Multiplexed Cancer Pathway Analysis

nCounter® PanCancer Pathways Panel *for* Gene Expression

Lucas Dennis, Patrick Danaher, Rich Boykin, Joseph Beechem

NanoString Technologies, Inc., Seattle WA 98109

v1.0 | MARCH 2014

Multiplexed Cancer Pathway Analysis Using nCounter® PanCancer Pathways Panel

Introduction

All cancers must evolve a means of sustaining self-sufficient growth and evading apoptosis[1,2]. This process typically occurs via the accumulation of mutational events that confer a growth advantage through deregulation of the molecular pathways controlling cell growth and cell fate[3]. Mutations in over 100 genes are known to drive tumorgenesis and within any given tumor there are between 2-8 mutated "driver genes" modulating the activity of critical molecular pathways[4]. Studying the deregulation of molecular pathways impacted by mutational events as well as monitoring expression of these driver genes is critical to gaining a complete understanding of the biology underlying cancer.

Molecular pathways are an attractive organizing principle for analysis of gene expression data as they provide a means to combine the noisy information in individual genes into stable and meaningful representations of fundamental biological processes[5]. Gene expression profiling has long been used within the cancer field to stratify cell populations and classify tumors[6–8]. This powerful ability is largely due to the fact that the gene expression state of a cell or tissue contains information about the biological processes occurring within a sample[9]. Pathway-based analyses provide a holistic view of the changes to fundamental biological processes allowing for deregulation of regulatory pathways to be linked back to "driver gene" status.

Pathways of the PanCancer Pathways Panel

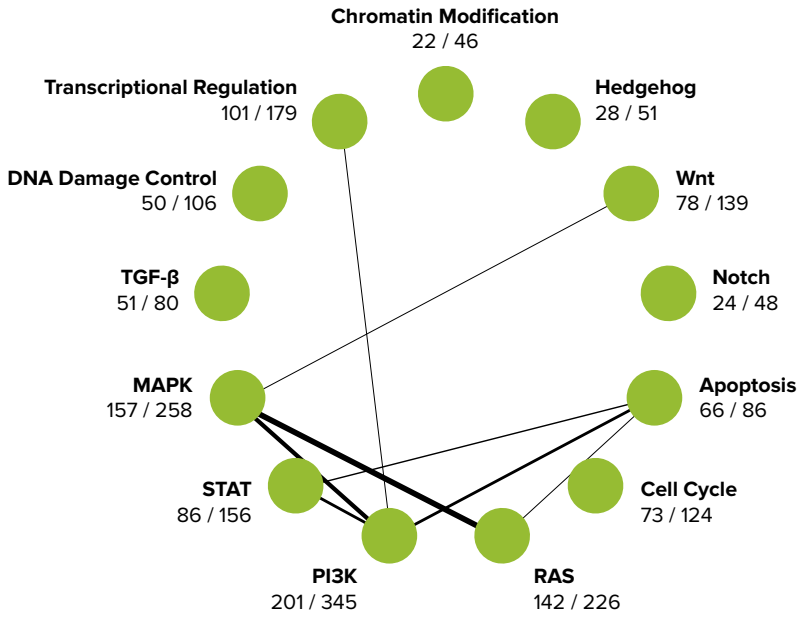


FIGURE 1: Pathways of the nCounter PanCancer Pathways Panel. Circles representing each of the 13 canonical pathways show the number of genes selected for the panel in relationship to the total number of known genes identified for each pathway as identified by KEGG. Lines have been drawn to show the relative overlap of genes that belong to multiple pathways, with thickness of line relating to number of shared genes.

In their seminal paper, Vogelstein *et al.*, argue that understanding the deregulation of pathways is integral to understanding the biology of any cancer. A growing number of studies have demonstrated that pathway based analysis of gene expression information provides a framework for understanding the discrete changes between the biology of different cancers and cancer subtypes[10–15]. To better understand the intricate network of pathways and interactions, NanoString has taken a biology-guided, data-driven approach to identify over 700 essential genes that capture the activity of 13 canonical cancer pathways* and associated driver genes (**FIGURE 1**). Each of the pathways was mapped to publicly available data-sources (KEGG **http://www.genome.jp/kegg/**, Reactome **http://www.reactome.org/**, GO **http://www.geneontology.org/**)[16–19] as described below (see Panel Design) in order to create a tool designed to enable a pathway-based approach to exploring the molecular mechanisms of cancer and cancer subtyping.

** For a detailed list and description of the 13 canonical cancer pathways, please visit www.nanostring.com/pancancer.*

PANEL DESIGN

Identification of Cancer Pathway Genes

Due to the coordinate nature of coexpression among genes within a pathway, a relatively small number of genes can capture the majority of the variability in a pathway's gene expression[19]. The 13 canonical cancer pathways described in the nCounter PanCancer Pathways panel contain a large number of genes. However, with a principled approach to gene selection, expression information for only 700 genes is sufficient to capture ~90% of the gene expression variability for all genes within the pathways (**FIGURE 2**).

