Journal of Biomedical Informatics 66 (2017) 32–41

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

kMEn: Analyzing noisy and bidirectional transcriptional pathway responses in single subjects

Qike Li[a,b,c,d,1], A. Grant Schissler[a,b,c,d,1], Vincent Gardeux[a,b,c], Joanne Berghout[a,b,c], Ikbel Achour[a,b,c], Colleen Kenost[a,b,c], Haiquan Li[a,b,c,*], Hao Helen Zhang[d,e,*], Yves A. Lussier[a,b,c,d,f,g,*]

[a] *Center for Biomedical Informatics and Biostatistics, The University of Arizona, Tucson, AZ 85721, USA*
[b] *Bio5 Institute, The University of Arizona, Tucson, AZ 85721, USA*
[c] *Department of Medicine, The University of Arizona, Tucson, AZ 85721, USA*
[d] *Graduate Interdisciplinary Program in Statistics, The University of Arizona, Tucson, AZ 85721, USA*
[e] *Department of Mathematics, The University of Arizona, Tucson, AZ 85721, USA*
[f] *University of Arizona Cancer Center, The University of Arizona, Tucson, AZ 85721, USA*
[g] *Institute for Genomics and Systems Biology, The University of Chicago, IL 60637, USA*

ARTICLE INFO

Article history:
Received 17 June 2016
Revised 28 October 2016
Accepted 16 December 2016
Available online 19 December 2016

Keywords:
k-means clustering
Single subject analysis
Precision medicine
Pathway analysis
N-of-1-pathways
HIV treatment response

ABSTRACT

Motivation: Understanding dynamic, patient-level transcriptomic response to therapy is an important step forward for precision medicine. However, conventional transcriptome analysis aims to discover cohort-level change, lacking the capacity to unveil patient-specific response to therapy. To address this gap, we previously developed two N-of-1-*pathways* methods, Wilcoxon and Mahalanobis distance, to detect unidirectionally responsive transcripts within a pathway using a pair of samples from a single subject. Yet, these methods cannot recognize bidirectionally (up and down) responsive pathways. Further, our previous approaches have not been assessed in presence of background noise and are not designed to identify differentially expressed mRNAs between two samples of a patient taken in different contexts (e.g. cancer vs non cancer), which we termed responsive transcripts (RTs).
Methods: We propose a new N-of-1-*pathways* method, k-Means Enrichment (kMEn), that detects bidirectionally responsive pathways, despite background noise, using a pair of transcriptomes from a single patient. kMEn identifies transcripts responsive to the stimulus through k-means clustering and then tests for an over-representation of the responsive genes within each pathway. The pathways identified by kMEn are mechanistically interpretable pathways significantly responding to a stimulus.
Results: In ~9000 simulations varying six parameters, superior performance of kMEn over previous single-subject methods is evident by: (i) improved precision-recall at various levels of bidirectional response and (ii) lower rates of false positives (1-specificity) when more than 10% of genes in the genome are differentially expressed (background noise). In a clinical proof-of-concept, personal treatment-specific pathways identified by kMEn correlate with therapeutic response (p-value < 0.01).
Conclusion: Through improved single-subject transcriptome dynamics of bidirectionally-regulated signals, kMEn provides a novel approach to identify mechanism-level biomarkers.

1. Introduction

Precision medicine requires a deep understanding of disease mechanisms at the level of the individual patient. However, the tools available for transcriptome analysis have lagged, relying on cohort-based statistics and averaged dynamic responses across multiple individuals. In addition, transcriptional signatures at the level of specific differentially expressed genes have often been difficult to reproduce or interpret [2]. Gene set (or pathway) oriented methods including gene set enrichment analysis (GSEA) and differentially expressed genes (DEG) followed by gene set enrichment (DEG + Enrichment) [3–5] have become popular ways to provide more robust and biologically interpretable disease mechanisms. However, these methods still rely on between-groups cohort-based calculations to provide an initial input list or ranking of genes prior to enrichment analyses. Nonetheless, defining gene

[*] Corresponding authors at: Center for Biomedical Informatics and Biostatistics, The University of Arizona, Tucson, AZ 85721, USA (H. Li and Y.A. Lussier). Graduate Interdisciplinary Program in Statistics, The University of Arizona, Tucson, AZ 85721, USA (H.H. Zhang).

E-mail addresses: haiquan@email.arizona.edu (H. Li), hzhang@math.arizona.edu (H.H. Zhang), yves@email.arizona.edu (Y.A. Lussier).

URL: http://www.lussierlab.org/publications/kMEn/ (Y.A. Lussier).

[1] These authors contributed equally to the work.

http://dx.doi.org/10.1016/j.jbi.2016.12.009
1532-0464/© 2016 Published by Elsevier Inc.

sets based on Gene Ontology (GO; [6]) provides a major advantage for mechanistic interpretation, while also reducing the number of studied features (pathways vs. gene level). To move from the 'average' responsive pathway in a cohort to patient-specific signals requires a new approach.

Several studies have begun to develop pathway-based analyses that apply to expression data derived from a single-patient sample. In the beginning of genome-wide expression analyses (early 2000s), we and others have tried to identify differentially expressed genes using simple gene expression fold change (FC) between two paired samples with an arbitrary cutoff. However, due to highly inaccurate measurements of expression platforms, particularly in low expression quantities, FC was found exceedingly uninformative in subsequent biological validation. Abandoning simplistic FC, Bottomly et al. compared a single transcriptome to a cohort reference transcriptome [7]. We conceived Functional Analysis of Individualized Microarray Expression (FAIME; Table 1) [8] to score pathways within a single sample, moving away from cohort-based approaches. Further, a new Gene Set Enrichment Analysis software was also designed to score pathways within a single sample (ssGSEA; Table 1), which is provided in the main GSEA portal (http://software.broadinstitute.org/gsea). ssGSEA remains a non-published, non-reviewee portal software, without formal evaluation accounts for its performance. However, these two "static expression" methods (FAIME, ssGSEA) were designed to make inferences from only one transcriptome and report expression of a pathway as compared to the background expression of the same sample (Table 1). In addition, the normal expression levels of some pathways are expected to be lower (or higher) than the average expression in a context-specific manner. These expression analyses of a static transcriptome cannot detect dynamic transcriptome changes such as those arising from treatment as measured from a change from baseline (control). While the DNA genome can be contrasted with a reference genome and yield meaningful interpretations for a single subject in the context of precision medicine (e.g. missense mutation associated to a Mendelian Disease), the transcriptome integrates the dynamic expression of the genome and epigenome over time and space. Therefore, for increased utility in precision medicine, new analytic frameworks providing meaningful interpretation of the dynamic changes of the transcriptome in the context of response to therapy or disease progression are required.

We conceived the N-of-1-*pathways* framework to analyze a pair of samples from a single patient [1,9–12] providing a personal transcriptome profile describing pathway-level responses. Under this framework, the response of a pathway is an accumulation of the gene level evidence, thereby mitigating the noise and artifacts inherent to the lack of replicates. Importantly, inferences are made based on the information from a single patient and thus are truly personalized. Current cohort-based methods (e.g. DEG + Enrichment and GSEA) require multiple replicates and therefore are not applicable in single-subject analysis when no intra-patient replicate is available. Existing N-of-1-*pathways* approaches can only detect concordant regulation of transcript expression between the two samples: the majority being either up- or downregulated within a pathway (Table 1).

