Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Olfactory receptor enrichment in fold-change rankings can arise without regulation: a null benchmark across 99 human transcriptomes

Kenji Ohe

oekenji@gmail.com

Fukuoka University https://orcid.org/0000-0002-1958-7572

Research Article

Keywords: olfactory receptors, ectopic expression, gene set enrichment, fold change ranking, detection above background, false positives, microarray, transcriptome reanalysis

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11719825/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.

Olfactory receptors top fold-change rankings by chance

Title page

Title: Olfactory receptor enrichment in fold-change rankings can arise without regulation: a null benchmark across 99 human transcriptomes

Running title: Spurious olfactory receptor enrichment in fold-change rankings

Author: Kenji Ohe

Department of Clinical Pharmacotherapy, Faculty of Pharmaceutical Sciences, Fukuoka University, Fukuoka, Japan

Correspondence: Kenji Ohe (ohekenji@fukuoka-u.ac.jp)

Keywords: olfactory receptors; ectopic expression; gene set enrichment; fold change ranking; detection above background; false positives; microarray; transcriptome reanalysis

Abstract

Background: Olfactory receptors are the largest gene family in the human genome and are essentially unexpressed outside olfactory sensory neurons. Enrichment analyses are often run on genes ranked by fold change, and a family that is uniformly undetected can occupy the extremes of such a ranking without any regulation having occurred. How often this happens, and whether it can be controlled for, has not been quantified.

Methods: We assembled 99 independent human transcriptome datasets profiled on one array platform (GPL6244; 6–14 samples each; 17,795–19,036 genes after collapsing to symbols, of which a median of 360 are olfactory receptors, 1.89% of the transcriptome). For each dataset we drew up to ten random balanced splits of the samples, giving 990 analyses in which no true group difference exists. In each we ranked genes by log2 fold change and counted olfactory receptors among the top 100. A two-component Gaussian mixture on mean log2 signal defined a detection boundary. As a control we drew 200 sets of non-olfactory genes matched on mean signal to within 0.05 log2 and counted their members in the same top 100. Two datasets with real two-group contrasts were examined in detail with 2,000 matched sets.

Results: Olfactory receptors fell below the detection boundary in a median of 99.7% of cases per dataset (interquartile range 99.4–100%), and in at least 98.6% in 88 of the 99 datasets. Under the null, the mean number among the top 100 was 1.54 against the 1.89

expected from their genome-wide share, but the distribution had a long tail: 13.3% of the 990 analyses placed four or more receptors in the top 100, 8.2% placed six or more and 3.3% placed nine or more. The excess tracked how far the ranking was dominated by undetected genes (Spearman ρ = 0.49). Where background-level genes made up at least 70% of the top 100, the mean receptor count was 5.58 and 61% of analyses reached four or more; where they made up less than 30%, the mean was 0.37 and 1% reached four. Expression-matched sets reproduced the observed counts closely (null mean 1.15), and the matched test rejected at 7.7% against a nominal 5%. In two datasets with real contrasts, receptors reached 4.8-fold and 4.2-fold their genome-wide share at the top of the ranking, and in both the matched test attributed this entirely to signal level (P = 0.30 and 0.50).

Conclusions: An apparent olfactory receptor signature at the top of a fold-change ranking is common when the ranking is dominated by undetected transcripts, and arises without regulation. Reporting the signal level of every gene named in an enrichment result, and testing family-level enrichment against expression-matched genes rather than the whole transcriptome, separates the two. The matched test is close to calibrated but slightly anti-conservative, so it supports rather than replaces direct measurement.

Introduction

Olfactory receptors form the largest gene family in the human genome, with roughly 400 intact members. Their expression is confined almost entirely to olfactory sensory neurons, and within those neurons each cell transcribes essentially one receptor allele, so even in olfactory epithelium the remaining several hundred receptors are silent [1]. That silence rests on chromatin rather than on the mere absence of an activating signal: the receptor clusters carry the hallmarks of constitutive heterochromatin, H3K9me3 and H4K20me3, and the silencing is laid down before any receptor is chosen rather than as a consequence of choosing one [2]. The same H3K9 methylation machinery governs receptor transcription in cells with no relation to olfaction [3]. A family held shut in this way is shut in every tissue, which is what makes it a clean test case for the question asked here. Reports of receptor expression in other tissues have accumulated, and some are supported by functional work, but transcript levels in those settings are low and the measurements are correspondingly delicate [4].

