

ROTS: reproducible RNA-seq biomarker detector—prognostic markers for clear cell renal cell cancer

Fatemeh Seyednasrollah^{1,2,*}, Krista Rantanen^{1,3}, Panu Jaakkola^{1,3,4} and Laura L. Elo^{1,2,*}

¹Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku, FI-20520, Finland,

²Department of Mathematics and Statistics, University of Turku, Turku, FI-20014, Finland, ³Department of Medical Biochemistry, University of Turku, Turku, FI-20014, Finland and ⁴Department of Oncology and Radiotherapy, Turku University Hospital, FI-20520 Turku, Finland

Received April 14, 2015; Revised July 23, 2015; Accepted July 28, 2015

ABSTRACT

Recent comprehensive assessments of RNA-seq technology support its utility in quantifying gene expression in various samples. The next step of rigorously quantifying differences between sample groups, however, still lacks well-defined best practices. Although a number of advanced statistical methods have been developed, several studies demonstrate that their performance depends strongly on the data under analysis, which compromises practical utility in real biomedical studies. As a solution, we propose to use a data-adaptive procedure that selects an optimal statistic capable of maximizing reproducibility of detections. After demonstrating its improved sensitivity and specificity in a controlled spike-in study, the utility of the procedure is confirmed in a real biomedical study by identifying prognostic markers for clear cell renal cell carcinoma (ccRCC). In addition to identifying several genes previously associated with ccRCC prognosis, several potential new biomarkers among genes regulating cell growth, metabolism and solute transport were detected.

INTRODUCTION

The recent comprehensive assessments of the RNA-seq technology provide important guidelines to produce high-quality RNA-seq data sets (1–4). The overall results from these community efforts demonstrate reproducibility of RNA-seq platforms and data analysis strategies for quantifying gene expression levels. Although these evaluations involved controlled experiments that are rather far from actual clinical applications, they strongly support that the

RNA-seq technology can produce data that is of sufficient quality to many biomedical applications.

In addition to producing accurate estimates of gene expression levels, the utility of the RNA-seq technology depends on the availability of rigorous tools for downstream analysis of these data such as quantifying differences between sample groups. This part still lacks well-defined best practices. Although a number of advanced statistical methods have been developed (e.g. edgeR (5,6), DESeq (7), baySeq (8), Cuffdiff2 (9)), several studies demonstrate that their performance depends strongly on the data under analysis and there is no ‘one fits all method’ that would always perform well (10–12). This compromises practical utility in real biomedical and clinical studies that aim to identify reliable biomarkers for diagnosis, prognosis or treatment of patients.

To address the challenge of selecting a suitable statistic, we propose to use a data-adaptive procedure, named ROTS (Reproducibility Optimized Test Statistic). It determines an optimal test statistic directly from the data by maximizing the reproducibility of the detections across bootstrap samples (refer to MATERIALS AND METHODS for details). The utility of reproducibility optimization in microarray studies of gene expression has been demonstrated (13,14). In this study the reproducibility optimization is shown to significantly improve the reliability of differential expression detection in RNA-seq data for the first time. An R-package implementing ROTS is available at <http://www.btk.fi/research/research-groups/elo/software/rots/>.

MATERIALS AND METHODS

Data sets

Spike-in data set. The spike-in data set was generated by Rapaport et al. (11) and the expression files were downloaded from GEO with the accession number GSE49712. The selected samples were part of SEQC (MAQC-III)

*To whom correspondence should be addressed. Tel: +358 2 333 8009; Fax: +358 2 333 8009; Email: laliel@utu.fi
Correspondence may also be addressed to Fatemeh Seyednasrollah. Tel: +358 2 333 8697; Fax: +358 2 333 8009; Email: fatsey@utu.fi

project extracting from Stratagene Universal Human Reference RNA (UHRR) and Ambion Human Brain Reference RNA (UBRR). The samples were divided into two distinct experimental groups *A* and *B* with five technical replicates per group. All the replicates were enriched with 92 synthetic polyadenylated oligonucleotides introduced by the External RNA Control Consortium (ERCC) (15) to validate the differential expression findings. The ERCC spike-in controls were spiked to have 0.5-, 0.67-, 1- or 4-fold changes between the mixture groups *A* and *B*. All the samples were sequenced using Illumina HiSeq2000 platform and produced paired-end reads of length 100 bp. The reads were mapped and assembled using TopHat (v.2.0.3) (16) and UCSC hg19 as the genome reference. HTSeq (v.0.5.p3) was applied for gene expression abundance estimation (17).

TCGA clear cell renal cell carcinoma (ccRCC) data set. The ccRCC data set was published by The Cancer Genome Atlas (TCGA) (18). All the biospecimens were collected from patients with ccRCC diagnosis and kidney as primary site before any treatment procedures. The mRNA expression files were downloaded from the TCGA website with IlluminaHiSeq.RNASeqV2 platform code (see <https://tcga-data.nci.nih.gov/tcga/>). The demographics data and clinical features were available in the supplementary tables of the original study (18). In total, 448 patients had mRNA expression data available in TCGA, among which 442 patients had the necessary clinical values available and were used in the analysis.

ccRCC validation data set. The ccRCC prognostic findings were validated in an independent RNA-seq data set published by Sato et al. (19). The mRNA libraries had been sequenced into 100 bp paired-end reads using Illumina HiSeq2000 platform. The alignment files were downloaded from European Genome-phenome Archive (EGA) using the accession number EGAS00001000509 (Data Set ID: EGAD00001000597). The expression levels were estimated using HTSeq (v.0.6.1) package and UCSC hg19 genome reference. The clinical features were obtained from the original publication (19). Those 93 samples which included complete clinical data were used in the analysis.

