Dispersion analysis of PoTRA ranked mRNA mediated dysregulated pathways in Breast Invasive Cancer from a TCGA Pan-Cancer study

Margaret K Linan [Corresp., 1, 2], **Valentin Dinu** [1, 2]

[1] Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ, United States

[2] Department of Biomedical Informatics, Arizona State University, Tempe, AZ, United States

Corresponding Author: Margaret K Linan
Email address: mlinan@asu.edu

Background. Our publication of the new pathways of topological rank analysis (PoTRA) algorithm demonstrated a novel approach for using the Google Search PageRank algorithm to analyze gene expression networks to identify biological pathways significantly disrupted in hepatocellular carcinoma. In order to apply the PoTRA algorithm to analyze other cancer gene expression data sets, of various sizes and normal:tumor ratio composition, two important questions must be answered: 1. What is the optimal normal:tumor sample ratio?; and 2. What is the minimum number of samples that should be used for PoTRA analysis? To address these questions, the average standard deviation (SD) in PoTRA-ranked mRNA mediated dysregulated pathways was studied using randomly sampled data sets with various normal:tumor ratios and sizes drawn from the TCGA Breast Invasive Carcinoma (TCGA-BRCA) project.

Methods. To identify the optimal normal:tumor sample ratios, the SD analysis used random combinations of 1:N unbalanced normal:tumor data sets: (1:1, 1:2, 1:3, 1:5, 1:7, 1:9). To identify the minimum sample size, random resampling of normal and tumor samples of various sizes are used: (3 vs 3), (5 vs 5), (10 vs 10), (25 vs 25), (50 vs 50), (75 vs 75), (100 vs 100), and (113 vs 113).

Results. This analysis suggests that the 1:1 ratio achieves the lowest average rank variation and that the minimum sample size of 50 normal and 50 tumor samples reaches a steady state in the average rank variation.

Conclusion. In conclusion, future applications of the PoTRA algorithm to analyze gene expression data sets such as TCGA should use balanced data sets as well as a minimum sample size of 50 for both normal and tumor to ensure the most robust performance.

Dispersion Analysis of PoTRA Ranked mRNA Mediated Dysregulated Pathways in Breast Invasive Cancer from a TCGA Pan-Cancer Study

Margaret K. Linan[1,2], Valentin Dinu[1,2]

[1]Department of Health Sciences Research, Mayo Clinic, Scottsdale, Arizona, United States of America
[2]Department of Biomedical Informatics, Arizona State University, Tempe, Arizona, United States of America
Corresponding author:
Margaret K. Linan[1,2]

Mayo Clinic, Department of Health Sciences Research, Scottsdale, Arizona 85259
Email address: mlinan@asu.edu

Abstract

Background. Our publication of the new pathways of topological rank analysis (PoTRA) algorithm demonstrated a novel approach for using the Google Search PageRank algorithm to analyze gene expression networks to identify biological pathways significantly disrupted in hepatocellular carcinoma. In order to apply the PoTRA algorithm to analyze other cancer gene expression data sets, of various sizes and normal:tumor ratio composition, two important questions must be answered: 1. What is the optimal normal:tumor sample ratio?; and 2. What is the minimum number of samples that should be used for PoTRA analysis? To address these questions, the average standard deviation (SD) in PoTRA-ranked mRNA mediated dysregulated pathways was studied using randomly sampled data sets with various normal:tumor ratios and sizes drawn from the TCGA Breast Invasive Carcinoma (TCGA-BRCA) project.
Methods. To identify the optimal normal:tumor sample ratios, the SD analysis used random combinations of 1:N unbalanced normal:tumor data sets: (1:1, 1:2, 1:3, 1:5, 1:7, 1:9). To identify the minimum sample size, random resampling of normal and tumor samples of various sizes are used: (3 vs 3), (5 vs 5), (10 vs 10), (25 vs 25), (50 vs 50), (75 vs 75), (100 vs 100), and (113 vs 113).
Results. This analysis suggests that the 1:1 ratio achieves the lowest average rank variation and that the minimum sample size of 50 normal and 50 tumor samples reaches a steady state in the average rank variation.
Conclusion. In conclusion, future applications of the PoTRA algorithm to analyze gene expression data sets such as TCGA should use balanced data sets as well as a minimum sample size of 50 for both normal and tumor to ensure the most robust performance.

Introduction

The Cancer Genomics Cloud (CGC) platform was developed by Seven Bridges and funded by the National Cancer Institute so that the large scale analyses of open and controlled cancer genomics data can be executed at little or no-cost (Lau et al., 2017). Multi-omics repositories such as The Cancer Genome Atlas (TCGA) make available large scale cancer genomics data as unbalanced sets of normal and tumor (Weinstein et al., 2013). A class imbalance is defined as a set of data with unequal numbers of samples in each class and thus results in a majority class and minority class (He and Ma, 2013). In the field of data mining, this imbalance impacts the accuracy and error rate of classifiers (He and Ma, 2013). Similarly in the field of bioinformatics, and as seen in this work, a computational tools that are applied to unbalanced data sets will have more variation in its results. Therefore, different sizes of the balanced data set must be used with the computational tool to determine its threshold for robustness (i.e., the size of the balanced data set that results in the least variation in the reported results). There are several methods in the field of data mining that can be used address the imbalance problem, such as sampling and skew-insensitivity (He and Ma, 2013).