This study introduces a novel method within the N-of-1-*pathways* framework using k-Means clustering [13] of transcript fold change (FC) followed by gene set Enrichment (kMEn) analysis. We demonstrate that kMEn enables bidirectional response detection as well as unidirectional pathway responses while remaining robust against overall transcriptome variability (background noise) (Table 1). kMEn outperforms the other N-of-1-*pathways* methods in two simulation studies. Then, using a clinical case study on publicly available data, we applied kMEn to identify patient-level transcriptional pathway response to antiretroviral therapy in 20 HIV-infected individuals.

2. Methods

Fig. 1 and Table 2 present an overview of the kMEn approach and the list of acronyms used in this study, respectively.

2.1. Datasets

Transcriptome datasets. Simulation studies were based on RNA-sequencing data from seven biological replicates of the MCF7 breast cancer cell line (Gene Expression Omnibus, GSE51403; [15]), which allowed us to estimate the expression level and variation of each gene. These seven biological replicates were sequenced by Illumina HiSeq 2000. The clinical case study was performed on microarray data from peripheral blood mononuclear cells (PBMCs) isolated from 20 HIV-infected patients before and 48-weeks after antiretroviral treatments (Gene Expression Omnibus, GSE44228) [16]. 12 patients were treated with non-nucleoside reverse transcriptase inhibitor (NNRTI) and 8 with protease inhibitor (PI). An additional 12 patients treated by both medications were not included. This dataset also included the peripheral $CD4^+$ T-cell counts for each patient and timepoint.

Knowledge-base datasets. GO was used to provide functional annotations of genes into gene sets (pathways). Biological process

Table 1
Methods comparison to the application of single-subject pathway analysis.

		Bidirectional response detection	Background noise adjustment[b]	Discovery of DEGs	Transcriptome dynamics from a pair of samples	Precision recall	Original publication
Single-subject pathway analysis	N-of-1-*pathways* kMEn	Yes	Yes	Yes	Yes	+++	Current Manuscript
	N-of-1-*pathways* MD	No	No	No	Yes	++	Yes
	N-of-1-*pathways* Wilcoxon	No	No	No	Yes	++	Yes
	ssGSEA[a]	No	Yes	No	No	N/A	No
	FC + ssGSEA[a]	No	Yes	No	Yes	+	No
	FAIME	No	Yes	No	No	N/A	Yes
Cohort-based Pathway analyses	DEG + Enrichment	N/A	N/A	N/A	N/A	N/A	N/A
	GSEA	N/A	N/A	N/A	N/A	N/A	N/A

N/A indicates not applicable; cohort-based methods, such as DEG + Enrichment and GSEA, cannot be applied to single-subject analysis, and therefore all assessments are not applicable. ++ indicates moderate accuracy by the measure of precision-recall;+ indicates low accuracy by the measure of precision-recall. +++ indicates high accuracy by the measure of precision recall. Note: empty cells imply the lack of the corresponding feature.

[a] FC + ssGSEA is a new application of ssGSEA using the fold change expression of a gene across two paired samples, rather than static gene expression on one sample (as intended by ssGSEA authors). We previously conceived FC + ssGSEA, and we have shown that is has lower accuracy than N-of-1-*pathways* Wilcoxon [10]. ssGSEA as described on the GSEA portal is not applicable to paired samples (of note, ssGSEA was never formally published or evaluated).

[b] Background noise adjustment is achieved by genome-wide competitive modeling [14].

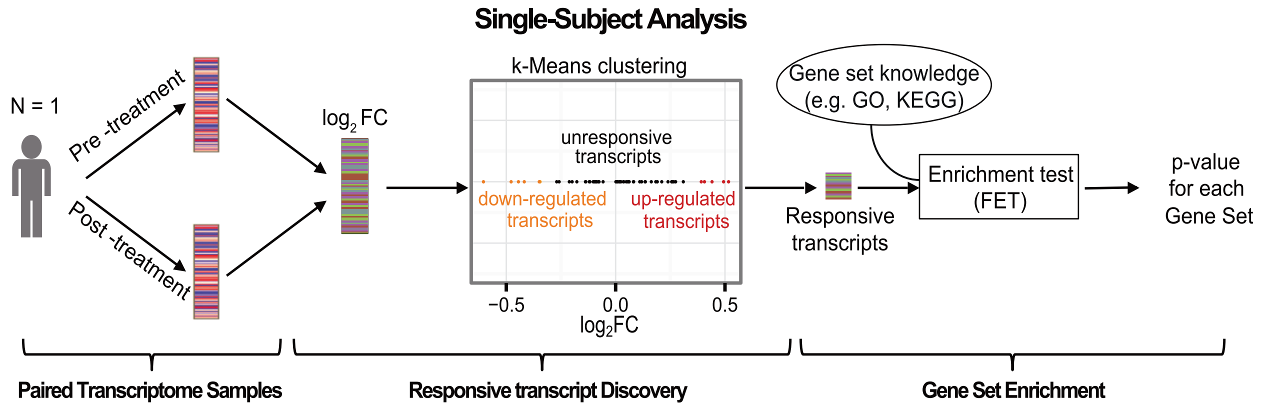

Fig. 1. N-of-1-*pathways* kMEn overview. The transcript expression measurements of each single-subject paired sample are used to calculate the fold change (FC) between two samples. The k-means-based clustering of the FC values was then used to partition transcripts into responsive (either up or down) versus nonresponsive. An enrichment test was subsequently applied on the responsive transcripts within each pathway using Fisher's Exact Test, controlling for multiple comparisons. The term "responsive transcripts" (RTs) refers to the transcripts changing across conditions but derived from single-subject analysis, while "differentially expressed genes (transcripts)" (DEGs) pertains to those derived from analysis of a cohort.

Table 2
Acronyms and definitions.

Acronym	Description
AUC	Area Under the precision-recall Curve
CD4$^+$ FC	CD4$^+$ T cell counts (Fold change of)
Cohort Expectation Standard	Proxy Gold Standard derived from cohort statistics to evaluate single subject analyses
DEG	Differentially Expressed Gene (transcripts)
FC	*Fold change*
FDR$_{BH}$	Controlled False Discovery Rate (FDR) using Benjamini-Hochberg procedure
FDR$_{BY}$	Controlled FDR using Benjamini-Yekutieli procedure
FET	Fisher's Exact Test
GO	Gene Ontology
GO-BP	Gene Ontology Biological Processes
HIV	Human Immunodeficiency Virus
kMEn	N-of-1-*pathways* kMeans Enrichment
MD	N-of-1-*pathways* Mahalanobis Distance
NNRTI	Non-Nucleoside Reverse Transcriptase Inhibitor
NRTs	*Unresponsive transcripts*
OR	Odds Ratio
PBMCs	Peripheral Blood Mononuclear Cells
PCA	Principal Component Analysis
PI	Protease Inhibitor
RTs	*Responsive transcripts (differentially expressed between two samples of a single subject)*
SAM	Significant Analysis of Microarrays
Wilcoxon	N-of-1-*pathways* Wilcoxon

(GO-BP) containing 15 to 500 genes annotations were included to afford comparison to previous N-of-1-*pathways* methods. Files were downloaded in March 2013 using R Bioconductor org.Hs.eg.dg package [17]. Note, we use GO-BP to define gene sets when testing N-of-1-*pathways* methods, and the term 'gene set' and 'pathway' are used interchangeably.

