Generalized Reporter Score-based Enrichment Analysis for Diverse Omics Data

Chen Peng[1,2], Qiong Chen[1,2], Shangjin Tan[4,5], Chao Jiang[1,2,3*]

[1]Zhejiang Provincial Key Laboratory of Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang 310058, China

[2] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310009, China

[3]Center for Life Sciences, Shaoxing Institute, Zhejiang University, Shaoxing, Zhejiang 321000, China

[4]BGI Research, Wuhan, Hubei 430074, China

[5]BGI Research, Shenzhen, GuangDong 518083, China

Abstract

Enrichment analysis contextualizes biological features in pathways to facilitate a systematic understanding of high-dimensional data and is widely used in biomedical research. The emerging method known as the reporter score-based analysis (RSA) shows more promising sensitivity, as it relies on *p-values* instead of raw values of features. However, RSA can only be applied to two-group comparisons and is often misused due to the lack of a convenient tool. We propose the Generalized Reporter Score-based Enrichment Analysis (GRSA) method for enrichment analysis of multi-group and longitudinal omics data. The GRSA is implemented in an R package, ReporterScore, integrating a powerful visualization module and updatable pathway databases. A comparison with other common pathway enrichment analysis methods, such as Fisher's exact test and GSEA, reveals that GRSA exhibits increased sensitivity across multiple benchmark datasets. We applied GRSA to the microbiome, transcriptome, and metabolome data to show its versatility in discovering new biological insights in omics studies. Finally, we showcased the applicability of the GRSA method beyond functional enrichment using a custom taxonomy database. We believe the ReporterScore package will be an invaluable tool for broad biomedical research fields. The ReporterScore and a complete description of the usages are publicly available on GitHub (https://github.com/Asa12138/ReporterScore).

Introduction

Functional enrichment analysis is an essential bioinformatic method that helps understand the biological significance of large omics datasets, such as transcriptomic, metagenomic, and metabolic datasets, and formulate hypotheses for downstream experimental investigations[1]. By identifying enriched functional categories, such as gene ontology terms or biological pathways, we can gain insights into the underlying biological processes and functions.

Methods for functional enrichment analysis can be roughly divided into three categories based on underlying statistical methods: (i) overrepresentation analysis (ORA), (ii) functional class scoring (FCS), and (iii) pathway topology-based (PT)[2]. Common enrichment analysis methods in omics research are shown in Table 1. The algorithm of reporter score-based analysis (RSA) was originally developed by Patil and Nielsen in 2005 to identify metabolites associated with the metabolic network's regulatory hotspots[3]. The RSA has regained popularity in recent years due to its extended application in functional enrichment analysis in microbiome research, which can help identify microbial functional pathways that undergo significant changes in different conditions[4]. RSA is an FCS method based on parsing the *p-values* of the differential abundance or correlation analyses. The rationale is that the *p-value* can be considered as a standardized statistic that reflects the differences between different genes, regardless of the mean expression values. The pathways with significantly lower *p-values* than the background *p-value* distribution will be enriched[3].

However, RSA is often misused due to the lack of specific tools and systematic understanding of the algorithm[5]. In addition, the sign (plus or minus) of the reporter score of each pathway in classic RSA does not represent the increasing or decreasing trend of the pathway expression; rather, any reporter scores (including negative scores) less than a specified threshold simply indicates that the corresponding pathway is not significantly enriched. This often leads to misinterpretations of the results.

Inspired by the classic RSA, we developed the improved Generalized Reporter Score-based Enrichment Analysis (GRSA) method, implemented in the R package ReporterScore, along with comprehensive visualization methods and pathway databases. GRSA is a threshold-free method that works well with all types of biological features, such as genes in the transcriptome, compounds in the metabolome, and species in the metagenome. GRSA works in two modes: classic RSA (the mixed mode) and enhanced RSA (the directed mode). The enhanced RSA uses signs of the reporter score to distinguish up-regulated or down-regulated pathways, which is more intuitive. Importantly, the GRSA supports multi-group and longitudinal experimental designs, as we have included multi-group compatible statistical methods for calculating *p-values* (for a full list of supported methods, please see Table S1). Additionally, the ReporterScore package also supports custom hierarchical and relational databases, providing extra flexibility for advanced users. In this study, we described the general utility of GRSA, benchmarked GRSA against other most commonly used enrichment methods on six omics datasets, and demonstrated the applications of GRSA on diverse omics datasets in four case studies.

Table 1: Methods of common enrichment analysis.

Category	Method	Tools	Notes	Reference
ORA	Hypergeometric test / Fisher's exact test	DAVID (website), clusterProfiler (R package)	The most common methods used in enrichment analysis.	[6,7]
FCS	Gene set enrichment analysis (GSEA)	GSEA (website), clusterProfiler (R package)	A computational method that determines whether a set of genes shows statistically significant and concordant differences between two biological states.	[8,9]
FCS	Generalized Reporter Score-based analysis (GRSA/RSA)	**ReporterScore (R package developed in this study)**	Find significant metabolites (first report), pathways, and taxonomy based on the *p-values* for multi-omics data.	[3,5]
PT	Reporter feature analysis	/	Integrates bio-molecular network topology with transcriptome data to identify the key biological features.	[10]
PT	Topology-based pathway enrichment analysis (TPEA)	TPEA (R package)	Integrates topological properties and global upstream/downstream positions of genes in pathways.	[11]

Result

Workflow overview

The ReporterScore package has built-in KEGG pathway, module, gene, compound, and GO databases and provides a function for customizing databases, so it is compatible with feature abundance tables from diverse omics data. Importantly, the input data should not be filtered to preserve the background *p-value* distribution.

For the transcriptomic, scRNA-seq, and related gene-based omics data, a gene abundance table can be used. For the metagenomic and metatranscriptomic data, which involve many different species, a KO abundance table can be used, which is generated using Blast, Diamond, or KEGG official mapper software[12] to align the reads or contigs to the KEGG database[13] or the EggNOG database[14]. For the metabolomic data, an annotated compound

abundance table can be used, but compound ID conversion according to the database (e.g., convert compound names to C numbers in the KEGG database) is required.

The workflow of GRSA in the ReporterScore package is shown in Figure 1, using metagenomic data as an example. The KO abundance table (rows are KOs and columns are samples) and metadata table (rows are samples and columns are experimental design groups) were used as the input for GRSA. First, the *p-values* for all KOs were calculated by a selected statistical method (Figure 1A). Then, in the classic mode, the *p-values* were directly converted to Z-scores (Figure 1B [1]). In the directed mode, the *p-values* were divided by 2, converted to Z-scores, and assigned plus or minus signs, denoting up- and down-regulated KOs (Figure 1B [2-4]). Next, the Z-score of the pathway j (Z_{path_j}) was calculated by summing the Z scores of KOs within the pathway j, and divided by the square root of the number of KOs (k_j) in the pathway j (Figure 1C [1]). The Z_{path_j} is further standardized by the background pathway Z-score distribution, generated by randomly sampling k_j KOs from the total KO pool (Figure 1C [2]). The standardized pathway Z-score is henceforth referred to as the reporter score of a pathway ($ReporterScore_j$). The details of the GRSA algorithm are described in the Method section.

We designed the ReporterScore package to be user-friendly. The function reporter_score calculates the reporter scores for a matching feature abundance table and metadata in one step. The included assorted visualization methods can be used to explore the entire pathways and features within pathways (Figure 1D). An example code tailored for a KO abundance table is as follows.

```
library(ReporterScore)
# Load the KO abundance table
KO_abundance <-read.table("ko_abundance.tsv", header = TRUE, sep = "\t
")
# Get the sample metadata
metadata <-read.table("sample_metadata.tsv", header = TRUE, sep = "\t")
# Run RSA analysis
reporter_score_res = reporter_score(KO_abundance, "Group", metadata,
                                    mode="directed", type = "pathway")
# Visualization
plot_report(reporter_score_res, rs_threshold = c(-3,3))
plot_report_circle_packing(reporter_score_res, rs_threshold = c(-3,3))
plot_KOs_in_pathway(reporter_score_res, map_id = "map00780")
plot_KOs_heatmap(reporter_score_res, map_id = "map00780")
plot_KOs_network(reporter_score_res, map_id = c("map05230","map04922"))
```

Next, we collected several benchmark datasets (Table S2) to investigate the performance of GRSA and compare the GRSA with other commonly used enrichment analysis methods.