These sampling methods are standard techniques for improving classification accuracy and include the random under- and oversampling of the majority and the minority classes by a factor chosen by the user (Chawla et al, 2008). In the case of bioinformatics, such sampling techniques could potentially bias the resulting metrics of any computational tool. Thus, a better approach would be to randomly resample each class in their entirety while making sure that both classes are equally represented in multiple balanced data sets. Similarly, the skew-insensitivity techniques that use machine learning algorithms would not be an ideal or cost-effective solution for balancing large scale and labeled genomic data sets because these algorithms build predictive models (He and Ma, 2013).

We recently published the Pathways of Topological Rank Analysis (PoTRA) algorithm (Li, Liu and Dinu, 2018), which demonstrated a novel approach for using the Google Search PageRank algorithm (Page et al., 1999) to analyze gene expression networks to identify biological pathways significantly disrupted in hepatocellular carcinoma. In order to apply the PoTRA algorithm to analyze other cancer gene expression data sets, of various sizes and normal:tumor ratio composition, two important questions must be answered: 1. What is the optimal normal:tumor sample ratio?; and 2. What is the minimum number of samples that should be used for PoTRA analysis?

In the present work, to address these questions, the average standard deviation (SD) in PoTRA-ranked mRNA mediated dysregulated pathways was studied using randomly sampled data sets with various normal:tumor ratios and sizes drawn from the TCGA Breast Invasive Carcinoma (TCGA-BRCA) project. Sample permutation and random resampling without replacement were used for the creation of test sets. These test sets were used to determine the robustness threshold

of the PoTRA algorithm (Li, Liu and Dinu, 2018). Determining the robustness threshold for this tool is important because it helps reduce the variation in the aggregated pathways ranks and thus improves PoTRA's accuracy.

Materials & Methods

The CGC platform (Lau et al., 2017) and Docker (Merkel, 2014) were utilized in the creation of containers for multiple data management and analysis computational tools, leveraging the PoTRA algorithm (Li, Liu and Dinu, 2018). Rabix composer was used to port these tools to the CGC. The HTSeq-FPKM normalized protein-coding mRNA data from the Breast Invasive Cancer TCGA project (TCGA-BRCA) was extracted from the CGC's TCGA GRCh38 repository. The data set consisted of 113 normal and 1102 tumor samples. These data were analyzed in the CGC with the application of R scripts (R Core Team, 2013) for the principal components analysis (PCA), random resampling, PoTRA, permutation and standard deviation analyses (Figure 1).

PCA analysis was performed on the CGC platform with a docker container that included the R libraries ggplot2, ggpubr, ggfortify (Wickham, 2016; ggpubr; Tang, 2018). The aim of the PCA analysis was to explore the distributions of the normal and tumor TCGA-BRCA data sets. The gene expression patterns of normal and tumor samples are often expected to be distinct, in some cases when the normal sample is located in affected non-tumor tissue, the expression patterns can overlap those of the tumor sample. In both cases, PoTRA was used to further determine if the normal and tumor tissue samples had detectable differences (P-value < 0.05) in the PageRank detected hub genes of 301 KEGG pathways.

Additionally, this data set was divided into the following combinations of normal and tumor to further study the robustness of the PoTRA pathway analysis algorithm: 1. Sample size analysis: (3 vs 3), (5 vs 5), (10 vs 10), (25 vs 25), (50 vs 50), (75 vs 75) and (100 vs 100), 2. Ratio analysis: (113 vs 113), (113 vs 226), (113 vs 339), (113 vs 565), (113 vs 791), (113 vs 1017) with 200 datasets for each. Random resampling was used to randomly choose samples from the normal and tumor data for each of the 1:1 and the 1:N subsets. All ratios (1:1, 1:2, 1:3, 1:5, 1:7, 1:9) of the data were permuted by a factor of 20 using a docker container that included the R libraries dplyr, modelr and purrr (dplyr; modelr; purr) on the CGC. Then the PoTRA algorithm was applied to detect significantly dysregulated pathways and to rank these pathways. The standard deviation algorithm was applied to the rank data to determine the minimum sample size and ratio of normal and tumor data that are associated with the most robust performance of the PoTRA algorithm (i.e., lowest average rank variation).

Results

TCGA Samples

Among the 17 TCGA cancer types (Table 1) that had HTSeq FPKM normalized data sets and more than 3 normal samples, the breast invasive cancer project (TCGA-BRCA) had the most tumor samples (n=1,102).

The top ranked dysregulated pathways (Table 2) that resulted from the PoTRA analysis of the BRCA-TCGA datasets had the following in common, they had the lowest average Fisher's Exact (FE) test p-values, average variability and average rank.

Standard Deviation Analyses

Impact of Sample Size on Pathway Rank Variability

Among the sample sizes that were analyzed using a 1:1 normal:tumor ratio, the lowest average standard deviation for the ranks of the dysregulated pathways detected in the HTSeq normalized mRNA data were for sample sizes 50, 75, 100 and 113 (Figure 3). The average standard deviations for these four sample sizes are 35, 35, 36 and 33. The smaller sample sizes have the highest variability in pathway ranks. Therefore, a minimum of sample size 50 is recommended for both phenotypes.

Impact of Normal:Tumor Ratio on Pathway Rank Variability

For the 1:N ratio analysis in the non-permuted data (Figure 3), the ratio 1:9 achieves the lowest average standard deviation. Furthermore, to determine if this conclusion remains true after permuting the data, the data sets for the 1:N ratios were permuted by a factor of 20, and had their average standard deviation compared to the non-permuted data (Figure 4). The non-permuted ratios 1:7 and 1:9 no longer had the lowest average standard deviation. Rather, the permuted ratio of 1:1 had the lowest average standard deviation.