Preprocessing procedures: TMM normalization and Voom transformation

The necessity of normalization has been proved in RNA-seq studies (20,21). Accordingly, the Trimmed mean of M values (TMM) method implemented in Bioconductor edgeR package was used to normalize the expression levels. The Voom transformation implemented in Bioconductor Limma package was applied to transform the expression levels suitable for further differential expression testing, following the recommendation by Law et al. (22).

ROTS differential expression testing

Making *a priori* assumptions about data set distributions contradicts the observed biological variation in real RNA-seq experiments. To eliminate biases, we propose to learn an appropriate test statistic directly from the data, building

on our data-adaptive reproducibility optimization procedure ROTS (Reproducibility Optimized Test Statistic) (13). The input of ROTS is a count matrix with genomic features as rows and samples as columns. Genomic feature can refer to a gene, a transcript or an exon but it is called gene throughout this manuscript for convenience. The aim of ROTS is to rank the genes according to their differential expression. For each data set, the ranking statistic is determined by maximizing the reproducibility of the gene rankings in bootstrapped data sets.

Let us denote by x_{gi}^j the normalized read count of gene *g* in sample *i* from condition *j*. The mean and variance of gene *g* within each condition is defined as

$$\bar{x}_g^j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{gi}^j \text{ and } (s_g^j)^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (x_{gi}^j - \bar{x}_g^j)^2,$$

where n_j is the number of samples in condition *j*. The gene rankings are estimated using a family of modified *t*-statistics:

$$d_\alpha(g) = \frac{|\bar{x}_g^1 - \bar{x}_g^2|}{\alpha_1 + \alpha_2 s_g}, \quad \alpha_1 \in [0, \infty), \quad \alpha_2 \in \{0, 1\}.$$

Here, α_1 and α_2 are common to all genes and

$$s_g = \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1 - 1)(s_g^1)^2 + (n_2 - 1)(s_g^2)^2}{n_1 + n_2 - 2}}$$

is the pooled standard error of gene *g* across the conditions. Specific choices of the parameters α determine the ROTS statistic. For instance, the special case of $\alpha_1 = 0$ and $\alpha_2 = 1$ corresponds to the ordinary *t*-statistic. The other special cases include signal log-ratio ($\alpha_1 = 1$ and $\alpha_2 = 0$) or SAM-statistic (23) ($\alpha_2 = 1$ and α_1 a percentile of the standard deviations). In ROTS the parameters α are not predefined but they are determined by a reproducibility optimization procedure. The aim is to ensure appropriate accuracy of variance estimation, which is considered as the main challenge in RNA-seq data analysis.

The optimization of $d_\alpha(g)$ is based on maximizing its reproducibility through bootstraps. Pairs of bootstrapped data sets D_1^b and D_2^b are sampled from the source data set *D* preserving the same sample size and sample labels (sampling with replacement within the groups). The reproducibility is then computed as the average overlap of the *k* most top ranked genes ordered by applying the test statistic d_α across *B* pairs of bootstrap data:

$$R_k(d_\alpha) = \frac{1}{B} \sum_{b=1}^B R_k^b(d_\alpha).$$

For the optimization, a *z*-type statistic is used defined as

$$Z_k(d_\alpha) = \frac{R_k(d_\alpha) - R_k^0(d_\alpha)}{s_k(d_\alpha)},$$

where the denominator is the standard deviation of the bootstrap distribution of $R_k(d_\alpha)$ and $R_k^0(d_\alpha)$ represents the null reproducibility in *B* random permutations across the whole data set. Specifically, ROTS maximizes the reproducibility statistic $Z_k(d_\alpha)$ over a dense lattice of parameters

α where $\alpha_1 \geq 0$ and $\alpha_2 \in \{0, 1\}$ and various numbers of top ranked genes between 5 and G , where G denotes the total number of genes in the experiment.

The output of ROTS is the optimized Z-score, reproducibility and ROTS-statistic for each gene together with false discovery rate (FDR) estimate to assess the significance of differential expression. In RNA-seq studies, small numbers of replicate samples and large biological variation remain the main challenges which call the efficiency of statistical methods under question. With ROTS, the optimized Z-score and reproducibility are the main indicators to decide the success of differential expression detection. As a rule of thumb, reproducibility Z-scores below 2 indicate that the data or the statistics are not sufficient for reliable detection.

