Supplementary Information 2: Figures S1–S9

Figure legends

Fig. S1 Epithelial annotation evidence and donor composition

Marker-expression dot plot, cluster-to-state evidence and donor/state composition. Cell-weighted marker summaries are displays, not inferential tests. Donor-dominated uncertain clusters are flagged

Fig. S2 Complete donor-level primary-screen audit

Minimum P/FDR values, tested-gene counts and donor eligibility for all four modeled epithelial states. Zero genes passed the FDR threshold

Fig. S3 Six post-hoc candidate modules and selection steps

All six modules passing the recorded exploratory filter are shown, including Stem_progenitor_SB_M12. Stage-label-blind covariance construction is separated from later stage-aware, paired-FAP and leave-one-donor-out filters

Fig. S4 M02 gene composition, annotation overlap and leave-one-gene influence

The canonical 36 genes, descriptive categories, selection-affected gene contributions and overlap with epithelial-state annotation evidence. No g:Profiler term was significant after g:SCS correction. The figure spans two pages to maintain readable lettering

Fig. S5 GSE161277 patient-level results for all six candidate modules

Patient differences and six-module adjusted results in the three matched donors. M02 is not presented as independently supported

Fig. S6 Bulk cohort details, standardized sensitivity and cancer-only context

Platform-specific bulk effects and leave-one-accession-out summaries are separated from GSE132465 and TCGA cancer-only context panels. Sample-scoped arrays are not relabeled as patients

Fig. S7 Stool-RNA model degeneration and one-time test

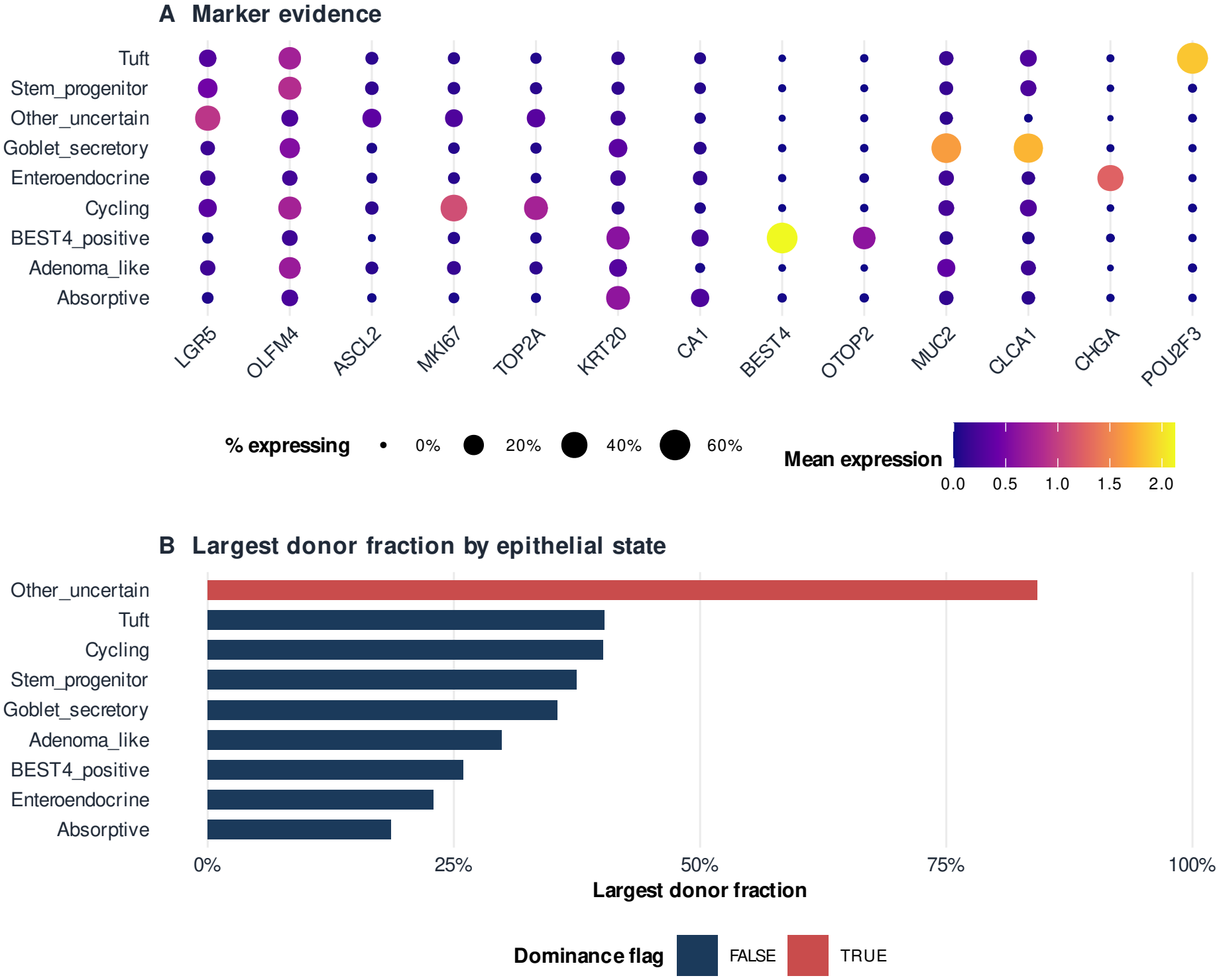
Selected-feature counts, coefficient structure, number of unique predicted probabilities and frozen test AUCs. The test split was used once and retired; no direct 36-gene test was added