These results demonstrate that the lowest average standard deviation can be achieved by the ratio 1:1. This means that the ranks of the dysregulated pathways from the PoTRA algorithm will be more consistently reported when a ratio of 1:1 is used to create the balanced data sets that are then analyzed with PoTRA.

The top 10 dysregulated pathways for the TCGA-BRCA project was further explored (Figure 5) to determine how much the average ranks of these pathways were affected by the increasing ratio size. Interestingly, these pathways can be grouped by the changes in average rank as the ratio increases. In the first group the cAMP and PI3K-Akt signaling pathways, the human papillomavirus infection and the proteoglycans in cancer pathway have similar changes in their average ranks, with the cAMP signaling pathway being the most affected by the increasing ratio size. In the second group, the Ras and cGMP-PKG signaling pathways have similar changes in their average ranks. In the third group, the Rap1 signaling and Regulation of actin cytoskeleton pathways have very similar changes in their average ranks. In the fourth group, Pathways in

cancer and MAPK signaling pathways have similar changes in average rank, with the MAPK signaling pathway being the least affected by the increasing ratio size.

Discussion

Multiple pathway analysis algorithms have been created to analyze and rank pathways associated with disease (Subramanian et al., 2005; Mi et al., 2013; Li, Liu and Dinu, 2018, Panther). Each algorithm takes different approaches to determining the robustness and accuracy of their pathway ranks. They also take into consideration different types of information to help differentiate or confirm the biological importance of the resulting ranked pathways such as stratifying the pathways by survival outcomes to using multiple public resources such as GSEA and EnrichNet to validate the algorithm's ranked pathways (Verbeke et al., 2015; Liu, Wei and Ruan, 2017).

In an associated work, the ranked dysregulated pathways from a TCGA pan-cancer analysis using the PoTRA algorithm were validated by cross-referencing the highest ranked pathways against the KEGG database (Linan M, Wang J, Dinu V). In the present work, the variation in the ranked pathway results is not overlooked but instead studied so that the root cause of the variation can be found and minimized. Indeed the variation in the ranked pathways is due to the unbalanced nature of the HTSeq FPKM normalized mRNA data from the TCGA-BRCA project. The unbalance is due to higher number of tumor vs. normal samples and is common in cancer research, including the TCGA project, as illustrated in Table 1.

In this work, we investigated the effect of normal:tumor ratio composition and sample size on the variability of pathway ranks in the PoTRA analysis tool. We concluded that the 1:1 ratio achieved the lowest average pathway rank variation by comparing using a range of non-permuted and permuted normal:tumor data sets, (1:1, 1:2, 1:3, 1:5, 1:7, 1:9). By using different sample sizes of 1:1 balanced data sets (3, 5, 10, 25, 50, 75, 100 and 113), we concluded that the minimum size for the sample data set should be 50 normal and 50 tumor samples. This will ensure the most robust detection of mRNA-mediated dysregulated pathways with the PoTRA program. To further explore the robustness of the PoTRA tool, additional analyses could be performed by clustering the tumor mRNA data by gene expression values to identify potentially distinct disease subsets or by taking into account additional clinical phenotype data, such as survival information. Overall, the present work informs users how to minimize the amount of variation in the pathway rankings of their PoTRA results and how to potentially test and improve the robustness of other biological pathway analysis tools.

The present work also demonstrates how pathway ranks are changed by data set size. Interestingly, the MAPK pathway had the least variation in the different ratios of normal:tumor, perhaps because this pathway is very active in breast invasive carcinoma. In contrast, the cAMP signaling pathway had the greatest variability, perhaps because this pathway is associated with

tumor progression and therefore targeted by chemotherapies that are prescribed to the BRCA patients. In fact, the pathways with greatest variability (cAMP signaling, Human Papillomavirus infection, PI3K-Akt signaling pathway, Proteoglycans in cancer) also had no detectable differences (FE Test P-value > 0.05) in hub genes between normal and control mRNA networks in ratios 1:7 and 1:9. This may indicate that PoTRA can be used to measure the efficacy of a chemotherapy that target genes in particular pathways.

Conclusions

Using a 1:1 ratio of normal and tumor sample as well as minimum of 50 samples per phenotype reduces the variability in mRNA mediated dysregulated pathways detected by the PoTRA algorithm. The use of this ratio and minimum sample size will ensure the most robust performance of the PoTRA algorithm.

Acknowledgements

We would like to thank the Cancer Genomics Cloud technical support team.

References

Chawla NV., Cieslak DA., Hall LO., Joshi A. 2008. Automatically countering imbalance and its empirical relationship to cost. *Data Mining and Knowledge Discovery* 17(2):225-252. DOI: 10.1007/s10618-008-0087-0.

dplyr: A Grammar of Data Manipulation 2018. *Available at https://CRAN.R-project.org/package=dplyr* (accessed September 21, 2018).

ggpubr: "ggplot2" Based Publication Ready Plots 2018. *Available at https://CRAN.R-project.org/package=ggpubr* (accessed September 21, 2018).

He H., Ma Y. (eds.) 2013. Imbalanced Datasets: From Sampling To Classifiers. In: *Imbalanced Learning: Foundations, Algorithms, and Applications*. John Wiley & Sons, Inc, 43–59.