This creates a specific analytical hazard. On a whole-transcriptome array, probes for an absent transcript return background signal, and a ratio formed from two background measurements has much larger variance than a ratio formed from two abundant ones. Genes that are not expressed therefore carry the largest apparent fold changes. When a gene list is built by ranking on fold change, or when a ranked list is passed to a gene set tool, a family that is uniformly absent in the tissue under study will be over-represented at the extremes of that ranking. The enrichment tool sees only the list and its annotations; it

has no access to the signal levels behind them, and it will return the family's Gene Ontology terms — for olfactory receptors, sensory perception of smell and G-protein-coupled receptor signalling — whether or not any transcript was detected. That a gene's chance of entering such a list depends on properties unrelated to regulation, and that enrichment computed against an unmatched background inherits that dependence, is established for sequencing data, where transcript length and read depth bias gene ontology results [5].

The mechanism is not in doubt, but its practical importance is an empirical question that has not been answered. How often does an olfactory receptor signature appear at the top of a fold-change ranking when nothing has been done to the cells? Does it depend on properties of the dataset that an analyst can check beforehand? And is there a control that separates a genuine family-level effect from this behaviour?

We addressed these questions with a null benchmark. We assembled 99 independent human transcriptome datasets on a single array platform, drew random balanced splits of the samples within each dataset so that no true group difference exists, and asked how often olfactory receptors reach the top of the resulting fold-change rankings. We then tested whether comparing the family against non-olfactory genes matched on signal level recovers the correct answer, and we examined two datasets with real experimental contrasts in which receptors did reach the top. Our aim is a check that can be applied to any receptor-family result, including results already in the literature, using only deposited data.

Methods

Dataset assembly

We queried the NCBI Gene Expression Omnibus on 9 November 2026 for human series on platform GPL6244 (Affymetrix Human Gene 1.0 ST) with 6 to 14 samples, and retrieved the deposited series matrices for 107 of them. Restricting to a single platform fixes the probe set, the olfactory receptor content and the summarisation method, so that differences between datasets reflect the data rather than the array design. GPL6244 carries probes for 360 olfactory receptor genes after collapsing to gene symbols, 1.89% of the measured transcriptome, which is close to the family's share of annotated human genes.

A dataset entered the analysis if its matrix held at least 5,000 rows and 6 to 14 samples, at least 8,000 rows survived the join to gene symbols, the fitted mixture assigned between 20% and 95% of genes to the expressed component, and at least 200 olfactory receptor genes were present. Ninety-nine datasets met these criteria. Values above 100 were taken as linear and log2 transformed with a floor of 1. Probes were mapped to gene symbols using the GEO platform annotation, probes mapping to several symbols were

discarded, and for each symbol the probe with the highest mean signal was retained. Values deposited for this platform are usually produced by robust multi-array average summarisation [6], but processing is at the depositor's discretion and is not uniform across series, which matters for the behaviour reported below.

Null splits

For each dataset we enumerated all balanced splits of the samples into two groups of floor(n/2) and drew up to ten at random without replacement, giving 990 analyses in total. Because the split is random with respect to the experimental design, no systematic difference between the groups is expected, and any structure in the resulting ranking is a property of the measurements rather than of a treatment. In each split we computed the log2 fold change per gene as the difference of group means and ranked genes from the largest increase downwards.

Detection boundary

Mean log2 signal across all samples was computed per gene, and a two-component Gaussian mixture was fitted to that distribution with scikit-learn using ten initialisations and a fixed seed. Genes were assigned to the expressed component where its posterior probability exceeded 0.5, and the smallest mean signal meeting that criterion defined the detection boundary for that dataset. This is a distribution-based substitute for probe-level detection-above-background, which requires raw CEL files and the antigenomic probe set [7]; it uses only deposited values and can therefore be applied to any series matrix.

The separation between the two fitted component means was recorded for every dataset as a check on the fit, since a boundary taken from a distribution that is not bimodal carries no information.

Expression-matched control

For each olfactory receptor gene we drew one non-olfactory gene at random from those whose mean log2 signal lay within 0.05 of it, giving a control set of the same size and the same signal distribution. We verified that a partner was available for every receptor gene in every dataset analysed in detail. The number of control-set members among the top 100 was counted, and the procedure repeated 200 times for the null splits and 2,000 times for the worked examples. A one-sided P value was computed as (k + 1)/(N + 1), where k is the number of control sets reaching at least the observed receptor count.

Two further properties of the receptor set were compared with matched sets in the same way. The first was the standard deviation of the per-gene log2 fold change across ten random balanced splits, as a measure of how variable a gene is independently of its mean.

