https://doi.org/10.1038/s41587-026-03379-8

A unified and practical taxonomy for false discovery rate control in omics research

Thomas Burger, Xinzhou Ge & Jingyi Jessica Li

Check for updates

False discovery rate (FDR) control is central to omics research, yet the diverse procedures used rely on different assumptions. This Comment proposes a 2 × 2 matrix that helps sequencing-based and mass-spectrometry-based omics researchers choose the appropriate family of FDR procedures, or alternative methods that may be more rigorous, depending on data and study design.

The FDR, introduced three decades ago by Benjamini and Hochberg[1], is the expected proportion of false discoveries among findings labeled as significant. FDR control has become essential in modern omics research, particularly for high-throughput screens that test thousands of biomolecules simultaneously. Its importance is particularly evident in transcriptomic data analysis, where researchers identify differentially expressed genes (DEGs) between conditions. Transcriptomic data typically consist of N biological samples collected under more than one condition, such as from cancer patients receiving a specific treatment versus those who are not. A goal of DEG analysis is to identify genes whose expression levels differ significantly between conditions, as these may serve as potential biomarkers for further study. This approach involves screening thousands of genes to identify a small subset of promising candidates for experimental validation. Similar goals arise, for instance, when searching for biomarker candidates, such as plasma-circulating proteins, using mass spectrometry. For all of these screening experiments, given the high financial and logistical costs of downstream validation, it is crucial to control false discoveries among the retained features (including genes and proteins). This is precisely the goal of FDR control.

In practice, today's omics workflows employ a variety of seemingly unrelated FDR control procedures, ranging from traditional multiple-testing adjustments to proteomics-inspired target–decoy competitions and, additionally, Bayesian approaches, which are not covered here (Supplementary Information). This diversity, combined with increasingly complex experimental designs, can be confusing for omics researchers and bioinformatics software developers seeking appropriate FDR control procedures.

Here, we propose a 2 × 2 matrix that organizes existing FDR control procedures along two orthogonal axes. We refer to features that do not carry true biological signals, such as non-DEGs or random spectrum–biomolecule pairs, as null features, meaning features whose measurements follow the null hypothesis. The horizontal axis shows the theoretical versus numerical null representation for these features, whereas the vertical axis shows whether the null reference is shared across features or defined separately for individual features (feature-pooled versus feature-specific). Additionally, we map the FDR control procedures commonly used for major omics tasks within this matrix, clarifying which family each procedure belongs to. Finally, we provide decision rules to help computational biologists and bioinformatics tool developers choose, implement and sanity-check FDR control in their own workflows.

We do not aim to comprehensively catalog every FDR control procedure developed across omics and statistics; rather, our primary goal is to bridge perspectives from sequencing-based and mass-spectrometry-based omics, which have historically developed different FDR control conventions, and to organize the procedures that they commonly rely on into a shared, practical taxonomy.

FDR control in sequencing-based and mass-spectrometry-based omics

Achieving rigorous FDR control in high-throughput screening requires statistical theory. In sequencing-based transcriptomics or microbiome studies, where thousands of genes or taxa (that is, features) are tested for differential abundance across conditions, one common FDR-control workflow uses feature-level P values followed by the Benjamini–Hochberg procedure[1]. A P value can be calibrated against either a theoretical or a numerical null; in standard sequencing-based applications, it is commonly obtained from a theoretical null, corresponding to our traditional family. The Benjamini–Hochberg procedure then uses all feature-level P values to choose a threshold that controls the FDR at a prespecified target level (for example, 5%). The core principle is that P values for null features should be valid and, ideally, uniformly distributed between 0 and 1, whereas P values for truly differential features should be enriched near 0 (Fig. 1). Under these assumptions, one can estimate the expected proportion of null features among discoveries (that is, the FDR) from the empirical P-value distribution. A more detailed description of traditional FDR control procedures is provided in the Supplementary Information.