Lau JW., Lehnert E., Sethi A., Malhotra R., Kaushik G., Onder Z., Groves-Kirkby N., Mihajlovic A., DiGiovanna J., Srdic M., Bajcic D., Radenkovic J., Mladenovic V., Krstanovic D., Arsenijevic V., Klisic D., Mitrovic M., Bogicevic I., Kural D., Davis-Dusenbery B., Seven Bridges CGC Team 2017. The Cancer Genomics Cloud:

Collaborative, Reproducible, and Democratized-A New Paradigm in Large-Scale Computational Research. *Cancer Research* 77:e3–e6. DOI: 10.1158/0008-5472.CAN-17-0387.

Li C., Liu L., Dinu V. 2018. Pathways of topological rank analysis (PoTRA): a novel method to detect pathways involved in hepatocellular carcinoma. *PeerJ* 6. DOI: 10.7717/peerj.4571.

Liu L., Wei J., Ruan J. 2017. Pathway Enrichment Analysis with Networks. *Genes* 8:246. DOI: doi: 10.3390/genes8100246.

Merkel D. 2014. Docker: lightweight Linux containers for consistent development and deployment. *Linux Journal*.

Mi H., Muruganujan A., Thomas PD. 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research* 41:D377–D386. DOI: 10.1093/nar/gks1118.

modelr: Modelling Functions that Work with the Pipe. 2018. *Available at https://CRAN.R-project.org/package=modelr* (accessed September 22, 2018)

Page L., Brin S., Motwani R., Winograd T. 1999. The PageRank citation ranking: bringing order to the web.

PANTHER: A Library of Protein Families and Subfamilies Indexed by Function 2018. *Available at https://genome.cshlp.org/content/13/9/2129.full* (accessed September 23, 2018).

purrr: Functional Programming Tools 2018. *Available at https://CRAN.R-project.org/package=purrr* (accessed September 23, 2018).

R Core Team 2013. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* 3:201.

Subramanian A., Tamayo P., Mootha VK., Mukherjee S., Ebert BL., Gillette MA., Paulovich A., Pomeroy SL., Golub TR., Lander ES., Mesirov JP. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 102:15545–15550. DOI: 10.1073/pnas.0506580102.

Tang Y., Horikoshi M., Wenxuan L. 2016. ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. *The R Journal 8.2*:478–489.

Verbeke L., Van den Eynden J., Fierro A., Demeester P., Fostier J. 2015. Pathway Relevance Ranking for Tumor Samples through Network-Based Data Integration. *PLOS ONE* 10. DOI: https://doi.org/10.1371/journal.pone.0133503.

Weinstein JN., Collisson EA., Mills GB., Shaw KM., Ozenberger BA., Ellrott K., Shmulevich I., Sander C., Stuart JM. 2013. The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nature genetics* 45:1113–1120. DOI: 10.1038/ng.2764.

Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Figure 1

The CGC workflow for the testing of PoTRA's robustness threshold.

This workflow is used for the detection of dysregulated pathways and the standard deviation analyses. 1) and 3) Both phenotypes are merged and random resampling is used to extract samples from both phenotypes. 5) The unbalanced data sets with ratios 1:7 and 1:9 are permuted by a factor of 20. 2),4),6) The PoTRA algorithm detects the dysregulated pathways and ranks them. 7) The standard deviation of the dysregulated pathway ranks are averaged and plotted.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27306v1 | CC BY 4.0 Open Access | rec: 30 Oct 2018, publ: 30 Oct 2018