The second was the proportion of variance explained by the first principal component of the per-gene centred expression matrix, as a measure of whether the set moves coherently across the arrays of a dataset. For each dataset the observed value was compared with 100 expression-matched sets.

Worked examples

Two datasets with unambiguous two-group designs were analysed with their true group labels: GSE69911 (H9 embryonic stem cells, non-silencing siRNA n = 3 versus TCF7L1 siRNA n = 3) and GSE106234 (Huh-7.5 hepatoma cells, mock n = 2 versus calcitriol n = 2). Gene-wise statistics used moderated t-statistics with empirical Bayes variance shrinkage [8] and Benjamini–Hochberg control of the false discovery rate [9]. Neither study reports any claim about olfactory receptors; they are used here only because their rankings place receptors near the top.

Software

Analyses used Python 3.13 with pandas, NumPy, SciPy, statsmodels, scikit-learn and Matplotlib. Scripts and the list of accessions are archived at Zenodo, doi:10.5281/zenodo.24080022.

Results

Olfactory receptors are undetected in essentially every dataset

Across the 99 datasets the mixture placed a background component below an expressed component, the background component accounting for a median of 50% of genes (range 11–80%). The two component means were separated by a median of 2.74 log2 units, but by less than 0.1 in two datasets (GSE65632, GSE119577), where the signal distribution is not bimodal and the fitted boundary is not interpretable. Olfactory receptors fell below the boundary in a median of 99.7% of instances per dataset (interquartile range 99.4–100%), and in at least 98.6% in 88 of the 99; a median of 360 receptor genes was measured per dataset (range 313–360). The premise of the analysis therefore holds in the great majority of datasets: these are cultured human cells and tissues in which the family is not expressed, and any apparent regulation of it is a statement about measurements near the noise floor.

Under a null, one analysis in eight places four or more receptors at the top of the ranking

In the 990 random balanced splits, no true group difference exists by construction. The mean number of olfactory receptors among the 100 genes with the largest increase was 1.54, slightly below the 1.89 expected from the family's genome-wide share. The mean is not the relevant quantity, however, because the distribution is strongly right-skewed (Fig. 1A). Four or more receptors — at least twice the expected share — appeared in 13.3% of analyses, six or more in 8.2% and nine or more in 3.3% (Table 1). In 59.3% of analyses none appeared at all.

An analyst who finds nine olfactory receptors among their top 100 genes is therefore looking at something that happens in roughly one null analysis in thirty, and four or more at something that happens in one in eight.

Table 1. Olfactory receptors among the top 100 genes by log2 fold change in 990 null analyses

Receptors in the top 100	Percentage of analyses	Cumulative at or above
0	59.3%	100%
1	16.4%	40.7%
2	5.5%	24.3%
3	5.6%	18.8%
4	3.1%	13.3%
5	2.0%	10.2%
6	2.0%	8.2%
7	1.1%	6.2%
8	1.7%	5.1%
9 or more	3.3%	3.3%

Expected from the family's genome-wide share: 1.89 genes.

These rates do not depend on the datasets in which the mixture separated poorly. Excluding the two datasets whose component means differed by less than 0.1 log2 units leaves 970 splits with a mean of 1.57 receptors in the top 100 and 13.6% reaching four or more; excluding the four separated by less than 1 log2 unit leaves 950 splits, a mean of 1.55 and 13.7%. The association with the background fraction of the ranking strengthens slightly under both exclusions (Spearman $\rho = 0.52$ and 0.53).

The excess depends on how far the ranking is dominated by undetected genes

The fraction of the top 100 made up of background-level genes varied widely across analyses, and it predicted the receptor count (Spearman ρ = 0.49; Fig. 1B). Where background-level genes made up at least 70% of the top 100 (144 analyses), the mean receptor count was 5.58 and 61% of analyses reached four or more. Where they made up less than 30% (322 analyses), the mean was 0.37 and 1% reached four or more.

This gives an analyst a diagnostic that requires no extra computation. If the genes at the top of a fold-change ranking are mostly undetected, the ranking is reporting noise, and an enrichment computed from it will reflect whichever families happen to be absent in that tissue.

