

Single-cell references can reverse, inflate and blur the associations inferred from bulk deconvolution

Yunhao Jiang

biojiang777@gmail.com

University of Edinburgh <https://orcid.org/0009-0004-5954-7315>

Research Article

Keywords: deconvolution, single-cell reference, cell-type proportions, association analysis, donor sampling, MuSiC

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11472895/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Bulk deconvolution yields cell-type proportions that are tested against age, disease and treatment as if the single-cell reference were exact. The reference can decide the answer. In 981 blood donors with labelled truth, the CD8 T-cell age association reversed sign in all 80 six-type references, because CD8 cells shift towards effector-memory states with age and a reference that averages over states cannot follow them. In 48 islet donors, a non-significant alpha-cell trend with HbA1c became significant in 54–100% of references, depending on their design, because the reference error rose with HbA1c. Both effects recurred with three further estimator classes (unweighted NNLS, the CIBERSORT v-SVR procedure and the InstaPrism algorithm), so they belong to the reference, not to one fitter; the regression estimators inflated the islet error and the Bayesian one attenuated it. Accuracy does not flag the problem: 15-state references with 30 donors were more accurate than those with 100, yet only the latter gave weighted MuSiC the right sign. The systematic part of the error is removed only by a reference that resolves the shifting states; the random part, which rivals the bulk-side standard error in small references, can be measured by a donor bootstrap and folded into the confidence interval. A random-effects model of donor sampling covers the removable part: most of 226 public references hold cell types whose median donor contributes fewer than 10 cells, below the 9–17 cells at which a donor profile becomes mostly noise, and a precision weight derived from the model removes about half of the error such donors add to MuSiC.

KEY POINTS

- The single-cell reference can decide the conclusion: in 981 blood donors with labelled truth, the CD8 T-cell age association reversed sign in all 80 six-type references, and in 48 islet donors a borderline alpha-cell association became significant in 54–100% of references.
- The error is reference-borne: the reversal and the CD4 inflation recurred with four estimator classes. Its direction is set by the estimator: three regression classes inflated the islet alpha-cell slope (1.1–2.1-fold), the InstaPrism algorithm attenuated it (0.59-fold).
- Accuracy does not flag it: state-resolved references with 30 donors had lower error than those with 100 (2.91 vs 3.26 points), yet under weighted MuSiC only the latter recovered the CD8 sign.
- The systematic part is removed only by a reference that resolves the shifting states; the random part is measured by a donor bootstrap within one reference and folded into the confidence interval.
- Donor sampling is the removable part: in 60% of 226 public datasets a cell type's median donor contributes fewer than 10 cells, below $n^* = 1/\rho$ (9–17 cells in blood), which is estimable from the reference; a precision weight removes about half of MuSiC's allocation penalty, and many donors with few cells beat few donors with many.

INTRODUCTION

Reference-based deconvolution estimates cell-type proportions in bulk RNA-seq samples from single-cell profiles (1–3), and the estimates are then analysed as phenotypes, as when beta-cell proportions were related to HbA1c (1). The reference enters this chain once, when many cells and donors are collapsed into one profile per cell type, and it is then treated as exact. Benchmarks have varied which reference is used, its size, its match to the mixture, its platform and the states it resolves (5,6,11–13,30–33), and have scored the result by accuracy. None has asked whether the reference changes the conclusion that is finally drawn, and none has varied how a reference's cells are spread over its donors, although public references are often thin at that level.

Two kinds of reference error can reach a conclusion. A systematic error that moves with the covariate under study shifts the slope of an association; a random error from the particular donors sampled widens its true uncertainty beyond the confidence interval computed from the bulk samples. We previously found the second kind in a controlled setting (14): with donors and total cells fixed, a 10:1:1 allocation of cells among three donors raised MuSiC's error by 0.62 points at 60 cells per type but not at 300, and only when minor donors contributed ten or fewer cells.

Here we ask four questions. Can the reference reverse or create an association, and by which route? Is that error a property of the reference or of the fitter, which collapses the reference by equal-donor averaging (MuSiC (1)) or by pooling cells (CIBERSORT-type signatures (2,4)) and then fits by regression or by a Bayesian posterior (7–10)? Which part of the error can be removed and which can be measured? And how should the donors and cells of a reference be chosen (15)? We answer them in two cohorts in which labelled single-cell data give ground truth for proportions and their associations, 981 PBMC donors (OneK1K (17)) and 48 islet donors with HbA1c (18), with four estimator classes, a census of 226 public references and a random-effects model of donor sampling whose estimator is inserted into unchanged MuSiC 1.0.0 and tested on the frozen grids of our previous study (14). The main findings are:

- **The reference can reverse a conclusion.** The CD8 T-cell age association reversed in every six-type reference because CD8 cells shift towards effector-memory states with age; in islets an alpha-cell error that rose with HbA1c made a borderline association significant in a design-dependent share of references. Missing cell types reverse signs by another route (16,34); these errors arise in complete references.
- **The reference, not the fitter.** The reversal and the CD4 inflation recurred in four estimator classes. The islet error changed direction with the class, so the same reference gives a significant association under one estimator and none under another.
- **Accuracy does not flag it.** State-resolved references with 30 donors were more accurate than those with 100, yet only the latter gave weighted MuSiC the right CD8 sign; pooled-cell estimators recovered it with 30.
- **Two parts, two remedies.** The systematic part is removed only by a reference that resolves the shifting states. The random part, which rivalled the bulk-side standard error in small references, is reproduced by a donor bootstrap within one reference and can be folded into the interval.

- **Donor sampling is the removable part.** Equal-donor and pooled-cell averaging are the two limits of one precision-weighted estimator, indexed by the ratio ρ of between- to within-donor variance; below $n^* = 1/\rho$, 9–17 cells in PBMC, a donor profile is mostly noise, and most public references contain such donors. The precision weight removes about half of MuSiC's allocation penalty (45% observed, 48% predicted), and at a fixed budget many donors with few cells beat few donors with many for three of the four classes.

MATERIALS AND METHODS

Full procedures, settings and fit counts are given in Supplementary Methods; this section states what was done.

Fitting and error. All fits used `music.iter` from MuSiC 1.0.0 with the settings of `music_prop`; both outputs were recorded, the variance-weighted proportions ("weighted MuSiC") and the non-negative least-squares proportions ("NNLS"). Only the mean-profile step was changed. Every comparison with the previous study (14) was gated by a control in which the equal-donor basis reproduced the stored predictions (maximum difference 5.6×10^{-16}). MAE is the mean absolute error over cell-type proportions in percentage points; the allocation penalty is $I(B) = \text{MAE}(10:1:1) - \text{MAE}(\text{balanced})$ at B cells per type (19).

OneK1K. 981 donors with Azimuth level-2 labels (24) mapped to six types (14); whole sequencing pools assigned to 597 target and 384 reference donors; target pseudobulks summed raw counts, truth was the labelled proportion. Fixed-budget references used $D \in \{3, 5, 10, 15, 30, 50, 100\}$ donors with 300/D cells per type, all-cells references $D \in \{10, 30, 100\}$ with every cell up to 300 per donor and type, eight donor sets per design (80 references); state-resolved references used the 15 level-2 states as reference types with $D \in \{30, 100\}$ (16 references), summed back to six types.

Islets. Targets were the 48 donors of (18) (eight types; HbA1c, age and sex). External references came from the 11 control donors of the HPAP dataset of Fasolino et al. (25–28) in fixed-budget, 10:1:1 and all-cells designs; within-cohort references were cross-fitted between two disease-stratified folds (142 references; 117 reference sets).

Associations. OneK1K proportions were regressed on age, sex and sequencing pool; islet proportions on HbA1c, age and sex; $|t| > 1.96$ (OneK1K) or 2.015 (islets) was called significant. Reference-side variability is the SD of a slope across the donor sets of a design divided by the mean bulk-side SE (OneK1K) or by the SE of the labelled slope (islets). The scaling κ of deconvolved on labelled proportion separates rescaling from covariate-dependent error (Proposition 4). The GSE50244 islet bulk cohort (29) was fitted with the 92 external references.

Other estimators. Two pooled-cell estimators were fitted to the same targets with the same cell draws as every reference: the CIBERSORT v-SVR procedure as implemented in EpiDISH 2.26.0 (CBS mode; linear v-SVR with $v \in \{0.25, 0.5, 0.75\}$, lowest reconstruction RMSE kept) on a per-reference pooled-cell

signature of the top-50 markers per type, with a Python replica of the recipe of (14) run alongside (Supplementary Table S23), and the InstaPrism algorithm (initial module of InstaPrism 0.1.6, a fixed-point implementation of the BayesPrism posterior (9), transcribed into Python from the released source with its defaults and no reference update). The replica and the InstaPrism algorithm were also run with equal-donor averaging of the same cells. Four classes are compared: MuSiC weighted (MW), MuSiC NNLS (MN), SVR signature (SV) and the InstaPrism algorithm (IP); the reporting rules were fixed before the fits (Supplementary Tables S18–S22).

Donor bootstrap. For three OneK1K designs and each of their eight donor sets, 20 replicates drew donors with replacement from the set; the bootstrap SD of the age slope was compared with the SD across the eight sets, and the reference-resampling interval is the point estimate $\pm 1.96 \times \sqrt{(\text{SE}^2 + \text{SD}_{\text{boot}}^2)}$ (300,888 fits).

Model and estimator. For gene g , donor d and cell c of one type, $x_{\text{gdc}} = \mu_g + b_{\text{gd}} + e_{\text{gdc}}$ with $b_{\text{gd}} \sim (0, \tau^2_g)$ and $e_{\text{gdc}} \sim (0, \sigma^2_g)$. The donor profile, MuSiC's library-size-weighted mean over the donor's cells, has variance $\tau^2_g + \sigma^2_g/n_d$, with n_d the Kish effective cell number. Proposition 1: among unbiased combinations of donor profiles the variance is minimised by weights $w_d \propto 1/(\tau^2 + \sigma^2/n_d)$; $\tau^2 \rightarrow \infty$ gives MuSiC's equal weights and $\tau^2 \rightarrow 0$ the cell-proportional weights of pooled signatures. With $\rho = \tau^2/\sigma^2$ and $n^* = 1/\rho$, a donor's weight relative to a fully sampled donor is $n_d/(n_d + n^*)$. Proposition 2, that correcting Σ for its sampling component improves MuSiC's gene weights, was tested and falsified. Proposition 3: for N cells split equally among D donors, the variance of the equal-donor mean, $\tau^2/D + \sigma^2/N$, decreases in D , whereas the expectation of Σ , $\tau^2 + \sigma^2 D/N$, increases. Proposition 4 gives the three ways reference error reaches an association slope (Supplementary Note 1). ρ_k was estimated per cell type as $\Sigma_g \hat{\tau}^2_g / \Sigma_g \hat{\sigma}^2_g$ over genes expressed in the bulk, and weights $w_{\text{dk}} \propto 1/(\rho_k + 1/n_{\text{dk}})$ were normalised over the donors with cells of type k .