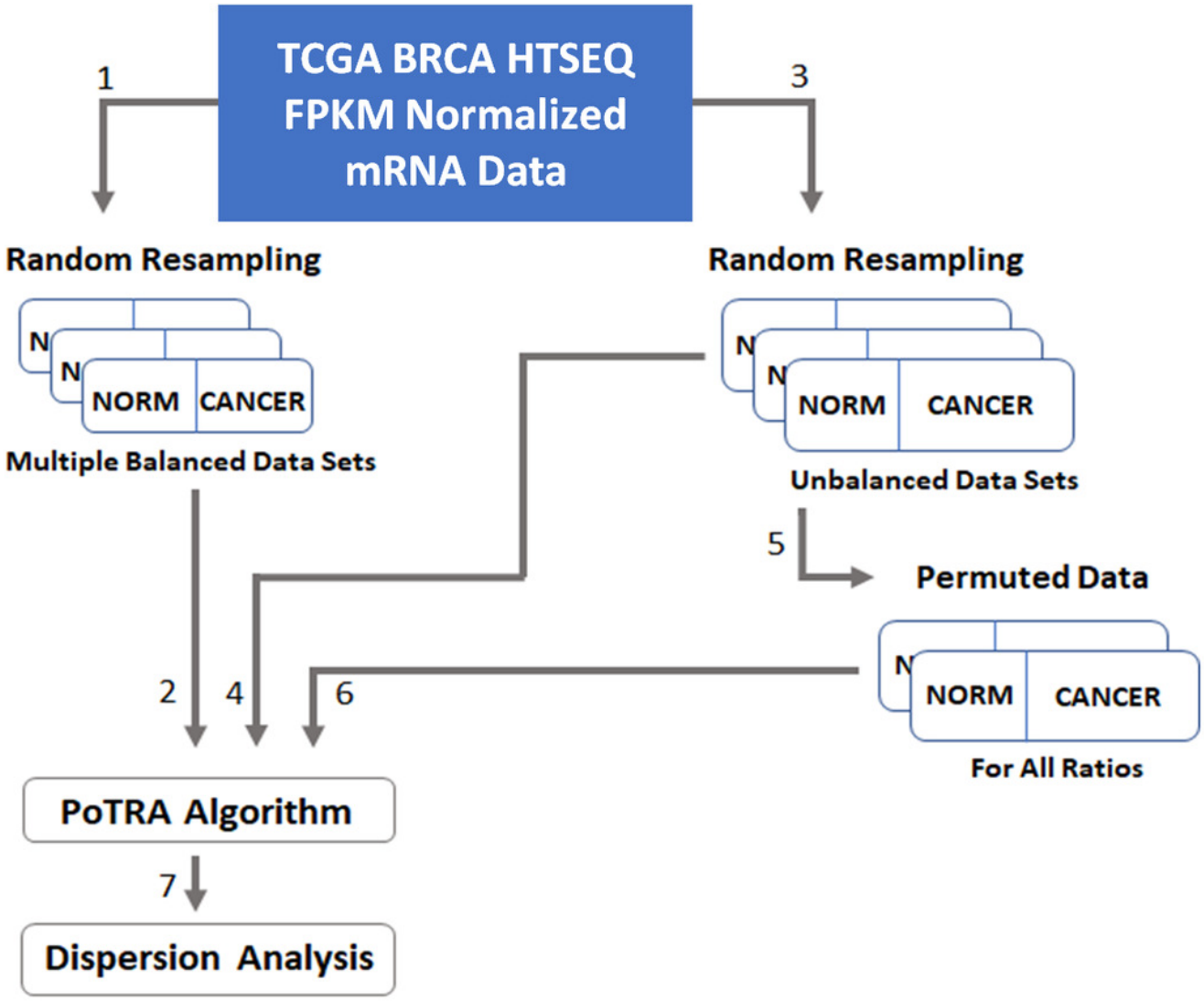

Figure 1 **The CGC workflow for the testing of PoTRA's robustness threshold.**

This workflow is used for the detection of dysregulated pathways and the standard deviation analyses. 1) and 3) Both phenotypes are merged and random resampling is used to extract samples from both phenotypes. 5) The unbalanced data sets with ratios 1:7 and 1:9 are permuted by a factor of 20. 2),4),6) The PoTRA algorithm detects the dysregulated pathways and ranks them. 7) The standard deviation of the dysregulated pathway ranks are averaged and plotted.

Figure 2

PCA Analysis of the Breast Invasive Cancer (BRCA) Project Data.

Normal and tumor BRCA HTSeq-FPKM normalized protein coding mRNA gene expression data.