What decides whether undetected genes reach the top

Whether background-level genes reach the top of a ranking is a property of the deposited values rather than of the tissue. For each dataset we drew ten further random balanced splits and regressed the log standard deviation of the per-gene log2 fold change on mean log2 signal. In 24 of the 99 datasets the slope was negative, meaning the fold-change noise was largest among the genes with the least signal; in the remaining 75 it was flat or positive. The two groups behave differently. With a negative slope the mean receptor count in the top 100 was 3.09 and 35.4% of splits reached four or more; elsewhere the mean was 0.96 and 7.2% reached four or more (Spearman ρ between slope and receptor count = −0.43). Ranking the datasets instead by the ratio of fold-change standard deviation in background-level genes to that in expressed genes gives the same picture (Table 3). The overall rate of 13.3% is the mixture of these two populations: 24/99 × 35.4% + 75/99 × 7.2% ≈ 14%.

The proportion of the dataset that is background-level does not predict the receptor count (Spearman ρ = 0.09). What matters is not how many undetected genes a dataset contains but whether their fold changes are large enough to reach the top of the ranking, and that is set by how the deposited values were normalised and summarised. This is why the effect appears in a minority of datasets rather than in all of them, and it gives a second diagnostic: plot the per-gene fold-change spread against mean signal, and treat a ranking from a dataset whose spread rises as signal falls as uninformative at its extremes.

Table 3. Olfactory receptors in the top 100 by the mean–variance behaviour of the dataset. Datasets are in quartiles of the ratio of fold-change standard deviation among background-level genes to that among expressed genes, from ten random balanced splits per dataset.

Expression-matched sets reproduce the receptor counts

Sets of non-olfactory genes matched on mean signal to within 0.05 log2 placed a mean of 1.15 genes in the same top 100, close to the 1.54 observed for receptors, and their distribution tracked the receptor distribution across the whole range (Fig. 1A). Taking the matched distribution as the null, the test rejected at the 5% level in 7.7% of the 990 analyses. The control is therefore approximately calibrated but mildly anti-conservative, which is expected because matching is on mean signal alone and does not equalise probe-level variance or genomic position.

Among the analyses that reached four or more receptors, 41.7% still rejected at $P < 0.05$ against matched sets. Matching therefore removes most but not all of the excess, and a matched P value below 0.05 should be read as a reason to measure the transcripts directly rather than as evidence of regulation.

Two datasets with real contrasts

Two datasets placed receptors far above their expected share when analysed with their true group labels (Fig. 2). In GSE69911, comparing non-silencing with TCF7L1 siRNA in H9 embryonic stem cells, nine receptors appeared among the top 100 against 1.9 expected, a 4.8-fold enrichment, with 78% of the top 100 at background level. In GSE106234, comparing mock with calcitriol in Huh-7.5 hepatoma cells, eight appeared against 1.9 expected, a 4.2-fold enrichment, with 98% of the top 100 at background level.

In both, expression-matched sets reproduced the counts almost exactly: a matched mean of 7.26 against nine observed in GSE69911 ($P = 0.30$) and 7.64 against eight observed in GSE106234 ($P = 0.50$). The apparent enrichment in these experiments is fully accounted for by the signal level of the family. Neither study makes any claim about olfactory receptors, and we draw none; the datasets serve to show that the pattern appears in real analyses and that the matched control resolves it.

Table 2. Summary of the benchmark

Quantity	Value
Datasets	99
Null analyses (random balanced splits)	990
Genes per dataset after collapsing to symbols	17,795–19,036

Quantity	Value
Olfactory receptor genes per dataset	313–360 (median 360)
Olfactory receptor share of the transcriptome	1.89%
Receptors below the detection boundary	Median 99.7% per dataset (IQR 99.4–100%)
Mean receptors in the top 100, observed	1.54
Mean receptors in the top 100, expression-matched sets	1.15
Null analyses with 4 or more receptors in the top 100	13.3%
Null analyses with 9 or more	3.3%
Spearman ρ, receptor count against background fraction of the top 100	0.49
Matched test rejection rate at nominal 5%	7.7%

Discussion

In 990 analyses of real human transcriptome data in which no group difference existed, olfactory receptors reached at least twice their genome-wide share at the top of the fold-change ranking in 13.3% of cases, and at least nine of the top 100 genes in 3.3%. The family is undetected in essentially all of these datasets, so none of this reflects regulation. It follows that finding a handful of olfactory receptors among the most changed genes, or a Gene Ontology result for sensory perception of smell from a ranked list, is by itself uninformative about whether anything happened to the cells.

The behaviour is not uniform, and that is useful. It concentrates in analyses whose rankings are dominated by undetected genes: where background-level genes made up at least 70% of the top 100, three analyses in five reached four or more receptors, against one in a hundred where they made up less than 30%. The fraction of the top 100 that is undetected is computable from any deposited expression matrix and takes seconds. We suggest it as a routine diagnostic, reported alongside any enrichment derived from a ranked list.