2.2. N-of-1-pathways kMEn algorithm

kMEn identifies responsive transcripts and prioritizes pathways as follows:

- Using the absolute value of the log-transformed expression fold change ($|\log_2(\mathrm{FC})|$) between two samples (e.g. paired samples, before and after treatment, from the same subject), every transcript was clustered into two groups (k = 2), 'biologically responsive' and 'biologically unaltered', by the nonparametric clustering algorithm k-Means (Eq. (1)) [13]. The k-Means algorithm minimizes the within-cluster differences while maximizing the cross-cluster differences via the objective function (Eq. (1)),

$$\arg\min_{G} \sum_{k=1}^{2} \sum_{X_i \in G_k} \|X_i - \mu_k\|^2 \tag{1}$$

where $argmin_G$ finds the partition G that minimizes the objective function, $\|X_i - \mu_k\|$ is the Euclidean distance, $X_i = |\log_2(\mathrm{FC})_i|$ of transcript$_i$, and μ_k is the arithmetic mean of $|\log_2(\mathrm{FC})|$ in transcript cluster G_k. The transcripts cluster with the highest mean is defined as 'responsive transcripts' (RTs) and the other cluster as 'unreponsive transcripts' (NRTs).

- Within the 'responsive' cluster, transcripts with positively-signed $\log_2(\mathrm{FC})$ are annotated as upregulated and similarly transcripts with negative $\log_2(\mathrm{FC})$ as downregulated (Fig. 1).
- Each gene set is tested for enrichment of responsive transcripts (using R function *fisher.test*). Specifically, a Fisher's Exact Test (FET) [18] is conducted on a 2 × 2 contingency table of genes (responsive or unresponsive vs. in the pathway or not). The test results in a nominal p-value which is corrected for multiple comparisons via Benjamini-Yekutieli (FDR$_{BY}$) [19].

2.3. Simulation study: comparing N-of-1-pathways methods in the presence of bidirectional response and background noise in synthetic pathways (Figs. 2 and 3)

Two simulation studies – a precision-recall assessment (Simulation 1) and a false positive rate comparison (Simulation 2) - were designed to evaluate and compare the accuracy of N-of-1-*pathways* kMEn to that of two previously published single-subject transcriptome response methods. The Wilcoxon signed-rank test (Wilcoxon) was designed to determine a difference in central tendency of transcript expression within a given pathway and successfully predicted cancer survival outcomes [10]. The Mahalanobis Distance (MD) overcame statistical shortcomings of the Wilcoxon approach while providing a clinically relevant metric of pathway response in breast cancer [12]. We could not compare results to cohort-based transcriptome analyses such as GSEA [4], as they require calculations from multiple distinct subjects and were not applicable to the single-subject design of this study.

The Negative Binomial distribution, NB(μ_g, ϕ_g), was used to simulate the expression counts of each transcript (noted "g") [20,21]. The mean (μ_g) and overdispersion (ϕ_g) were estimated from the seven biological replicates of breast cancer (Methods Section 2.1;

Table 3
Parameters for generating simulated transcriptomes. The terms of 'gene set' and 'pathway' were used interchangeably referring to biological mechanisms.

Parameter	Description	Parameter values
path.size	Number of transcripts per pathway	{5, 10, 15 to 490 by steps of 25, 500}
path.pct.dys	% of responsive transcripts in the target pathway	{5% to 100% by steps of 5%}
path.FC	FC expression of responsive transcripts within a pathway	{1.5, 2, 4}
pct.up	% of upregulated transcripts among responsive transcripts	{0% to 50% by steps of 10%}
pct.bkgrd.noise	Background noise of the whole transcriptome	{1%, 5%, 10%, 20%}
FC.bkgrd	Transcript FC of the background between two simulated sets	{1.5, 2, 4}

[15]). In total, there are 8280 scenarios for the bidirectional response simulation (Methods Section 2.3.1) and 299 scenarios pertaining to a background transcript noise (prevalence of responsive genes in the whole transcriptome; Methods Section 2.3.2) simulation. These scenarios explore distinct settings of parameters used in the simulation, as each simulated scenario corresponds to a unique combination of parameters (Table 3).

2.3.1. Simulation study 1: exploring the impact of bidirectional pathway responses

The first simulation study assesses precision and recall of the three N-of-1-*pathways* methods under bidirectionally responsive pathways. 8280 distinct scenarios resulting from all combinations of the first four parameters in Table 3 were explored. Explicitly, the simulation replicates were generated as follows:

Step 1. Estimate parameters μ_g and ϕ_g for the Negative Binomial (NB) distribution of each transcript *g*, using a breast cancer cell line dataset via method of moments. Since each of the seven replicates in the dataset has a library size approximately equal to 25.5×10^6, library size normalization was not necessary.
Step 2. Generate a pair of transcriptomes by randomly generating two realizations of NB(μ_g, ϕ_g) for every transcript *g*. One transcriptome is defined as baseline and the other as case.
Step 3. Fix a combination of parameters (*n* = *path.size*, *p* = *path.pct.dys*, *f* = *path.FC*, *u* = *pct.up*).
Step 4. Randomly select without replacement *n* transcripts to synthetically create a pathway.
Step 5. Randomly select without replacement *p* percent of the *n* transcripts to designate the responsive transcripts within a pathway when comparing the two transcriptomes. Let $m = p \times n$ be the number of responsive transcripts.
Step 6. Randomly select without replacement *u* percent of the *m* responsive transcripts to designate the transcripts that will be upregulated (i.e., expression will be multiplied by the factor *f*). The remaining (100 - *u*)% responsive transcripts are designated as downregulated and, thus, multiplied by 1/*f*.
Step 7. Replace the expression of responsive transcripts in the case transcriptome with $f \times$ (baseline transcript expression) and $1/f \times$ (baseline transcript expression), for up- and downregulated transcripts, respectively.
Step 8. Repeat Step 4 to Step 7 one hundred times to produce 100 simulation replicates.
Step 9. Randomly select without replacement *n* transcripts to serve as a negative control pathway. Retain the paired transcript expression for these *n* transcripts from the values generated in Step 2. Repeat this procedure 100 times to create a balanced number of positive and negative cases of pathway response. This produces 200 simulated replicates for this combination of parameters *n*, *p*, *f*, and *u*.
Step 10. Repeat Step 3 to Step 9 for each of 8280 distinct combinations of parameters.

The Area Under the precision-recall Curve (AUC) was calculated for each of the 8280 distinct combinations of parameters by varying the p-value cutoffs from 0 to 1 produced by each N-of-1-*pathways* method (R function provided: http://www.lussierlab.org/publications/kMEn/). For MD, we used sampling with replacement of the transcripts within a pathway to create a bootstrapped distribution [12]. Using the R *ggplot2* package [22], box plots of the combined AUCs were plotted (Fig. 2).