Frozen grids. The frozen inputs of the previous study (14,20,21): three-donor PBMC references at 60 and 300 cells per type, balanced or 10:1:1 (full grid, 322,560 fits; diagnostic subset of 336 references), a single-type imbalance design and a four-donor pancreas design.

Census. The 226 human, primary, 10x CELLxGENE Discover datasets with at least three donors (22), tabulated by donor and author-labelled type (median cells per donor, fraction of donors below 10 cells, Kish effective donor number (23); Supplementary Table S1).

Software. MuSiC 1.0.0 in R 4.5.2; Python 3.11 with numpy, pandas, scipy, statsmodels, scikit-learn, h5py and matplotlib. All data are public and de-identified.

RESULTS

Most public references are thin at the donor level

Of 226 CELLxGENE datasets from all tissues with at least three donors (median 15.5 donors, range 3–981), 60% contained a cell type whose median donor contributed fewer than 10 cells and 76% one below 25. Across 3,615 dataset–type pairs, 24% had a median donor below 10 cells (Figure 1A); the Kish effective donor number was 55% of the nominal number for the median type of the median dataset and 28% for its sparsest type (Figure 1B). Sparse pairs occurred at every cohort size and in every tissue group: in 75% of the 60 blood and immune datasets and 54% of the 166 others (Figure 1C, D; Supplementary Table S1). The regime of our previous study (14), a few donors with 5–10 cells of a type, is the common case, and dataset totals do not show it.

Reference error that moves with the covariate reverses the CD8 T-cell age association

We regressed deconvolved proportions on age and sex with sequencing pool as a fixed effect and compared the slopes with those of the labelled proportions of the same 597 donors (Table 1). Every sex contrast kept its sign in all 80 references. For age, five types kept their sign, but CD8 T cells reversed in every reference: the labelled slope was -0.072 points per year ($t = -2.6$) and the deconvolved slopes $+0.04$ to $+0.24$ (96% with $|t| > 1.96$; Figure 2A). The CD4 T-cell slope was inflated 5- to 8-fold (median 6.3), so a non-significant trend ($t = -1.6$) became significant in every reference. Neither donor number, quota scheme nor basis changed these outcomes (Supplementary Table S9).

The reversal follows a state shift that a donor-averaged profile cannot track. The deconvolved CD8 proportion correlated with the labelled one ($r = 0.82$), but its bias grew by 0.24 points per year of age, three times the magnitude of the true slope. In the labelled data, the effector-memory share of CD8 T cells rose by 0.72 points per year ($t = 15$) and the naive share fell by 0.70 ($t = -16$); the CD8 bias correlated 0.71 with a donor's effector-memory share, and adjusting for that share reduced its age dependence from 0.239 to 0.064 points per year (Figure 2B, C; Supplementary Table S10).

Table 1. Age and sex associations in OneK1K: labelled truth versus deconvolved proportions (weighted MuSiC, equal-donor basis, 80 six-type references).

Cell type	True age slope, points/year [t]	True sex contrast F – M, points [t]	Median ratio deconvolved/true age slope	References with wrong age sign	References with wrong sex sign
B cells	–0.005 [–0.25]	+1.67 [2.58]	3.5 (true $\approx$ 0)	0/80	0/80
CD14+ monocytes	+0.023 [2.81]	–0.60 [–2.46]	1.25	0/80	0/80
CD4 T cells	–0.051 [–1.57]	+3.98 [4.11]	6.3	0/80	0/80
CD8 T cells	–0.072 [–2.62]	–2.12 [–2.57]	–2.1	80/80	0/80
FCGR3A+ monocytes	+0.014 [3.86]	–0.34 [–3.14]	1.18	0/80	0/80
NK cells	+0.091 [4.47]	–2.59 [–4.22]	1.6	0/80	0/80

State-resolved references (15 level-2 states, 100 donors, eight donor sets): CD8 T –0.037 (wrong sign 0/8, significant 0/8); CD4 T –0.057 (ratio 1.13, significant 0/8); B –0.015; NK +0.070 (ratio 0.77).

Covariate-dependent error makes a borderline islet association significant in a design-dependent share of references

In the 48 islet donors (17 normoglycaemic, 14 prediabetic, 17 with type 2 diabetes; HbA1c 4.6–13.1%), the labelled beta-cell share fell by 3.90 points per HbA1c unit ($t = -2.91$, adjusted for age and sex), the decline MuSiC's authors reported from deconvolved islet bulk (1); the alpha-cell share rose non-significantly (+2.03, $t = 1.86$). Across 117 reference sets (92 external, 25 cross-fitted), the beta-cell decline kept its sign and remained significant in all 117 with weighted MuSiC and in 114 with NNLS (Table 2; Figure 3A); its slope was inflated (medians by design –4.5 to –5.7). The alpha-cell slope was inflated 1.2- to 1.7-fold (medians by design), which made the non-significant trend significant in every within-cohort and all-cells reference, in 83–88% of three-donor external references with 300 cells per type, and in 62% (balanced) and 54% (10:1:1) of those with 60 cells (Figure 3B; Supplementary Table S13). Whether an alpha-cell association is reported depends on how the reference is built.

The inflation was not a scaling of the proportions. Regressing deconvolved on labelled proportion gave a median κ of 1.09 for beta and 1.02 for alpha cells, but the slope ratios were 1.33 and 1.43, exceeding κ in 116 and 115 of 117 sets. The excess came from the residual after scaling, which moved with HbA1c (–0.82 and +0.96 points per unit) and contributed 0.21 (beta) and 0.47 (alpha) to the ratio, against 0.09

and 0.02 from $\kappa - 1$ (Figure 3C, D; Supplementary Table S14). As for CD8 T cells, the error tracked the covariate under study.

Real bulk data could not extend this test. In the islet bulk cohort in which the beta-cell decline was reported (GSE50244; 89 samples, 77 with HbA1c) (1,29), every external HPAP reference assigned beta cells a mean share of 0.26% or less, so no beta-cell slope was significant and alpha cells absorbed the endocrine signal (median share 43%). INS accounted for 37% of beta-cell UMIs in the 3' reference but 0.067% of reads in the deposited bulk counts, and three other beta-cell genes (G6PC2, PCSK1, SLC30A8) made up an 8- to 10-fold smaller share of reference than of bulk reads (Figure 3E, F; Supplementary Table S15). Real bulk therefore tested how well reference and bulk profiles matched, not donor design.

Table 2. Beta- and alpha-cell associations with HbA1c in 48 islet donors (weighted MuSiC, equal-donor basis; selected designs, all designs in Supplementary Table S13). Labelled slopes: beta -3.90 ($t = -2.91$), alpha $+2.03$ ($t = 1.86$) points per HbA1c unit.

Reference design	Reference sets	Beta: median slope	Beta: same sign / significant	Alpha: median slope	Alpha: significant
Within cohort, 3 donors $\times$ 100 cells	8	-5.48	8/8 / 8/8	$+2.78$	8/8
Within cohort, 12 donors $\times$ 25 cells	8	-5.49	8/8 / 8/8	$+2.69$	8/8
External, 3 donors, 60 cells per type, balanced	8	-4.52	8/8 / 8/8	$+2.63$	5/8
External, 3 donors, 60 cells per type, 10:1:1	24	-4.65	24/24 / 24/24	$+2.54$	13/24
External, 3 donors, 300 cells per type, balanced	8	-5.23	8/8 / 8/8	$+3.39$	7/8
External, 3 donors, 300 cells per type, 10:1:1	24	-5.00	24/24 / 24/24	$+3.25$	20/24
External, 6 donors, 300 cells per type, balanced	8	-4.96	8/8 / 8/8	$+3.29$	7/8
External, 6 donors, all cells (≤ 300)	8	-5.37	8/8 / 8/8	$+3.25$	8/8
External, 11 or 19 donors, all cells	2	-5.25	2/2 / 2/2	$+3.3$	2/2

The reversal and the inflation are reference-borne; the direction of the islet error depends on the estimator class

To ask whether these errors belong to the reference or to MuSiC, every reference was refitted with two pooled-cell estimators, the CIBERSORT v-SVR procedure as implemented in EpiDISH (SVR) and the InstaPrism algorithm, alongside MuSiC's weighted and NNLS outputs (Figure 4; Supplementary Tables S18–S22).

The CD8 T-cell age slope was positive in 80 of 80 six-type references for MuSiC weighted, MuSiC NNLS and SVR and in 78 of 80 for the InstaPrism algorithm (median slopes +0.154, +0.236, +0.128 and +0.055 points per year against -0.072), and significant in 77, 76, 78 and 64 (Figure 4A). The CD4 slope was inflated 6.3-, 7.3-, 6.1- and 4.4-fold and significant in all 80 references for every class (Figure 4B).

In the islets, the beta-cell decline kept its sign and significance in 108–117 of 117 sets for all classes, but the alpha-cell association separated them: the three regression-type estimators inflated the slope (1.11-, 1.43- and 2.06-fold for NNLS, weighted MuSiC and SVR) and made it significant in 46, 95 and 117 of 117 sets, whereas the InstaPrism algorithm attenuated it (0.59-fold; significant in 6 of 117) (Figure 4D). The same reference yields a significant alpha-cell association under one estimator and none under another: the error is reference-borne, its direction is set by the estimator.