The breakthrough of the Benjamini–Hochberg procedure has yielded many popular refinements. Notably, the q-value procedure, which estimates the proportion of null features from the data and uses this estimate to make the Benjamini–Hochberg threshold less conservative, is common in genomics software (Table 1). Likewise, the Benjamini–Hochberg procedure was originally justified assuming that features, such as genes, behave largely independently of one another. In practice, biological features are often correlated (for example, co-regulated genes), which can affect the validity of FDR control and has motivated statistical refinements[2].

Mass spectrometry (MS)-based omics methods, such as proteomics, peptidomics, metabolomics and lipidomics, face analogous needs for FDR control when assigning large numbers of experimental spectra to candidate biomolecules. A false discovery corresponds

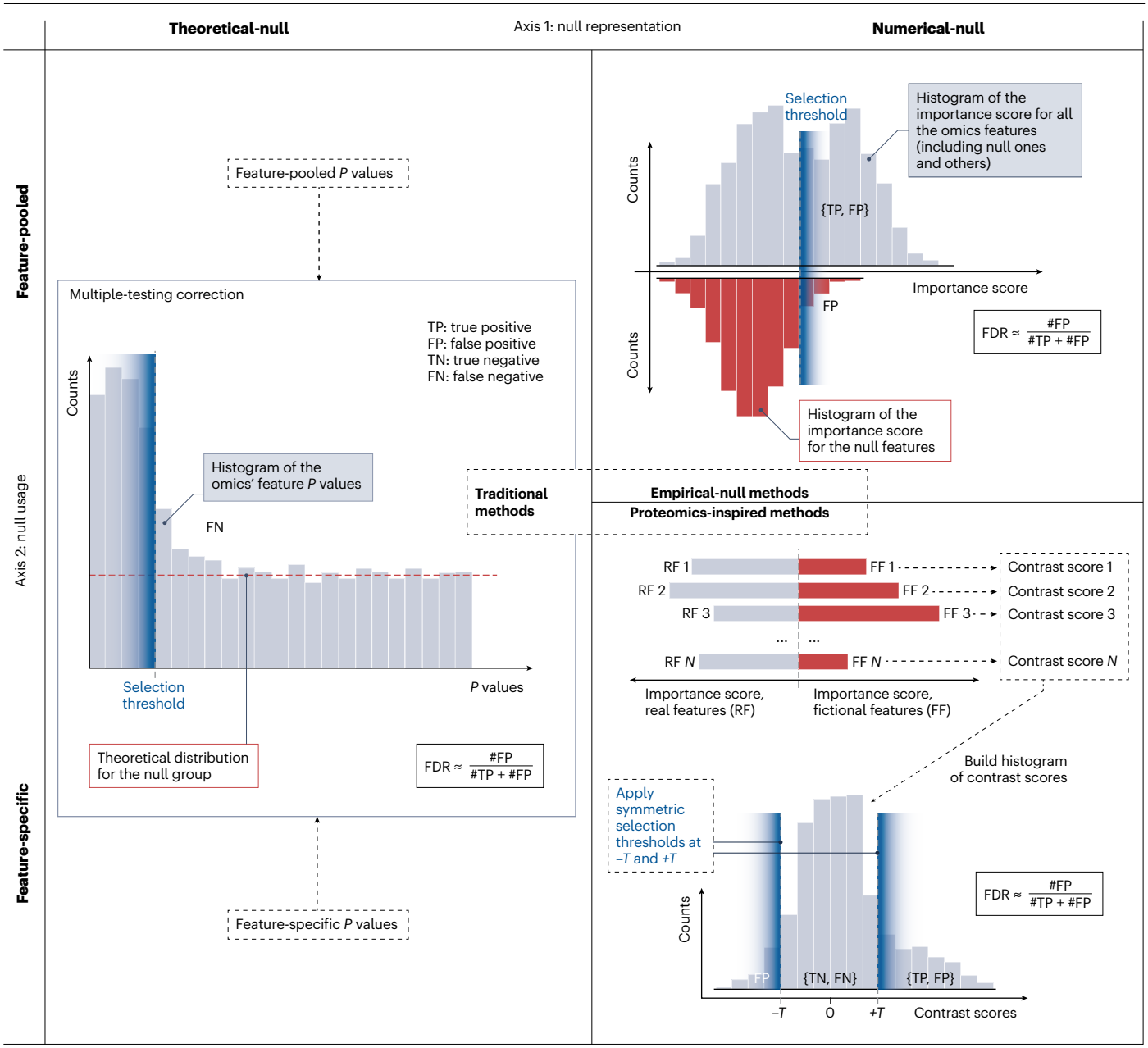


Fig. 1 | Schematic representation of the proposed taxonomy. The taxonomy classifies procedures according to how the null reference used to calibrate feature-level evidence is represented: theoretical versus numerical and feature-pooled versus feature-specific. This classification yields three families: traditional methods (left), empirical-null methods (top right) and proteomics-inspired methods (bottom right). Each panel illustrates one example of how the null representation can be converted into feature-level evidence and subsequently into an FDR-controlled discovery set. These downstream steps are not necessarily unique to the corresponding family. The traditional methods panel illustrates theoretically calibrated *P* values followed by a *P*-value-based FDR control procedure such as the Benjamini–Hochberg procedure. The empirical-null panel illustrates direct estimation of false discoveries using a pooled numerical null, whereas the proteomics-inspired panel illustrates feature-specific real-versus-fictional comparisons summarized by contrast scores and a competition-based threshold. For empirical-null methods, the pooled numerical null may alternatively be used to construct empirical *P* values that are subsequently supplied to a *P*-value-based FDR-control procedure. Thus, the form of the feature-level evidence or downstream thresholding rule does not by itself determine the family. Measured