2.3.2. Simulation study 2: studying the false positive rate in the presence of background noise

299 pathway dysregulation scenarios were used to compare the false positive rate of the three methods in the presence of background noise. The three dysregulation parameters *path.pct.dys*, *path.FC*, and *pct.up* are set to zero and varied by path.size and by the two noise parameters (*pct.bkgrd.noise*, *FC.bkgrd*; Table 3). The simulated replicates are generated as follows:

Step 1 & 2. Same as the Step 1 and 2 in **2.3.1** respectively.
Step 3. Fix a combination of parameters (*n* = *path.size*, *b* = *pct.bkgrd.noise*, *f* = *FC.bkgrd*).
Step 4. Randomly select without replacement *b* percent transcripts to designate responsive transcripts.
Step 5. Replace the expression of the responsive transcripts in the case transcriptome with $f \times$ (baseline transcript expression). This guarantees the simulated fold change (background noise) is exactly *f*.
Step 6. Randomly select without replacement *n* transcripts to synthetically create a pathway.
Step 7. Repeat Step 4 to Step 6 one hundred times to produce 100 simulation replicates.
Step 8. Repeat Step 3 to Step 7 for all 299 distinct combinations of parameters.

As no pathway-specific fold change is induced, any pathway found responding by an analytical method is considered a false positive result. In Fig. 3, the false positive rate (1-specificity) was calculated using the proportion of the falsely identified pathways (p-value < 0.05; except MD which uses a conservative decision rule: proxy p-value = 0% as described by Schissler et al. [12]). Note, *pct.up* was set to equal 0 to avoid confounding effects from bidirectional responses.

2.4. Clinical case study: the individualized response to therapy in HIV-infected subjects using kMEn

N-of-1-*pathways* kMEn was applied to analyze the transcriptomic profile change of 20 HIV-infected patients who had been treated with antiretroviral therapy [16]. Patients were adult men (19–58 years old) with varied ethnicity who had not been on any antiretrovirals prior to this study; see reference by Massanella et al. for full clinical and demographic descriptions. RNA was prepared from PBMCs collected prior to treatment and 48 weeks later when all exhibited successful viral suppression and some degree of T-cell recovery. 8 patients were treated with nucleoside reverse transcriptase inhibitors (NRTIs) in combination with a protease inhibitor (PI) and 12 patients were treated with NRTIs in combina-

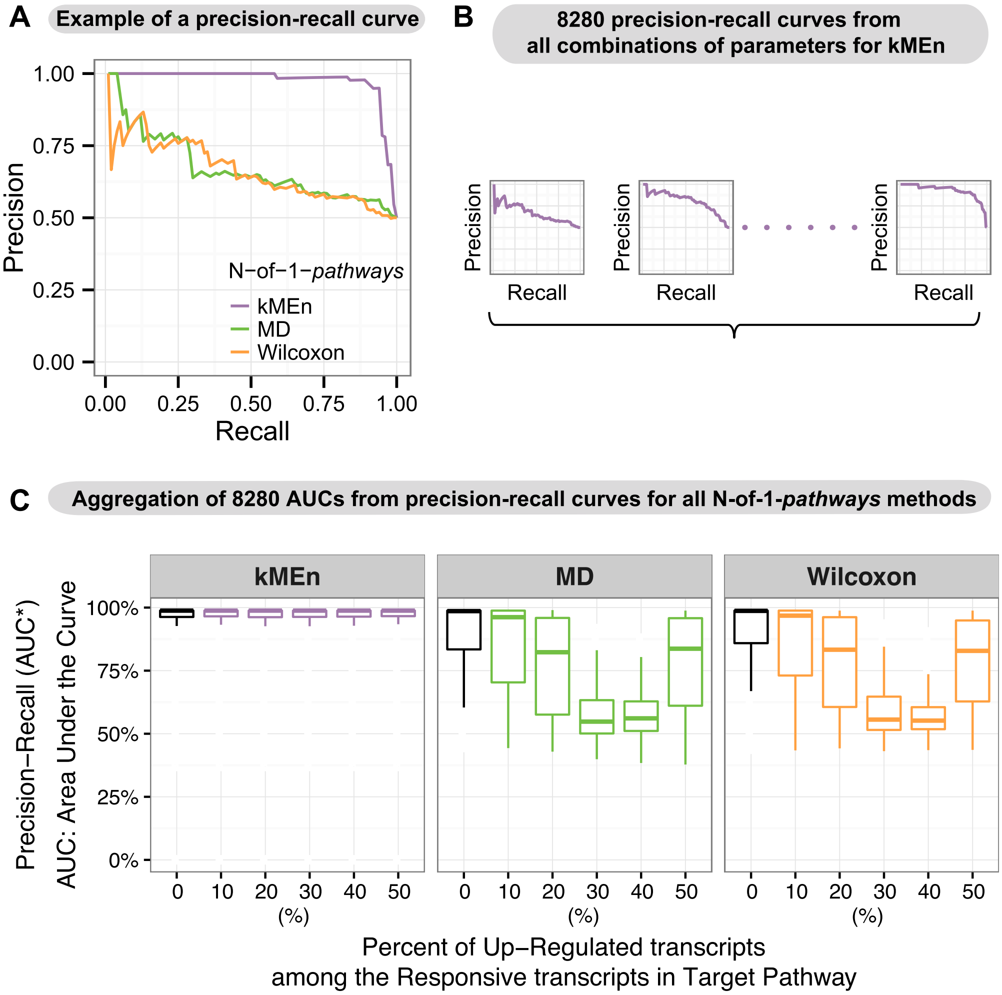

Fig. 2. N-of-1-*pathways* kMEn detects both unidirectionally and bidirectionally responsive pathways. Datasets were simulated using 8280 combinations of parameters. (A) Example of precision-recall curve of the three N-of-1-*pathways* methods is presented (*path.size* = 115 *transcripts*, *path.FC* = 1.5, *path.pct.dys* = 10%; Table 3). (B) This panel illustrates precision-recall curves resulting from 8280 distinct combinations of parameters (Methods Section 2.3) for kMEn; each provides an AUC that can be aggregated into boxplots in Panel C. (C) Box plots show the uniformly superior performance of kMEn especially when the pathways were simulated with bidirectional responsive transcripts (colored boxplots).

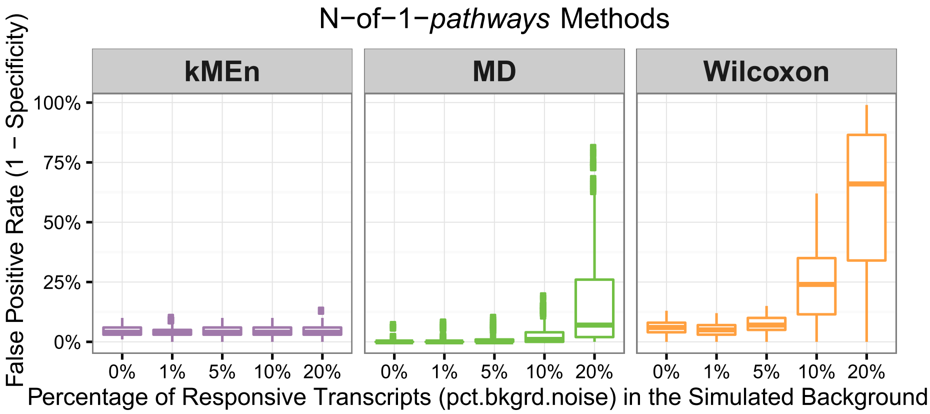

Fig. 3. kMEn is resistant to background noise. The box plot shows that the false positive rate of kMEn (Methods Section 2.3) is unaffected by the level of background noise, in contrast to Wilcoxon and MD. 299 different simulation cases were considered for each level of background noise.

tion with a non-nucleoside reverse transcriptase inhibitor (NNRTI). RNA samples were hybridized to Illumina HumanWG-6v3 Expression BeadChip microarrays and transcriptomes were deposited in Gene Expression Omnibus under accession GSE44228 where we retrieved them. Gene expression values in the HIV dataset were $\log_2$ transformed and normalized using robust spline normalization [23].