The systematic part is removed only by a reference that resolves the shifting states, and the donors it needs depend on the fitter

A reference that resolves the states removes the artefact. With the 15 Azimuth level-2 states as reference types, summed back to six, error fell from 4.4–4.5 to 2.9–3.3 points (Supplementary Table S11). With 100 donors the CD8 slope was -0.037 points per year (sign correct in 8 of 8) and the CD4 slope -0.057 (true -0.051; not significant in 8 of 8, as in the labelled data): reversal and inflation disappeared, although the CD8 slope was attenuated rather than recovered (Figure 2D). With 30 donors the CD8 slope was ≈ 0 (sign correct in 4 of 8) and the CD4 slope remained inflated 1.8-fold. The CD8 slope of a state-resolved reference is the sum of its naive, TCM and TEM slopes: -0.153, +0.005 and +0.150 points per year with 30 donors, cancelling, against -0.185, +0.010 and +0.139 with 100 (labelled -0.176, -0.011 and +0.115), each with a 1.5- to 2.8-fold smaller SD across donor sets (Figure 2E; Supplementary Table S12). In all three states the median donor lies above n^* (Figure 2F), so the gain came from the donor term τ^2/D , not from cell counts. Only CD4 CTL and NK CD56-bright had a median donor below n^* ; both were over-estimated, as were the rare states above n^* (CD4 TEM and Treg; Supplementary Table S12). State-resolved references were more accurate with 30 than with 100 donors (2.91 vs 3.26 points): the more accurate reference gave the wrong conclusion.

The other estimator classes needed fewer donors: with the same 15-state references, the two pooled-cell estimators recovered the sign of the CD8 slope in 8 of 8 references at 30 donors (significant in 6 and 8), whereas MuSiC weighted recovered it only at 100 donors and never its significance (Figure 4C). The state shift is a property of six-type references; how many donors a state-resolved reference needs is a property of the fitter.

The random part rivals the bulk-side standard error in small references and can be measured by a donor bootstrap

Beyond the systematic error, the spread of a slope across random donor sets is uncertainty that a confidence interval computed from the bulk samples ignores. In OneK1K it exceeded the bulk-side standard error for CD8 T cells (1.9-fold) and NK cells (1.3-fold) with three reference donors, was 0.5–0.8 times it with 15–30 donors, the size of a typical public reference, and fell to 0.2 or less with 100 donors contributing all their cells (Figure 5A; Supplementary Table S16). In the islets it was 0.15–0.63 times the standard error of the labelled slope (Figure 5B). The spread was not a MuSiC artefact: at three donors it exceeded the bulk-side standard error for MuSiC NNLS (2.8-fold) and the InstaPrism algorithm (2.1) as well; SVR was less sensitive to the donor draw (0.9) while carrying the largest systematic bias (Figure 4G).

This spread can be estimated from a single reference. Resampling donors with replacement within each reference of three designs gave a bootstrap SD of the age slope that was 0.77–1.49 times the SD across the eight independent donor sets (median 1.07; Figure 5C; Supplementary Table S17). The reference-resampling interval, $\pm 1.96 \times \sqrt{(\text{SE}^2 + \text{SD}_{\text{boot}}^2)}$, widened the bulk-side interval by a median factor of 1.23 for CD8 T cells with 15 donors (Figure 5D) and covered the point estimates of the other seven donor sets in 99.9% of pairs (bulk-side interval alone 99.5%). It did not cover the labelled CD4 and CD8 T slopes in any of the 24 references of either type, because that error is systematic (Figure 5E).

Donor-sampling noise is the removable part: one estimator with two limits, an estimable n^* , a weight for MuSiC and a design rule

The random part has a second source besides which donors were drawn: how many cells each donor contributed. Under a random-effects model of reference sampling, the minimum-variance combination of donor profiles weights donor d by $1/(\tau^2 + \sigma^2/n_d)$ (Materials and Methods). With $\rho = \tau^2/\sigma^2$, a donor with n cells carries weight $n/(n + n^*)$ relative to a fully sampled one, where $n^* = 1/\rho$ is the count at which cell-sampling variance equals donor variance (Figure 6A). Equal donor weights (MuSiC) and cell-proportional weights (pooled signatures) are the limits $\tau^2 \rightarrow \infty$ and $\tau^2 \rightarrow 0$ (Figure 6B): the first gives a

profile from a few cells the same influence as one from many, the second concentrates on the dominant donor and keeps its donor effect however many cells are added. p can be estimated from the reference (Figure 6C): in 336 three-donor PBMC references (14), median p_k ranged from 0.058 to 0.117 by type, so n^* was 9–17 cells, matching the ten-cell threshold observed previously; OneK1K references gave similar medians (0.06–0.13), most islet types much higher p (Supplementary Table S2).

Inserted into unchanged MuSiC and refitted on the frozen grids of our previous study (14), the precision-weighted basis lowered the 10:1:1 penalty of weighted MuSiC at 60 cells per type from +0.591 to +0.322 points, a 45% reduction, improved 12 of 14 held-out donors, left the penalty at 300 cells unchanged and changed balanced references by at most 0.003 points (Table 3; Figure 7A). The model predicted a 48% reduction (41–58% by type), and p_k estimated in OneK1K, a different cohort, predicted 43–46% (Figure 6D; Supplementary Table S2). With imbalance confined to one type, 69–71% of the penalty was removed at minor-donor quotas of 2, 5 and 10 cells (Figure 7B; Supplementary Table S3); gene-level weights gave the same result as donor-level weights (Supplementary Table S4).

The weight helps only MuSiC's weighted output. The NNLS penalty rose from +0.061 to +0.239 points (Table 3), and a scan of a common p shows why (Figure 7C): cell-sampling noise is independent across genes and averages out in a least-squares fit over thousands of genes, whereas donor effects shift many genes together, so plain regression is dominated by the donor term, for which equal weights are near optimal, while MuSiC's iterative gene weighting lets gene-wise noise back in. Correcting Σ for its sampling component (Proposition 2) reduced the penalty only from +0.752 to +0.635 points and raised the error of balanced references by 0.12–0.23 points (Figure 7D–F). Reweighting also failed when every type had two-cell donors (pancreas, +0.760 $\rightarrow$ +0.769; Supplementary Table S5) and when p was high: in islets, six of the eight types had p of 0.29–0.93, so n^* was 1–3 cells, below the five cells of each minor donor, and the three-donor 10:1:1 penalty (+0.50 points at 60 cells per type) stayed at +0.55 (Figure 7G; Supplementary Table S6). The residual penalty is information the unequal design did not collect.

Reweighting cannot recover cells that a design did not collect, so we asked how a fixed budget should be spread (15). In OneK1K, weighted-MuSiC error at 300 cells per type fell from 4.42 points with 3 donors of 100 cells to 3.77 with 15 donors of 20 and 3.41 with 100 donors of 3, for CD4 T, CD8 T and NK cells as well as for the sparse monocytes (Figure 8A, B; Supplementary Table S7); in the islets the error of cross-fitted references fell from 2.79 points with 3 donors to 2.27 with 23–24 (Figure 8C). With ten or more donors the equal-donor and precision-weighted bases differed by at most 0.03 points, although natural imbalance was large: with 100 donors contributing all their cells, the Kish effective donor number was 41 for CD14+ and 45 for FCGR3A+ monocytes against 100 for CD4 T cells (Figure 8D; Supplementary Table S8). Allocation matters for the small references the census shows to be common, not for cohort-scale atlases.

The design rule held for three of the four classes: SVR error also fell with donors (5.91 $\rightarrow$ 4.37 points), whereas the InstaPrism algorithm was flat (3.97 $\rightarrow$ 3.89), its error set by the state-shift bias at every D (Figure 4E). The weight, by contrast, is MuSiC's alone: replacing pooled-cell by equal-donor averaging of

the same cells changed the error of a Python replica of the SVR recipe by less than 0.05 points and improved the InstaPrism algorithm by 0.04–0.32 points, as much with equal quotas as with natural imbalance (Figure 4F). Donor reweighting cures MuSiC's allocation penalty; it does not touch the errors that changed conclusions.

Table 3. Allocation penalties on the full grid (12 replication draw blocks, 840 targets; percentage points).

Output	Basis	I(60)	I(300)	I(300) - I(60) [MCSE]	Held-out donors with I(300) < I(60)	Balanced MAE, 60 / 300 cells
Weighted MuSiC	Equal-donor	+0.591	-0.073	-0.665 [0.049]	13/14	5.396 / 5.329
Weighted MuSiC	Precision-weighted	+0.322	-0.063	-0.386 [0.049]	13/14	5.399 / 5.329
NNLS	Equal-donor	+0.061	-0.089	-0.149 [0.102]	10/14	8.632 / 8.723
NNLS	Precision-weighted	+0.239	-0.043	-0.282 [0.092]	14/14	8.639 / 8.726

DISCUSSION

A deconvolved proportion is a measurement made through a reference, and an association drawn from it inherits the reference's errors. Of the three ways reference error can reach a slope (Supplementary Note 1, Proposition 4), the benign ones, a constant offset and a scaling of the proportion, changed no conclusion here. The errors that did change conclusions moved with the covariate: the effector-memory shift of CD8 T cells with age reversed an association in every six-type reference, and the alpha-cell residual that rose with HbA1c made a borderline association significant. Missing cell types reverse signs by another route (16,34); these errors arise in complete references, and they are not visible in the fit.

They are reference-borne, not fitter-borne. The reversal and the CD4 inflation appeared with four estimator classes, two regression fitters, a marker-based SVR and a Bayesian fixed point, so no choice of fitter protects a six-type reference from a within-type state shift. The fitter does set the direction of the islet error: three regression estimators inflated the alpha-cell slope and the InstaPrism algorithm attenuated it, so the same reference makes an association significant under one estimator and not under another. Two consequences follow. An association drawn from deconvolved proportions should be shown under at least two estimator classes, and agreement between them should not be read as evidence that the reference is adequate. And accuracy is not that evidence either: the 15-state references with 30 donors had lower error than those with 100, yet only the latter gave weighted MuSiC the right sign.

The two components of reference error call for different remedies. The systematic component is removed only by a reference that resolves the states that move with the covariate, and the donors it needs depend on the fitter: weighted MuSiC recovered the CD8 sign only at 100 donors and never the significance, whereas the two pooled-cell estimators recovered both at 30. The random component, which bulk-side confidence intervals ignore and which rivalled the bulk-side standard error in small references, can be measured by a donor bootstrap and folded into the interval; methods that carry proportion uncertainty into downstream estimates (35) could take this component as input.