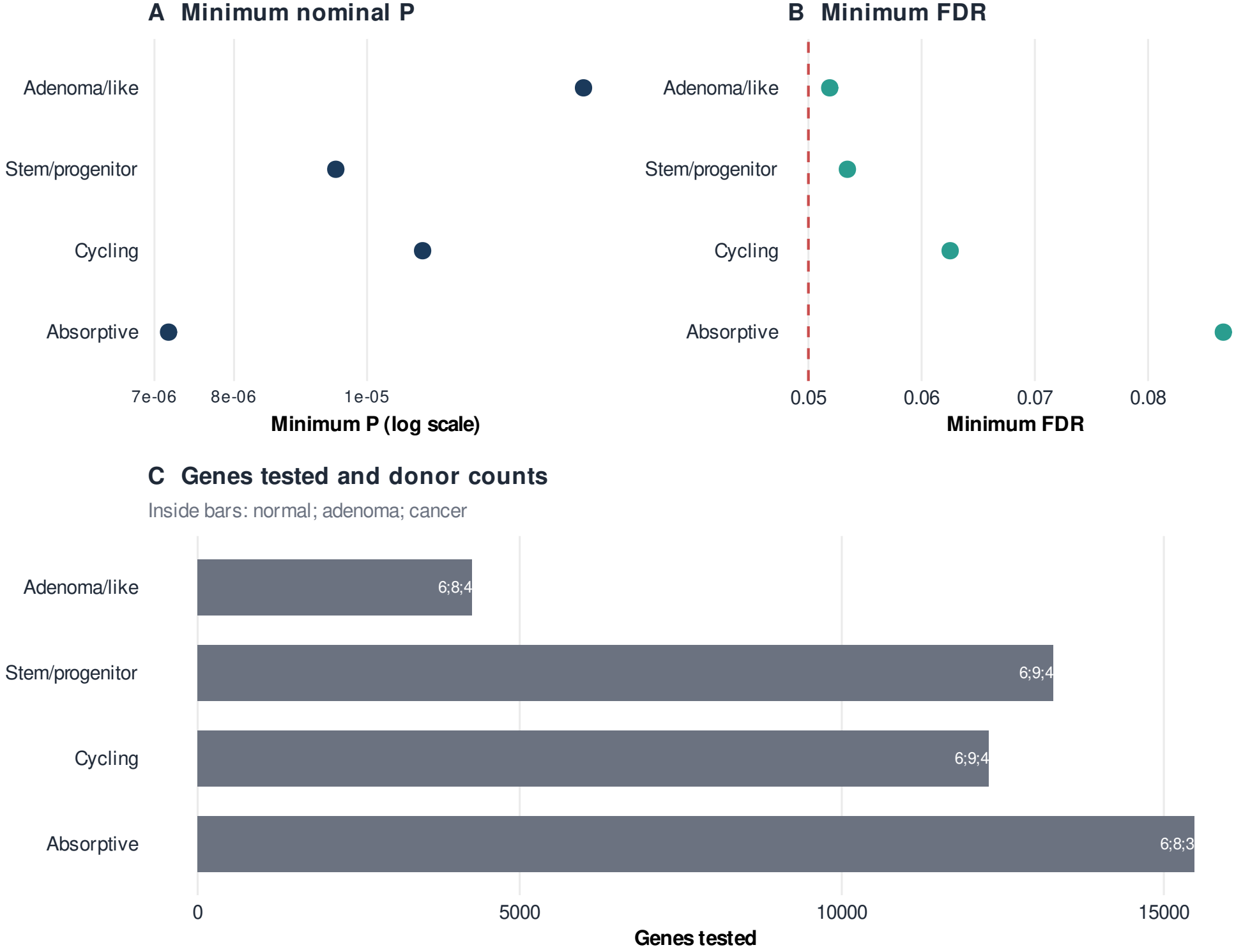
Fig. S8 Figshare feature mapping and endpoint non-estimability

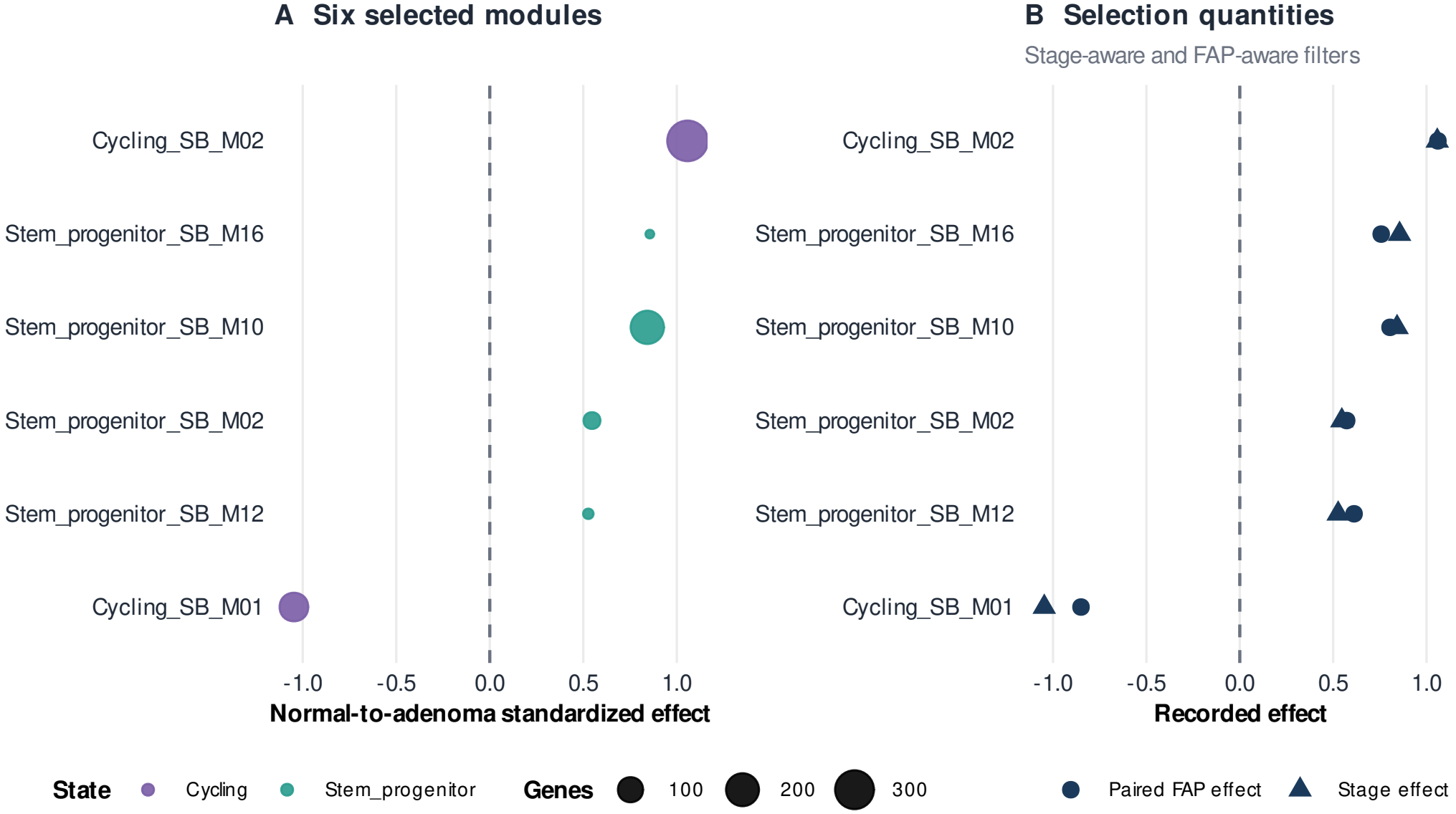
The mapping path from 36 canonical genes to 33 available features is shown. The historical “29/36” name represented a minimum threshold, while all 33 mapped genes were averaged. The canonical 36-gene endpoint was not estimable