N-of-1-*pathways* kMEn was performed on each pair of PBMC transcriptomes (pre- vs. post-treatment) to identify responsive transcripts and GO-BP pathway terms. Pathway p-values were converted into 'Z-scores' using Eq. (2), where ϕ^{-1} is the inverse function of the standard normal cumulative distribution function. As the gene set p-values provided by kMEn are generated from Fisher's Exact Test, Z-scores are determined by the odds ratio (OR; Eq. (3)) of responding pathway genes after adjusting for sampling error and are clinically relevant.

$$Z_{pathway} = -\Phi^{-1}\ (p\text{-}value) \tag{2}$$

$$OR_{pathway} = \frac{a \times d}{b \times c} \tag{3}$$

where a = number of responsive transcripts within the pathway, b = number of nonresponsive transcripts within the pathway, c = number of responsive transcripts outside the pathway, and d = number of nonresponsive transcripts outside the pathway.

2.4.1. Validation of identified responsive pathways from kMEn by comparison to a conventional cohort-based approach and visualization in a similarity Venn Diagram (Fig. 4)

A real gold standard is unfeasible as it would require biological testing of all pathways to infer true positives as well as true negatives. Here, we propose an alternative: cohort-expectation standard. We generated the cohort expectation standard list of GO terms for each treatment (pre- vs. post-treatment for NNRTI; pre- vs. post-treatment for PI) using a conventional cohort-based DEG + Enrichment approach. We used the Significance Analysis of Microarrays (SAM) algorithm, adjusted for multiple comparison via Benjamini-Hochberg (FDR_{BH}), to identify differentially expressed transcripts (i.e., DEGs; $FDR_{BH} < 5\%$) between the pre-treatment and post-treatment samples [24]. Fisher's Exact Test [18] was used to determine enrichment of DEGs among GO-BPs using $FDR_{BY} < 5\%$ a [19]. GO terms identified by kMEn were compared to those found by SAM + Enrichment.

Since distinct GO-BP terms may share genes and proximity in the GO hierarchical classification, GO-BP functional similarity was also employed to unbiasedly identify relatedness of findings between the cohort-expectation standard and results [25]. The GO-BP functional similarity was quantified by information theoretic similarity (ITS). ITS was calculated on each distinct pair among the 3219 GO terms with ⩾15 and ⩽500 gene annotations, leading to 5,182,590 pairs of which 31,117 (≈6 out of 1000) have an ITS ⩾ 0.7, an *a priori* cutoff for significance previously described [25–29].

2.4.2. Correlations between GO-BP scores and CD4$^+$ T cell fold changes and single GO-BP investigation at transcript level correlations (Fig. 5)

Treatment-specific responsive pathways (i.e., 'differentially responsive') were calculated based on the Z-scores of each GO-BP (Eq. (2)) generated by kMEn applied to pre-treatment and post-treatment samples of each patient. The Z-scores of prioritized pathways were assessed for differences between NNRTI-treated and PI-treated groups of HIV-infected patients by Welch's *t*-test. Pathways with a nominal p-value <5% were regarded as treatment-specific pathways. Spearman's rank correlation [30]

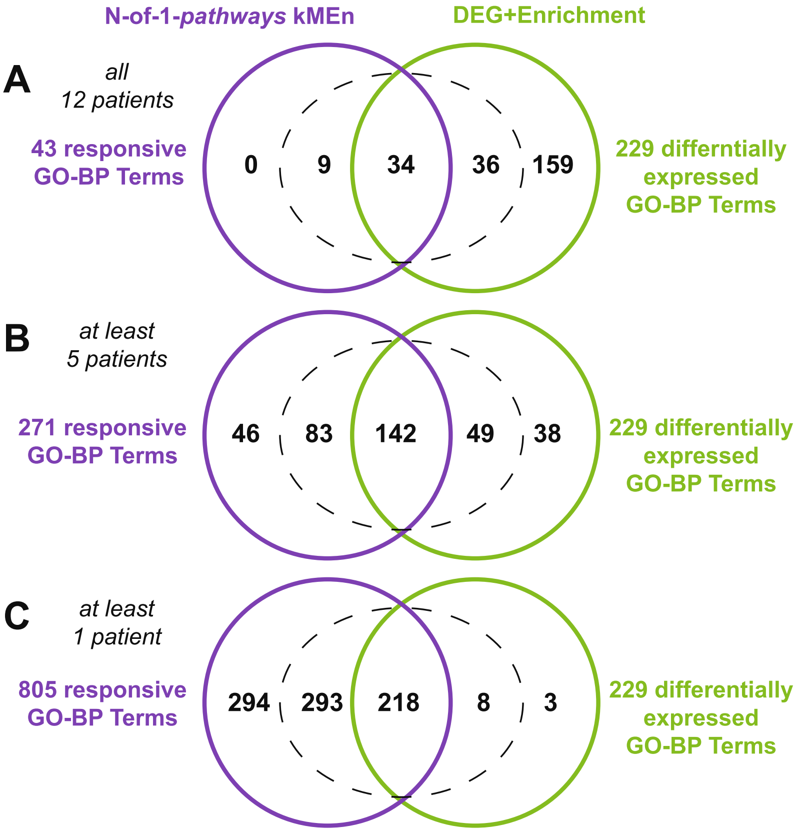

Fig. 4. N-of-1-*pathways* kMEn recovers pathways of the cohort-expectation standard. Information-theoretic-adjusted Venn diagrams were used to visualize the overlap and the similarity between GO-BP terms found by single-subject analytic kMEn and those found by cross-patient analyses of the 12 NNRTI patients (DEG + Enrichment, cohort-expectation standard, see Methods Section 2.4.1). The dashed oval indicates GO-BP terms that are highly similar assessed by ITS, but do not overlap. For example, in Panel A, there are 36 GO-BP terms found by DEG + Enrichment that were not found by kMEn; however, these 36 pathways are functionally similar to at least one of the 43 pathways found by kMEn. (A) All 43 pathways consistently found in every patient by kMEn are similar to terms in the cohort-expectation standard (similarity-based [1] OR = 5.5, $p < 10^{-10}$). (B) Over 80% (191/229) of the cohort-expectation standard is recovered by kMEn when exploring the pathways commonly responsive in at least 5 patients (similarity-based OR = 4.4, $p < 10^{-10}$). (C) Over 98% (226/229) of the cohort-expectation standard is recovered when investigating GO-BP terms found responsive in at least one patient by kMEn (similarity-based OR = 2.9, $p < 10^{-10}$).

coefficients were calculated between the treatment-specific pathway Z-scores and the fold changes of CD4+ T cell count (CD4+ FC, Eq. (4); Sup. Table 1).

$$[CD4^{+}FC]_i = \frac{[CD4^{+}\text{ cell count after treatment}]_i}{[CD4^{+}\text{ cell count before treatment}]_i} \tag{4}$$

Principal component analysis (PCA) was conducted on the pathway scores of the treatment-specific responsive GO-BPs (R function *prcomp* from stats package) [31], and Spearman's rank correlation coefficients were calculated between the FC of CD4+ T cell count from each patient and the projection of each patient GO-BP scores on the first principal component (Fig. 5). Finally, the top five positive and top five negative correlations between GO-BPs and CD4+ FC are reported in Suppl. Table 1, sorted by the pathway contribution to the principal component (PCA loading).