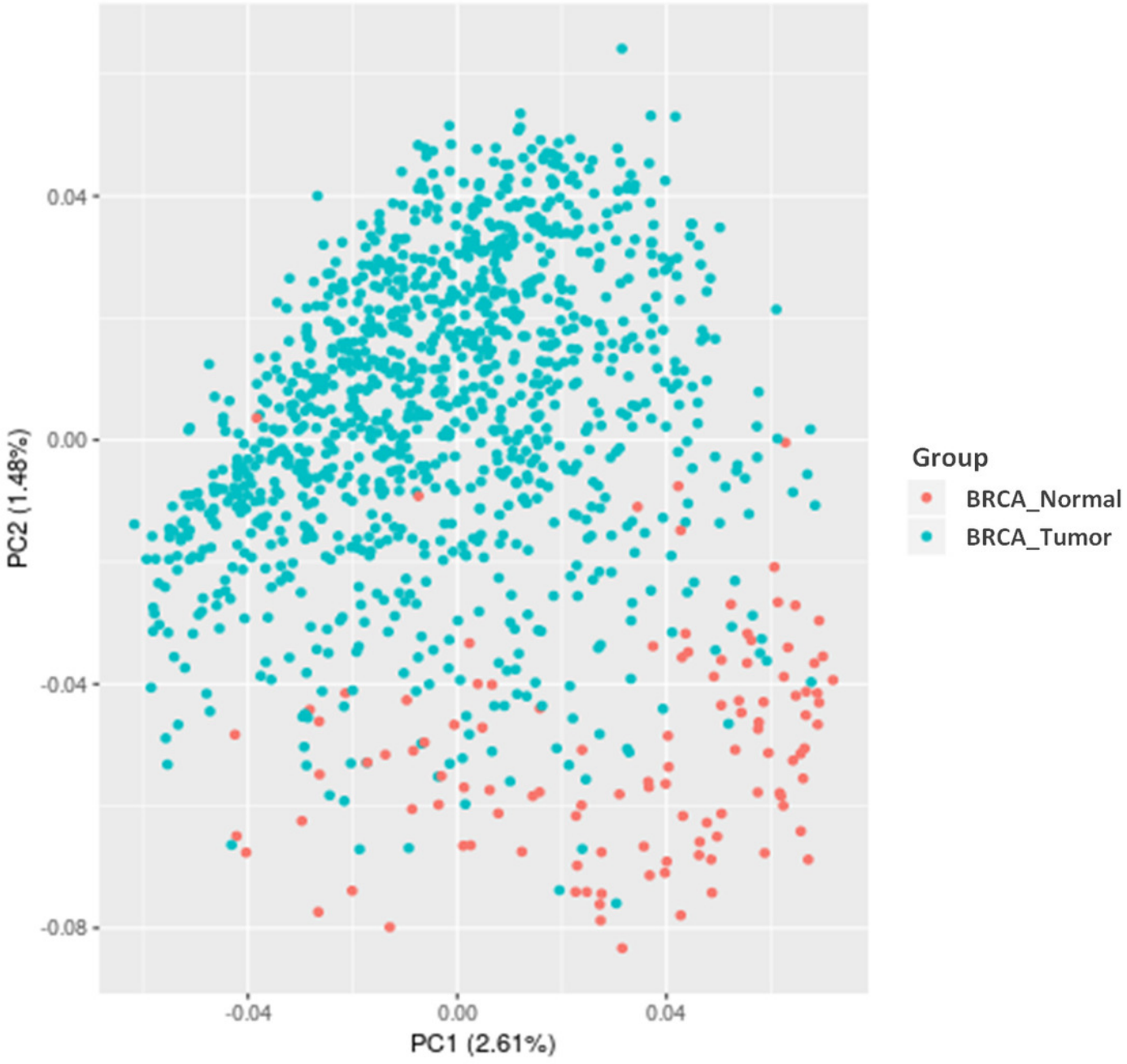

Figure 2 PCA Analysis of the Breast Invasive Cancer (BRCA) Project Data.
Normal and tumor BRCA HTSeq-FPKM normalized protein coding mRNA gene expression data.

Figure 3

Average Standard Deviation of Non-Permuted Ranks.

Line plot of the average standard deviation by the sample size of each phenotype (normal and tumor). The average standard deviation decreases as the sample size increases for both phenotypes. The sample size 50, is the minimum sample size needed per phenotype for the PoTRA algorithm to yield pathway ranks with the least variation.

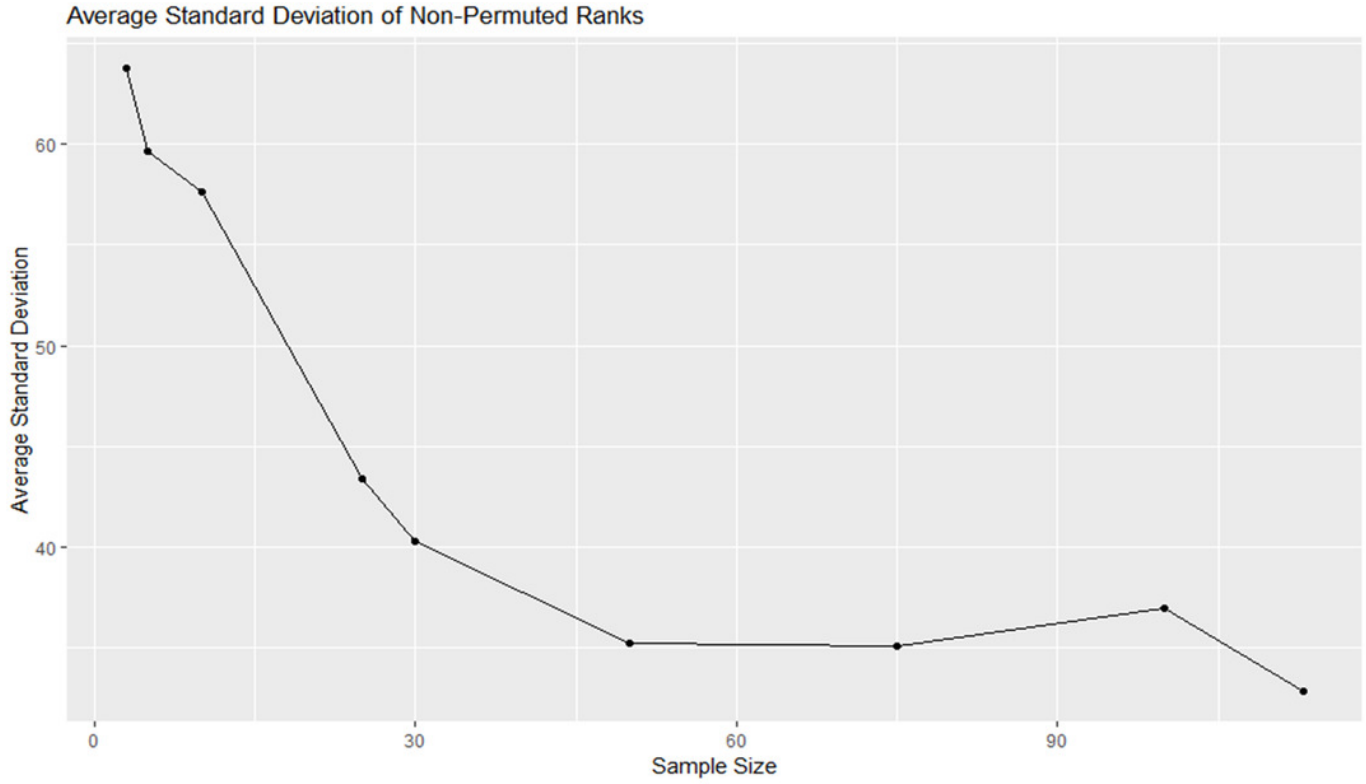

Figure 3 Average Standard Deviation of Non-Permuted Ranks.
Line plot of the average standard deviation by the sample size of each phenotype (normal and tumor). The average standard deviation decreases as the sample size increases for both phenotypes. The sample size 50, is the minimum sample size needed per phenotype for the PoTRA algorithm to yield pathway ranks with the least variation.

Figure 4

Average Standard Deviation of Permuted Ranks.