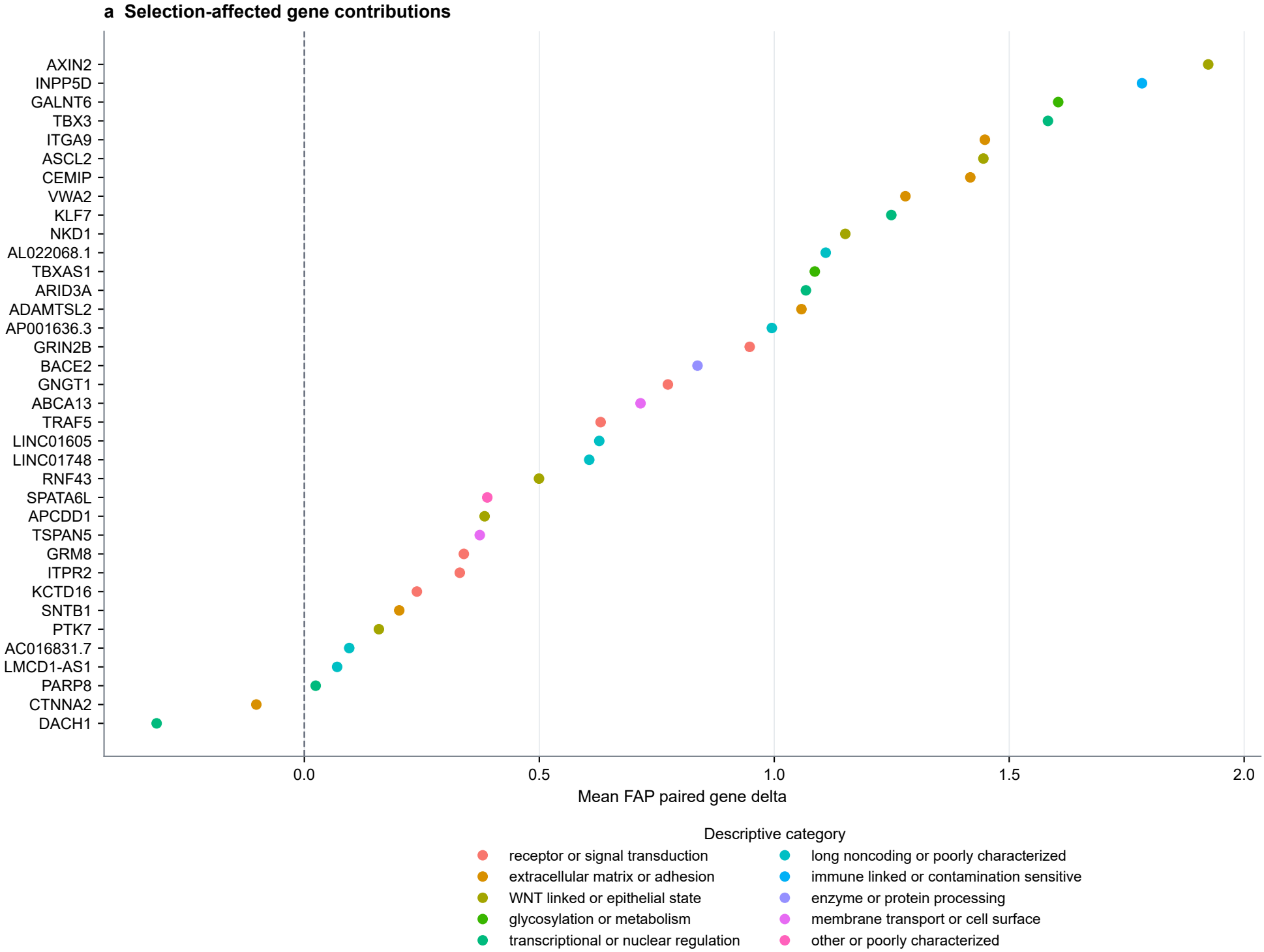
Fig. S9 Simulation-based spatial recoverability and region-registration feasibility

Pseudo-spot score recovery is labeled as simulation-based technical performance. Region-registration agreement was assessed without M02 values but lacked an independent pathologist gold standard. Only case4 passed. Its primary and most sensitivity scores were negative, whereas gene-z was effectively zero; the frozen description was method dependent. Patient-level spatial analysis was not estimable


A Marker evidence
Tuft
Stem_progenitor
Other_uncertain
Goblet_secretory
Enteroendocrine
Cycling
BEST4_positive
Adenoma_like
Absorptive
LGR5
OLFM4
ASCL2
MKI67
TOP2A
KRT20
CA1
BEST4
OTOP2
MUC2
CLCA1
CHGA
POU2F3
% expressing
0%
20%
40%
60%
Mean expression
0.0
0.5
1.0
1.5
2.0
B Largest donor fraction by epithelial state
Other_uncertain
Tuft
Cycling
Stem_progenitor
Goblet_secretory
Adenoma_like
BEST4_positive
Enteroendocrine
Absorptive
0%
25%
50%
75%
100%
Largest donor fraction
Dominance flag
FALSE
TRUE