features are shown in gray and fictional null features in red; blue shading indicates which values of the feature-level evidence are deemed significant. For simplicity, the schematic shows false discovery rate (FDR) ≈ #FP/(#TP + #FP), where #FP and #TP denote number of false and true positives, respectively, to illustrate the intuition behind FDR control. Strictly, this ratio is the realized false discovery proportion, whereas the FDR is its expectation over repeated realizations of the data and procedure. Step-by-step descriptions of these families of methods are provided in the Supplementary Information.

Table 1 | The three steps of FDR control procedures in the three families of methods

	Traditional (theoretical-null, feature-pooled or feature-specific)	Empirical-null (numerical-null, feature-pooled)	Proteomics-inspired (numerical-null, feature-specific)
Step 1. Null representation (defines the family)	Theoretical null distribution, either shared across features or feature-specific.	Numerical null distribution pooled across features (obtained, for example, from permutations, negative controls, model-based simulations or empirical-null estimation).	Numerical null counterpart constructed separately for each feature (for example, a decoy sequence, a knockoff feature or a permutation-generated counterpart).
Step 2. Feature-level evidence (typical, not defining)	A feature-level statistic is calibrated against a theoretically derived null distribution. This step typically yields a *P* value.	A pooled numerical null is used to calibrate feature statistics. This step may yield empirical *P* values.	Each real feature is compared with a feature-specific fictional counterpart. The comparison produces feature-level evidence such as a contrast or competition score.
Step 3. Discovery-set threshold (typical, not defining)	Often the Benjamini–Hochberg procedure, *q* value or another *P*-value-based FDR procedure.	Empirical *P* values may be passed to the Benjamini–Hochberg or another *P*-value-based procedure. Alternatively, the pooled numerical null may be used directly to estimate the expected number of false discoveries above a cutoff (rescaled by the number of null statistics), and hence the FDR.	Typically, a competition-based thresholding rule that aggregates real-versus-fictional comparisons across features.