Other differential expression analysis tools

All the statistical analyses in this manuscript were performed using R version 3.0.2. For comparing ROTS with other available methods, we used edgeR (version 3.2.4), DESeq (version 1.12.1), DESeq2 (version 1.0.19), Cuffdiff 2.0.2, Limma (version 3.16.7), baySeq (version 1.12.0), NOISeq (version 2.12.0) and PoissonSeq (version 1.1.2).

ccRCC prognosis analysis

Unsupervised clustering was performed by applying the R *hclust* function using the Ward method and Manhattan distances. The patient-specific risk scores were calculated similarly as by Shaughnessy et al. (24). Specifically, the risk scores were defined as the difference between the \log_2 -transformed expression levels of the up- and down-regulated genes in the prognostic signature of 152 ROTS detections. Next, the scores were clustered into four groups using K-means clustering method. Finally, the Kaplan–Meier analysis was performed to compare the survival of the ccRCC patients in the four risk categories. The significance of the differences between the categories was tested using the log rank test.

RESULTS AND DISCUSSION

ROTS outstanding performance: highest accuracy and lowest false discovery rate

The improved sensitivity and specificity of ROTS over state-of-the-art methods was demonstrated in a controlled spike-in study. Spike-in data sets are benchmarks to investigate the strengths and weaknesses of new computational methods. The spiked data set presented by Rapaport et al. (11), previously used to evaluate different methods for RNA-seq differential expression analysis, was analyzed using ROTS and a number of state-of-the-art methods of the field, including edgeR (5,6), DESeq (7), DESeq2 (25), Limma (26,27), Cuffdiff2 (9), PoissonSeq (28), NOISeq (29) and baySeq (8). This data set includes technical replicates of human whole body ($n = 5$) and human brain samples ($n = 5$) spiked with 92 synthetic oligonucleotides provided by the External RNA Controls Consortium (ERCC) (refer to MATERIALS AND METHODS for details). The pre-determined

fold-changes of the synthetic RNAs (4, 1, 0.67 and 0.5) enable measurement of true positive (sensitivity) and true negative detection rates (specificity) of different gene ranking methods.

The Receiver Operating Characteristic (ROC) analysis showed how ROTS outperforms the other methods in terms of sensitivity and specificity (Figure 1A; area under the curve AUC = 0.941; DeLong's test $P < 0.01$ compared to all other methods except for baySeq for which $P = 0.077$). The second best method was baySeq (AUC = 0.891), whereas NOISeq showed the lowest AUC value (AUC = 0.704). This comparison was repeated for smaller sample sizes ($N = 2, 3$ and 4) with selected methods (edgeR, DESeq, Limma and ROTS). Again ROTS performed significantly better than the other methods with sample sizes larger than 2 (Wilcoxon rank sum test P -value < 0.06 ; see Supplementary Figure S1). To further investigate the ability of the methods to control type I error rate (i.e. to avoid false positive detections) the approach of the original paper (11) was followed. Specifically, the false discovery rate (FDR) values of the spiked controls (fold-change 1) were examined. They represent non-differentially expressed genes whose FDR values should be high. Strikingly, ROTS showed an outstanding outcome in this comparison (Figure 1B). Among the 23 non-differentially expressed RNA controls, ROTS detected only one false positive at FDR < 0.05 , whereas most of the other methods detected at least ten false positives. PoissonSeq presented the worst performance in this comparison with as many as 18 false positive detections. NOISeq was not included in this step as it does not report FDR values for the technical replicates. Our results in the spike-in data set strongly support the significant advantage of ROTS over the current widely used statistical methods.

ROTS defines prognostic signature of renal cell cancer using two independent data sets

The efficiency of ROTS was then investigated in real patient data involving large biological heterogeneity to take a step toward clinical applications of RNA-seq. The aim was to identify prognostic markers for clear cell renal cell carcinoma (ccRCC), which is an important clinical problem. The ccRCC accounts for the majority of cases of kidney cancer (30). The clinical course of ccRCC is heterogeneous as is the mutational profile. Current prognostic post-nephrectomy markers (e.g. UICC and SSIGN) are based on clinicopathological features such as grade and TNM classification, but ccRCC lacks widely accepted genetic markers for prognosis (31,32). In addition to understanding ccRCC molecular characteristics, accurate biomarkers are required to stratify the disease for selecting patients for adjuvant trials and close surveillance. Recent data from two large published ccRCC studies (18,19), which included both RNA-seq measurements as well as corresponding clinical information, were analyzed. Data from the study by The Cancer Genome Atlas (TCGA) (18) was utilized to detect candidate markers associated with patient outcome (poor or better prognosis) and the findings were verified in a completely independent data by Sato et al. (19) (referred to as validation data in the following).