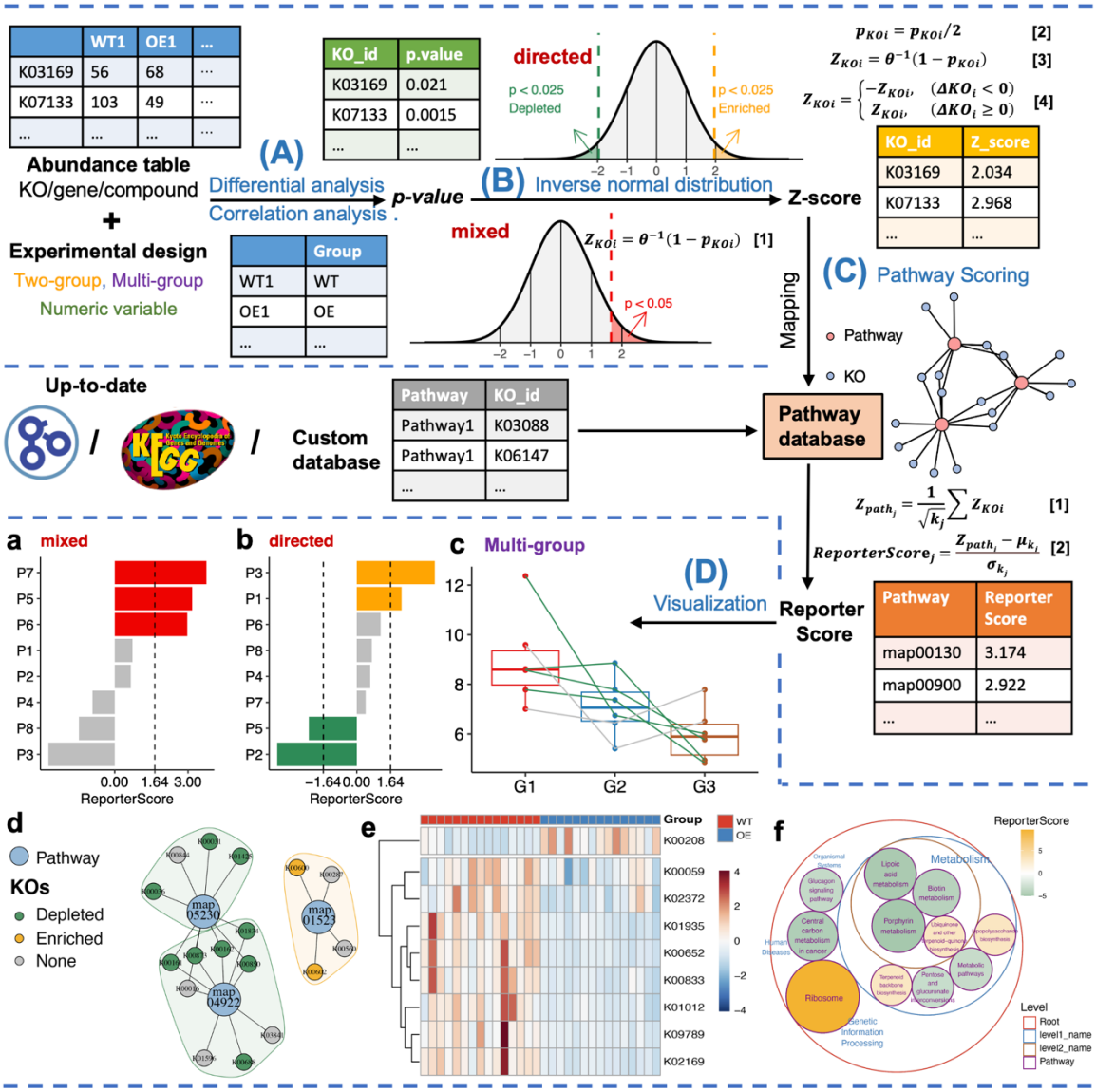

Figure 1: The overall workflow of GRSA in the ReporterScore package. GRSA mainly consists of four parts: (A) Calculation of the *p-value* for each KO between two or multiple groups by various statistical methods. (B) Conversion of the *p-value* of a KO to Z-score by inverse normal distribution and assignment of a plus or minus sign to each Z-score in the directed mode. (C) Mapping KOs to annotated pathways and calculating the reporter score for each pathway. KO_i represents a certain KO; p_{KO_i} is the *p-value* of KO_i; Z_{KO_i} is the Z-score transformed from p_{KO_i}; ΔKO_i is the abundance difference of between groups. A total of k_j KOs were annotated to the corresponding pathway. μ_{k_j} and σ_{k_j} are the mean and the standard deviation of the background Z-score distribution Z_{path_null}, respectively. (D) The ReporterScore package provides various visualization methods for the GRSA result: (a) The bar chart shows reporter scores of pathways in the mixed mode. The red color indicates significantly enriched pathways, with reporter scores greater than 1.64, corresponding to a *p-value* of 0.05. (b) The bar chart shows reporter scores of pathways in the directed mode. The orange and green colors indicate up-regulated and

down-regulated pathways with absolute reporter scores greater than 1.64. (c) The box chart shows the pattern of a selected pathway in the directed mode with a multi-group design, each line represents the trend of the average abundance of one KO. Line colors indicate whether the KO is significantly enriched (orange), depleted (green), or neither (grey). (d) The network plot shows the KOs present in selected pathways; some KOs can be shared by several pathways. Big dots represent pathways, and small dots represent KOs. The colors of small dots represent the trend of KOs. The colors of the shades encircling pathways denote whether the pathway is overall up-regulated (orange) or down-regulated (green). (e) The heatmap displays the abundance of each KO in a pathway for different samples (columns). (f) The circular packing plot shows the hierarchical relationship of selected pathways; the size of the circle indicates the absolute value of the reporter score, and the color of the circle indicates that the pathway is overall up-regulated (orange) and down-regulated (green).

Applying GRSA to multi-group and longitudinal omics data

An important feature of GRSA is the newly developed directed mode. The key difference between the directed mode and the mixed mode (classic RSA) is that in the directed mode, the plus or minus sign of the reporter score indicates the increasing or decreasing trend of the pathway (Figure 1B). However, in the mixed mode, the signs of the reporter score do not indicate the trends of the pathways.

We performed GRSA on the public ex_KO_profile dataset (Table S2) in two modes (Figure S1A). For each pathway enriched in the directed mode, most KOs within the pathway share the same trend (Figure S1B, blue and red boxes). If KOs within a pathway had opposing trends, the signed Z-scores of these KOs would cancel either other, leading to an insignificantly enriched pathway in the directed mode (Figure S1B, orange box). In comparison, in the mixed mode, the trend of the enriched pathway cannot be determined (Figure S1C). Therefore, the directed mode helps find pathways with consistently changing KOs. For simplicity, we use GRSA in the directed mode henceforth.

Another major advantage of GRSA is the full support of multi-group and longitudinal omics data. The ReporterScore package calculates the *p-value* for each feature between groups using differential abundance analysis ("T-test", "Wilcoxon rank-sum test", "Kruskal-Wallis test", "ANOVA") and correlation analysis ("Pearson", "Spearman", "Kendall"). The Kruskal-Wallis test or ANOVA assesses if the feature abundance varies significantly across multiple groups. The default correlation analysis treats group assignments as ordinal (e.g., groups "G1", "G2", and "G3" will be converted to 1, 2, 3), so the correlation analysis would test if the feature abundance linearly increases or decreases over a series of time points. Moreover, the ReporterScore package also supports the definition of any specified patterns (e.g., groups "G1", "G2", and "G3" can be set as 1, 10, 100 when an exponentially increasing trend is expected).

We applied GRSA with different statistical methods on multiple datasets. For the classic two-group design, the difference in results mainly stems from the parametric methods versus the non-parametric methods (Figure S2A). The Jaccard similarity exceeded 0.84 for parametric methods and 0.78 for non-parametric methods (Figure S2A). For the multi-group data, differential abundance analyses and correlation analyses performed differently.

Specifically, we compared pathways enriched with ANOVA and Pearson correlation methods (parametric methods) by identifying clusters of KOs within the significantly enriched pathways using fuzzy c-means. In the ANOVA-based results, four patterns were present. In the Pearson correlation-based results, only two patterns were observed: increasing and decreasing (Figure S2B), presumably because correlation analysis detects linear patterns by default. The comparison of non-parametric differential abundance and correlation analyses showed highly similar results (Figure S2C). As a general rule, the users need to make sure the statistical methods are reasonable for the datasets and experimental designs[15].

Lastly, GRSA also supports other statistical tests, such as "DESeq2", "Edger", "Limma", "ALDEX", "ANCOM"[16], to calculate the reporter scores as follows.

```
#1. Use specific statistical test method to get the p-value
ko_pvalue=your_method(KO_abundance)
#2. Transfers the p-value of KOs to the Z-score (select mode: mixed, di
rected)
ko_stat=pvalue2zs(ko_pvalue, mode=choice_of_mode)
#3. Calculate the reporter score of each pathway.
reporter_s=get_reporter_score(ko_stat)
```

GRSA shows higher sensitivity than other most commonly used enrichment analysis methods

We next compared GRSA against other most commonly used enrichment analysis methods. Fisher's exact test is one of the most common functional enrichment analyses, which relies on an arbitrary cutoff of fold change and/or significance. GSEA is a classic functional class scoring method and analyzes all features based on their differential expression rank without prior feature filtering.