Every procedure can be divided into three steps. Only step 1 (null representation) determines the family in our taxonomy. Steps 2 and 3 are shown because families are often conflated at these stages. In particular, *P* values can be obtained from either a theoretical or a numerical null, therefore their use does not by itself make a procedure traditional. A *P*-value-based rule such as the Benjamini–Hochberg procedure can be applied downstream of either null representation. Representative examples are given in Table 2.

to a random match between a spectrum measured in an experiment and a biomolecule in a database. Because the quality of such matches is difficult to calibrate using standard statistical tests, the standard theoretically calibrated *P*-value-based framework does not easily apply. Instead, proteomics researchers developed the target–decoy competition (TDC) strategy[3], which has since been generalized to other MS-based omics methods[4]. In TDC, each potential peptide sequence (target) is paired with a fictional sequence that cannot exist in the sample (decoy), which is generated by reversing or shuffling the amino acids of the sequence. Each experimental spectrum is scored against both targets and decoys, and only the higher score is retained. Under the assumption that incorrect target and decoy matches are equally likely, the number of decoy identifications above a score cutoff estimates the expected number of incorrect target identifications above the same cutoff, thus leading to an FDR estimate (Fig. 1). A more detailed description of competition-based FDR control procedures is provided in the Supplementary Information.

Since the introduction of TDC, FDR control in MS-based omics has evolved along two main lines. On the analytical side, MS workflows have become increasingly quantitative. In addition to biomolecule identification, researchers now perform differential abundance analyses of biomolecules across conditions, which are typically handled with multiple-testing procedures similar to those used in sequencing-based transcriptomics and microbiome studies. On the theoretical side, the statistical foundation of TDC has been clarified. Nearly a decade after its introduction, independent theoretical work established that competition between real and fictional features can validly control the FDR under appropriate assumptions[5]. Such competition-based procedures apply beyond biomolecule identification and are also valid for differential abundance analysis, where it is no longer necessary to compute a *P* value for each feature[6]. Any suitably constructed 'importance score' measuring the feature's interest is sufficient to enable a competition-based FDR control (Supplementary Information).

Together, these developments have led to the coexistence of several distinct FDR control procedures, each with its own assumptions, and some applicable beyond their original domain.

A practical taxonomy of FDR control procedures

We organize existing FDR control procedures along two axes (Fig. 1). The horizontal axis shows how the null reference for each feature is represented: via either a theoretical null distribution or a numerical null, as detailed below. The vertical axis shows whether a common null is shared across features (feature-pooled) or whether a separate null is constructed specifically for each feature (feature-specific).

The horizontal axis distinguishes FDR control procedures according to how features are represented under the null hypothesis. Theoretical-null procedures calibrate a feature-level statistic that quantifies its evidence against the null using a theoretically derived null distribution. For each feature, they evaluate its statistic against a null distribution based on assumptions about the data-generating mechanism of the feature measurements, such as a normal or negative-binomial distribution. We consider a null distribution theoretical if it is obtained, analytically or by exact computation, from an assumed model of the data-generating mechanism, irrespective of whether its parameters are exactly known (such as the degrees of freedom of a *t* or χ^2 distribution) or estimated from the data (such as a gene-specific dispersion parameter in a negative-binomial model). We do not distinguish between these two cases in our taxonomy. Numerical-null procedures, in contrast, construct a numerical representation of the null distribution of the feature-level statistic from generated, resampled or control data, or estimate a null distribution from observed statistics. In our taxonomy, we use theoretical-null procedures to define what we term the 'traditional family' of FDR procedures and numerical-null procedures to define the 'empirical-null' and 'proteomics-inspired' families.

The vertical axis distinguishes numerical-null procedures by whether the null representation is shared by all features (feature-pooled) or is specific to each feature (feature-specific; Fig. 1).