Donor sampling is the part of the random component that a reference's construction controls. Equal-donor and pooled-cell averaging are limits of one estimator, and the ratio ρ that places a reference between them can be estimated from the reference itself; the ten-cell threshold of the previous study (14) is not a constant but $n^* = 1/\rho$, 8–17 cells in two PBMC atlases and 1–3 cells for most islet types, and donors below it are present in most public references of every tissue. References could report n^* and the Kish effective donor number per cell type alongside their totals. The remedies here are specific to the averaging rule: precision weighting suits MuSiC's weighted output with small, unbalanced references and costs nothing when quotas are equal, whereas unweighted regression is dominated by the coherent donor component, for which equal weights are near optimal, and neither pooled-cell estimator gained from donor reweighting of the same cells. For a fixed budget, both cohorts favoured spreading cells over many donors, in line with power analyses of multi-sample single-cell designs (15), for every class but the InstaPrism algorithm, whose error was set by the state shift. The weights act within a dataset and can precede the dataset-level weighting of SCDC (12); Bisque addresses platform differences rather than donor sampling (8).

Both cohorts are pseudobulks that share platform and annotation with their references, which understates systematic bias, and the OneK1K truth inherits Azimuth label noise. In the one real bulk cohort we tried, mismatched reference and bulk profiles left beta cells unidentifiable before donor design could matter, so an independent flow-cytometry or sorted-population truth with a matched reference is the next test. Targets contained only the analysed types, so incomplete references were not examined. The model describes cell-sampling noise under exchangeable donor effects and correct labels; it did not hold when every type had two-cell donors, and it does not explain the over-estimation of rare states or the source of the islet residual. The SVR class was run with EpiDISH's implementation of the CIBERSORT procedure rather than CIBERSORTx itself (a Python replica of the recipe agreed with it in the sign of every CD8 and CD4 slope; Supplementary Table S23), and the InstaPrism algorithm was transcribed from its released source without the optional reference update and without a check against the compiled package, so the four classes bracket the field's estimators rather than enumerate its software.

An association drawn from deconvolved proportions should therefore carry a reference-resampling interval, come from a reference that resolves the states known to shift with the covariate, be shown under more than one estimator class, and be checked against labelled or sorted samples where these exist.

Declarations

DATA AVAILABILITY

OneK1K, the two islet datasets and the 226 census datasets (listed in Supplementary Table S1) are available from CELLxGENE Discover; HPAP data are subject to the HPAP data-use terms and are not redistributed. The GSE50244 islet bulk RNA-seq data are available from GEO, and the count object used here from the MuSiC website (<https://xuranw.github.io/MuSiC/data/>). The frozen inputs of the previous study (14) are on Zenodo (doi:10.5281/zenodo.23012998). Code and all derived result tables are on Zenodo (doi:10.5281/zenodo.23535596) and in the Supplementary Data.

SUPPLEMENTARY DATA

Supplementary Note 1 (model, propositions and proofs) and Supplementary Tables S1–S23, indexed in Supplementary Table S0, are available as supplementary files of this preprint, together with Supplementary Methods.

ACKNOWLEDGEMENTS

This manuscript used data acquired from the database (<https://hpap.pmacs.upenn.edu/>) of the Human Pancreas Analysis Program (HPAP; RRID:SCR_016202; PMID: 31127054; PMID: 36206763). HPAP is part of a Human Islet Research Network (RRID:SCR_014393) consortium (UC4-DK112217, U01-DK123594, UC4-DK112232, and U01-DK123716).

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study analysed only publicly available, de-identified count data and coded annotations from the OneK1K, islet and census single-cell datasets and the GSE50244 bulk RNA-seq dataset. The original studies obtained ethics approval and participant consent as described in their publications (17,18,25,29). No new participants were recruited and no new samples were collected. Based on advice from the author's academic supervisor, additional ethics approval was not required for this secondary analysis.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

None declared.

References

1. Wang X, Park J, Susztak K, Zhang NR, Li M (2019) Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* 10:380
2. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA (2015) Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 12:453–457
3. Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P, De Preter K (2020) Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun* 11:5650
4. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D et al (2019) Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 37:773–782
5. Jin H, Liu Z (2021) A benchmark for RNA-seq deconvolution analysis under dynamic testing environments. *Genome Biol* 22:102
6. Avila Cobos F, Panah MJN, Epps J, Long X, Man T-K, Chiu H-S, Chomsky E, Kiner E, Krueger MJ, di Bernardo D et al (2023) Effective methods for bulk RNA-seq deconvolution using scRNA-seq transcriptomes. *Genome Biol* 24:177
7. Erdmann-Pham DD, Fischer J, Hong J, Song YS (2021) Likelihood-based deconvolution of bulk gene expression data using single-cell references. *Genome Res* 31:1794–1806
8. Jew B, Alvarez M, Rahmani E, Miao Z, Ko A, Garske KM, Sul JH, Pietiläinen KH, Pajukanta P, Halperin E (2020) Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.*, 11, 1971
9. Chu T, Wang Z, Pe’er D, Danko CG (2022) Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat Cancer* 3:505–517
10. Tsoucas D, Dong R, Chen H, Zhu Q, Guo G, Yuan G-C (2019) Accurate estimation of cell-type composition from gene expression data. *Nat Commun* 10:2975
11. Vallania F, Tam A, Lofgren S, Schaffert S, Azad TD, Bongen E, Haynes W, Alsup M, Alonso M, Davis M et al (2018) Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases. *Nat Commun* 9:4735
12. Dong M, Thennavan A, Urrutia E, Li Y, Perou CM, Zou F, Jiang Y (2021) SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references. *Brief Bioinform* 22:416–427
13. Maden SK, Kwon SH, Huuki-Myers LA, Collado-Torres L, Hicks SC, Maynard KR (2023) Challenges and opportunities to computationally deconvolve heterogeneous tissue with varying cell sizes using single-cell RNA-sequencing datasets. *Genome Biol* 24:288
14. Jiang Y (2026) Effects of unequal donor allocation in single-cell references on MuSiC deconvolution depend on reference-cell budget: a PBMC simulation study. *Res Square*. 10.21203/rs.3.rs-11156252/v1

15. Schmid KT, Höllbacher B, Cruceanu C, Böttcher A, Lickert H, Binder EB, Theis FJ, Heinig M (2021) scPower accelerates and optimizes the design of multi-sample single cell transcriptomic studies. *Nat Commun* 12:6625
16. Jiang H, Gao F, Liu P, Wu Y, Jie Y, Li Y, Jiang Y (2026) Incomplete references leave bulk deconvolution targets non-identifiable, but identification has a computable precision price. *bioRxiv*. 10.64898/2026.09.10.750775
17. Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, Lu Q, Rowson A, Taylor TRP, Clarke L et al (2022) Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* 376:eabf3041
18. Bandesh K, Motakis E, Nargund S, Kursawe R, Selvam V, Ansarullah, Bhuiyan RM, Eryilmaz GN, Willett AM, White JK et al (2026) Deep single-cell decoding of human pancreatic islets reveals T2D β -cell gene expression defects. *EMBO J* 45:3978–4005
19. Morris TP, White IR, Crowther MJ (2019) Using simulation studies to evaluate statistical methods. *Stat Med* 38:2074–2102
20. Perez RK, Gordon MG, Subramaniam M, Kim MC, Hartoularos GC, Targ S, Sun Y, Ogorodnikov A, Bueno R, Lu A et al (2022) Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science*, 376, eabf1970
21. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, Ryu JH, Wagner BK, Shen-Orr SS, Klein AM et al (2016) A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* 3:346–360e4
22. CZI Cell Science Program, Abdulla S, Aevermann B, Assis P, Badajoz S, Bell SM, Bezzi E, Cakir B, Chaffer J, Chambers S et al (2025) CZ CELLxGENE Discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *Nucleic Acids Res* 53:D886–D900
23. Kish L (1965) *Survey Sampling*. Wiley, New York
24. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M et al (2021) Integrated analysis of multimodal single-cell data. *Cell* 184:3573–3587e29
25. Fasolino M, Schwartz GW, Patil AR, Mongia A, Golson ML, Wang YJ, Morgan A, Liu C, Schug J, Liu J et al (2022) Single-cell multi-omics analysis of human pancreatic islets reveals novel cellular states in type 1 diabetes. *Nat Metab* 4:284–299
26. Kaestner KH, Powers AC, Naji A, Consortium HPAP, Atkinson MA (2019) NIH Initiative to Improve Understanding of the Pancreas, Islet, and Autoimmunity in Type 1 Diabetes: The Human Pancreas Analysis Program (HPAP). *Diabetes* 68:1394–1402
27. Shapira SN, Naji A, Atkinson MA, Powers AC, Kaestner KH (2022) Understanding islet dysfunction in type 2 diabetes through multidimensional pancreatic phenotyping: The Human Pancreas Analysis Program. *Cell Metab* 34:1906–1913
28. Patil AR, Schug J, Naji A, Kaestner KH, Faryabi RB, Vahedi G (2023) Single-cell expression profiling of islets generated by the Human Pancreas Analysis Program. *Nat Metab* 5:713–715

29. Fadista J, Vikman P, Laakso EO, Mollet IG, Esguerra JL, Taneera J, Storm P, Osmark P, Ladenvall C, Prasad RB et al (2014) Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc. Natl. Acad. Sci. U.S.A.*, 111, 13924–13929
30. Xu X, Li R, Mo O, Liu K, Li J, Hao P (2024) Cell-type deconvolution for bulk RNA-seq data using single-cell reference: a comparative analysis and recommendation guideline. *Brief Bioinform* 26:bbaf031
31. Zhao Y, Vestal BE, Moore CM (2026) Evaluating reference-mixture matching in cell-type deconvolution with single-cell RNA-seq references. *Brief Bioinform* 27:bbag184
32. Hu M, Chikina M (2024) Heterogeneous pseudobulk simulation enables realistic benchmarking of cell-type deconvolution methods. *Genome Biol* 25:169
33. Singh A, Cakmak P, Lun JH, Macas J, Plate KH, Reiss Y, Schupp J, Imkeller K (2026) Benchmarking cell-type deconvolution in cross-platform transcriptomic data. *Genome Biol* 27:265
34. Ivich A, Davidson NR, Grieshober L, Li W, Hicks SC, Doherty JA, Greene CS (2025) Missing cell types in single-cell references impact deconvolution of bulk data but are detectable. *Genome Biol* 26:86
35. Cheng Y, Lin C, Li H, Xu K, Zhao H (2026) UBD: incorporating uncertainty in cell type proportion estimates from bulk samples to infer cell-type-specific profiles. *Brief Bioinform* 27:bbaf711