GSEA calculates an Enrichment Score (ES) by moving through the ranked features list, increasing the ES if a feature is in the pathway, and decreasing the ES if not. These running sum values are weighted so that enrichment in the top- and bottom- ranking features is amplified, while enrichment in the moderate ranks are not amplified. The ES is normalized to pathway size, yielding a Normalized Enrichment Score (NES). Positive and negative NES indicate enrichment at the top and bottom of the feature list, respectively. Lastly, a permutation-based *p-value* is computed, and multi-test correction is applied, yielding a False Discovery Rate (FDR) or Q value from 0 (significant) to 1 (not significant)[8]. However, the GSEA cannot be directly applied to multi-group or longitudinal datasets.

PT-based methods may be better at identifying biologically meaningful pathways than non-PT-based methods in some scenarios[17]. However, PT-based methods require pathways with comprehensive topological structure, while most pathways don't apply, limiting the versatility of PT-based methods[18]. So, here we only compared GRSA against most commonly used non-PT enrichment analysis methods: fisher.test, enricher and GSEA, using the identical pathway database.

We compared the performance of GRSA with fisher.test (Fisher) provided by the R base package, enricher provided by the clusterProfiler package (CP; an improved fisher.test), and

gene set enrichment analysis (GSEA) on the same six datasets. In each pair-wise comparison, we characterized the proportions of pathways identified by the GRSA, the competing tool, and both, using all significant pathways as the denominator (Figure 2A). GRSA consistently identified a larger proportion of pathways than Fisher and CP, and largely overlapped with GSEA, indicating the higher sensitivity of threshold-free methods.

In five out of six cases, GRSA identified more enriched pathways than GSEA, GRSA-specific pathways are shown in Table S3. For example, in the colorectal cancer datasets (GSE41011 and GSE33126 datasets), pathways related to fat digestion[19], carbohydrate metabolism[20], and chemical carcinogenesis were only enriched by GRSA (Figure S3A). In the myocardial infarction dataset (GSE141512), GRSA identified the NF-kappa B signaling pathway and apoptosis pathways, which were shown to be involved in the pathological characteristics of myocardial infarction[21] (Figure S3B). Therefore, GRSA can identify additional pathways biologically relevant to the studied diseases, which may be neglected by other tools.

In addition to enrichment analyses, some studies directly added the abundance of features within a pathway as the pathway abundance and performed differential abundance analyses at the pathway level (DAP)[22]. In DAP, increased and decreased features cancel each other. We compared DAP to GRSA with various tests ("T-test", "ANOVA", "Wilcoxon rank-sum test", "Kruskal-Wallis test") and found that the DAP method could identify more differential pathways than GRSA (Figure 2B). However, many DAP-specific pathways, such as "map00627" in KEGG database included only one significant feature, and high-abundance features will always mask the dynamic changes of low-abundance features in DAP (Figure 2C, Figure S4). In contrast, GRSA-enriched pathways showed many KOs with consistent small changes, such as "map03430" (Figure 2D). For the overlapping pathways of the two methods, most of the KOs shared the same significant trend, such as "map00785" (Figure 2E).

Next, we showcased the versatile utility of GRSA in multiple types of omics data.

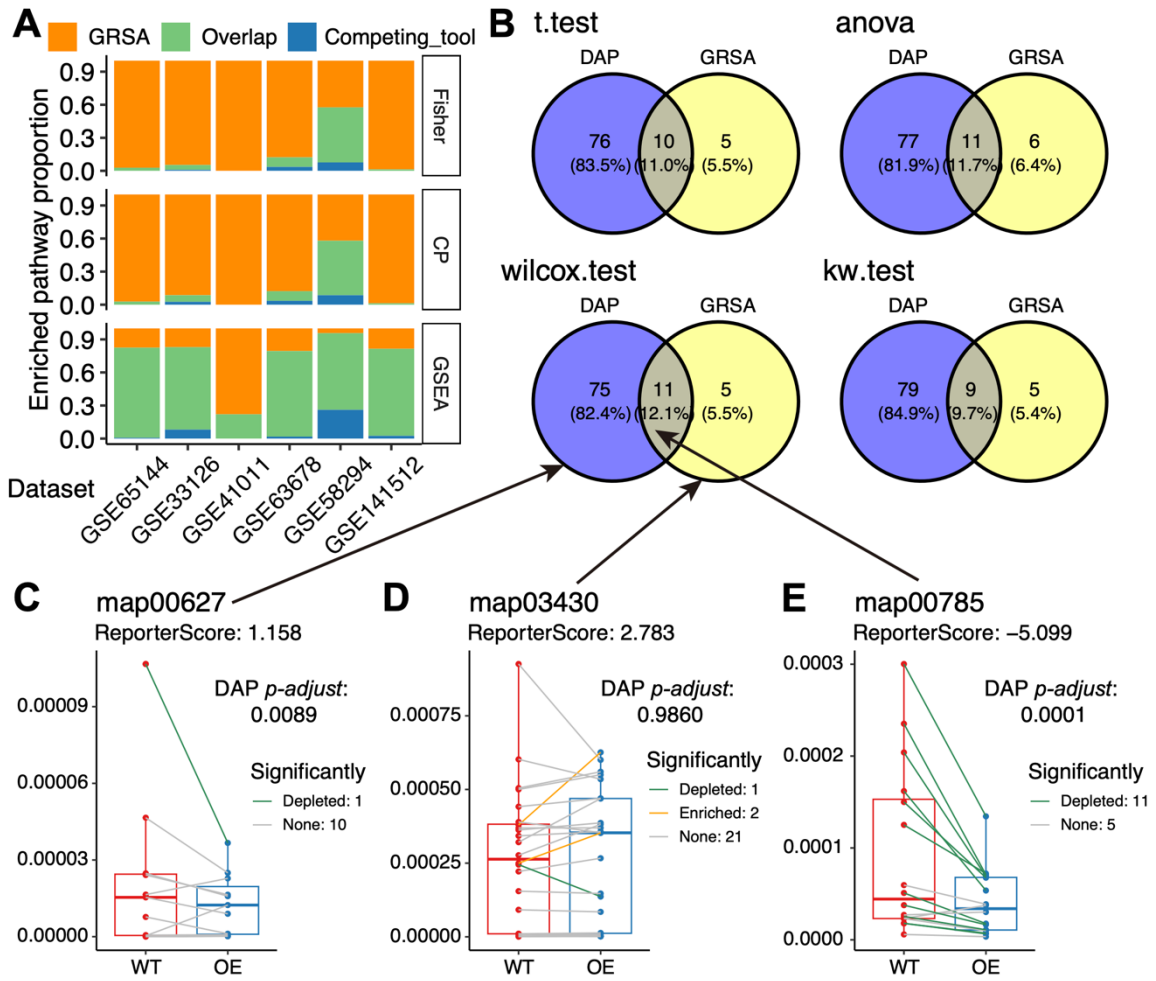

Figure 2: Comparisons of GRSA and other methods of enrichment and differential abundance analyses. (A) Proportion of enriched pathways in GRSA and other enrichment analysis methods based on 6 benchmark datasets. Orange pathways were only identified by GRSA, blue pathways were only identified by the competing tool and green pathways were identified by GRSA and the comparing methods. Fisher: fisher.test; CP: improved fisher.test used by clusterProfiler; GSEA: gene set enrichment analysis by clusterProfiler. (B) Comparison between GRSA (directed mode) and DAP using 4 statistical test methods. (C-E) Box charts of DAP-specific (C), GRSA-specific (D), and shared pathways (E) using 3 KEGG pathways in Wilcox.test as examples.

Case study 1: The functional analysis and age-related dynamics of the skin microbiota

For microbiome data, we collected the KO profile of the IHSMGC (integrated Human Skin Microbial Gene Catalog) dataset published by Wang et al.[23] and re-analyzed the data using the GRSA method. In the previous study, they used DAP approach instead of enrichment analysis to identify the most significantly differential abundant functional modules related with cutotypes. We applied the GRSA to the datasets to find the functional differences between the two cutotypes. The results were consistent with the previous study. As an

example, the vitamin biosynthesis-related function showed a difference, as modules related to the biosynthesis of thiamine, phylloquinone, and cobalamin were enriched in the *M-cutotype*; while functions related to tetrahydrofolate, menaquinone, pantothenate, and ubiquinone were enriched in the *C-cutotype* (Figure 3A). In addition, the *M-cutotype* was enriched with a large number of modules related to the metabolism of sulfur, phenylacetate (aromatic compound), and amino acids, while the *C-cutotype* was enriched with modules related to carbohydrate metabolism (Figure S5). Importantly, GRSA also identified pathways not found in the previous study. The *M-cutotype* was enriched with modules related to nucleotide metabolism, such as the degradation and de novo biosynthesis of purine (Figure S5), indicating that the *M-cutotype* microbiota has a higher nucleotide turnover rate and stronger proliferation[24].