Next, we identified the three patients with the smallest CD4+ T-cell fold change in response to NNRTI treatment and the three patients with the largest CD4+ T-cell fold change. In these diametric extremes, we further explored the transcripts annotated to kMEn-identified candidate GO-BP terms to determine if these can potentially serve as a group of biomolecular markers for predicting drug response. Transcript responses were compared according to expression fold change, directionality, and concordance between samples in each category.

3. Results and discussion

3.1. Simulation studies confirm increased accuracy of N-of-1-pathways kMEn

Via simulation, we compared the accuracy of the three N-of-1-*pathways* methods, kMEn, Wilcoxon, and MD, in the cases of unidirectional response or bidirectional response (Fig. 2). In addition, we tested the robustness of the three N-of-1-*pathways* methods against background noise (Fig. 3).

N-of-1-*pathways* kMEn outperforms Wilcoxon and MD in the simulation studies. As hypothesized, kMEn more accurately detects pathway response in the presence of bidirectional response of transcripts, and it is also more robust against overall transcriptome background noise than Wilcoxon or MD (Fig. 3). When the pathway is unidirectionally dysregulated, kMEn slightly outperforms the other two methods (Fig. 2). Also, the false positive rates of the three methods are comparable when the background noise is absent (Fig. 3). As MCF7 is reported to carry 69 chromosomes, which may affect gene expression, a smaller, secondary simulation study was conducted on kMEn alone. Here, we explored the robustness of kMEn's performance on a different generative dataset using the estimates of gene-specific negative binomial parameters obtained from RNA-seq datasets derived from 13 biological replicates of healthy brain tissue downloaded from GTEx (Suppl. Fig. S1). The results from this simulation closely mirror the results from the larger scale study (Figs. 2 and 3).

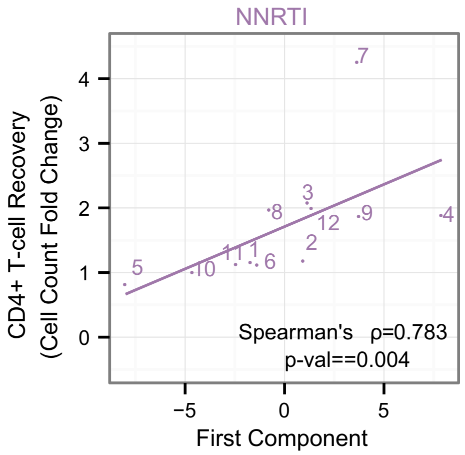

Fig. 5. HIV treatment-specific mechanisms derived from kMEn single-subject metrics correlate with response to therapy measured by Fold Change of CD4+ T cell count. Treatment-specific GO-BP mechanisms were discovered independently of the CD4+ FC biomarker of HIV response to therapy. The first principal component of kMEn pathway scores correlates with CD4+ T cell recovery (cell count FC) (Methods Section 2.4.2) for NNRTI treated-patients (above). This study was repeated for both treatments considered together (Suppl. Fig. S4). FC = cell count fold change (Eq. (4)).

N-of-1-*pathways* Wilcoxon and MD methods were designed to detect the unidirectional pathway expression change, and as a consequence, pathways with both upregulated and downregulated transcripts may result in no overall expression change. Conversely, signals of upregulated and downregulated transcripts are additive in kMEn. In addition, N-of-1-*pathways* kMEn belongs to the class of competitive gene set tests [14,32], which defines the pathway response relative to the genome background while Wilcoxon and MD are self-contained gene set tests that only analyze transcript expression within pathways. Together, these explain why background noise significantly elevates the false positive rates for Wilcoxon and MD (Fig. 3) relative to kMEn. Resistance to background noise is important in applications, such as cancer studies, since cancer genomes can acquire a large amount of passenger mutations, which do not present a cancer cell growth advantage [33].

Investigating further the testing operating characteristics (empirical false positives and power), a study of GO-BP terms identified as dysregulated among pairs of the seven MCF7 cell line biological replicates (used in the simulation study) was conducted (Suppl. Fig. S2). While both MD and kMEn both detect less false discoveries than expected at 5% FDR_{BY} (and Wilcoxon for the majority of pairs), N-of-1-*pathways* kMEn detects very few false discoveries overall. Notably, there were no common falsely discovered pathways among independent pairs as identified by kMEn. This indicates that no GO-BP term is inherently susceptible to biological-variation-induced dysregulation for these data. We also investigated AUC (Fig. 2) stability by repeating the simulation 100 times for four representative sets of parameters. The goal is to study the empirical variation in the AUC metric for kMEn. Two of the four AUC distributions are centered near 0.999 (as often occurs for kMEn) and have negligible variability across the 100 replicates of the experiment (standard error of the AUC statistic <0.002). The remaining two distributions of AUC are centered at 0.91 and 0.94 with standard errors of 0.0189 and 0.0153, respectively. As such, it appears that variation in the simulated AUC increases as it tends downward from 1 and may approach a non-negligible amount of variation for relatively small AUC. Thus the aggregation of AUC across the simulation configurations we performed could possibly be enhanced by a variation-based weights in future studies.

3.2. A clinical case study of kMEn: interpreting individualized drug response of HIV-infected patients

3.2.1. Personal pathway responses detected by N-of-1-pathways kMEn confirmed in a cohort-expectation standard (Fig 4)

As a proof-of-concept, we applied kMEn to real patient data in order to assess the relevance of kMEn-discovered pathways (GO-BP) associated with the antiretroviral treatment in HIV patients. After calculating significant GO-BP terms associated with drug response according to kMEn in each patient and DEG + Enrichment across the cohort ('cohort-expectation standard') (Methods Section 2.4.1), we compared pathways identified by both kMEn and the cohort-expectation standard (Fig. 4). We use the term cohort-expectation standard for two reasons: (i) creating a biological gold standard for thousands of pathways is biologically unfeasible

(Methods Section 2.4.1) and (ii) a real gold standard for a single subject would require a gold standard specific to each subject, which is feasible at high cost and rate limiting, as it would require patient genotype-specific biological replicates to use cohort statistics (unheard of in GEO). These are illustrated using a "Similarity Venn Diagram", a visualization method we previously developed [1]. As in traditional Venn diagrams, the Similarity Venn Diagram has two solid circles to represent the responsive GO-BP sets identified by each of the two methods with the overlapping area presenting the exact GO-BP matches contained in the two sets. The Similarity Venn Diagram also presents two counts describing the GO-BPs within one set that are similar (ITS $\geqslant$ 0.7) but not identical to those of the other set in the two regions adjacent to the middle region (area within the dotted lines). Numbers in the outmost regions represent the number of GO-BPs, which are not shared or similar (ITS < 0.7) to any GO-BPs in the opposing set.

As illustrated, kMEn found 43 responsive pathways that are functionally similar to the cohort-expectation standard in every one of the 12 NNRTI patients. This result supports the notion that pathways detected as significant by kMEn in single subjects and also consistently found in all subjects may have minimal 'false positives' or as in this case study, none. Moreover, kMEn demonstrates remarkable sensitivity recovering over 80% of the cohort-expectation standard pathways among the 271 pathways commonly responsive in at least five patients. In a less conservative analysis with 805 pathways considered as responsive in at least one patient, over 98% of responsive pathways identified by conventional methods were also identified by kMEn. Only 63% of the responsive pathways found by kMEn in a single patient could be recovered by conventional cohort-based methods.