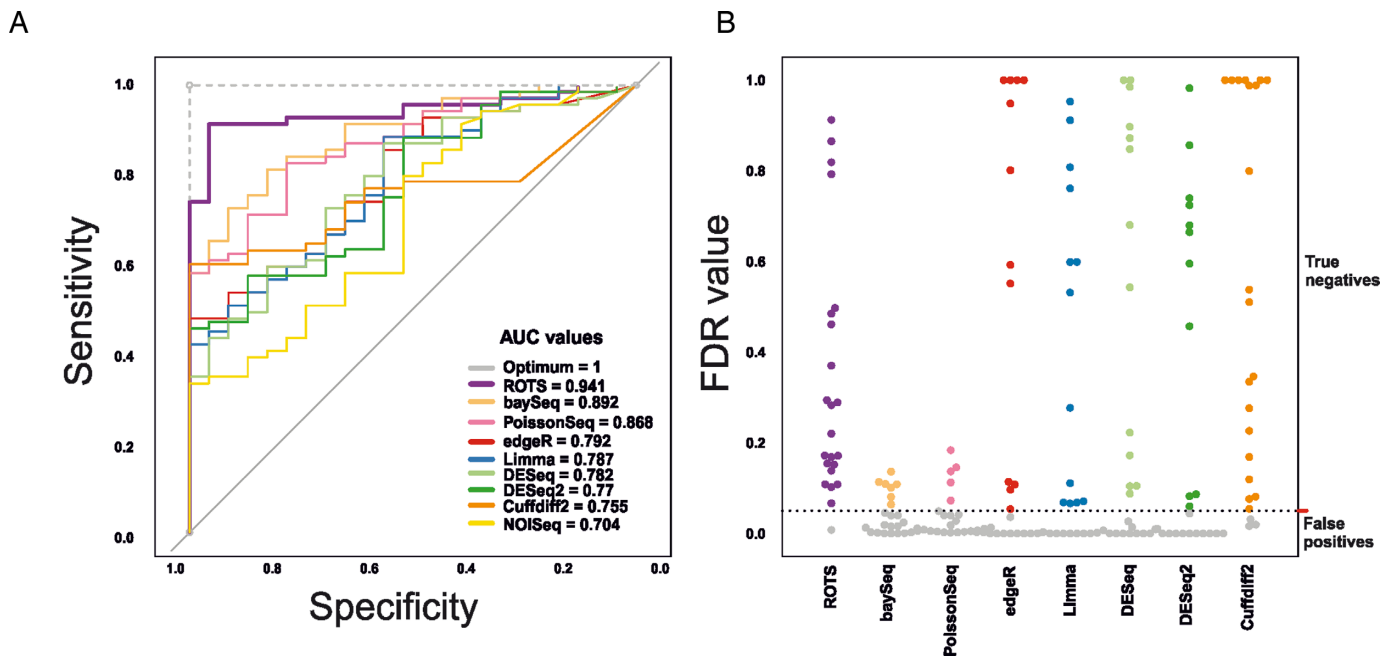


Figure 1. Efficient detection of spiked material. (A) Receiver Operating Characteristic (ROC) curves of the different statistical methods in the spike-in data together with the areas under the curves (AUC). (B) False discovery rate (FDR) values of the non-differentially expressed spiked controls. The gray dots correspond to false positive detections at $FDR < 0.05$.

In the TCGA data, ROTS detected 2208 differentially expressed genes at $FDR < 0.05$ between 40 patients ($\sim 10\%$) with the longest survival time (>60 months) and 40 patients with the shortest survival time (<12 months); see Supplementary Table S1 for the characteristics of the groups. The ROTS reproducibility values indicated appropriate reproducibility of the results ($R = 0.57$, $Z = 5.27$). To focus on the most promising candidate markers for ccRCC prognosis, those 152 genes that showed \log_2 fold-change above 1.6 (~ 3 -fold change) and average expression above the lowest 30% were retained for subsequent analysis, similarly as in the RNA-seq assessment studies (1,3–4) (Supplementary Table S2). For comparison, Limma detected 130 genes with the same criteria, all of which were among the ROTS detections. The genes detected exclusively by ROTS included, for instance, *EPO*, *REN*, *FABP1* and *IGFBP1* that are regulated by the pVHL-HIF pathway, which is commonly over-activated in ccRCC. This supports the potential relevance of the additional ROTS findings.

To assess the utility of the 152 ROTS detections as prognostic markers, a risk score was defined for each of the 442 patients in the complete TCGA data in terms of signal log-ratio of the up-regulated versus the down-regulated genes (24,33) (refer to MATERIALS AND METHODS for details). Investigation of the expression levels of the detected genes revealed four clusters of ccRCC patients (Figure 2A). Accordingly, four risk score categories were defined (Figure 2B). This revealed a highly significant association between the risk scores and the survival of the patients (Figure 2C; log rank test $P < 10^{-15}$). While the 5-year survival in the best group (blue) was $\sim 80\%$, it decreased below 20% in the worst survival group (red); in the two intermediate groups the 5-year survival was 50–60%. This supports the prognos-

tic value of the detected genes beyond the 80 samples used for detecting the markers.

To avoid over-fitting to a single study, further validation of the markers and risk score model was carried out in a completely independent data set of 100 ccRCC patients (19). The signal log-ratios of the detected genes between the best and poorest survival patients were highly correlated across the TCGA and validation data when patients with comparable survival times (<12 months, >60 months) were considered (Figure 2D; Pearson correlation 0.796, $P < 10^{-15}$). Over 90% of the detected 152 genes showed a concordant change to the same direction in both data sets (Supplementary Table S2, Supplementary Figures S2 and S3). Importantly, a highly significant association between the risk scores and survival verified the risk score model developed using the TCGA data (Figure 2E; log rank test $P < 10^{-4}$). These results confirm the capability of ROTS to robustly identify reproducible candidate markers well as the potential of the risk score model to identify especially poor prognosis ccRCC patients.