The previous study divided all samples into 5 age groups and found the prevalence of the *M-cutotype* significantly increased with age. However, they did not perform age-related functional analysis. We re-analyzed the multi-group data using GRSA based on Pearson correlation analysis to explore the functional dynamics related to aging. The larger positive reporter scores indicate that the module has an overall increasing trend with respect to age, such as "M00866", related to lipid A biosynthesis (Figure 3B), while modules with negative reporter scores show an overall decreasing trend with respect to age, such as "M00061", related to D-Glucuronate degradation (Figure 3C). We next analyzed the chronological trend of the functional modules at the KEGG level B (Figure 3D), which better reflects the overall metabolic activities of the microbiome. We found that the carbohydrate metabolism activity of the skin microbiota decreases with aging, while the lipid, amino acid, and nucleotide metabolism activity increases with aging. These results suggest that the energy sources of the skin microbiota significantly change with aging.

The vitamin biosynthesis-related functional modules also showed differences with respect to aging (Figure 3D). For the glycan metabolism-related functional modules, biosynthesis of KDO2-lipid A and CMP-KDO increased with aging. KDO2-lipid A is an essential component of lipopolysaccharide (LPS) in most gram-negative bacteria, which has endotoxin activity and stimulates host immune responses through Toll-like receptor 4 (TLR4)[25]. CMP-KDO is an important intermediate in the synthesis of KDO2-lipid A, and CMP-KDO synthesis is the key rate-limiting step for introducing KDO into LPS[26]. These results suggest that microbiota of aging skins likely accumulate endotoxins and stimulate host inflammation. In addition, we found degradation pathways of several sulfated glycosaminoglycans (chondroitin sulfate, dermatan sulfate, and keratan sulfate) decreased in aging skin. Sulfated glycosaminoglycans play a key role in regulating skin physiology, and there is ample evidence that their properties and functions change over time and with extrinsic skin aging[27 28]. Total sulfated glycosaminoglycan abundance was reduced in aging skin[29], which may lead to the decreased degrading ability of the skin microbiota for the sulfated glycosaminoglycans.

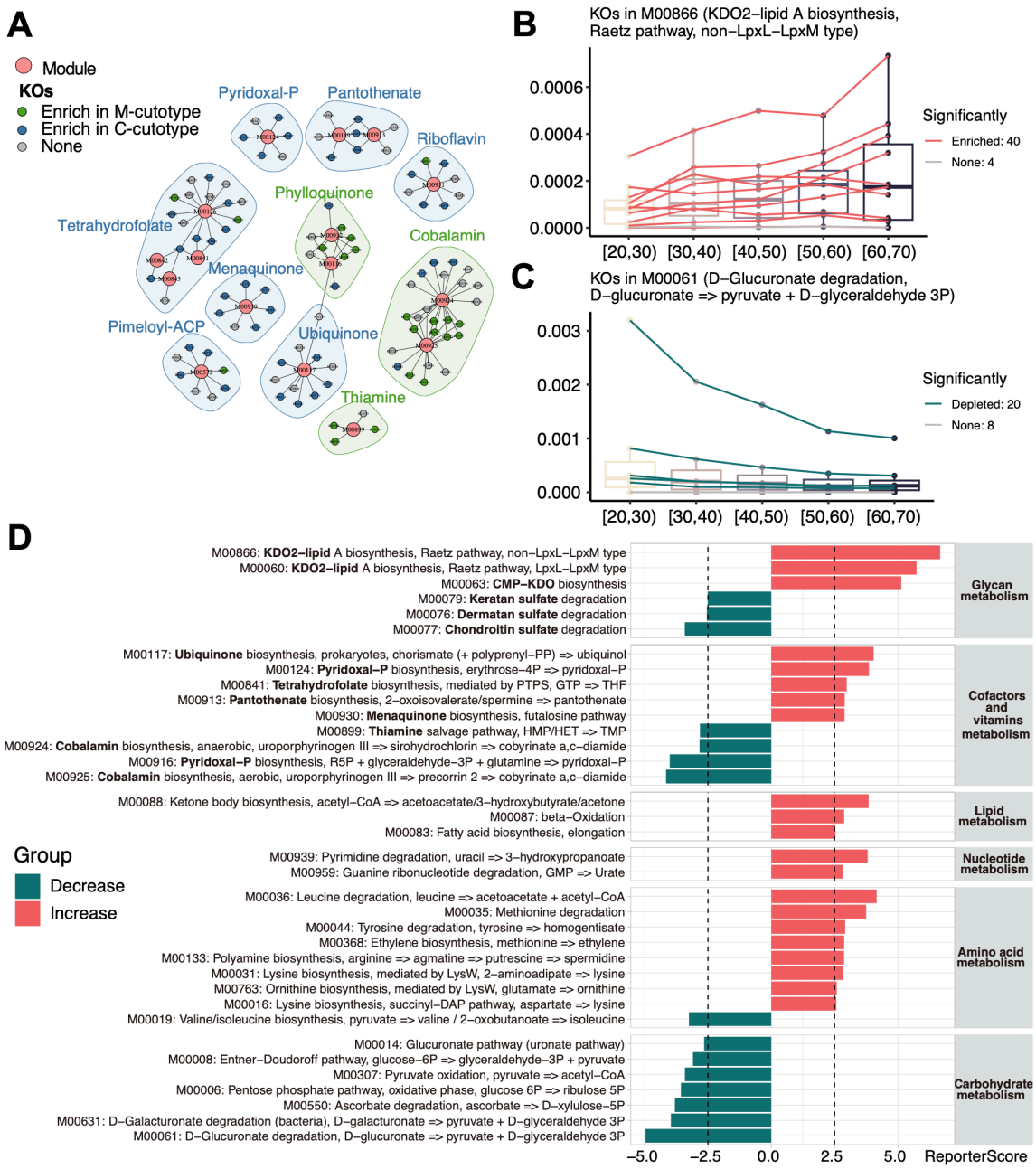

Figure 3: Application of GRSA to the skin microbiome of the IHSMGC dataset. (A) The network of KO-Module enriched in the *M-cutotype* (green) and *C-cutotype* (blue). Only modules related to vitamin biosynthesis were shown. Big dots represent modules; small dots represent KOs. The colors of small dots represent cutotypes. Shades indicate modules involved in the biosynthesis of the same vitamin. The colors of shades denote modules enriched in the *M-cutotype* (green) or enriched in the *C-cutotype* (blue). (B-C) The Box charts of modules "M00866" and "M00061" across ages. The colors of the lines represent the trend of KOs relative abundance in the module. "M00866" had the biggest positive reporter score (increasing), while "M00061" had the biggest absolute value of negative score (decreasing). (D) The Bar chart shows significantly enriched modules over

320 ages; the reporter score threshold of 2.5 corresponds to a confidence of about 0.995, and
321 these modules are grouped based on the KEGG level B. Colors denote the modules that
322 were up-regulated (red) or down-regulated (green) with aging.

323 Case study 2: The functional transcriptional dynamics during 324 cardiomyocyte differentiation

325 We applied GRSA to the transcriptomic dataset published by Liu et al. in 2017[30]. The study
326 used the WGCNA method to analyze the temporal transcriptomic changes during the
327 differentiation of cardiomyocytes from 2 hiPSC lines and 2 hESC lines at 4 timepoints
328 (pluripotent stem cells at day 0, mesoderm at day 2, cardiac mesoderm at day 4, and
329 differentiated cardiomyocytes at day 30). Significant changes were observed in the four
330 stages of differentiation among all cell lines. For example, genes in module 1 were highly
331 expressed only in differentiated cardiomyocytes (stage CM), and their enriched Gene
332 Ontology (GO) terms of Biological Process (BP) were related to heart functions, such as
333 regulation of cardiac contraction and muscular system processes. However, WGCNA did
334 not assume the patterns to be linear so that genes can be only highly expressed at day 2
335 during mesoderm development, for example.

336 In addition to linearly increasing or decreasing patterns, GRSA allows users to specify any
337 expected patterns for enrichment analysis. To start, we used the fuzzy C-means clustering
338 method to identify the main gene expression patterns (Figure 4A), and then used these
339 patterns for GRSA to obtain significantly enriched pathways in each pattern (using the
340 RSA_by_cm function in the ReporterScore package). For example, "Heart process
341 (GO:0003015)" was a significantly enriched GO term for Cluster 6, which was highly
342 expressed only in stage CM (day 30). We identified many genes consistent with the
343 expression pattern of Cluster 6 (Figure 4B).