A Minimum nominal P
Adenoma/like
Stem/progenitor
Cycling
Absorptive
7e-06
8e-06
1e-05
Minimum P (log scale)
B Minimum FDR
Adenoma/like
Stem/progenitor
Cycling
Absorptive
0.05
0.06
0.07
0.08
Minimum FDR
C Genes tested and donor counts
Inside bars: normal; adenoma; cancer
Adenoma/like
6;8;4
Stem/progenitor
6;9;4
Cycling
6;9;4
Absorptive
6;8;3
0
5000
10000
15000
Genes tested


A Six selected modules
Cycling_SB_M02
Stem_progenitor_SB_M16
Stem_progenitor_SB_M10
Stem_progenitor_SB_M02
Stem_progenitor_SB_M12
Cycling_SB_M01
-1.0
-0.5
0.0
0.5
1.0
Normal-to-adenoma standardized effect
B Selection quantities
Stage-aware and FAP-aware filters
Cycling_SB_M02
Stem_progenitor_SB_M16
Stem_progenitor_SB_M10
Stem_progenitor_SB_M02
Stem_progenitor_SB_M12
Cycling_SB_M01
-1.0
-0.5
0.0
0.5
1.0
Recorded effect
State
Cycling
Stem_progenitor
Genes
100
200
300
Paired FAP effect
Stage effect

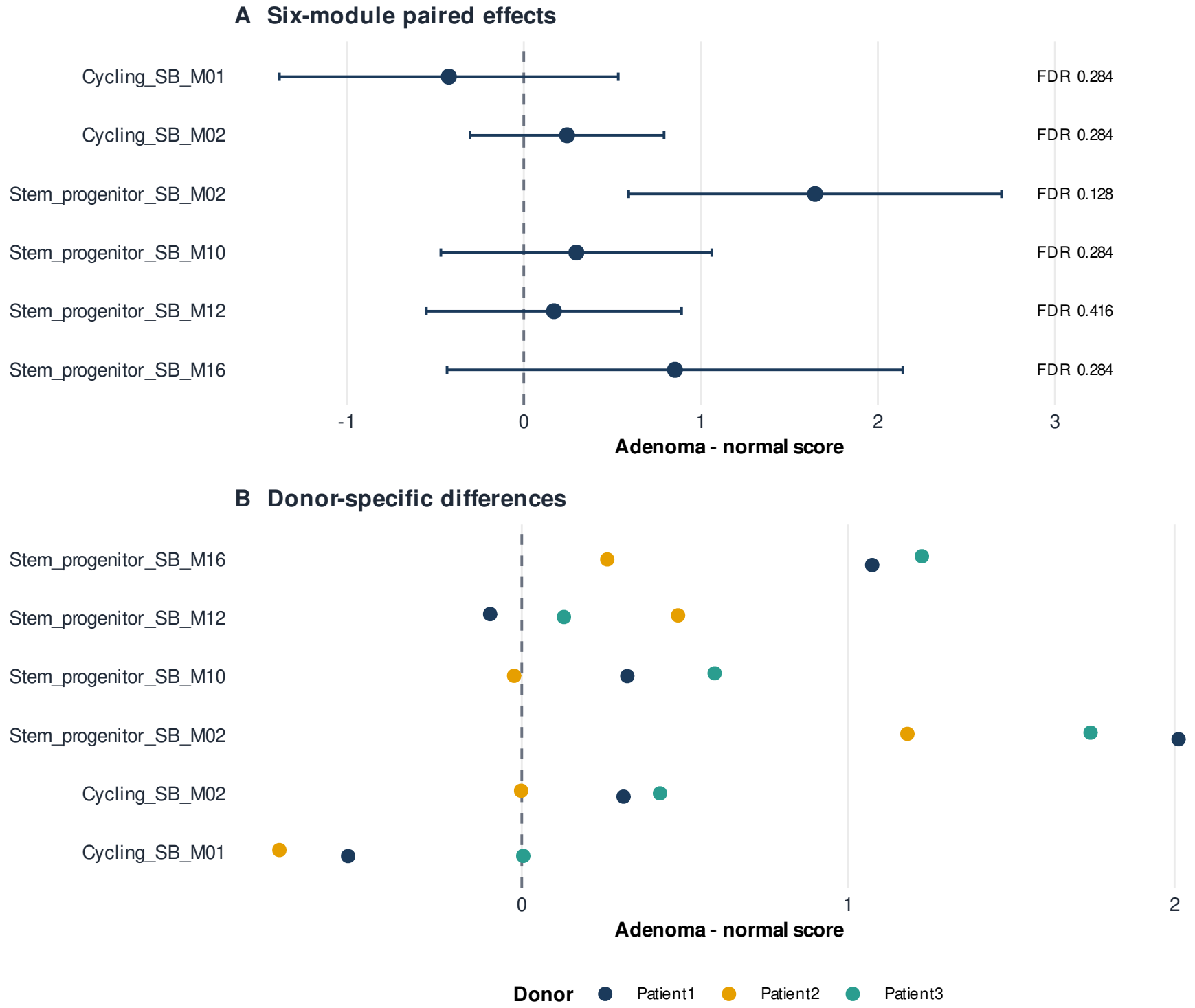

a Selection-affected gene contributions
AXIN2
INPP5D
GALNT6
TBX3
ITGA9
ASCL2
CEMIP
VWA2
KLF7
NKD1
AL022068.1
TBXAS1
ARID3A
ADAMTSL2
AP001636.3
GRIN2B
BACE2
GNGT1
ABCA13
TRAF5
LINC01605
LINC01748
RNF43
SPATA6L
APCDD1
TSPAN5
GRM8
ITPR2
KCTD16
SNTB1
PTK7
AC016831.7
LMCD1-AS1
PARP8
CTNNA2
DACH1
0.0
0.5
1.0
1.5
2.0
Mean FAP paired gene delta
Descriptive category
receptor or signal transduction
extracellular matrix or adhesion
WNT linked or epithelial state
glycosylation or metabolism
transcriptional or nuclear regulation
long noncoding or poorly characterized
immune linked or contamination sensitive
enzyme or protein processing
membrane transport or cell surface
other or poorly characterized