Biological and technical evaluation of detected prognostic biomarkers and utilized method

Analysis of the detected genes using the Ingenuity Pathway Analysis (IPA) tool suggests three major functional groups: molecular transport, small molecule biochemistry, and amino acid and lipid metabolism (Supplementary Table S3; $P < 0.05$). Categorization of the genes into five main biological function groups and four additional biochemical function groups implies major involvement of metabolism ($\sim 18\%$ of the genes), particularly glucose metabolism, as found also in the original TCGA study (Figure 3A, Supplementary Table S4). Several previously reported markers

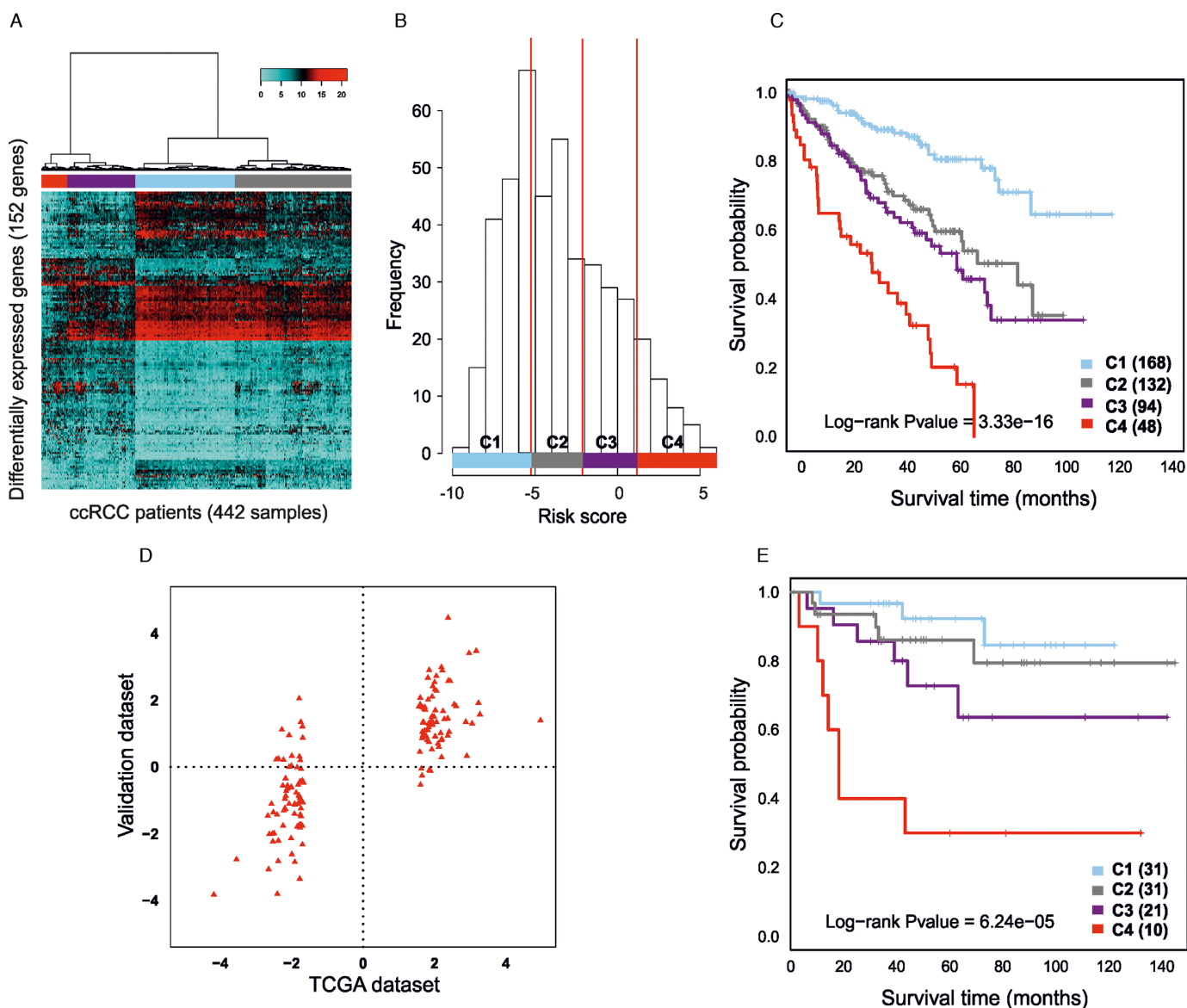


Figure 2. Novel prognostic markers for ccRCC. **(A)** Unsupervised clustering of the patients (columns) across the genes detected as differentially expressed by ROTS (rows). **(B)** Distribution of the risk scores across the patients in the TCGA data. The red vertical lines show the cutoffs used to define the four risk categories ($-5.10, -2.17, 0.94$). The risk categories significantly overlapped with the similarly colored clusters in panel A (69, 92, 78 and 92 percent of overlap with C1, C2, C3 and C4 survival groups respectively; Fisher's exact test P -value < 0.01). **(C)** Kaplan–Meier curves comparing the survival of the ccRCC patients in the four risk categories in the TCGA data. The numbers in parentheses indicate the numbers of patients in the different risk categories. The colors correspond to the colors in panels A and B. **(D)** Correlation of the signal log-ratios of the differentially expressed genes (red triangles) between the best and poorest survival patients in the TCGA (x -axis) and validation data (y -axis). Patients with comparable survival times (< 12 months, > 60 months) were considered in the analysis. A highly significant Pearson correlation of 0.796 was observed ($P < 10^{-15}$). **(E)** Kaplan–Meier curves comparing the survival of the ccRCC patients in the four risk categories in the validation data. The highly significant association between the risk score categories and survival verify the risk score model developed using the TCGA data.