Line plot of the average standard deviation of permuted pathway ranks for the ratios 1:N (N=1,2,3,5,7,9) for each phenotype (normal:tumor). The permuted samples achieve the lowest average standard deviation for the Ratio 1:1. In conclusion, data sets that have a ratio of 1:1 are associated with the lowest average standard deviation of ranks.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27306v1 | CC BY 4.0 Open Access | rec: 30 Oct 2018, publ: 30 Oct 2018

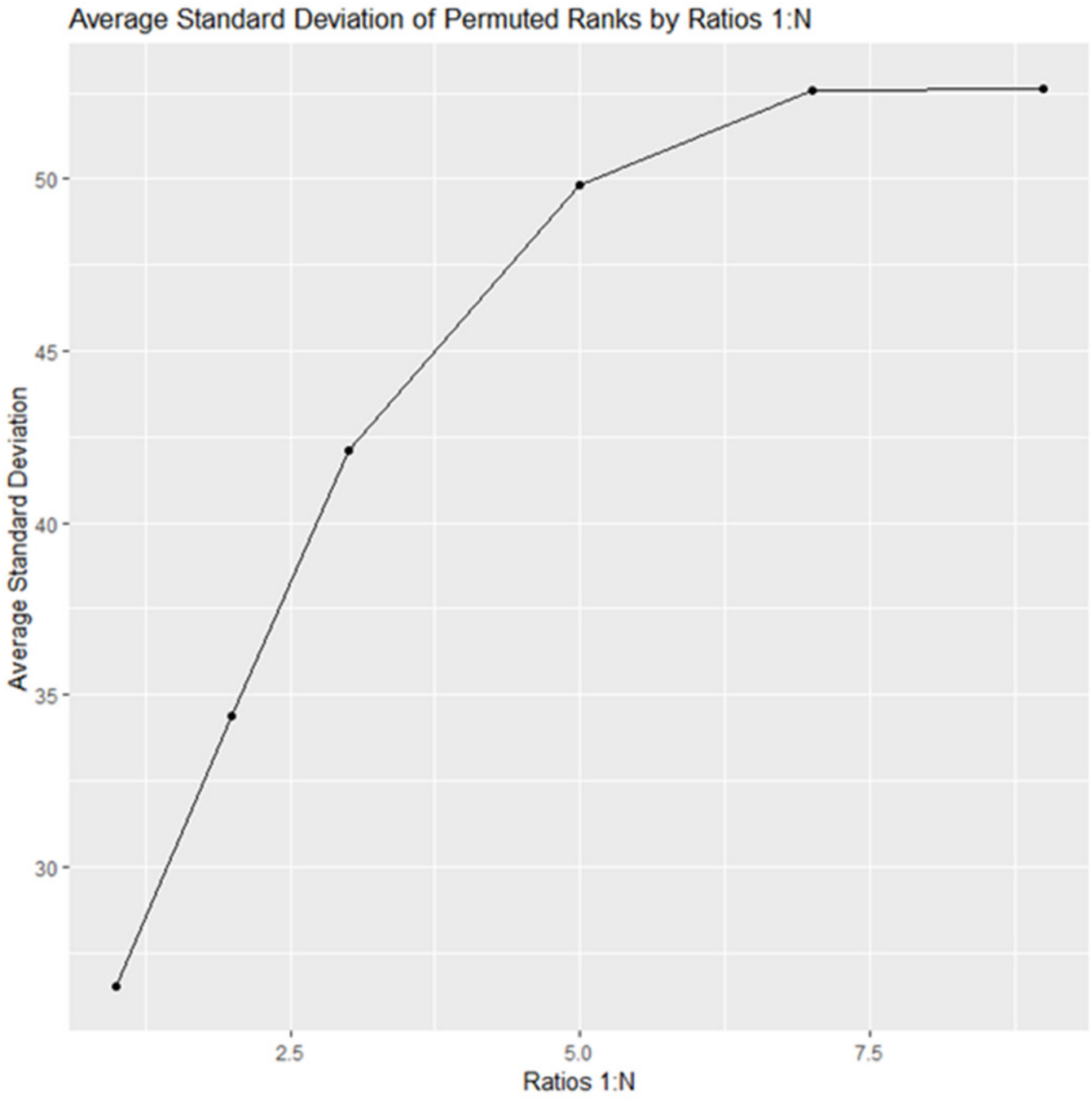

Figure 4 **Average Standard Deviation of Permuted Ranks.**
Line plot of the average standard deviation of permuted pathway ranks for the ratios 1:N (N=1,2,3,5,7,9) for each phenotype (normal:tumor). The permuted samples achieve the lowest average standard deviation for the Ratio 1:1. In conclusion, data sets that have a ratio of 1:1 are associated with the lowest average standard deviation of ranks.

Figure 5

Comparison of Permuted Average Ranks for the Top 10 Dysregulated Pathways.

The line plot compares the permuted and averaged ranks of the top 10 dysregulated pathways for the TCGA-BRCA project. The averaged ranks increase in value as the number of samples in the tumor data increases in the unbalanced data set, specifically where ratios 1:N (N=1,2,3,5,7,9) represents 113 normal samples and 113*N tumor samples in an unbalanced data set. The cAMP signaling pathway was most affected by the increasing ratio size, while MAPK signaling pathway was the least affected.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27306v1 | CC BY 4.0 Open Access | rec: 30 Oct 2018, publ: 30 Oct 2018