Figures

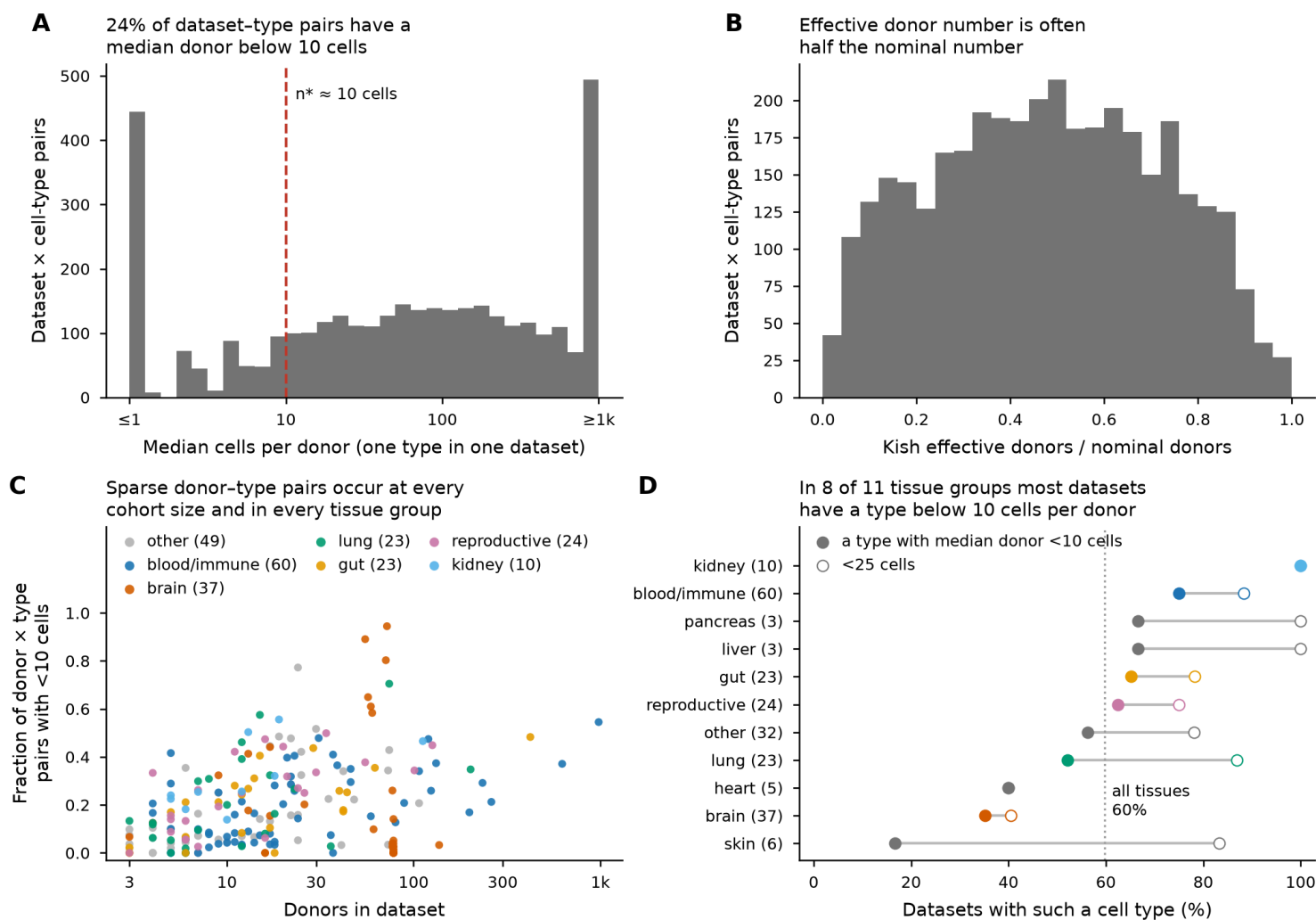


Figure 1

Most of 226 public CELLxGENE references from all tissues are thin at the donor level. (A) Median cells per donor for each of 3,615 dataset–cell-type pairs (x axis clipped to 1–1,000; the first bin holds 362 pairs whose median donor has no cell of the type); dashed line, $n^* \approx 10$ cells, the PBMC estimate. (B) Kish effective donor number divided by the nominal donor number for each pair. (C) Fraction of donor–type pairs with fewer than 10 cells against donors per dataset; points, datasets coloured by tissue group (number of datasets in parentheses). (D) Share of datasets in each tissue group with at least one cell type whose median donor has fewer than 10 (filled) or 25 (open) cells; dotted line, all 226 datasets. Types with fewer than 50 cells in a dataset were dropped.

Alt text: Four panels showing that most public single-cell datasets contain cell types whose donors contribute fewer than 10 cells, in every tissue group and at every cohort size.

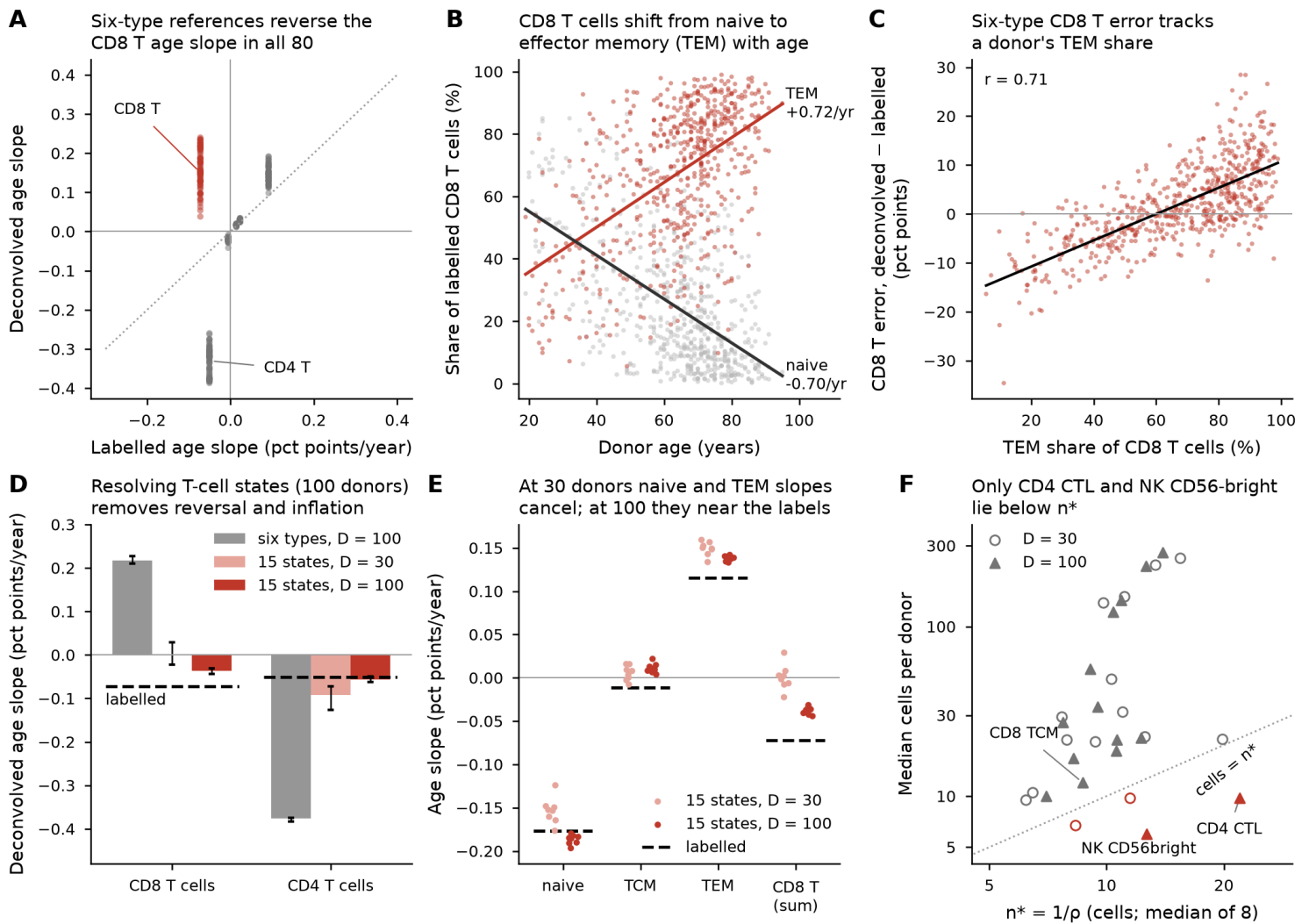


Figure 2

A within-type state shift reverses the CD8 T-cell age association in OneK1K; resolving states removes it. (A) Deconvolved against labelled age slope for six types in 80 six-type references (weighted MuSiC, equal-donor basis; 597 donors; proportions regressed on age, sex and sequencing pool). (B) Effector-memory (TEM, red) and naive (grey) shares of each donor's labelled CD8 T cells against age, with OLS slopes. (C) CD8 T-cell error of six-type references (deconvolved minus labelled proportion, averaged over the 80 references) against the donor's TEM share. (D) Age slopes of CD8 and CD4 T cells (median and range over eight donor sets) with six-type references of 100 donors and with references resolving 15 level-2 states (30 or 100 donors); dashed lines, labelled slopes. (E) The CD8 T-cell slope of the 15-state references decomposed into its naive, TCM and TEM state slopes (points, the eight references with 30 or 100 donors); dashed lines, labelled slopes. (F) Median cells per donor against $n^* = 1/p_k$ for the 15 states (medians over eight references); open circles, 30 donors; triangles, 100 donors; red, states whose median donor lies below n^* ; dotted line, equality.

Alt text: Six panels showing that the deconvolved CD8 T-cell age slope has the wrong sign, that this error follows the shift of CD8 T cells to effector memory with age, and that state-resolved references with 100

donors remove it.