were confirmed such as the key glucose metabolism regulators *ALDOB*, *G6PC* and *PKLR* (Figure 3B). Additionally, several new metabolism regulating markers were detected that have been missed by previous large-scale studies. These include, for instance, the glucose transporter/sensor *SLC2A2*, and the central gluconeogenesis regulator *PCK1* (Figure 3B).

The largest proportion ($\sim 26\%$) of the prognostic genes were from the cellular transporter and solute carrier groups (Figure 3A). The solute carrier family genes were highly en-

riched; 26 ($\sim 9\%$) out of ~ 300 family members were among the prognostic genes (Fisher's exact test $P < 0.05$). Some of them, such as members of solute carrier family 16, were detected by the TCGA study. Additionally, ROTS detected 19 previously undetected solute carrier genes, including for example *SLC38A5*, which transports glutamate, an essential nitrogen donor for cancer cells to build amino acids and to maintain mTOR activity (Figure 3B).

The better prognosis genes included most of the detected glucose metabolism and organic anion/cation transporters.

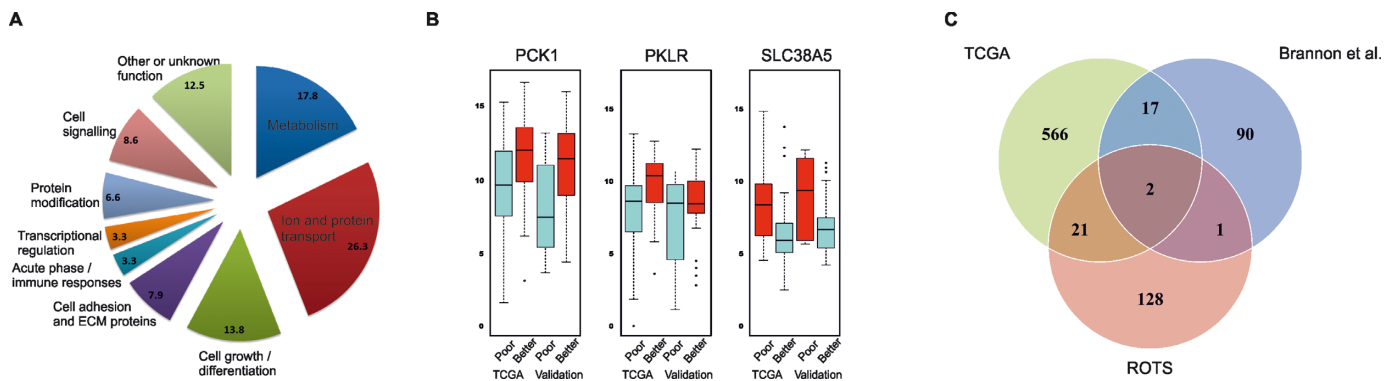


Figure 3. Biological insights from ccRCC prognostic markers. **(A)** Functional groups of the differentially expressed genes detected by ROTS. **(B)** Examples of the detected differentially expressed genes. The boxes show the median and the interquartile range (IQR) of the expression levels of the poor and better prognosis patients in the TCGA and the validation data, the whiskers indicate their range and the points correspond to extreme observations with values greater than 1.5 times the IQR. The boxplots for all the differentially expressed genes are shown in Supplementary Figures S2 and S3. **(C)** Venn diagram summarizing the overlap between the prognostic genes reported by the previous studies (18,36) and ROTS (see Supplementary Table S2 for detailed information).

Increased expression of six glycolytic genes were categorized into good prognosis group (ALDH1L1, ALDOB, G6PC, PCK1, PKLR, SLC2A2). The glycolytic genes are direct transcriptional targets of HIF-1 α , which is known to function as a tumor suppressor in kidney cancer in contrast to several other cancer types (34,35). The finding is in line with previously reported metabolic shift in RCC and increased feed of the TCA cycle in good prognosis group (18). (Supplementary Figure S4). Network analysis of the ROTS detections in the manually curated Ingenuity Knowledge Base revealed several interactions between the better prognosis genes linked to cellular metabolism (Supplementary Figure S5). These observations are in line with the earlier view (18). The poor prognosis genes were enriched in a variety of cell growth signaling molecules (e.g. phosphatases), extracellular matrix and remodeling proteins (collagens, metalloproteins) and acute phase/immune response genes (CRP, SAA family) (Supplementary Figure S4).