The expression-matched control is the corresponding test. Matched sets reproduced the receptor counts closely and rejected at 7.7% against a nominal 5%, so the procedure is close to calibrated. The residual can be located. Across the 99 datasets the fold-change standard deviation of receptor genes exceeded that of expression-matched non-

olfactory genes by 3.9% on average, significantly in 46 of the 99 against the 5 expected by chance. Matching on mean signal therefore leaves a small excess in variability, and that excess is enough to make the test mildly anti-conservative. Coherent movement of the family, which the clustering of these genes in the genome and the sequence similarity between paralogues would produce, is not the explanation: the variance explained by the first principal component of the receptor set exceeded that of matched sets in only 44 of the 99 datasets, so the receptors do not move together more than genes of the same signal level. Among null analyses that reached four or more receptors, two in five still rejected at $P < 0.05$. A matched P value below 0.05 is therefore a reason to measure the transcripts directly, not a licence to interpret a receptor family result as biology.

The two worked examples show both halves of this in a real setting. In H9 embryonic stem cells and in Huh-7.5 hepatoma cells, rankings placed receptors at 4.8-fold and 4.2-fold their expected share, which read on its own would look like a coordinated effect. The matched control returned $P = 0.30$ and 0.50, attributing both to signal level. Neither of the original studies examined olfactory receptors, so no interpretation was at stake in either case; what the examples establish is that the pattern occurs in ordinary experiments and that the control recovers the right answer.

We have used olfactory receptors because they are the clearest case: a large, well-annotated family with a known tissue restriction, so the truth is known in advance. The mechanism is not specific to them. Any family that is uniformly undetected in the tissue under study — taste receptors, keratin-associated proteins, testis-restricted genes, many G-protein-coupled receptors — will behave the same way, and the same diagnostic and control apply. What makes olfactory receptors worth singling out is that the biological claim they invite, receptor-mediated signalling in a non-olfactory tissue, is one that readers may find plausible.

Recommended practice

1. Report the mean signal of every gene named in an enrichment result, alongside its fold change. A reader can then see which calls rest on detected transcripts.
2. Report the fraction of any ranked gene list that falls below a detection boundary. Above roughly 70%, treat family-level enrichment from that list as uninterpretable without further evidence.
3. Test family-level enrichment against genes matched on expression level, not against the whole transcriptome. Report the matched null distribution, not only a P value.
4. Treat a surviving matched result as a hypothesis for direct measurement — quantitative RT-PCR with cycle thresholds, melt curves, no-reverse-transcription controls and efficiency reported — rather than as a finding.

Limitations

- The benchmark uses one array platform. The quantitative rates will differ on platforms whose deposited values are background-subtracted and floored, where undetected probes carry little variance and the effect is suppressed; we observed this informally on Illumina BeadChip data but did not benchmark it.
- The detection boundary is derived from the distribution of deposited values rather than from probe-level detection-above-background, which requires raw files. Datasets whose depositors applied a detection filter before deposit will show a smaller background component and a correspondingly smaller effect.
- The two-component mixture assumes the signal distribution is bimodal. In two of the 99 datasets the fitted component means differed by less than 0.1 log2 units, so no boundary could be placed and the proportion of receptors called undetected is meaningless there; in two further datasets the separation was below 1 log2 unit. The gap between component means should be reported alongside any boundary derived this way, and a boundary from a fit that does not separate should not be used.
- Random balanced splits remove systematic group differences but not batch structure within a dataset. Some splits will coincide with a processing batch, which is part of what an analyst faces in practice but means the null is not purely technical.
- The two worked examples were selected because they showed the pattern. They illustrate the control; they are not an unbiased estimate of how often real contrasts produce it.
- We did not evaluate RNA sequencing, where undetected transcripts give zero or near-zero counts and the relevant failure mode differs.

Declarations

Data availability: All data are public in GEO. The 99 accessions analysed, and the two examined in detail (GSE69911, GSE106234), are listed in Supplementary Table S1. Analysis scripts are archived at Zenodo, doi:10.5281/zenodo.24080022. No new experiments were performed.

Author contributions: K.O. is the sole author. K.O. conceived the study, assembled the datasets, performed all analyses, and wrote the manuscript.

Use of AI tools: Data retrieval, statistical scripts and a first draft of the text were prepared with assistance from an AI system (Claude, Anthropic). The authors checked all analyses and take full responsibility for the content. [Keep, edit or remove per author decision and journal policy.]