344 GRSA results for all clusters were shown in the Figure 4C. Cluster 2 was highly expressed
345 only at day 0 and its enriched GO terms was mainly related to the mitotic cell cycle, which
346 was expected for stem cell self-renewal processes. Cluster 5 had the highest expression
347 level on day 2 and was mainly enriched in various transcription and translation processes.
348 Many known transcription factors such as EOMES, MIXL1, and WNT3A were assigned
349 to this cluster and play important roles in mesoderm development. Cluster 4 was highly
350 expressed at day 0 and day 2 and showed a gradually decreasing trend, its function
351 overlapped with Clusters 2 and 5. Cluster 1, highly expressed at day 4, was related to
352 mesoderm formation such as morphogenesis and organ development. Clusters 3 and 6 were
353 primarily up-regulated in differentiated cardiomyocytes (CM stage), and they were related
354 to heart functions, such as regulation of heart contraction and muscle system processes,
355 similar to module 1 in the previous study. Interestingly, the biological processes of
356 hiPSCs/hESCs at day 2 (Cluster 5) focused on various RNA-related metabolisms, which
357 was not found in the previous study, indicating that complex transcriptional regulations are
358 involved for further mesoderm formation. Therefore, using the identified expression
359 patterns across groups, we successfully identified pathways and modules essential to
360 different stages of the cardiomyocyte differentiation processes.

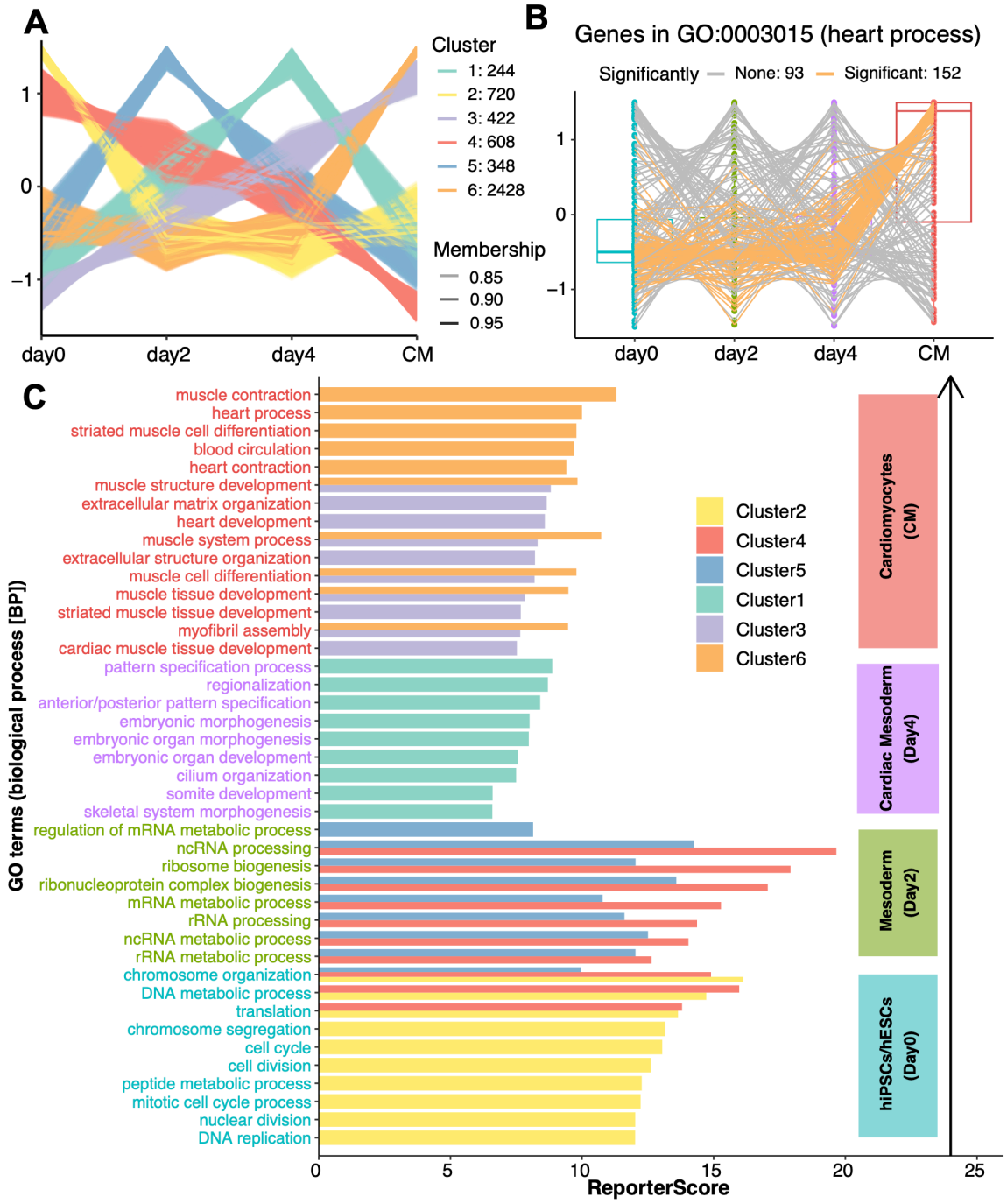

Figure 4: Application of GRSA to the transcriptomic dataset on the cardiomyocyte differentiation processes. (A) C-means clustering result of gene abundance profiles across four differentiation stages. The genes with membership scores greater than 0.8 were displayed. The alpha (transparency) of each line was related to the value of its membership score, and the abundance was standardized. (B) The box chart of "GO:0003015" (heart process) across four time points, the colors of the lines represent the correlative significance of each gene with Cluster 6 within the GO term. "GO:0003015" is a representative term of Cluster 6. (C) The bar chart shows

370 significantly enriched GO terms for each clustering pattern corresponding to the
371 differentiation stages. The colors of the bars represent the cluster information, and the
372 representative GO terms with the highest reporter scores in each cluster were shown. The
373 text labels on the left were colored according to the stages with the highest expression. In
374 general, Cluster 2 corresponds to day 0, Clusters 4 and 5 correspond to day 2, Cluster 1
375 corresponds to day 4, and Clusters 3 and 6 correspond to CM. Note that only pathways
376 with significant positive scores are shown. The negative score of specified patterns would
377 indicate anti-correlative patterns, which should have already been identified by c-means
378 analysis, such as cluster 3 vs. 4.

379 Case study 3: The systematic maternal metabolomic changes correlate 380 with gestational age

381 We next applied GRSA to metabolomic data from a Danish pregnancy cohort in which
382 female participants had blood drawn weekly from pregnancy to the postpartum period for
383 untargeted metabolomics analysis[31]. Using gestational age as the study variable, they
384 modeled a metabolic clock and found that several marker metabolites increased linearly
385 with gestational age.

386 We performed GRSA using gestational age (a numeric variable) and enriched for pathways
387 that were significantly up- or down-regulated with respet to gestational age in weeks. We
388 found several important pathways upregulated with gestational age: steroid hormone
389 biosynthesis, cortisol synthesis and secretion, and Oocyte meiosis (Figure 5A). Multiple
390 steroid hormones were up-regulated with increasing gestational age (Figure 5B), including
391 progesterone that interacts with the hypothalamic-pituitary-adrenal axis (HPA axis)[32] and
392 estriol-16-glucuronide produced by the placenta[33]. At the same time, two steroid hormones
393 related to androgen were down-regulated: dehydroepiandrosterone sulfate and
394 androsterone 3-glucuronide, as the concentration of androgens plays important
395 physiological functions during pregnancy[34]. We also found that pathways related to the
396 metabolism of aromatic amino acids were down-regulated with increasing gestational age
397 (Figure 5A), which has been reported[35].

398 Importantly, we identified several up-regulated pathways related to human diseases that
399 were not mentioned in the previous study. Cushing syndrome happens when the body has
400 too much of the hormone cortisol for a long time, which could be induced by healthy
401 pregnancy[36]. The up-regulation of pathways related to breast cancer were also noticeable
402 as Pregnancy-Associated Breast Cancer (PABC) accounts for 7% of all breast cancer in
403 young women[37]. Importantly, more potential discoveries can be made if the metabolite-
404 pathway database can be improved.