We found similar patterns when comparing our results with kMEn to the pathways recorded from this data by Massanella [16]. Several of the pathways they highlighted were too broad to compare directly (>500 gene annotations, e.g. "cellular process"), as we decided *ab initio* to filter out such pathways from our predictions (Methods Section 2.1). kMEn was able to recover exact or highly similar (ITS $\geqslant$ 0.7) matches in at least 30% of patients for all of their cohort-level findings that did fit our parameters, additional closely related biological processes (ITS $\geqslant$ 0.7), along with GO-BP terms that may describe more individual responses (observed in <30% of patients) (Suppl. Fig. S3).

This observed low sensitivity of the conventional method to identify responsive pathways in a few patients may indicate a lack of statistical power due to small sample size. Further, these methods are not designed to identify the responsive pathways specific to each patient, though these individual differences may explain diverse response to therapy due to personal genetic and epigenetic architecture. Altogether the results show that kMEn captures pathways found by conventional methods while discovering additional patient-specific information.

3.2.2. kMEn enables discovery of treatment response mechanisms at the single-subject level

We hypothesized that treatment-specific pathways (GO-BPs) identified by kMEn would inform on patient response to HIV therapy and reduce the number of features (gene sets) to investigate. We focused on pathways that responded differently between NNRTI and PI therapies (Methods Section 2.4.2) as they target distinct HIV enzymes (reverse transcriptase vs. protease), have known different side effects, and likely impact differently on the host biology. We compared the pathway scores between 12 NNRTI-treated and 8 PI-treated patients, from which we identified the 53 GO-BP referred to as *treatment-specific mechanisms*.

To determine the clinical relevance of these treatment-specific mechanisms, the abundance of CD4$^+$ cells were examined within patients. As T helper cells expressing the glycoprotein CD4 are known to be a target of HIV infection, they have been used previously as a reliable biomarker to monitor and assess response to anti-HIV therapeutics [34]. For each subject, we calculated the fold change of CD4$^+$ T cell recovery before and after treatment (Methods; Eq. (4)) and conducted a principal component analysis of GO-BP scores (Methods Section 2.4.2). kMEn-generated pathway scores, projected on the first principal component, correlated well with CD4$^+$ T cells FC in HIV-infected patients ($p = 0.021$, Fig. 5; Suppl. Fig. S4). This correlation is evident when considering both treatments together ($n = 20$, Suppl. Fig. S4) and NNRTI therapy alone ($n = 12$, Fig. 5). In contrast, the Wilcoxon and MD pathway scores projected to the first principal component only correlated with CD4$^+$ T cell response to both treatments and not to NNRTI alone (Methods Section 2.4.2; Suppl. Fig. S5). Notably, the projected values of all transcripts or the 106 DEGs identified by cohort statistics did not correlate with CD4$^+$ T cell count FC (Methods Section 2.4.2; Suppl. Fig. S5). The clinical case study results are consistent with those of the simulation studies: kMEn pathway scores provided more comprehensive prediction and assessment of antiretroviral therapy response compared to our previous N-of-1-*pathways* and conventional cohort-based transcript analysis methods. In addition, among the 53 treatment-specific pathways, we observed that five GO-BPs (Suppl. Table S1) correlated with CD4$^+$ FC among NNRTI-treated subjects (nominal $p < 0.05$, Methods Section 2.4.2).

Finally, we investigated how the responsive transcripts (before and after treatment in each patient) contribute to the observed correlation between the GO-BP mechanism scores and the CD4$^+$ FC found in patients treated with NNRTI. The result showed that kMEn successfully detects responsive pathways with bidirectional transcript signals (Suppl. Fig. S6). Additionally, looking into the GO-BP "mRNA catabolic process", we identified 22 transcripts whose expression fold changes are all lower/higher in one diametric extreme group than the opposing group, based on the individual CD4$^+$ T-cell recovery. These transcripts include *AUH*, *EXOSC5*, *CNOT4*, *PPP2R2A*, *RPS16*, *RPL13A*, *RPL37*, *EIF4B*, *RPL7*, *RPS23*, etc. Individually, none of these transcripts could be correlated with the CD4$^+$ FC. One interpretation is that the summative effect of transcripts on a biological process is consistent within groups and distinct between groups while their individual genetic and epigenetic architecture is distinct. It has also been shown repeatedly that very few single gene-level biomarkers are reproducibly predictive of therapeutic response [2]. kMEn provides the opportunity of discovering compound biomarkers representative of the response for predicting disease progression.

3.3. Limitations and future studies

Together, we have shown that under certain conditions, kMEn outperforms Wilcoxon and MD N-of-1-*pathways* methods. One caveat should be mentioned regarding the simulation study: Genes are often linked within regulatory networks and may be encoded within the same genomic locus, which may lead to gene co-expression. However, in the simulation study, for simplicity, we simulated pathways in which gene expression values are independent from each other. This situation was not evaluated by the current simulation and will be addressed in future studies. We note that as a competitive gene set model, kMEn compares pathway-level transcriptional response to that of the background, i.e., transcripts not in the pathway [14,32]. This approach affords kMEn greater accuracy by utilizing comprehensive information and ostensibly provides resistance to unsuccessful cross-sample normalization or other technical variation (e.g., cDNA amplification biases) and warrants future study. Since Wilcoxon and MD are self-contained [14,32] (i.e., only require the expression values of the pathway), they may have a broader and more affordable application to clinical practice. This would allow for small-scale testing of gene sets at the single-patient level, perhaps to validate responsive pathways using real-time quantitative polymerase chain reaction or other targeted RNA-sequencing. An issue to consider that may affect the accuracy of kMEn is the dependence on correct k-Means clustering. At low transcript expression of the same transcript in both samples, high fold changes of transcript expression can be introduced by noise and experimental variation, which could mistakenly be identified as responsive by kMEn; although making inferences on gene sets may mitigate this problem as shown in Suppl. Fig. S7. To advance the framework of gene clustering followed by enrichment test, other clustering techniques may also be explored to improve the accuracy of DEG calling. The cohort-expectation standard is derived from cohort-based statistics; the development of new biological and computational standards is required for studying altered genomics in single subjects and has been partially addressed in this study. Notably, kMEn is not restricted to transcriptomics data; it is applicable to other 'omics scales alone or integrated in multi-omics measures, as long as the data can be clustered into unaltered and aberrant groups. Lastly, CD4$^+$ FC is a good predictor of therapeutic response in HIV and provided a reliable phenotype to demonstrate the proof-of-concept in our study.

4. Conclusion

The simulations and clinical case study show that N-of-1-*pathways* kMEn method provides personal transcriptome profiles via the assignment of responsive pathways more accurately than previous single-subject methods, under bidirectional transcript responses and robust to background noise. In the case study, kMEn-scored gene sets enable the assessment of antiretroviral therapy response and correlation with cellular count profile of CD4$^+$ T cells. While T cell counts are an affordable and convenient measure of therapeutic response, these pathway-level and transcript-level correlations suggest the method could scale for predicting therapeutic outcomes from the peripheral blood in other immune-mediated diseases (e.g. asthma). This methodology innovates in that it identifies bidirectionally responsive transcripts within a pathway using dynamic changes derived from 'omics measures with resolution at a single-subject pair of samples. Further, we provide a framework for single-subject studies of response to therapy, a challenging phenotype to predict [35]. Broadly, N-of-1-*pathways* kMEn enables clinical interpretation of personal transcriptome dynamics, which has a direct application to precision therapeutics.