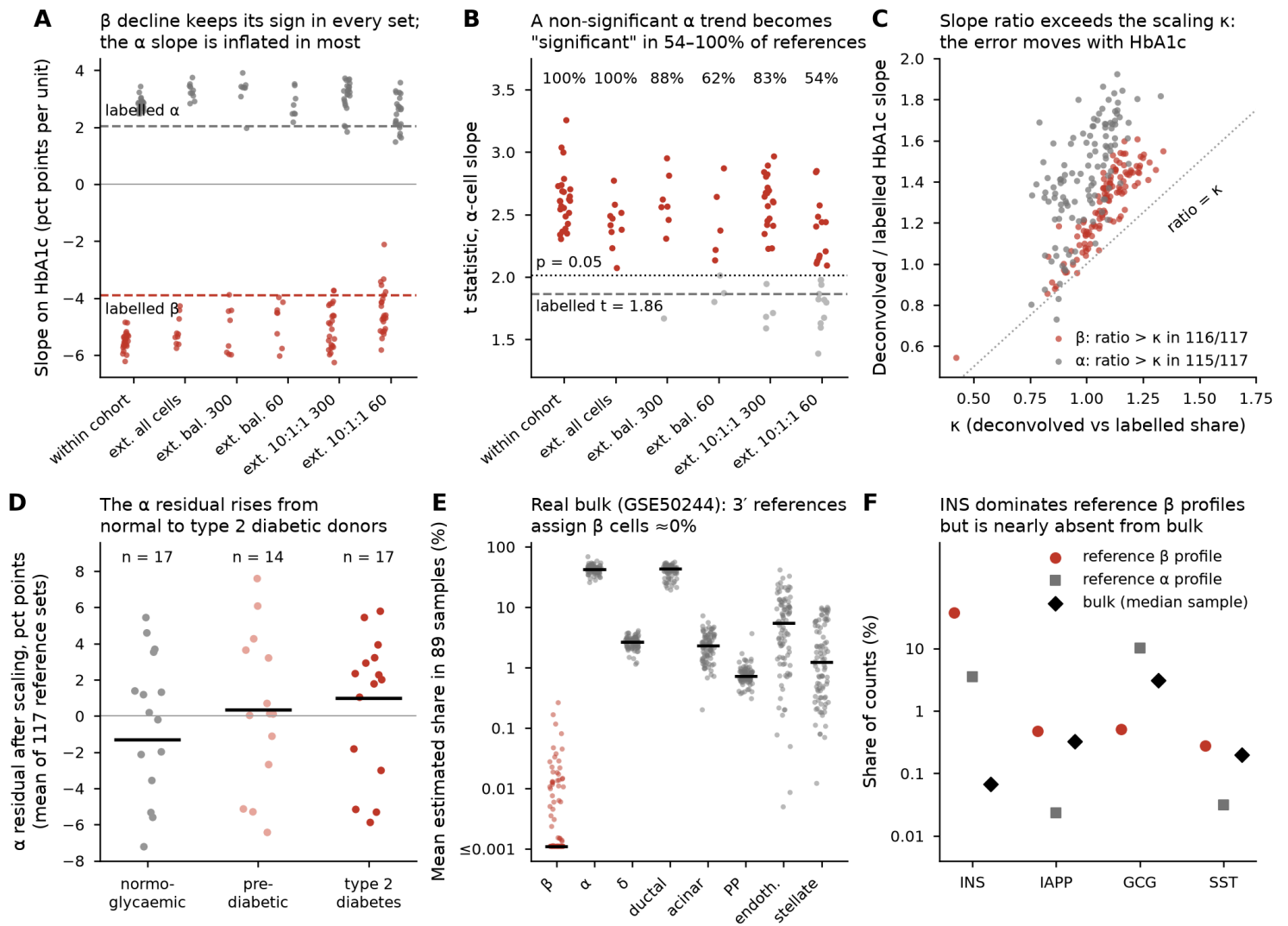


Figure 3

In islets, error that moves with HbA1c makes a borderline association significant; in real bulk, mismatched reference and bulk profiles prevent the test. (A) Slopes of beta-cell (red) and alpha-cell (grey) proportion on HbA1c (adjusted for age and sex) for each reference set; dashed lines, labelled slopes; external fixed-quota columns show three-donor references; "within cohort" pools 3–24 donors and "ext. all cells" 6, 11 and 19 (all designs in Supplementary Table S13). (B) t statistic of the alpha-cell slope; red, $|t| > 2.015$; percentages, share of reference sets in which the non-significant labelled trend ($t = 1.86$) became significant. (C) Ratio of deconvolved to labelled HbA1c slope against κ , the slope of deconvolved on labelled proportion across the 48 donors, for the 117 reference sets; dotted line, ratio = κ (pure scaling). (D) Alpha-cell residual after scaling (deconvolved minus fitted, κ -scaled labelled proportion), averaged over the 117 reference sets, for each donor by disease state; black lines, means. (E) Real islet bulk (GSE50244, 89 samples): mean estimated proportion of each type for the 92 external HPAP references (weighted MuSiC, equal-donor basis; values $\leq 0.001\%$ shown at the axis floor). (F) Share

of counts from four hormone and marker genes in the HPAP reference beta- and alpha-cell profiles and in the median bulk sample.

Alt text: Six panels showing inflated islet association slopes across references, a scaling analysis showing that the excess comes from error that moves with HbA1c, and a real-bulk test in which beta cells are not detected.

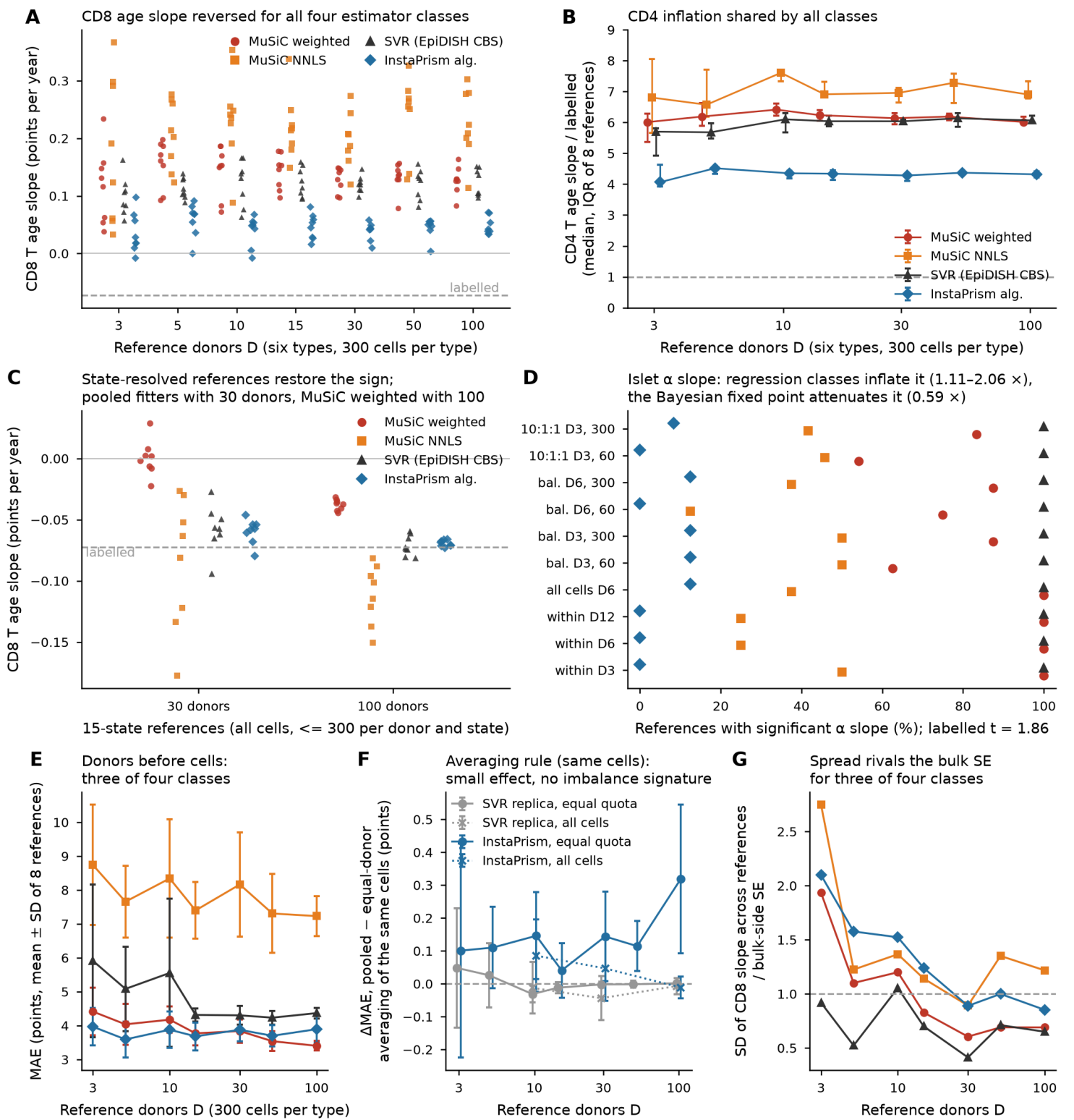


Figure 4

The effects recur across four estimator classes. MW, MuSiC weighted; MN, MuSiC NNLS; SV, CIBERSORT v-SVR procedure (EpiDISH CBS); IP, InstaPrism algorithm (Python transcription of the initial module). (A) CD8 T-cell age slope of each six-type OneK1K reference (300 cells per type, eight donor sets per D); dashed line, labelled slope. (B) CD4 T-cell age slope divided by the labelled slope (median and IQR of eight references). (C) CD8 T-cell age slope of the 15-state references with 30 and 100 donors (six-type sums of the state estimates). (D) Share of islet reference sets with a significant alpha-cell–HbA1c slope ($|t| > 2.015$), by design and estimator (labelled $t = 1.86$; D, reference donors; 60 or 300 cells per type; within, cross-fitted halves of the target cohort); the title gives the median slope ratio to the labelled slope. (E) MAE at a fixed budget of 300 cells per type against donors (mean $\pm$ SD of eight references). (F) Change in MAE when the same cells are averaged per donor and then equally over donors instead of pooled, for a Python replica of the SVR recipe and for IP, in equal-quota (solid) and all-cells (dotted) designs. (G) Standard deviation of the CD8 slope across the eight references of a design divided by the mean bulk-side standard error.