Comparison of the ROTS detections to two previous studies (18,36) of prognostic markers for ccRCC revealed a statistically significant but moderate overlap (Figure 3C and Supplementary Table S2; Fisher's exact test $P < 0.05$). For instance, the overlapping genes included the *ACE2* and *NPR3* that regulate blood pressure by the renin-angiotensin system. Additionally, new genes within this pathway were detected, including *REN*, the primary regulator of the pathway secreted by kidney cells. To further validate the method we studied the association of several known genes mutated in RCC with our prognostic groups (Supplementary Figure S6). In line with previous reports ROTS predicted *PBRM1* mutations to correlate well with good prognosis ($P < 0.01$) (18,37). Also, *BAP1* mutations revealed statistically significant correlation with poorer prognosis ($P < 0.01$). Several other mutated genes were also tested for prognostic correlation (e.g. *TP53*, *CDKN2A* and *PIK3CA*) but their mutation rate was too low to reveal any significant correlation. Moreover, our algorithm predicted a strong trend toward better prognosis in the *VHL* mutated group ($P < 0.05$). Although studies attempting to correlate *VHL* mutation with patient

prognosis have been somewhat conflicting, correlation of wild-type *VHL* with poor prognosis has been reported earlier (38,39).

Taken together, our results demonstrate the utility of the RNA-seq technology to detect reproducible markers when an appropriate test statistic is applied. Systematic tools are needed for unbiased and effective analysis of RNA-seq data to fulfill the high promise posed by the technology. Instead of developing new variants of different statistical tests, users require practical tools to choose an optimal method for their own data. The reproducibility optimization procedure enables this. Additionally, it provides information about the quality of the detections; low reproducibility values indicate that the data or the test statistics are not sufficient for reliable detection.

The high validation percentage of the detected ccRCC prognostic markers in independent data supports the general potential of ROTS in clinical RNA-seq studies. Although the main focus was on reliable marker detection, the developed ccRCC risk score model illustrates the use of the detections for disease signatures. The identified novel candidate genes serve as good starting points for further validation studies to confirm their utility in the clinic as support tools to predict ccRCC prognosis or revealing novel potential targets for ccRCC treatment. Our results suggest that ROTS enables stratification of patients in prognostic groups that can help to select patients for future RCC adjuvant trials and closer post-nephrectomy follow-up to timely reveal metastatic disease. The identified markers imply high potential of genes regulating cell growth and metabolism but also ion transport apart from glucose transport. It is also noteworthy that four genes regulating blood pressure through the renin-angiotensin system were identified as good prognosis markers, as blood pressure is commonly elevated in RCC patients and functions as a surrogate marker for tyrosine kinase inhibitor treatment efficacy.

AVAILABILITY

ROTS package is implemented with R and the source code together with Windows and Mac OS binaries are freely

available at <http://www.btk.fi/research/research-groups/elo/software/rots/>. In order to make the study fully reproducible, we provide all the codes to run the analyses in the Supplementary Code, which provides detailed information about the methods and parameters used in this study.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This study makes use of data generated by Department of Pathology and Tumor Biology, Kyoto University. The authors thank Dr Jukka Lehtonen, Dr Elizabeth Nguyen and Anna Pursiheimo for useful comments. We also gratefully acknowledge the use of bioinformatics infrastructure supported by Biocenter Finland and CSC IT Center for Science.

FUNDING

This study was supported by Sigrid Juselius Foundation, JDRF [grant number 2-2013-32], the Diabetes Research Foundation, and Biocenter Finland.

Conflict of interest statement. None declared.