405 Case study 4: The application of customized hierarchical relational 406 databases for GSRA

407 The algorithm of GRSA suggests that any features organized in a hierarchical relationship
408 can be used as an enrichment database. For example, if we wish to identify phylogenetic
409 groups of microbes with a specific abundance pattern among different groups of samples,
410 we can use the phylogenetic relationships of microbes, such as genus-species phylogenetic
411 data, for taxonomic enrichment analysis. To demonstrate, with the custom_modulelist

function, we used the species abundance table from the IHSMGC dataset and looked for genera enriched in the two cutotypes. We found that *Psychrobacter*, *Paracoccus*, *Chryseobacterium*, *Elizabethkingia*, *Deinococcus*, and *Microbacterium* were enriched in the *M-cutotype*, while *Acidipropionibacterium*, *Staphylococcus*, *Corynebacterium*, and *Cutibacterium* were enriched in the *C-cutotype* (Figure 5C), some of which were highly consistent with the differential species modules found by co-occurrence network in the previous study. However, we additionally found some genera such as *Brevundimonas* and *Rhodobacter* were enriched in the *M-cutotype*, while *Pahexavirus* (phages of *Propionibacterium* and *Cutibacterium*) was enriched in the *C-cutotype* (Figure 5C), probably due to the higher sensitivity of GRSA.

Two species, *Moraxella osloensis* and *Cutibacterium acnes*, were used to define the cutotype in the previous study. Interestingly, while the *Cutibacterium* genus was a good biomarker between cutotypes, the *Moraxella* genus was not as the included species did not share the same trend (Figure 5D). Therefore, in additional to functional enrichment analysis, given a custom hierarchical relational database, the GRSA can be extended to any type of enrichment analyses.

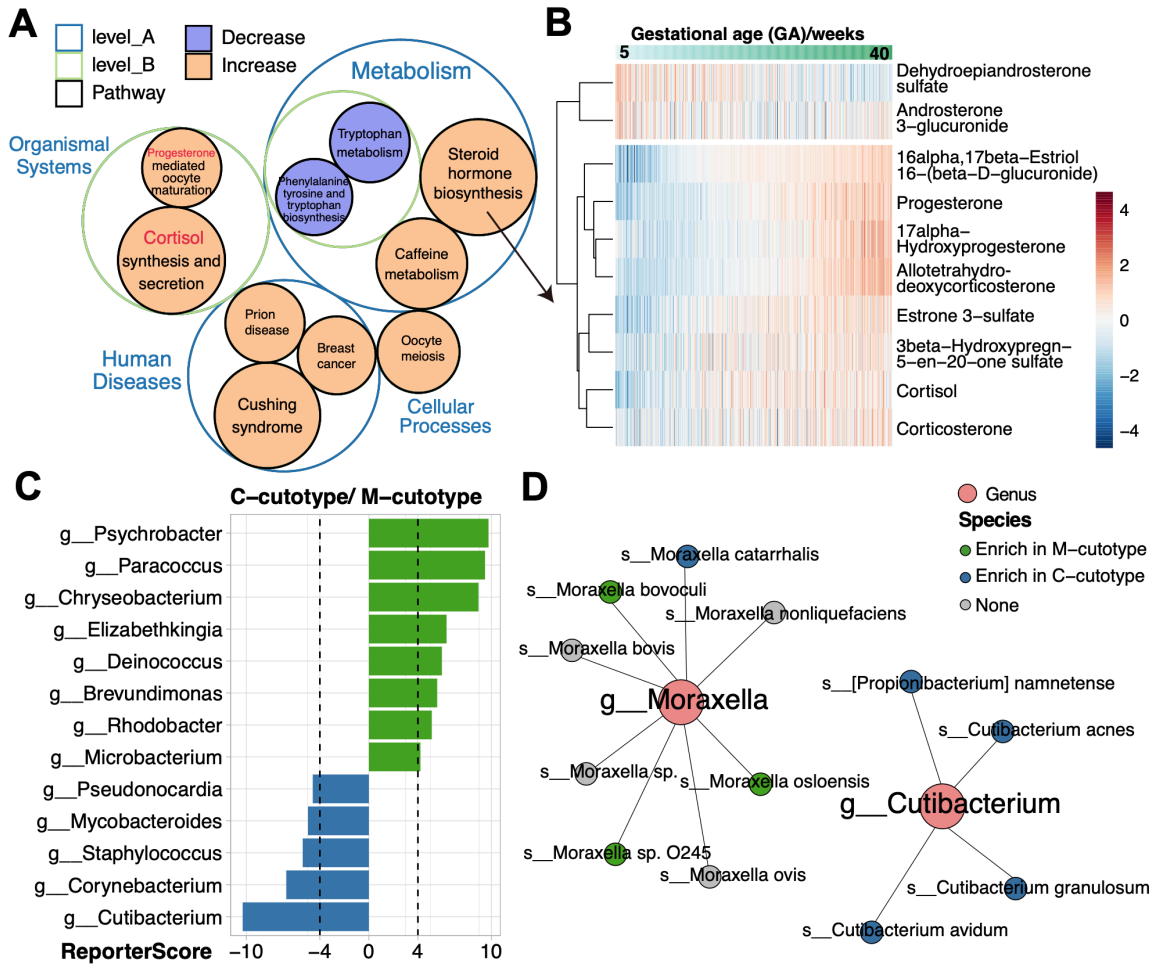

Figure 5: Application of GRSA in the metabolic data of the Danish pregnancy cohort and in the taxonomic enrichment analysis of the IHSMGC dataset. (A) The

circle packing chart shows the hierarchical relationships of significantly enriched pathways identified by GRSA in the metabolomic study. The size of the circle indicates the absolute value of the reporter score, and the color of the circle indicates the sign of reporter score. The positive reporter score indicates the pathway was increased (orange) while the negative reporter score indicates the pathway was decreased (purple). (B) The heatmap shows the abundance of metabolites present in the pathway "steroid hormone biosynthesis". The columns are samples ordered by the increasing gestational age. (C) The bar chart shows significantly enriched genera in the *C-cutotype* and *M-cutotype*. (D) The network plot shows the species in *g_Moraxella* and *g_Cutibacterium*, which are enriched in the *M-cutotype* (green) or *C-cutotype* (blue).

Discussion

We developed the ReporterScore package to democratize the GRSA method for enrichment analyses in the broad sense. We have improved upon the classic RSA method by introducing the directed mode for easy interpretation of the plus and minus signs of the results. More importantly, we introduced all the common statistical methods for differential abundance and correlation analyses, expanding the scope of GRSA from two-group comparison to multi-group comparison. With the support of numerical grouping variables and user-defined feature abundance patterns, we can apply GRSA to longitudinal data and perform enrichment analysis of features of specific patterns. We demonstrated these new applications with metagenomic, transcriptomic, and metabolomic data (Figures 4~6). Lastly, we show that the GRSA is not limited to functional enrichment analysis and can be easily applied for taxonomic enrichment analysis in microbiome studies. We also provided a rich repertoire of visualization methods (Figure 1D), facilitating quick communications between researchers. All the figures were generated using the visualization module in the ReporterScore package.

GRSA is a more sensitive enrichment method because it considers all KOs involved in the pathway compared to hypergeometric tests that only consider a pre-defined list (e.g. KO/gene with *p-value* < 0.05). Thus, GRSA can comprehensively assess the abundance differences in the pathway and not be affected by a priori cut-off of gene significance and the average abundance of a feature (Figure 2A). Compared with GSEA, which considers the magnitude of gene changes for ranking, GRSA uses *p-values* for ranking and permutation. However, more positive results are not always good, and we still need further experimental verification to illustrate the reliability of the enrichment results. In addition, we do not think DAP is an enrichment method because it does not consider the equal contribution of each gene in the pathway and completely ignores the functional background.

The GRSA applies to all omics datasets as long as a reference relational database is available. More importantly, we have acquired new biological insights in each of the case studies. Application of GRSA on the IHSMGC dataset suggested that aging skin microbiota may have different functional profiles than young skin microbiota. For example, biosynthesis of KDO2-lipid A and CMP-KDO increased in while degradation pathways of several sulfated glycosaminoglycans decreased in older skin microbiota, which may be

linked to changes in the skin's physiological properties. Further studies are needed to investigate the underlying mechanisms of these changes and their implications for skin health. Application of GRSA on the transcriptomic data of cardiomyocyte differentiations revealed that hiPSCs/hESCs at day 2 are specialized in various RNA-related metabolisms, suggesting the involvement of complex transcriptional regulation in further mesoderm formation. In addition, Application of GRSA on the metabolomic data from the Danish pregnancy cohort showed that several pathways related to human diseases were up-regulated with gestational age, including Cushing syndrome and PABC. It reminded us of some disease risks for normal pregnant women.

The GRSA offers the option for user-specified pattern for enrichment analysis, allowing for rapid testing of educated hypotheses in complex multi-group studies. This is demonstrated in our analysis of the transcriptomic data of cardiomyocyte differentiations.