Table 2 | Comparison of the three families of FDR control procedures

	Traditional	Empirical-null	Proteomics-inspired
Typical use cases	Differential analyses under standard experimental designs	Theoretical-null is hard to trust, but pooling features to construct a null distribution is plausible	Biomolecule identification Differential and association analyses for which no pooled null is plausible
Main strengths	Simple Fast Familiar Reproducible	Flexible calibration	Avoids pooled-null assumption Does not require a statistical test
Main risks and checks	Is the theoretical null appropriate for the data (are the *P* values valid)?	Are null statistics comparable across features?	Are real nulls and fictional counterparts exchangeable?
Example methods	Wilcoxon + Benjamini–Hochberg DESeq2 + Benjamini–Hochberg	Significance Analysis of Microarrays Gene Set Enrichment Analysis Nullstrap-DE	Target–decoy competition (TDC) Knockoff filters Clipper ClipperQTL

Empirical-null procedures pool statistics computed under the null into a common, feature-pooled null distribution shared across features[7,8]. In contrast, proteomics-inspired procedures generate one decoy feature for each target feature, providing a matched, feature-specific null counterpart. The proportion of null statistics exceeding a given cutoff can then be used to estimate the number of false discoveries in the original data above the same cutoff. A numerical null can be obtained in several ways: from negative control assays that contrast two groups of samples with no biological difference (experimental null), by randomly permuting sample labels across groups (permutation null)[8], by generating data from a fitted model that reflects the null hypothesis (model-based null), by randomly permuting sequences to generate decoy sequences (decoy null)[3,9] or by estimating the null distribution directly from the central bulk of observed statistics (empirical-null estimation)[7]. Because these approaches vary widely, our taxonomy focuses on how the resulting null is used—feature-specific versus feature-pooled—rather than on how the null is obtained.

The distinction between feature-specific and feature-pooled null distributions can also arise among theoretical-null procedures. For example, in DEG analysis, when the Wilcoxon rank-sum test is applied to each gene[10], the same theoretical null distribution is used across genes, corresponding to a feature-pooled null. In contrast, under a negative-binomial model with gene-specific dispersion parameters, the theoretical null distribution depends on the gene and is therefore feature-specific. In these examples, each feature-level statistic can be calibrated against its corresponding theoretical null distribution to obtain a *P* value, and this is followed by a multiple-testing correction, such as the Benjamini–Hochberg procedure. Because of the use of theoretical null distributions, we group these procedures within the traditional family without further distinguishing between them in our taxonomy.

Of note, our two axes describe how the null reference is represented for each feature to calibrate feature-level evidence, not the form of the resulting evidence or the downstream FDR control rule. A *P* value, the most used measure of evidence, can be obtained from either a theoretical or numerical null distribution; therefore, its use alone does not define a procedure as traditional. For example, if a feature-pooled numerical null is used to construct empirical *P* values and the Benjamini–Hochberg procedure is subsequently applied, we classify the procedure as empirical-null. Similarly, feature-specific numerical-null procedures ('proteomics-inspired' family) may summarize real-versus-fictional competition using contrast scores or other calibrated evidence. Regardless of how this evidence is represented (*P* values, contrast scores or other statistics), determining the final discovery set with FDR control requires aggregating information across features, because FDR is a property of the discovery set rather than of any individual feature. This may involve, for example, ranking *P* values across all features in traditional procedures and some empirical-null procedures, or considering all contrast scores in proteomics-inspired procedures. Some proteomics-inspired procedures further refine this aggregation step by pooling thresholding information only across similar features, but this reflects a refinement of the thresholding step rather than a change in how the null itself is constructed.

Guidelines for choosing FDR control procedures

Building on this taxonomy, we provide practical guidelines to help researchers choose appropriate FDR control procedures for common omics analyses. Table 1 breaks down the procedures of the three families into three steps and shows that only the first step, the null representation, determines the family. Table 2 summarizes practical guidance and representative examples, with references provided in the Supplementary Information.