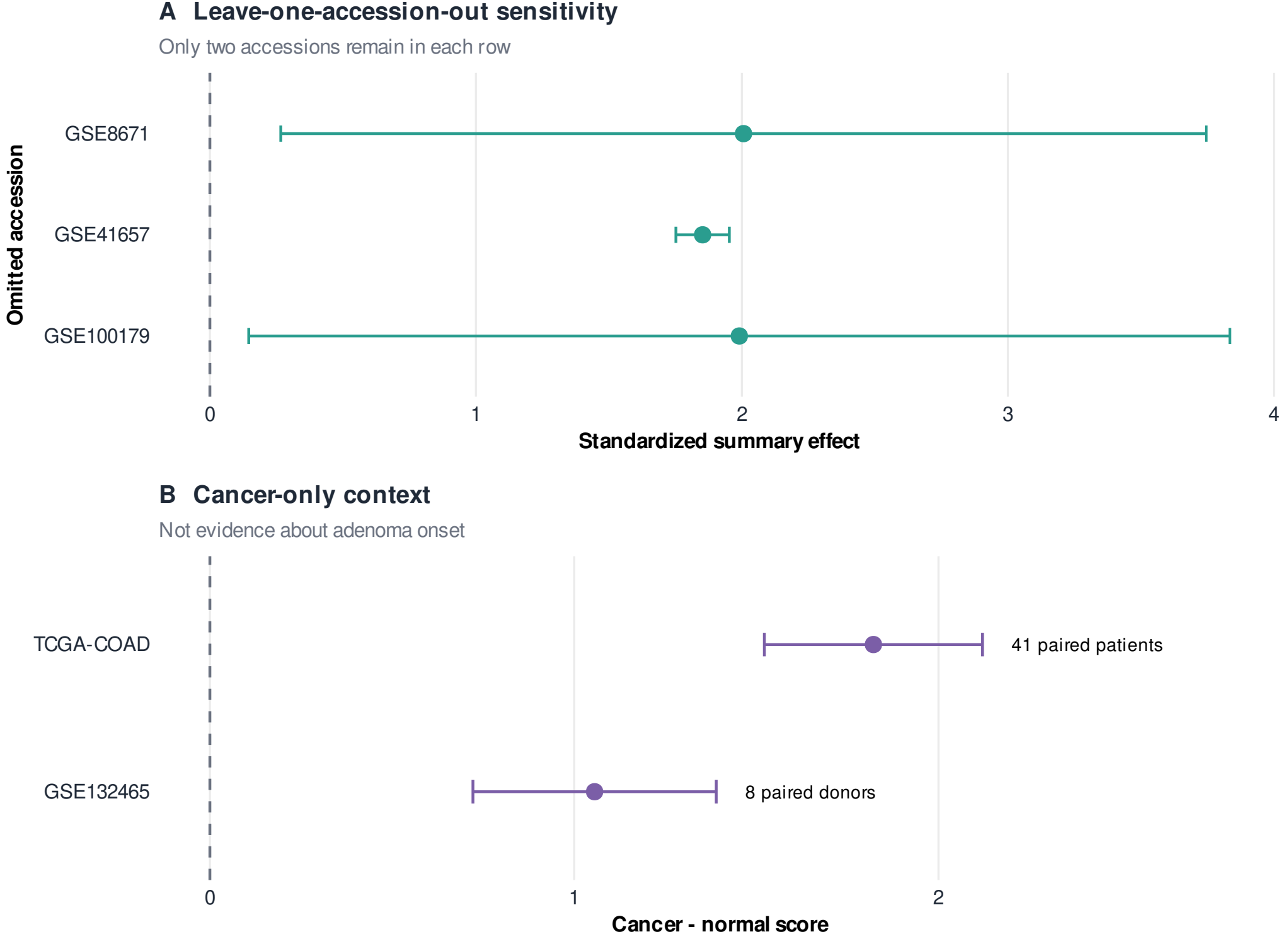
b Overlap with annotation evidence

Five recorded overlaps; module characterization is not independent of all state-label evidence

M02 gene	Evidence source	Label / cluster	Interpretive consequence
NKD1	cluster annotation	cluster 1	overlap disclosed; no exclusion from M02
APCDD1	cluster annotation	cluster 1	overlap disclosed; no exclusion from M02
NKD1	cluster annotation	cluster 8	overlap disclosed; no exclusion from M02
APCDD1	cluster annotation	cluster 8	overlap disclosed; no exclusion from M02
ASCL2	broad annotation evidence	Stem/progenitor	overlap disclosed; no exclusion from M02

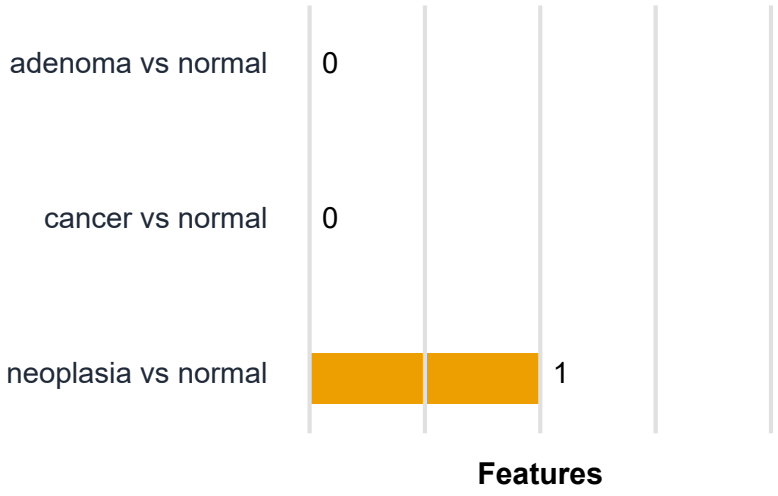
No g:Profiler term reached adjusted P < 0.05 against the 1,000-gene Stem/progenitor construction background