Finally, the extended application of GRSA in the taxonomic enrichment analysis of the IHSMGC dataset allowed us to identify key genera that significantly differed between the two cutotypes. The results were highly consistent with the microbial co-occurrence network analysis in the previous study, but performing GRSA by ReporterScore package was much faster and easier than the WGCNA network analysis.

In summary, we believe the GRSA method and the ReporterScore package can greatly facilitate the functional enrichment analyses of diverse omics data across fields of science, with higher sensitivity, compatibility with multi-group and longitudinal designs, and flexibility with customized databases for creative applications even beyond functional enrichment analyses.

Method

Algorithm

The algorithm of GRSA is described as follows, using metagenomic data as an example.

(1) Calculating the *p-values*

A specified method (the full list of supported statistical methods can be found in Table S1) was used to obtain the *p-value* of the abundance comparison of each feature between the experimental groups (i.e., p_{KO_i}, KO_i represents a certain KO; Figure 1A). We used KO to represent different features in the formulas for simplicity.

(2) Converting the *p-values* into Z-scores

For the classic mixed RSA, we used an inverse normal cumulative distribution function (θ^{-1}) to convert the *p-value* of each KO into Z-score (Z_{KO_i}). Thus, in the case of uniformly distributed *p-values* (random data assumption), the resulting Z-scores will follow a standard normal distribution (Figure 1B), the formula is:

$$Z_{KO_i} = \theta^{-1}\left(1 - p_{KO_i}\right)$$

For the new directed RSA, we first divided the *p-values* by 2, transforming the range of *p-values* from (0,1] to (0,0.5]:

$$p_{KO_i} = p_{KO_i}/2$$

Secondly, we used an inverse normal cumulative distribution function to convert the *p-value* of each KO into Z score (Z_{KO_i}). When the *p-value* is 0.5, the converted Z-score equals to 0. Since the above *p-values* are no greater than 0.5, all converted Z-scores will be greater than 0 (Figure 1B). The formula is:

$$Z_{KO_i} = \theta^{-1}\left(1 - p_{KO_i}\right)$$

We then determined if a KO is up-regulated or down-regulated and calculated the ΔKO_i.

In a differential abundance analysis (two-group design):

$$\Delta KO_i = \overline{KO_{i_{g1}}} - \overline{KO_{i_{g2}}}$$

$\overline{KO_{i_{g1}}}$ is the average abundance of KO_i in group1, and $\overline{KO_{i_{g2}}}$ is the average abundance of KO_i in group2.

In a correlation analysis (two-group, multi-group, and longitudinal design):

$$\Delta KO_i = \rho_{KO_i}$$

ρ_{KO_i} is the correlation coefficient between KO_i and the numeric variable to be examined.

Finally, assign a plus or minus sign to each Z-score:

$$Z_{KO_i} = \begin{cases} -Z_{KO_i}, & (\Delta KO_i < 0) \\ +Z_{KO_i}, & (\Delta KO_i \geq 0) \end{cases}$$

Therefore, a KO_i with a Z_{KO_i} greater than 0 is up-regulated, a KO_i with a Z_{KO_i} less than 0 is down-regulated.

(3) Scoring the pathway

We next used the Z-score of KOs to score the pathway. First, choose a pathway database as the reference. It is of particular interest to note any hierarchy relational table (e.g., KEGG, taxonomy database) can be used as a reference as long as the relationship between the upstream and downstream features (e.g., pathways and KOs) can be represented by a bipartite network (Figure 1C). For each pathway in the selected database, calculate the Z-score of pathway j (Z_{path_j}) as follows:

$$Z_{path_j} = \frac{1}{\sqrt{k_j}} \sum_{i=1}^{k_j} Z_{KO_i}$$

where Z_{KO_i} is the Z-score of KO_i within the $path_j$, and k_j denotes the total number of KOs in the $path_j$;

Next, we corrected Z_{path_j} using the background distribution of the Z-scores of all KOs ($Z_{KO_{all}} = \{Z_{KO_1}, Z_{KO_2}, \ldots, Z_{KO_j}\}$) to evaluate the significance of enrichment. Specifically, for a given pathway $path_j$ including k_j KOs, we randomly sampled the same number of KOs from the background $Z_{KO_{all}}$ and calculated the $Z_{path_null_j}$ for N times (N = 10,000 in this study[10]). We then standardized Z_{path_j} by subtracting the mean (μ_{k_j}) and dividing by the standard deviation (σ_{k_j}) of the $Z_{path_null_j}$ distribution. The standardized Z_{path_j} is the $ReporterScore_j$. The *p-value* of $ReporterScore_j$ is estimated by the above permutation. The formula for the reporter score and associated p-value are:

$$ReporterScore_j = \frac{Z_{path_j} - \mu_{k_j}}{\sigma_{k_j}}$$

$$p_value_j = \frac{\sum_{n=1}^{N} I\left(Z_{path_j}, Z_{path_null_{jn}}\right)}{N}$$

$$p_value_j = \begin{cases} p_value_j, & \left(ReporterScore_j < 0\right) \\ 1 - p_value_j, & \left(ReporterScore_j \geq 0\right) \end{cases}$$

where $Z_{path_null_j}$ have the same k to Z_{path_j}, μ_{k_j} is the mean of the randomly generated N $Z_{path_null_j}$, σ_{k_j} is the standard deviation of the randomly generated N $Z_{path_null_j}$. The $I(a, b)$ equals to 1 when $a > b$, else the $I(a, b)$ equals to 0.

Benchmark datasets

Benchmark datasets includes one example KO profile (ex_KO_profile downloaded from https://github.com/wangpeng407/ReporterScore) and 6 gene expression profiles of multiple human tissue types from the GEO database (https://www.ncbi.nlm.nih.gov/geo/).

We used these benchmark datasets with two-group or multi-group of experimental design to investigate the performance of GRSA, including similarities and differences between two working modes, between various statistical methods, and comparing GRSA with other commonly used enrichment analysis methods. Details of the datasets can be found in Table S2.

Case study datasets

Three case studies were re-analyzed using ReporterScore to demonstrate versatile applications to diverse circumstances, including the microbiome, transcriptome, and metabolome.

Skin microbiome data were generated by Wang et al. (2021)[23]. They sequenced 822 skin samples using the shotgun method and constructed the complete Human Skin Microbiome Gene Catalog (iHSMGC). A full KO profile based on KEGG database was provided at the website (https://ftp.cngb.org/pub/SciRAID/Microbiome/humanSkin_10.9M/AbundanceProfile/IHSMGC.KO.normalization.ProfileTable.gz). Metadata with details about including body

site, sex, age and cutotype was obtained via https://static-content.springer.com/esm/art%3A10.1186%2Fs40168-020-00995-7/MediaObjects/40168_2020_995_MOESM2_ESM.xlsx.

Transcriptomic data were extracted from the study by Liu et al. (2017)[30]. They investigated time-course transcriptomic profiling of cardiomyocyte differentiation derived from human hESCs and hiPSCs. The gene expression matrix is available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85331.

Metabolomic data were generated by Liang et al. (2020)[31]. They analyzed the untargeted mass-spectrum data of 784 samples from 30 pregnant women, and built a metabolic clock with five metabolites that time gestational age. The 264 identified level-1 and level-2 metabolites with HMDB IDs and their log2(intensity) can be found at https://ars.els-cdn.com/content/image/1-s2.0-S009286742030564X-mmc2.xlsx, while the original MS data is available at https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Project&ProjectID=PR000918.

Statistical analysis

All statistical analyses were done on the R 4.2.3 platform. The developed ReporterScore package (https://github.com/Asa12138/ReporterScore) was used for GRSA and visualization. Venn diagram and Venn network diagram were drawn by the pcutils package (https://github.com/Asa12138/pcutils/).

To compare the performance of different statistical methods in the two-group of experimental design, we defined a Jaccard similarity index:

$$Similarity = \frac{|method(i) \cap method(j)|}{|method(i) \cup method(j)|}$$

where method(i) (method(j)) is the number of significant pathways based on benchmark data sets enriched by different methods.

We used fuzzy c-means (FCM) clustering to explore the performance of different statistical test methods in multi-group experimental design (Figure S2D) and gene expression patterns in transcriptome (Figure 3A). FCM is an unsupervised machine-learning technique that partitions a population into groups or clusters[38]. Three methods (Elbow, Silhouette, and Gap statistic) were used to determine the optimal number of clusters. In FCM, the membership score is the probability that a feature belongs to a cluster. Each feature is assigned to a cluster based on its highest membership score.