For a default differential analysis, including searching for DEGs or differentially abundant molecules, traditional procedures are preferred (theoretical-null, either feature-specific or feature-pooled). For standard settings, such as contrasting conditions in a simple experimental design, a ready-to-use statistical test or generalized linear model is usually available, and procedures from the traditional family are computationally efficient. Moreover, given fixed software and data, they are reproducible, as the decision rule itself involves no randomness. In practice, this amounts to applying an appropriate test for calculating *P* values followed by a multiple-testing correction[1]. However, when statistical test assumptions are violated, *P* values can be invalid and FDR control can fail[10]; researchers should therefore check the assumptions on their data and use diagnostic visualizations to assess *P*-value validity (for example, whether the *P*-value histogram shows a modest enrichment near 0 on top of an approximately uniform background for null features; Supplementary Information).

If a test exists but its assumptions cannot be verified or are violated, empirical-null procedures (numerical-null, feature-pooled) should be used. When biases in data collection make it difficult to justify the assumed theoretical null (for example, when there are unmodeled confounders or technical artifacts that shift or overdisperse the null

distribution of test statistics), traditional procedures may no longer be trustworthy. Empirical-null approaches address this by generating or collecting null data (for example, via permutations or negative controls) or by estimating the null distribution directly from the central bulk of observed statistics[7]. In practice, several software tools implement empirical-null estimation internally. For customized implementation, because empirical-null procedures rely on a feature-pooled numerical null, researchers should check whether feature-level statistics have been appropriately standardized and whether it is reasonable to assume that null statistics from different features can be treated as coming from a single common distribution. In addition, randomness in empirical-null construction (for example, the finite number and random seed of permutations) can introduce modest variability in discoveries across repeated runs, so we recommend documenting these choices for reproducibility.

When experimental designs or analysis pipelines are too complex, or data too heterogeneous or unconventional, it can be unreasonable to assume that all null features follow the same pooled numerical-null distribution. In such settings, proteomics-inspired procedures (numerical-null, feature-specific) are a more appropriate option: they construct a competition between each real feature and its feature-specific fictional counterpart[6]. As with empirical-null procedures, careful construction and scrutiny of the null are essential. In particular, a key check is whether statistics for real null features are distributed similarly to those of their fictional counterparts, so that a true null feature and its fictional counterpart have exchangeable statistics. Notably, randomness in these pairwise competitions—driven, for example, by how fictional counterparts are randomly generated—can lead to greater variability in the set of discoveries across repeated runs than arises from empirical-null procedures that rely on a feature-pooled numerical null.

For biomolecule identification, proteomics-inspired procedures (numerical-null, feature-specific) are the most natural choice, especially for proteomics and peptidomics, because decoy generation and the TDC strategy are built into most peptide search engines. However, this apparent simplicity can hide two important difficulties.

First, the complexity of modern experimental workflows can create discrepancies between the idealized competition framework and its practical TDC implementation, leading to compromised FDR control. Such issues have been documented not only for post-processed identifications[11] but also for multistep identification pipelines (such as cascaded or multi-database searches in proteogenomics and metaproteomics), for cases in which many spectra map to the same peptide and for workflows sensitive to instrument tuning[12]. More specialized approaches, such as wide-tolerance ('open') searches or post-translational modification studies, may also be challenging. Problematic workflows can often be detected using entrapment experiments, whose theoretical foundations and limitations have recently been clarified[13].

Second, if spectrum–decoy pairs do not accurately represent random spectrum–molecule pairs, the resulting FDR estimates can be systematically biased. It is essential to check that decoys are constructed such that their matches are distributed similarly to incorrect target matches: that is, incorrect target matches and decoy matches should have approximately exchangeable statistics under the null, as in the proteomics-inspired procedures discussed in differential analysis. Tools exist to assess this property, but such checks are rare in published proteomics analyses. They are even more critical in peptidomics, where sequence combinatorics are greater.

Decoy generation and FDR control remain active research areas in metabolomics and lipidomics because their relevant chemical spaces lack the linear structure of amino acid sequences, and simple sequence shuffling cannot be directly transposed. Nevertheless, some target–decoy and related strategies are now mature enough to be used in practice[4]. The challenges highlighted above for proteomics and peptidomics—ensuring that decoys emulate random matches and that workflows remain compatible with the underlying FDR theory—remain largely unresolved in these domains and are arguably even more acute.