Alt text: Seven panels comparing four deconvolution estimators: all reverse the CD8 age slope and inflate the CD4 slope with six-type references; state-resolved references restore the CD8 sign for all; the islet alpha association is inflated by three regression estimators and attenuated by the Bayesian one; the fixed-budget donor gain holds for three of four estimators but not the InstaPrism algorithm, and equal-donor averaging of the same cells changes pooled-cell estimators little; reference spread exceeds the bulk standard error at three donors for three of four estimators.

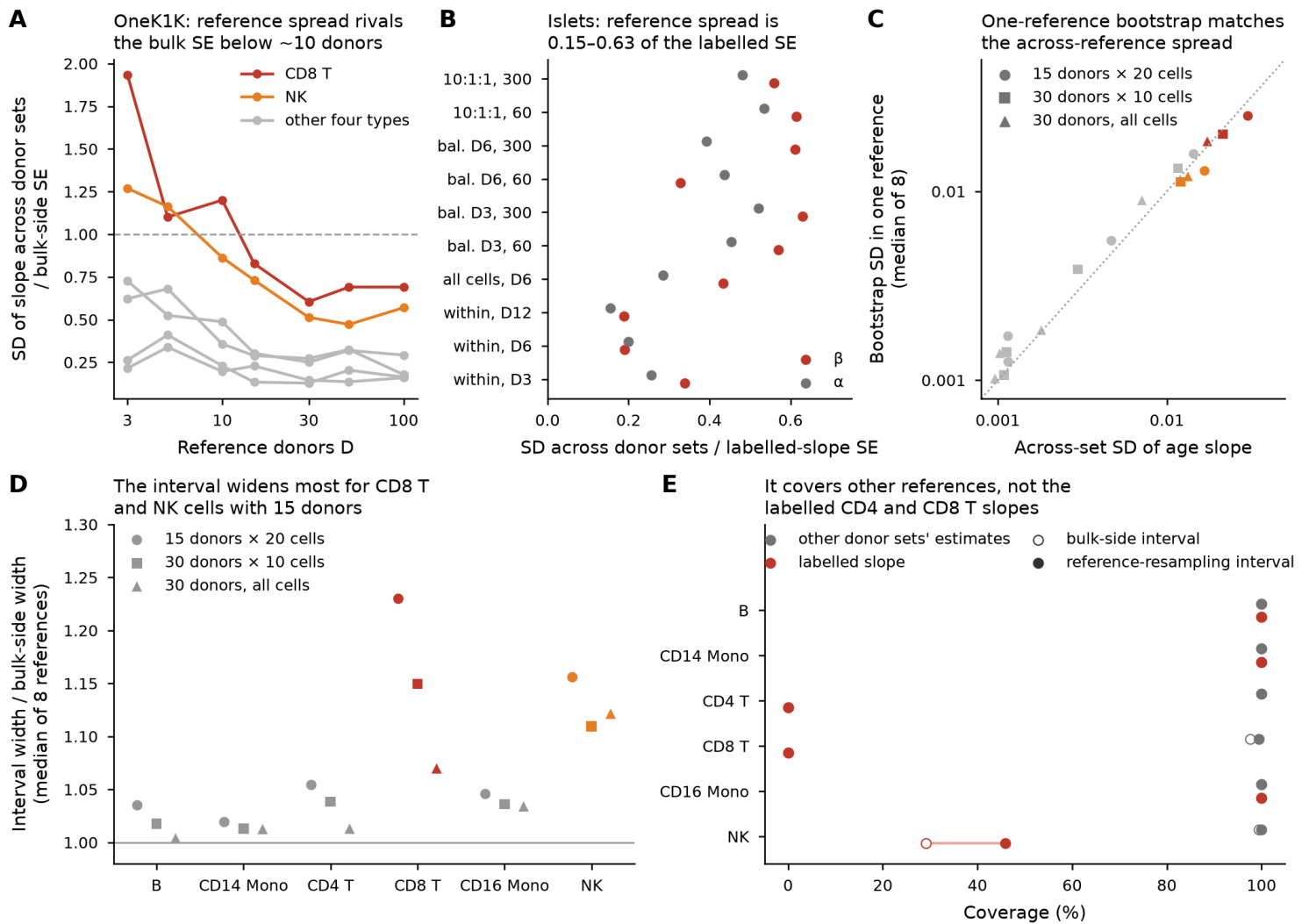


Figure 5

Reference-to-reference variability: its random part can be measured from one reference, its systematic part cannot. (A) OneK1K: standard deviation of the age slope across eight donor sets divided by the mean bulk-side standard error, for fixed-budget references. (B) Islets: standard deviation of the HbA1c slope across donor sets divided by the standard error of the labelled slope, by design (weighted MuSiC, equal-donor basis; D, reference donors; 60 or 300 cells per type). (C) Donor-bootstrap standard deviation of the age slope within one reference (20 replicates; median over eight donor sets) against the standard deviation across the eight donor sets, for six types and three designs; red, CD8 T cells; orange, NK cells; dotted line, equality. (D) Width of the reference-resampling interval, $\pm 1.96 \sqrt{(\text{SE}^2 + \text{SD}_{\text{boot}}^2)}$, relative to the bulk-side interval (median over eight references). (E) Coverage of the bulk-side (open) and reference-resampling (filled) intervals for the point estimates of the other seven donor sets of the same design (grey; mean over pairs) and for the labelled slope (red; 24 references per type).

Alt text: Five panels comparing reference-to-reference spread with bulk-side standard errors, showing that a donor bootstrap within one reference reproduces that spread and that the resulting interval covers other references but not the labelled CD4 and CD8 T-cell slopes.

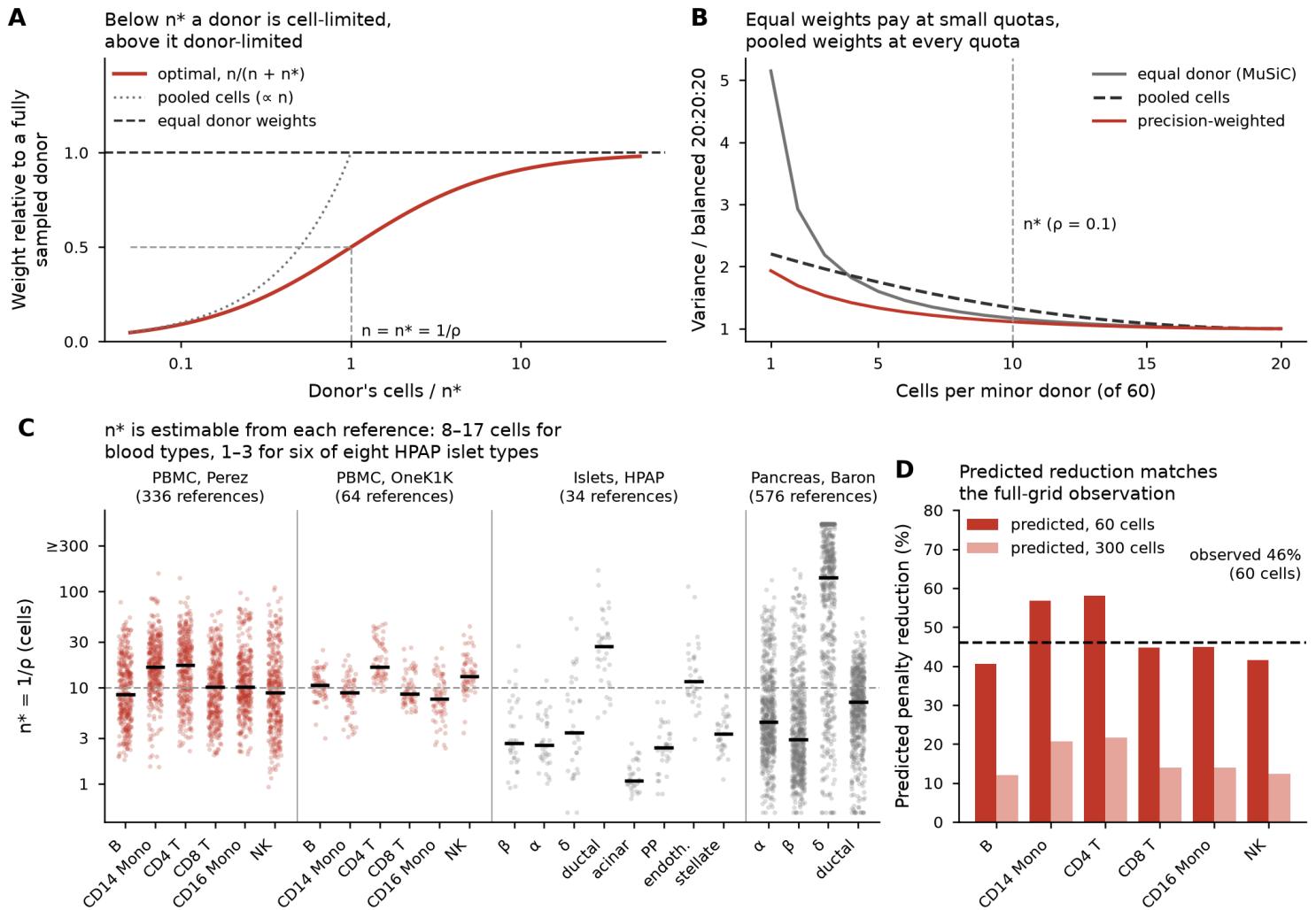


Figure 6

Equal-donor and pooled-cell averaging are two limits of one estimator, indexed by an estimable ratio.

(A) Optimal weight of a donor relative to a fully sampled donor, $n/(n + n^*)$, against its cell count in units of $n^* = 1/\rho$; pooled-cell weights ($\propto n$) and equal weights are the asymptotes. (B) Model variance of the equal-donor, pooled-cell and precision-weighted means for 60 cells split $(60 - 2q):q:q$ among three donors, relative to 20:20:20 ($\rho = 0.1$). (C) $n^* = 1/\rho_k$ estimated in each reference, by cell type and source: 336 three-donor references of the diagnostic subset (Perez), 64 OneK1K references with 10–100 donors, 34 external fixed-budget HPAP islet references and 576 references of the four-donor Baron pancreas design; points, references (values above 300 shown at 300); black lines, medians; dashed line, 10 cells. (D) Reduction of the 10:1:1 penalty predicted from the median ρ_k of the diagnostic subset at 60 and 300 cells per type; dashed line, reduction observed on the full grid at 60 cells per type.