A Six-module paired effects
Cycling_SB_M01
Cycling_SB_M02
Stem_progenitor_SB_M02
Stem_progenitor_SB_M10
Stem_progenitor_SB_M12
Stem_progenitor_SB_M16
FDR 0.284
FDR 0.284
FDR 0.128
FDR 0.284
FDR 0.416
FDR 0.284
-1
0
1
2
3
Adenoma - normal score
B Donor-specific differences
Stem_progenitor_SB_M16
Stem_progenitor_SB_M12
Stem_progenitor_SB_M10
Stem_progenitor_SB_M02
Cycling_SB_M02
Cycling_SB_M01
0
1
2
Adenoma - normal score
Donor
Patient1
Patient2
Patient3

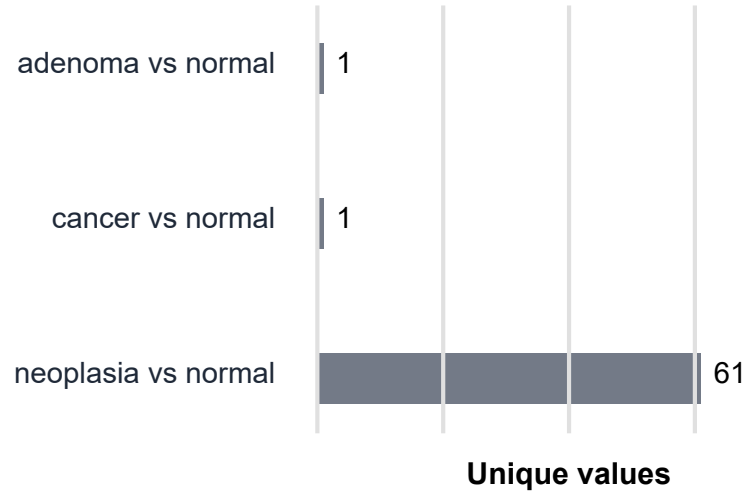

A Leave-one-accession-out sensitivity
Only two accessions remain in each row
Omitted accession
GSE8671
GSE41657
GSE100179
0
1
2
3
4
Standardized summary effect
B Cancer-only context
Not evidence about adenoma onset
TCGA-COAD
41 paired patients
GSE132465
8 paired donors
0
1
2
Cancer - normal score