Funding

The study was supported in part by the NIH grant K22LM008308, NSF DMS-1309507, DMS-1418172, and NCI P30CA023074 grant of the University of Arizona Cancer Center.

Conflict of interest

None declared.

Acknowledgements

We would like to thank Nima Pouladi for his helpful comments as well as Jacob Smith and Kyle Goble.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2016.12.009.

References

[1] V. Gardeux, A. Bosco, J. Li, M.J. Halonen, D. Jackson, F.D. Martinez, et al., Towards a PBMC "virogram assay" for precision medicine: concordance between ex vivo and in vivo viral infection transcriptomes, J. Biomed. Inform. 55 (2015) 94–103.

[2] C. Fan, D.S. Oh, L. Wessels, B. Weigelt, D.S. Nuyten, A.B. Nobel, et al., Concordance among gene-expression–based predictors for breast cancer, New Engl. J. Med. 355 (2006) 560–569.

[3] P. Khatri, M. Sirota, A.J. Butte, Ten years of pathway analysis: current approaches and outstanding challenges, PLoS Comput. Biol. 8 (2012) e1002375.

[4] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. USA 102 (2005) 15545–15550.

[5] T. Beissbarth, T.P. Speed, GOstat: find statistically overrepresented Gene Ontologies within a group of genes, Bioinformatics 20 (2004) 1464–1465.

[6] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, et al., Gene Ontology: tool for the unification of biology, Nat. Genet. 25 (2000) 25–29.

[7] D. Bottomly, P.A. Ryabinin, J.W. Tyner, B.H. Chang, M.M. Loriaux, B.J. Druker, et al., Comparison of methods to identify aberrant expression patterns in individual patients: augmenting our toolkit for precision medicine, Genome Med. 5 (2013) 103.

[8] X. Yang, K. Regan, Y. Huang, Q. Zhang, J. Li, T.Y. Seiwert, et al., Single sample expression-anchored mechanisms predict survival in head and neck cancer, PLoS Comput. Biol. 8 (2012) e1002350.

[9] V. Gardeux, A.D. Arslan, I. Achour, T.-T. Ho, W.T. Beck, Y.A. Lussier, Concordance of deregulated mechanisms unveiled in underpowered experiments: PTBP1 knockdown case study, BMC Med. Genomics 7 (2014) 1–13.

[10] V. Gardeux, I. Achour, J. Li, M. Maienschein-Cline, H. Li, L. Pesce, et al., 'N-of-1-pathways' unveils personal deregulated mechanisms from a single pair of RNA-Seq samples: towards precision medicine, J. Am. Med. Inform. Assoc. 21 (2014) 1015–1025.

[11] A.G. Schissler, Q. Li, J.L. Chen, C. Kenost, I. Achour, D.D. Billheimer, et al., Analysis of aggregated cell–cell statistical distances within pathways unveils therapeutic-resistance mechanisms in circulating tumor cells, Bioinformatics 32 (2016) i80–i89.

[12] A.G. Schissler, V. Gardeux, Q. Li, I. Achour, H. Li, W.W. Piegorsch, et al., Dynamic changes of RNA-sequencing expression for precision medicine: N-of-1-pathways Mahalanobis distance within pathways of single subjects predicts breast cancer survival, Bioinformatics 31 (2015) i293–i302.

[13] J.A. Hartigan, M.A. Wong, Algorithm AS 136: a k-means clustering algorithm, Appl. Stat. (1979) 100–108.

[14] J.J. Goeman, P. Bühlmann, Analyzing gene expression data in terms of gene sets: methodological issues, Bioinformatics 23 (2007) 980–987.

[15] Y. Liu, J. Zhou, K.P. White, RNA-seq differential expression studies: more sequence or more replication?, Bioinformatics 30 (2014) 301–304

[16] M. Massanella, A. Singhania, N. Beliakova-Bethell, R. Pier, S.M. Lada, C.H. White, et al., Differential gene expression in HIV-infected individuals following ART, Antiviral Res. 100 (2013) 420–428.

[17] M. Carlson, org.Hs.eg.db: Genome wide annotation for Human. R package version 3.2.3., ed2015.

[18] R.A. Fisher, Statistical methods for research workers, 1934.

[19] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, Ann. Stat. (2001) 1165–1188.

[20] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol. 11 (2010) R106.

[21] M.D. Robinson, D.J. McCarthy, G.K. Smyth, EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.

[22] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer, 2009.

[23] P. Du, W.A. Kibbe, S.M. Lin, Lumi: a pipeline for processing Illumina microarray, Bioinformatics 24 (2008) 1547–1548.

[24] V.G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, Proc. Natl. Acad. Sci. USA 98 (2001) 5116–5121.

[25] Y. Tao, L. Sam, J. Li, C. Friedman, Y.A. Lussier, Information theory applied to the sparse gene ontology annotation network to predict novel gene function, Bioinformatics 23 (2007) i529–i538.

[26] H. Li, I. Achour, L. Bastarache, J. Berghout, V. Gardeux, J. Li, et al., Integrative genomics analyses unveil downstream biological effectors of disease-specific polymorphisms buried in intergenic regions, NPJ Genom. Med. 1 (2016) 16006.

[27] M. Maienschein-Cline, Z.D. Lei, V. Gardeux, T. Abbasi, R.F. Machado, V. Gordeuk, et al., ARTS: automated randomization of multiple traits for study design, Bioinformatics 30 (2014) 1637–1639.

[28] K. Regan, K. Wang, E. Doughty, H. Li, J. Li, Y. Lee, et al., Translating Mendelian and complex inheritance of Alzheimer's disease genes for predicting unique personal genome variants, J. Am. Med. Inform. Assoc. 19 (2012) 306–316.

[29] H. Li, Y. Lee, J.L. Chen, E. Rebman, J. Li, Y.A. Lussier, Complex-disease networks of trait-associated single-nucleotide polymorphisms (SNPs) unveiled by information theory, J. Am. Med. Inform. Assoc. 19 (2012) 295–305.

[30] C. Spearman, The proof and measurement of association between two things, Am. J. Psychol. 15 (1904) 72–101.

[31] H. Abdi, L.J. Williams, Principal component analysis, Wiley Interdiscip. Rev.: Comput. Stat. 2 (2010) 433–459.

[32] D. Wu, G.K. Smyth, Camera: a competitive gene set test accounting for inter-gene correlation, Nucl. Acids Res. 40 (2012) e133.

[33] M.R. Stratton, P.J. Campbell, P.A. Futreal, The cancer genome, Nature 458 (2009) 719–724.

[34] B. Autran, G. Carcelain, T.S. Li, C. Blanc, D. Mathez, R. Tubiana, et al., Positive effects of combined antiretroviral therapy on CD4$^+$ T cell homeostasis and function in advanced HIV disease, Science 277 (1997) 112–116.

[35] F. Clavel, A.J. Hance, HIV drug resistance, New Engl. J. Med. 350 (2004) 1023–1035.