In practice, if it is suspected that TDC-based FDR control lacks rigor, two alternative strategies are available, at least for proteomics and peptidomics data. First, it is possible to use a search engine compatible with traditional FDR control (theoretical-null). A small number of tools produce peptide-to-spectrum match scores that can be converted to *P* values through a probabilistic model computed for each spectrum, without any decoy search, and then subjected to multiple-testing correction such as the Benjamini–Hochberg procedure[12]. Alternatively, it is possible to resort to target–decoy without competition. Target–decoy analyses that omit head-to-head competition between each target and its decoy use decoys to construct a pooled numerical null across features[9]. Formally, these procedures belong to the empirical-null family (numerical-null, feature-pooled): decoys are no longer used in a feature-specific manner through pairwise competition, but instead contribute to a shared null distribution. Sensitivity to decoy generation remains, but the absence of pairwise competition moderates its impact, because the influence of a few problematic decoys is diluted within the overall feature-pooled null.

Genetic association and enrichment analyses, such as genome-wide association studies (GWAS), quantitative trait locus (QTL) mapping and functional enrichment of discovery lists, also fit naturally into our taxonomy. Differential analysis is itself an association analysis, which relates each feature's abundance to a biological condition. GWAS and QTL mapping instead relate a trait or a feature's abundance to genetic variants, and enrichment analysis relates a discovery list to predefined functional categories. In these settings, statistical tests are typically applied feature-wise to produce *P* values, and traditional multiple-testing procedures (theoretical-null) are the default choice for FDR control. As in DEG analysis, however, violated model assumptions, approximations in test statistics and data-dependent preprocessing can distort theoretically derived *P* values and undermine FDR control.

When theoretically derived *P* values are doubtful or difficult to validate, empirical-null procedures that estimate the null distribution numerically, or proteomics-inspired competition-based procedures adapted to these contexts, can provide alternatives. For instance, gene set enrichment analysis (GSEA) uses phenotype-label permutations to generate null enrichment scores, normalizes the enrichment scores to account for gene-set-specific null behavior and estimates the FDR by comparing the observed and permutation-derived null distributions of normalized enrichment scores. In QTL mapping, ClipperQTL constructs feature-specific synthetic counterparts using permutations and compares each feature to its counterpart, allowing it to reduce the number of permutations dramatically while maintaining FDR control and yielding order-of-magnitude speed-ups[14]. In all cases, the same principles apply as in differential analysis: the plausibility of a pooled numerical null should be checked, feature-specific numerical-null approaches should be considered only when appropriate fictional counterparts can be constructed, and any randomness in null generation should be documented for reproducibility.

Conclusions

Three decades after the seminal work of Benjamini and Hochberg, FDR control in omics is undergoing a second revolution. Three main methodological frameworks are now available to practitioners: traditional multiple-testing procedures, which rely on theoretical representations of the null; proteomics-inspired methods, which leverage a competition-based framework with numerical, feature-specific nulls; and empirical-null methods, which stand in between, using numerical, feature-pooled nulls.

By bringing together perspectives from sequencing and MS-based omics and from modern statistical theory, our taxonomy highlights good practices for responsible data analysis, such as visually assessing *P*-value uniformity before applying traditional FDR procedures and benchmarking competition-based frameworks in MS workflows using entrapment experiments. For bioinformatic software developers focusing on either differential analysis or biomolecule identification, the information presented in Table 2 may not be sufficiently specific, given the diversity of experimental designs and computational workflows. A more detailed and technical discussion of the pros and cons of each family of methods is provided in the Supplementary Information.