A Selected features

Non-intercept terms

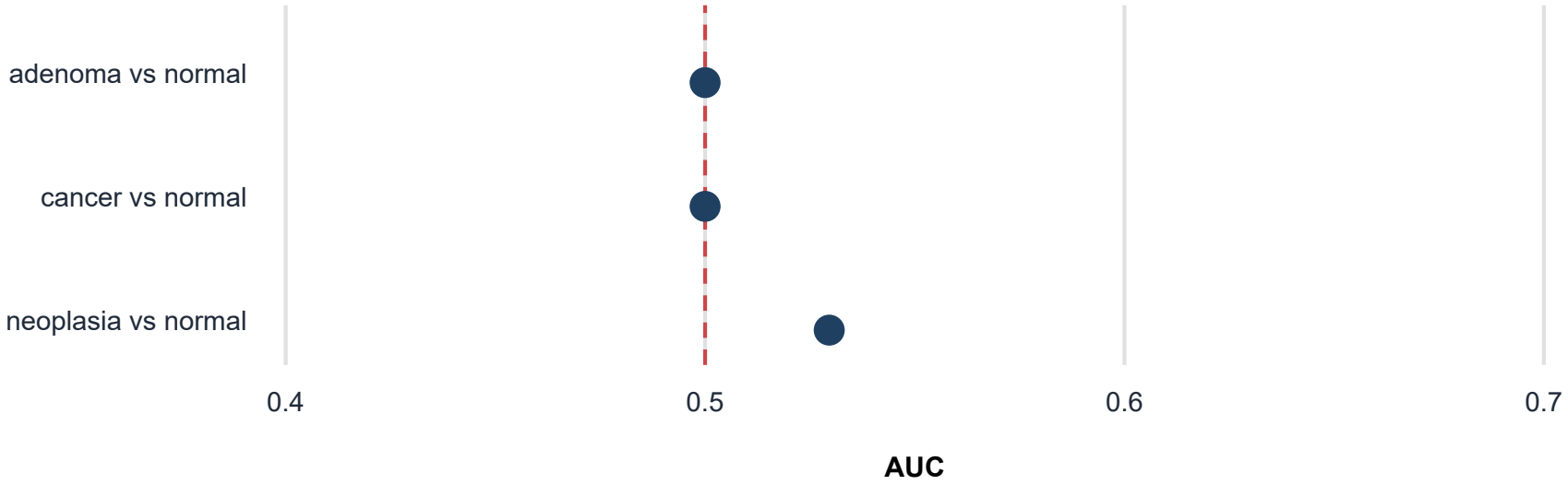


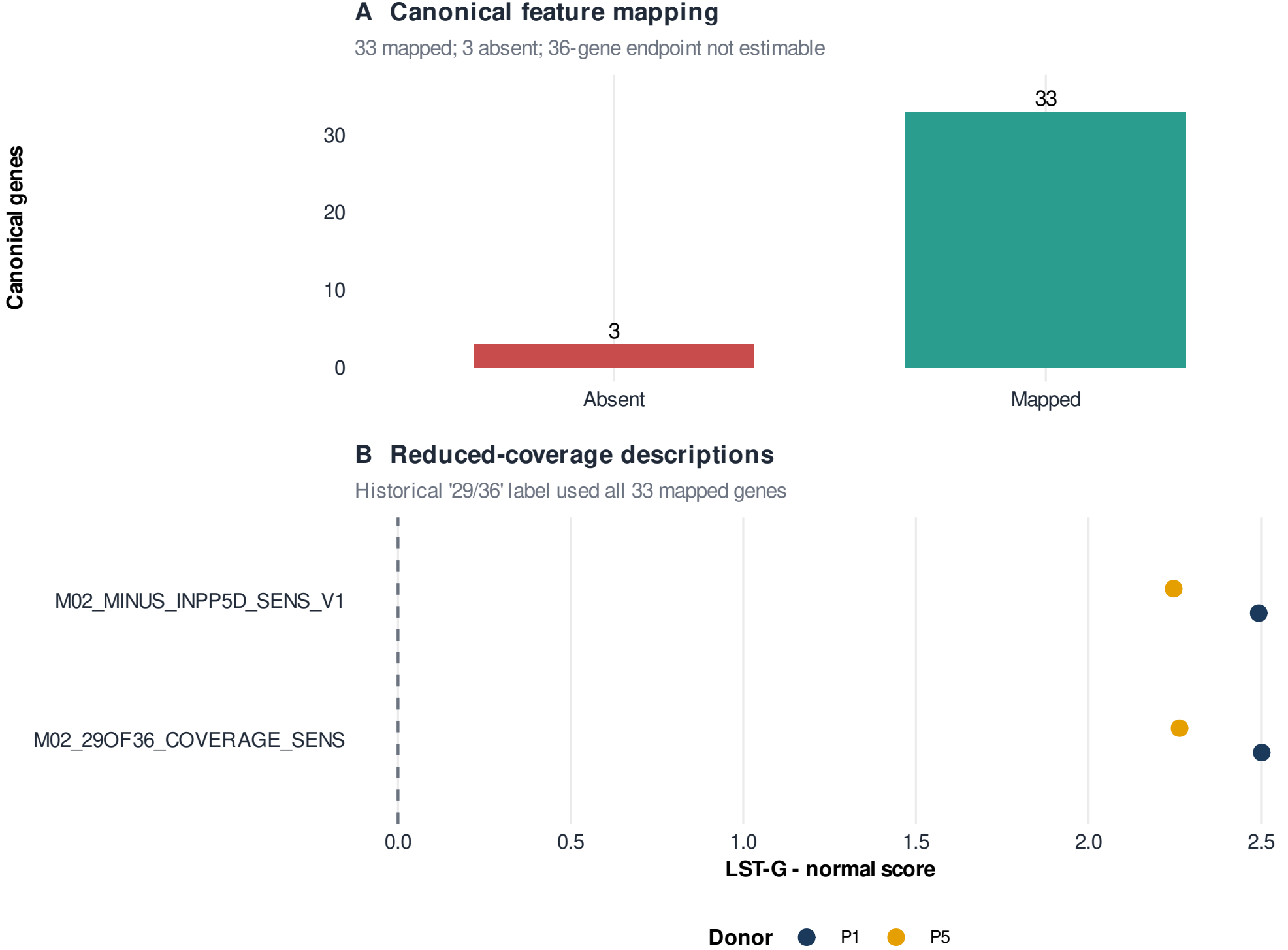
B Unique test probabilities

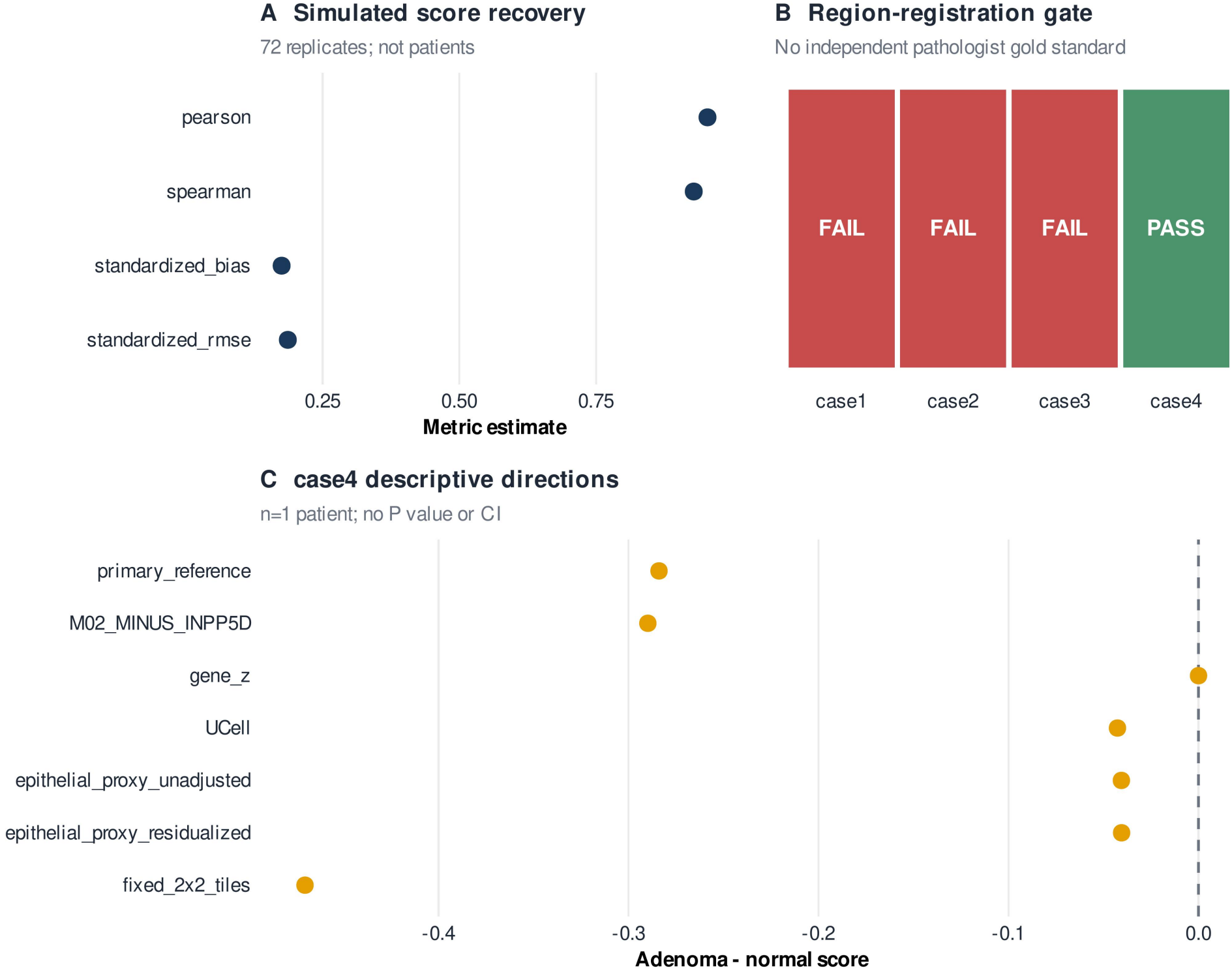


C One-time test AUC

Frozen test set




A Canonical feature mapping
33 mapped; 3 absent; 36-gene endpoint not estimable
Canonical genes
30
20
10
0
3
33
Absent
Mapped
B Reduced-coverage descriptions
Historical '29/36' label used all 33 mapped genes
M02_MINUS_INPP5D_SENS_V1
M02_29OF36_COVERAGE_SENS
0.0
0.5
1.0
1.5
2.0
2.5
LST-G - normal score
Donor
P1
P5