Funding: This research received no external funding.

Competing interests: The authors declare no competing interests. [confirm]

Figure legends

Figure 1. Olfactory receptors at the top of fold-change rankings under a null. (A) Distribution of the number of olfactory receptor genes among the 100 genes with the largest increase, across 990 random balanced splits of 99 human datasets (red), against expression-matched sets of non-olfactory genes drawn from the same datasets (grey). The dotted line marks the 1.89 genes expected from the family's genome-wide share. (B) Receptor count against the percentage of the same top 100 that falls below the detection boundary; each point is one split, jittered vertically, with the mean per 10% bin in red and the expected share dotted.

Figure 2. Two datasets with real experimental contrasts. (A, C) Distribution of mean log2 signal for all genes (grey) with the fitted two-component mixture (dashed) and olfactory receptors overlaid (red); the vertical line is the detection boundary. (B, D) Number of genes among the top 100 by log2 fold change for the olfactory receptor set (red line) against 2,000 expression-matched sets of non-olfactory genes (histogram), with the genome-wide share dotted. (A, B) GSE69911, H9 embryonic stem cells, non-silencing versus TCF7L1 siRNA, 3 versus 3. (C, D) GSE106234, Huh-7.5 hepatoma cells, mock versus calcitriol, 2 versus 2.

References

1. Chess A, Simon I, Cedar H, Axel R. Allelic inactivation regulates olfactory receptor gene expression. *Cell.* 1994;78:823–834.
2. Magklara A, Yen A, Colquitt BM, et al. An epigenetic signature for monoallelic olfactory receptor expression. *Cell.* 2011;145:555–570.
3. Jung H, Chae YC, Kim JY, Jeong OS, Kook H, Seo SB. Regulatory role of G9a and LSD1 in the transcription of olfactory receptors during leukaemia cell differentiation. *Sci Rep.* 2017;7:46182.
4. Flegel C, Manteniotis S, Osthold S, Hatt H, Gisselmann G. Expression profile of ectopic olfactory receptors determined by deep sequencing. *PLoS One.* 2013;8:e55368.
5. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 2010;11:R14.
6. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 2003;4:249–264.
7. Lockstone HE. Exon array data analysis using Affymetrix power tools and R statistical software. *Brief Bioinform.* 2011;12:634–644.

8. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 2004;3:Article 3.
9. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 1995;57:289–300.

Figures

Figure 1

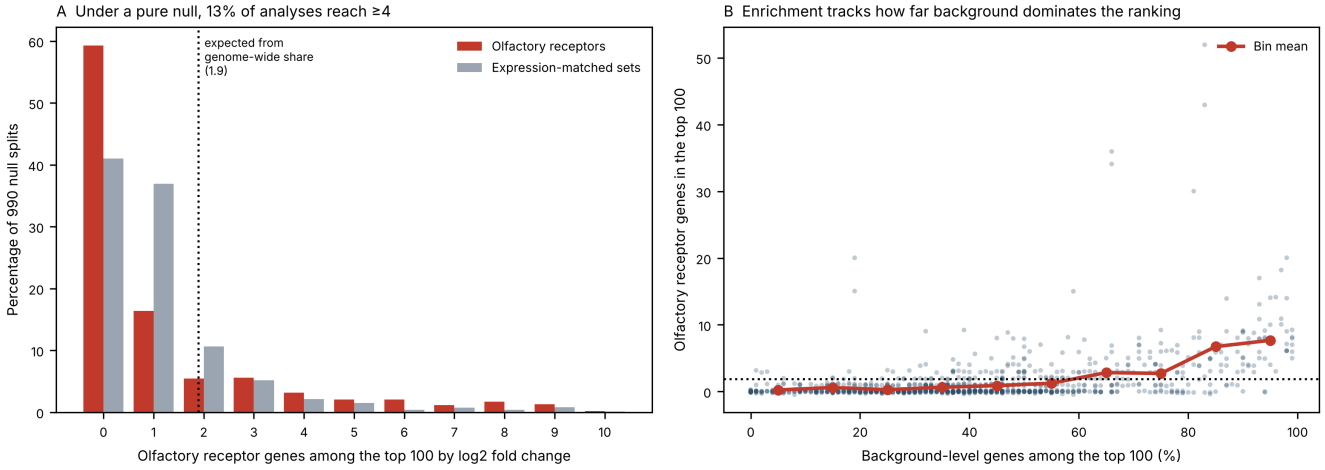


Figure 2