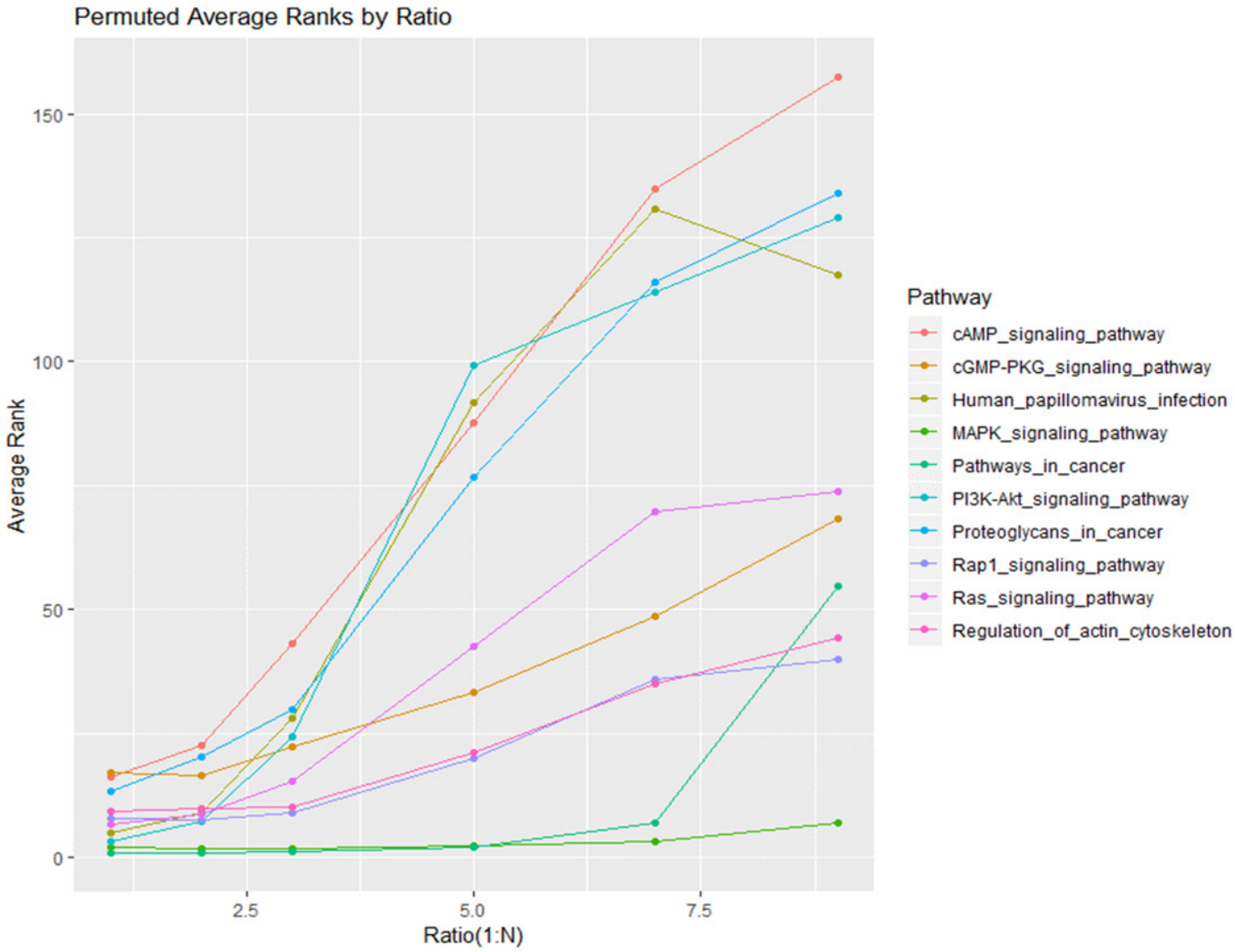

Figure 5 Comparison of Permuted Average Ranks for the Top 10 Dysregulated Pathways.
The line plot compares the permuted and averaged ranks of the top 10 dysregulated pathways for the TCGA-BRCA project. The averaged ranks increase in value as the number of samples in the tumor data increases in the unbalanced data set, specifically where ratios 1:N (N=1,2,3,5,7,9) represents 113 normal samples and 113*N tumor samples in an unbalanced data set. The cAMP signaling pathway was most affected by the increasing ratio size, while MAPK signaling pathway was the least affected.

Table 1(on next page)

The sample sizes for each phenotype by primary site.

Table 1 The sample sizes for each phenotype by primary site.

Primary Site of Cancer	Normal Samples	Tumor Samples
Adrenal Gland	3	257
Bile Duct	9	36
Bladder	19	414
Brain	5	667
Breast	113	1102
Cervix	3	304
Colorectal	51	644
Esophagus	11	161
Head and Neck	44	500
Kidney	128	891
Liver	50	371
Lung	108	1035
Pancreas	4	177
Prostate	52	498
Stomach	32	375
Thyroid	58	502
Uterus	35	607

1

Table 2(on next page)

Top 10 significantly dysregulated pathways.

Table 2 Top 10 significantly dysregulated pathways.

Pathways	Average Fisher's Exact P-Value	Variability	Average Rank
Pathways in cancer	9.97E-147	0	1.00
MAPK signaling pathway	4.52E-102	0	2.00
PI3K-Akt signaling pathway	1.17E-66	0	3.00
Ras signaling pathway	2.34E-45	0.90	5.30
Human papillomavirus infection	2.70E-43	1.60	5.20
Rap1 signaling pathway	1.57E-38	2.37	7.00
cAMP signaling pathway	5.17E-34	5.43	9.50
cGMP-PKG signaling pathway	1.16E-28	4.62	11.80
Proteoglycans in cancer	2.19E-29	6.29	13.00
Regulation of actin cytoskeleton	1.56E-27	7.52	14.10

1