Alt text: Model curves for donor weights and variance, the distribution of the estimated characteristic count n^* across four reference sources, and predicted penalty reductions by cell type.

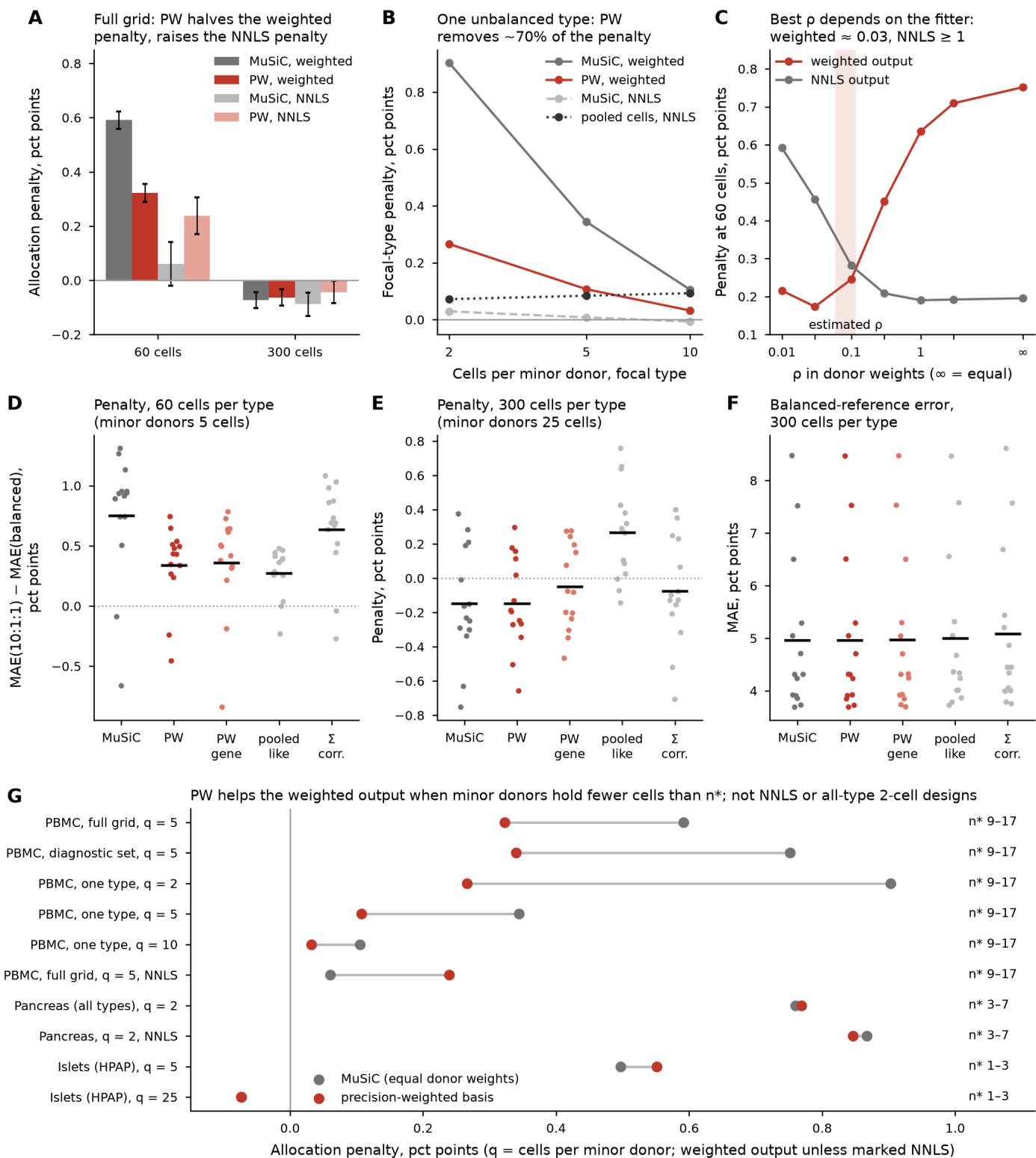


Figure 7

Precision weighting in unchanged MuSiC 1.0.0 on the frozen grids of the previous study: where it helps and where it does not. (A) Full grid (12 draw blocks, 840 targets; Table 3): allocation penalty of both outputs with the equal-donor (MuSiC) and precision-weighted (PW) bases at 60 and 300 cells per type; error bars, block MCSE. (B) Imbalance confined to one type (FCGR3A+ monocytes or NK cells; other types balanced at 300 cells): penalty of the focal type against cells per minor donor. (C) Diagnostic

subset: penalty at 60 cells per type against a common p used in the donor weights, for both outputs; shaded band, estimated p_k . (D–F) Diagnostic subset (14 held-out donors; points, donors; black lines, means): penalty at 60 (D) and 300 (E) cells per type and error of balanced references (F) for MuSiC, PW, its gene-level version, a median- p version that behaves like pooled weights, and MuSiC with sampling-corrected Σ . (G) Penalty with the MuSiC and PW bases in every setting tested; q , cells per minor donor; n^* , range of the median $1/p_k$ over cell types, without pancreatic delta cells (140) and islet ductal and endothelial cells (27 and 12); pancreas, four-donor Baron design with every type at 24 cells ($q = 2$); islets, three external HPAP donors at 60 or 300 cells per type. Weighted output unless marked NNLS.

Alt text: Seven panels comparing allocation penalties with equal-donor and precision-weighted bases across simulation designs, outputs and tissues.

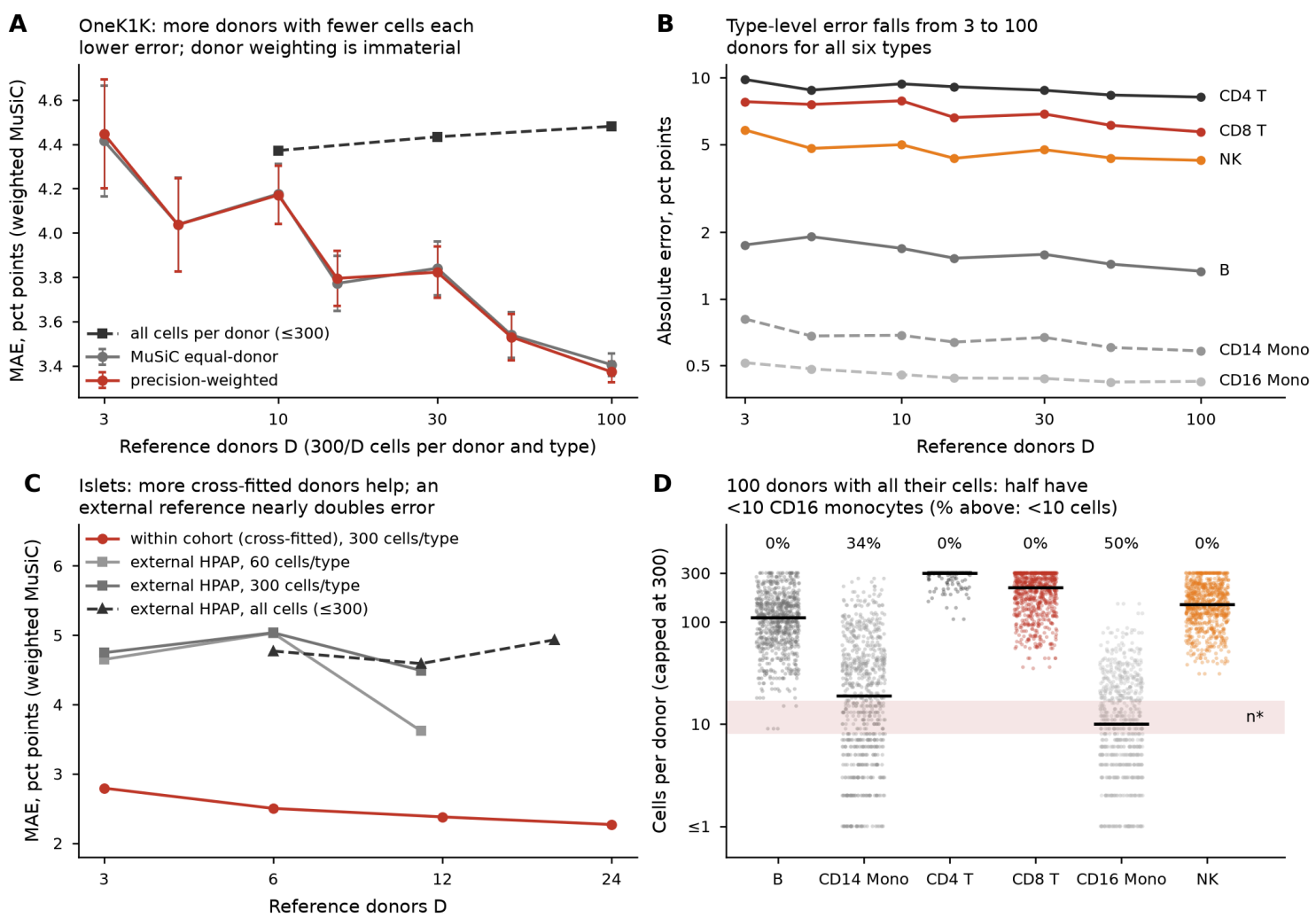


Figure 8

At a fixed cell budget, many donors with few cells each outperform few donors with many cells. (A) OneK1K, 597 target pseudobulks: MAE of weighted MuSiC against donor number for fixed-budget references (300 cells per type, 300/D per donor) with both bases, and for all-cells references (every cell up to 300 per donor and type); mean $\pm$ s.e.m. over eight donor sets. (B) Absolute error of each cell type

for the fixed-budget references (weighted MuSiC, equal-donor basis); dashed lines, monocytes, whose realised totals fell short of 300 at small D. (C) Islets, 48 target donors: MAE of weighted MuSiC for within-cohort cross-fitted references (300 cells per type) and external HPAP references (fixed budget at 60 or 300 cells per type, or all cells). (D) Cells per donor and type in the eight all-cells OneK1K references with 100 donors (800 donor entries, capped at 300); black lines, medians; shaded band, $n^* = 8-17$ cells; percentages, donors with fewer than 10 cells.

Alt text: Line plots showing that deconvolution error falls as cells are spread over more donors in two cohorts, and a strip plot of natural cells per donor by cell type.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1SupplementaryTables.zip](#)