In the comparison of GRSA with other commonly used enrichment methods, we first calculated the *p-values* of features by T-test and performed adjustment of the *p-values* using the Benjamini & Hochberg (BH) method to control for False Discovery Rate (FDR). A particular threshold of BH-adjusted *p-value* <0.05 and fold-change of feature abundance >2 were set to define the DEGs list. fisher.test (Fisher) was performed by the R base package. Improved fisher.test (CP) was performed by enricher function in the clusterProfiler package with the DEGs list. Gene set enrichment analysis (GSEA) was

performed by the GSEA function in clusterProfiler package, and the T-test statistic was used as the metric for ranking[39] of GSEA. GRSA was performed with the BH-adjusted *p-values* of features by T-test. For convenience, the ReporterScore package provides interface to the above mentioned enrichment methods: KO_fisher for fisher.test, KO_enrich modified from clusterProfiler based on fisher.test, and KO_gsea modified from GSEA in clusterProfiler. These enrichment methods also support custom databases and is compatible with the format of the input data for the reporter_score function in GRSA, making it easily to make cross-comparisons.

Acknowledgments

We are grateful to our colleagues at the core facility of the Life Sciences Institute, especially the NECHO high-performance computing cluster. This research was partly supported by grant from NSFC (82173645).

Author contributions

C.P. and C.J. conceived the study. C.P. developed the R package. C.J. and C.P. collected all datasets. C.P. completed the main benchmarking and case study analyses. Q.C. and S.T. contributed to the analyses. C.P. and C.J. drafted and revised the manuscript with input from other authors.

Data availability

The source data underlying all figures are available from GitHub (https://github.com/Asa12138/GRSA_figures).

Code availability

Code is available as an open-source R package 'ReporterScore', which can be downloaded from the GitHub (https://github.com/Asa12138/ReporterScore). The main analysis scripts (the Rmarkdown format) and source data are available from GitHub (https://github.com/Asa12138/GRSA_figures).

References

1. Zhao, K. & Rhee, S. Y. Interpreting omics data with pathway enrichment analysis. *Trends in Genetics* **39**, 308–319 (2023).

2. Mubeen, S., Tom Kodamullil, A., Hofmann-Apitius, M. & Domingo-Fernández, D. On the influence of several factors on pathway enrichment analysis. *Briefings in Bioinformatics* **23**, bbac143 (2022).

3. Patil, K. R. & Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2685–2689 (2005).

4. Zhu, F. *et al.* Metagenome-wide association of gut microbiome features for schizophrenia. *Nature Communications* **11**, 1612 (2020).

5. Liu, L., Zhu, R. & Wu, D. Misuse of reporter score in microbial enrichment analysis. *iMeta* **2**, e95 (2023).

6. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass.))* **2**, 100141 (2021).

7. Naylor, D. *et al.* Deconstructing the Soil Microbiome into Reduced-Complexity Functional Modules. *mBio* **11**, 10.1128/mbio.01349–20 (2020).

8. Reimand, J. *et al.* Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, cytoscape and EnrichmentMap. *Nature Protocols* **14**, 482–517 (2019).

9. Zhu, C.-M. *et al.* Alternate succession of aggregate-forming cyanobacterial genera correlated with their attached bacteria by co-pathways. *Science of The Total Environment* **688**, 867–879 (2019).

10. Oliveira, A. P., Patil, K. R. & Nielsen, J. Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Systems Biology* **2**, 17 (2008).

11. Yang, Q. *et al.* Pathway enrichment analysis approach based on topological structure and updated annotation of pathway. *Briefings in Bioinformatics* **20**, 168–177 (2019).

12. Kanehisa, M. & Sato, Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Science* **29**, 28–35 (2020).

13. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* **44**, D457–462 (2016).

14. Huerta-Cepas, J. *et al.* eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**, D309–D314 (2019).

15. Grech, V. & Calleja, N. WASP (Write a Scientific Paper): Parametric vs. Non-parametric tests. *Early Human Development* **123**, 48–49 (2018).

16. Nearing, J. T. *et al.* Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications* **13**, 342 (2022).

677 17. Nguyen, T.-M., Shafi, A., Nguyen, T. & Draghici, S. Identifying significantly
678 impacted pathways: A comprehensive review and assessment. *Genome Biology* **20**, 203
679 (2019).

680 18. Ma, J., Shojaie, A. & Michailidis, G. A comparative study of topology-based
681 pathway enrichment analysis methods. *BMC Bioinformatics* **20**, 546 (2019).

682 19. Ocvirk, S. & O'Keefe, S. J. D. Dietary fat, bile acid metabolism and colorectal
683 cancer. *Seminars in Cancer Biology* **73**, 347–355 (2021).

684 20. La Vecchia, S. & Sebastián, C. Metabolic pathways regulating colorectal cancer
685 initiation and progression. *Seminars in Cell & Developmental Biology* **98**, 63–70 (2020).

686 21. Zhang, Q. *et al.* Signaling pathways and targeted therapy for myocardial
687 infarction. *Signal Transduction and Targeted Therapy* **7**, 78 (2022).

688 22. Thomann, A. K. *et al.* Depression and fatigue in active IBD from a microbiome
689 perspectivea Bayesian approach to faecal metagenomics. *BMC Medicine* **20**, 366 (2022).

690 23. Li, Z. *et al.* Characterization of the human skin resistome and identification of two
691 microbiota cutotypes. *Microbiome* **9**, 47 (2021).

692 24. Diehl, F. F. *et al.* Nucleotide imbalance decouples cell growth from cell
693 proliferation. *Nature Cell Biology* **24**, 1252–1264 (2022).

694 25. Wang, X., Quinn, P. J. & Yan, A. Kdo2-lipid A: Structural diversity and impact
695 on immunopharmacology. *Biological Reviews* **90**, 408–427 (2015).

696 26. Heyes, D. J. *et al.* Structure-based Mechanism of CMP-2-keto-3-deoxymanno-
697 octulonic Acid Synthetase. *Journal of Biological Chemistry* **284**, 35514–35523 (2009).

698 27. Wang, S. T., Neo, B. H. & Betts, R. J. Glycosaminoglycans: Sweet as Sugar
699 Targets for Topical Skin Anti-Aging. *Clinical, Cosmetic and Investigational*
700 *Dermatology* **14**, 1227–1246 (2021).

701 28. Shin, J.-W. *et al.* Molecular Mechanisms of Dermal Aging and Antiaging
702 Approaches. *International Journal of Molecular Sciences* **20**, 2126 (2019).

703 29. Oh, J.-H., Kim, Y. K., Jung, J.-Y., Shin, J.-E. & Chung, J. H. Changes in
704 glycosaminoglycans and related proteoglycans in intrinsically aged human skin in vivo.
705 *Experimental Dermatology* **20**, 454–456 (2011).

706 30. Liu, Q. *et al.* Genome-Wide Temporal Profiling of Transcriptome and Open
707 Chromatin of Early Cardiomyocyte Differentiation Derived From hiPSCs and hESCs.
708 *Circulation Research* **121**, 376–391 (2017).

709 31. Liang, L. *et al.* Metabolic dynamics and prediction of gestational age and time to
710 delivery in pregnant women. *Cell* **181**, 1680–1692.e15 (2020).

711 32. Chrousos, G. P., Torpy, D. J. & Gold, P. W. Interactions between the
712 Hypothalamic-Pituitary-Adrenal Axis and the Female Reproductive System: Clinical
713 Implications. *Annals of Internal Medicine* **129**, 229–240 (1998).

714 33. Levitz, M., Kadner, S. & Young, B. K. Intermediary metabolism of estriol in
715 pregnancy. *Journal of Steroid Biochemistry* **20**, 971–974 (1984).

716 34. Makieva, S., Saunders, P. T. K. & Norman, J. E. Androgens in pregnancy: Roles
717 in parturition. *Human Reproduction Update* **20**, 542–559 (2014).

718 35. Cox, B. D. & Calame, D. P. Changes in Plasma Amino Acid Levels during the
719 Human Menstrual Cycle and in Early Pregnancy. A Preliminary Report. *Hormone and*
720 *Metabolic Research* **10**, 428–433 (1978).

721 36. Hamblin, R., Coulden, A., Fountas, A. & Karavitaki, N. The diagnosis and
722 management of Cushing's syndrome in pregnancy. *Journal of Neuroendocrinology* **34**,
723 e13118 (2022).

724 37. As, G. *et al.* Association with pregnancy increases the risk of local recurrence but
725 does not impact overall survival in breast cancer: A case-control study of 87 cases. *Breast*
726 *(Edinburgh, Scotland)* **30**, (2016).

727 38. Bezdek, J. C., Ehrlich, R. & Full, W. FCM: The fuzzy c-means clustering
728 algorithm. *Computers & Geosciences* **10**, 191–203 (1984).

729 39. Zyla, J., Marczyk, M., Weiner, J. & Polanska, J. Ranking metrics in gene set
730 enrichment analysis: Do they matter? *BMC Bioinformatics* **18**, 1–12 (2017).