REFERENCES

1. SeqC/MaQC-III Consortium. (2014) A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.*, **32**, 903–914.
2. Li, S., Labaj, P.P., Zumbo, P., Sykacek, P., Shi, W., Shi, L., Phan, J., Wu, P.-Y., Wang, M., Wang, C. *et al.* (2014) Detecting and correcting systematic variation in large-scale RNA sequencing data. *Nat. Biotechnol.*, **32**, 888–895.
3. Wang, C., Gong, B., Bushel, P.R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., Fang, H., Hong, H., Shen, J., Su, Z. *et al.* (2014) The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.*, **32**, 926–932.
4. Li, S., Tighe, S.W., Nicolet, C.M., Grove, D., Levy, S., Farmerie, W., Viale, A., Wright, C., Schweitzer, P.A., Gao, Y. *et al.* (2014) Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat. Biotechnol.*, **32**, 915–925.
5. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinform. Oxf. Engl.*, **26**, 139–140.
6. McCarthy, D.J., Chen, Y. and Smyth, G.K. (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.*, **40**, 4288–4297.
7. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
8. Hardcastle, T.J. and Kelly, K.A. (2010) baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*, **11**, 422.
9. Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L. and Pachter, L. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.*, **31**, 46–53.
10. Seyednasrollah, F., Laiho, A. and Elo, L.L. (2015) Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief. Bioinform.*, **16**, 59–70.
11. Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D. and Betel, D. (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol.*, **14**, R95.
12. Soneson, C. and Delorenzi, M. (2013) A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, **14**, 91.
13. Elo, L.L., Filen, S., Lahtesmaa, R. and Aittokallio, T. (2008) Reproducibility-optimized test statistic for ranking genes in microarray studies. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **5**, 423–431.
14. Elo, L.L., Hiissa, J., Tuimala, J., Kallio, A., Korpelainen, E. and Aittokallio, T. (2009) Optimized detection of differential expression in global profiling experiments: case studies in clinical transcriptomic and quantitative proteomic datasets. *Brief. Bioinform.*, **10**, 547–555.
15. Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R. and Oliver, B. (2011) Synthetic spike-in standards for RNA-seq experiments. *Genome Res.*, **21**, 1543–1551.
16. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.
17. Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
18. The Cancer Genome Atlas Research Network. (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*, **499**, 43–49.
19. Sato, Y., Yoshizato, T., Shiraishi, Y., Maekawa, S., Okuno, Y., Kamura, T., Shimamura, T., Sato-Otsubo, A., Nagae, G., Suzuki, H. *et al.* (2013) Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.*, **45**, 860–867.
20. Bullard, J.H., Purdom, E., Hansen, K.D. and Dudoit, S. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, **11**, 94.
21. Dillies, M.-A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J. *et al.* (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.*, **14**, 671–683.
22. Law, C.W., Chen, Y., Shi, W. and Smyth, G.K. (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.*, **15**, R29.
23. Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 5116–5121.
24. Shaughnessy, J.D., Zhan, F., Burington, B.E., Huang, Y., Colla, S., Hanamura, I., Stewart, J.P., Kordsmeier, B., Randolph, C., Williams, D.R. *et al.* (2007) A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood*, **109**, 2276–2284.
25. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
26. Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article3.
27. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47–e47.
28. Li, J., Witten, D.M., Johnstone, I.M. and Tibshirani, R. (2012) Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostat. Oxf. Engl.*, **13**, 523–538.
29. Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. and Conesa, A. (2011) Differential expression in RNA-seq: A matter of depth. *Genome Res.*, **21**, 2213–2223.
30. Rini, B.I., Campbell, S.C. and Escudier, B. (2009) Renal cell carcinoma. *Lancet*, **373**, 1119–1132.
31. Zigeuner, R., Hutterer, G., Chromceki, T., Imamovic, A., Kampel-Kettner, K., Rehak, P., Langner, C. and Pummer, K. (2010) External validation of the Mayo Clinic stage, size, grade, and necrosis (SSIGN) score for clear-cell renal cell carcinoma in a single European centre applying routine pathology. *Eur. Urol.*, **57**, 102–109.
32. Chin, A.I., Lam, J.S., Figlin, R.A. and Beldegrun, A.S. (2006) Surveillance strategies for renal cell carcinoma patients following nephrectomy. *Rev. Urol.*, **8**, 1–7.

33. Van Laar,R., Flinchum,R., Brown,N., Ramsey,J., Riccitelli,S., Heuck,C., Barlogie,B. and Shaughnessy,J.D. (2014) Translating a gene expression signature for multiple myeloma prognosis into a robust high-throughput assay for clinical use. *BMC Med. Genomics*, **7**, 25.
34. Raval,R.R., Lau,K.W., Tran,M.G.B., Sowter,H.M., Mandriota,S.J., Li,J.-L., Pugh,C.W., Maxwell,P.H., Harris,A.L. and Ratcliffe,P.J. (2005) Contrasting properties of hypoxia-inducible factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-associated renal cell carcinoma. *Mol. Cell. Biol.*, **25**, 5675–5686.
35. Shen,C., Beroukhi,R., Schumacher,S.E., Zhou,J., Chang,M., Signoretti,S. and Kaelin,W.G. (2011) Genetic and functional studies implicate HIF1 α as a 14q kidney cancer suppressor gene. *Cancer Discov.*, **1**, 222–235.
36. Brannon,A.R., Reddy,A., Seiler,M., Arreola,A., Moore,D.T., Pruthi,R.S., Wallen,E.M., Nielsen,M.E., Liu,H., Nathanson,K.L. *et al.* (2010) Molecular stratification of clear cell renal cell carcinoma by consensus clustering reveals distinct subtypes and survival patterns. *Genes Cancer*, **1**, 152–163.
37. Kapur,P., Peña-Llopis,S., Christie,A., Zhrebker,L., Pavía-Jiménez,A., Rathmell,W.K., Xie,X.-J. and Brugarolas,J. (2013) Effects on survival of BAP1 and PBRM1 mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation. *Lancet Oncol.*, **14**, 159–167.
38. Patard,J.-J., Rioux-Leclercq,N., Masson,D., Zerrouki,S., Jouan,F., Collet,N., Dubourg,C., Lobel,B., Denis,M. and Fergelot,P. (2009) Absence of VHL gene alteration and high VEGF expression are associated with tumour aggressiveness and poor survival of renal-cell carcinoma. *Br. J. Cancer*, **101**, 1417–1424.
39. Yao,M., Yoshida,M., Kishida,T., Nakaigawa,N., Baba,M., Kobayashi,K., Miura,T., Moriyama,M., Nagashima,Y., Nakatani,Y. *et al.* (2002) VHL tumor suppressor gene alterations associated with good prognosis in sporadic clear-cell renal carcinoma. *J. Natl. Cancer Inst.*, **94**, 1569–1575.