLECT: CELL-type Expression-specific integration for Complex Traits; LDSC: LD Score Regression; SNP: single nucleotide polymorphism; FDR: false discovery rate.