Our taxonomy also clarifies key open challenges, including whether decoy-generation schemes yield accurate null distributions. A related and still incompletely resolved challenge is how well each family of procedures tolerates correlation among biological features—a common situation in omics data, arising, for example, from co-regulated genes, shared genetic variants or peptides belonging to the same protein. For *P*-value-based thresholding rules such as the Benjamini–Hochberg procedure, the conditions under which correlated features still permit valid FDR control are relatively well characterized, along with adjustments for when these conditions do not hold[2]. Because these results concern the thresholding step rather than the null representation, they apply to both theoretically calibrated *P* values and empirical *P* values derived from a numerical null, provided that the *P* values satisfy the validity and dependence conditions required by the corresponding thresholding procedure. In contrast, comparable guarantees are more limited for empirical-null procedures that estimate the FDR directly from a pooled numerical null. For proteomics-inspired competition-based procedures, guarantees depend on fictional counterparts satisfying the required exchangeability and dependence conditions, which can be difficult to verify in practice. Recent work has shown that applying TDC naively to correlated or grouped features can fail[15], motivating ongoing methodological research. Addressing these issues should enable more reliable FDR control in increasingly complex experimental designs and, ultimately, the extraction of more biologically meaningful knowledge from high-throughput molecular screening technologies.

Thomas Burger [1] ✉, **Xinzhou Ge** [2] **& Jingyi Jessica Li** [3] ✉
[1]University Grenoble Alpes, CNRS, CEA, INSERM, UA13 BGE, FR2048 ProFI, Grenoble, France. [2]Department of Statistics and Data Science, Oregon State University, Corvallis, OR, USA. [3]Biostatistics Program, Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA.
✉e-mail: thomas.burger@cea.fr; lijy03@fredhutch.org

References

1. Benjamini, Y. & Hochberg, Y. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
2. Benjamini, Y. & Yekutieli, D. *Ann. Stat.* **29**, 1165–1188 (2001).
3. Elias, J. E. & Gygi, S. P. *Nat. Methods* **4**, 207–214 (2007).
4. Scheubert, K. et al. *Nat. Commun.* **8**, 1494 (2017).
5. Barber, R. F. & Candès, E. J. *Ann. Stat.* **43**, 2055–2085 (2015).
6. Ge, X. et al. *Genome Biol.* **22**, 288 (2021).
7. Efron, B. *J. Am. Stat. Assoc.* **99**, 96–104 (2004).
8. Tusher, V. G., Tibshirani, R. & Chu, G. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
9. Käll, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. *J. Proteome Res.* **7**, 29–34 (2008).
10. Li, Y., Ge, X., Peng, F., Li, W. & Li, J. J. *Genome Biol.* **23**, 79 (2022).
11. Danilova, Y., Voronkova, A., Sulimov, P. & Kertész-Farkas, A. *J. Proteome Res.* **18**, 2354–2358 (2019).
12. Etourneau, L. & Burger, T. *J. Proteome Res.* **21**, 2840–2845 (2022).
13. Wen, B., Freestone, J., Riffle, M., MacCoss, M. J., Noble, W. S. & Keich, U. *Nat. Methods* **22**, 1454–1463 (2025).
14. Zhou, H. J., Ge, X. & Li, J. J. *Genome Biol.* **26**, 207 (2025).
15. Freestone, J., Short, T., Noble, W. S. & Keich, U. *Bioinformatics* **38**, ii82–ii88 (2022). (Suppl. 2).

Acknowledgements

The authors are grateful to Changhu Wang and Chandra Sekhar Mukherjee at the Fred Hutchinson Cancer Center for their valuable feedback. T.B. is supported by the following grants from the French National Research Agency: ProFI project (ANR-10-INBS-08), GRAL CBH project (ANR-17-EURE-0003), France 2030 programme (ANR-19-P3IA-0003), PeptidOMS project (ANR-24-CE45-3296) and ProteoVir project (ANR-24-RRII-0001). J.J.L. is supported by the following grants from the US National Institutes of Health: NIGMS R35GM140888 and NHGRI R01HG014687.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-026-03379-8.