A Simulated score recovery
72 replicates; not patients
pearson
spearman
standardized_bias
standardized_rmse
0.25
0.50
0.75
Metric estimate
B Region-registration gate
No independent pathologist gold standard
FAIL
FAIL
FAIL
PASS
case1
case2
case3
case4
C case4 descriptive directions
n=1 patient; no P value or CI
primary_reference
M02_MINUS_INPP5D
gene_z
UCell
epithelial_proxy_unadjusted
epithelial_proxy_residualized
fixed_2x2_tiles
-0.4
-0.3
-0.2
-0.1
0.0
Adenoma - normal score

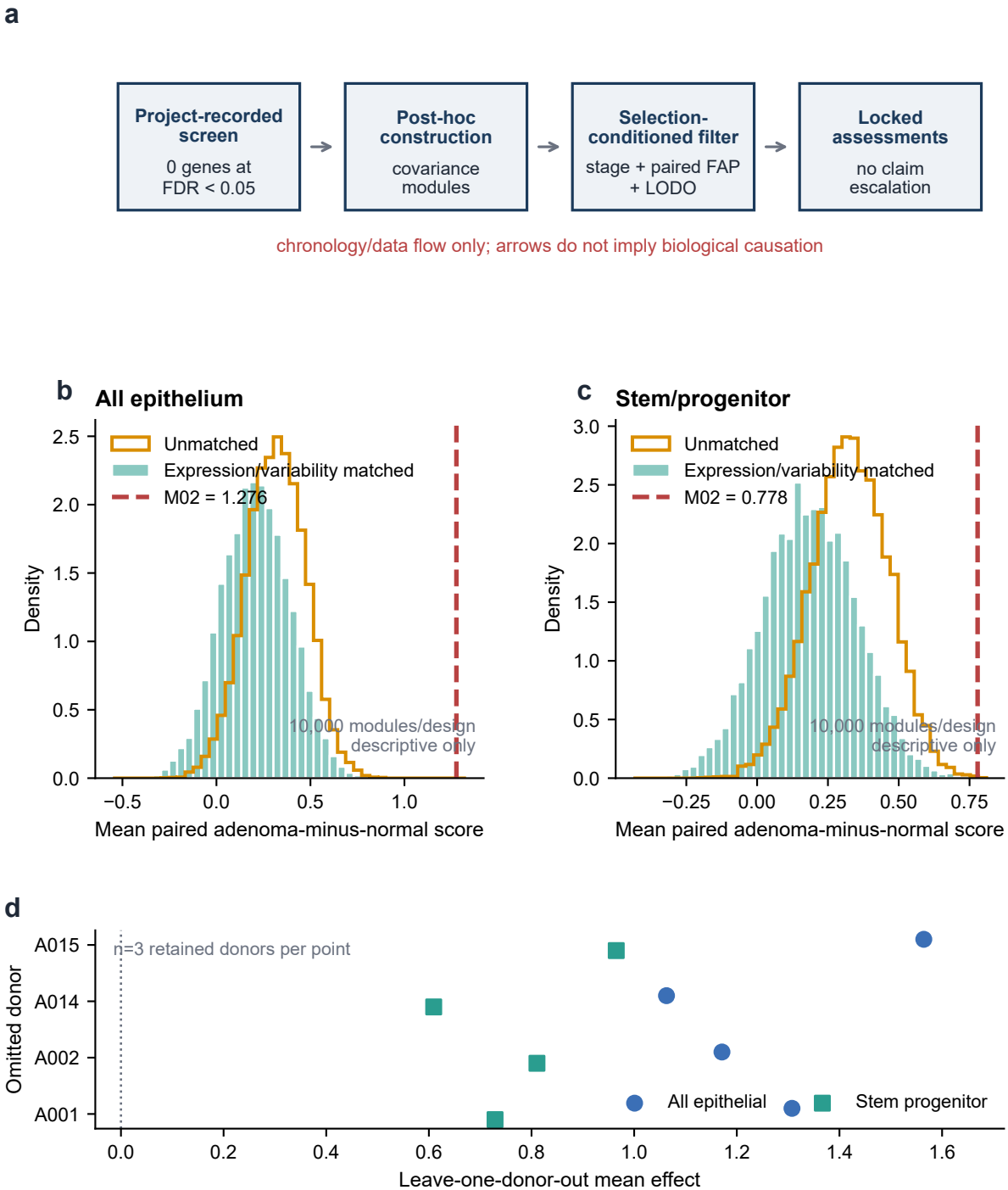

a
Project-recorded screen
0 genes at FDR < 0.05
Post-hoc construction
covariance modules
Selection-conditioned filter
stage + paired FAP + LODO
Locked assessments
no claim escalation
chronology/data flow only; arrows do not imply biological causation
b
All epithelium
Unmatched
Expression/variability matched
M02 = 1.276
10,000 modules/design
descriptive only
Density
Mean paired adenoma-minus-normal score
c
Stem/progenitor
Unmatched
Expression/variability matched
M02 = 0.778
10,000 modules/design
descriptive only
Density
Mean paired adenoma-minus-normal score
d
n=3 retained donors per point
Omitted donor
A015
A014
A002
A001
All epithelial
Stem progenitor
Leave-one-donor-out mean effect
Same four module-generating FAP donors; empirical tail fractions are not post-selection P values

Fig. S10 Selection chronology, random-gene-set reference and leave-one-donor-out audit

(a) Analysis chronology distinguishes the negative project-recorded screen, post-hoc module generation, candidate lock and later assessments. (b,c) M02 effects are overlaid on 10,000 expression- and variability-matched and 10,000 unmatched random 36-gene module distributions for all epithelium and Stem/progenitor profiles. (d) Four leave-one-donor-out means in each compartment. Tail fractions are descriptive and are not valid post-selection P values; all panels use the same four module-generating FAP donors.