In order to identify the maximally informative core set of genes that comprise the nCounter PanCancer Pathways Panel gene list, a variety of selection criteria were employed. We began by including 125 genes in which mutations are known to drive oncogenesis[3]. We further included 127 genes that appeared in three or more pathways and 50 genes that were consistently differentially expressed between tumor and healthy controls in The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/) datasets across a number of tissues. Finally, an analysis of the cancer literature identified another 75 genes as key members of the cancer landscape. Together these methods added 297 pathway genes central to the understanding of cancer biology to our panel.

Additional cancer pathway-associated genes were identified using a data-driven statistical approach designed to capture genes that explained the maximal amount of expression variability for each pathway. That is, we sought genes that were not just highly variable in their own expression, but whose expression state carried information about the expression level of other genes (see text box). After this data-driven selection, the resulting genes were cross-referenced against our analysis of the cancer literature before inclusion in the final nCounter PanCancer Pathways Panel gene list.

An annotated list of all genes in the nCounter PanCancer Pathways panel is available at www.nanostring.com. This interactive list contains information for each gene in the panel including its classification, i.e. driver gene, pathway member or housekeeping gene, accession, alternate names and probe target sequence. Additionally, the gene list contains information about each of the pathways along with database source information from KEGG, Reactome and GO.

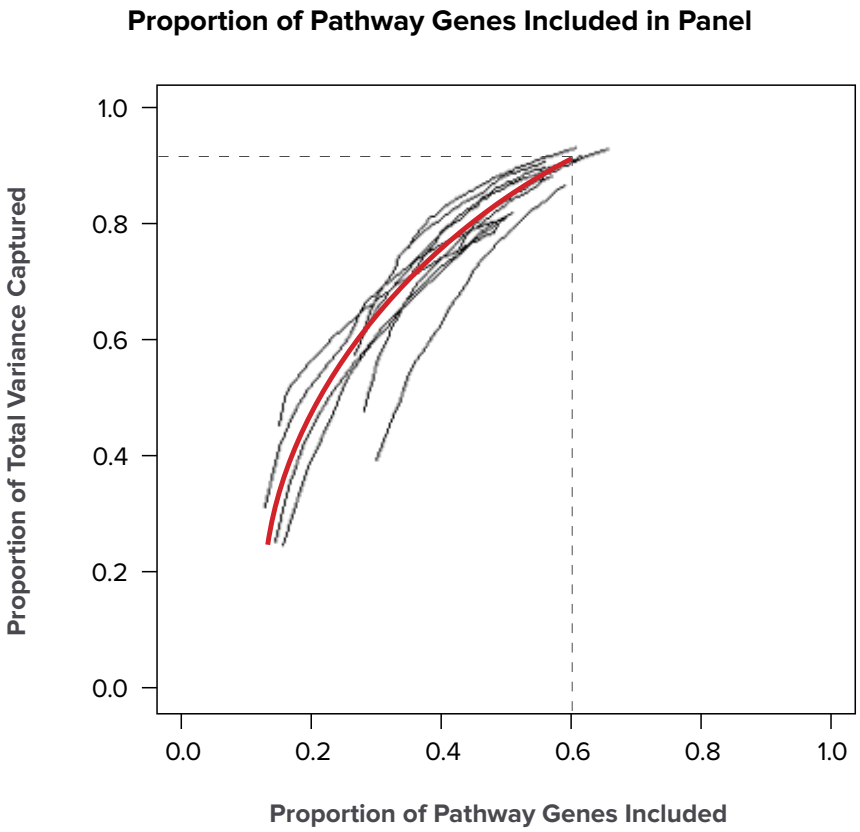


FIGURE 2: Proportion of total variance in pathways captured by the selected pathway genes. Individual black lines denote the proportion of pathway gene variance captured by the selected gene set as gene number increases. Each line represents a pathway, and each line's upper-right terminus corresponds to the number of genes ultimately selected for inclusion in the gene list for a given pathway. The thick red line denotes the proportion of pathway gene variance captured on average across all pathways and highlight that 60% of genes in a pathway are sufficient to capture 90% of the gene expression variance within a pathway.

Identifying Maximally Informative Pathway Genes

For each pathway and each TCGA dataset, we calculated a stable estimate of the pathway genes' covariance matrix using the graphical lasso[20]. Given a set of ***selected*** pathway genes, this covariance matrix allowed us to measure the proportion of total variability in the ***unselected*** pathway genes that could be captured by the ***selected*** genes. For each pathway, we initialized our selected gene set as the intersection of the pathway's genes and the 297 genes already included in our list. We next selected the gene that, in combination with the already selected genes, predicted the maximum amount of variance of the unselected genes across all the TCGA datasets. This process was iterated until the ***selected*** genes captured ~90% of the variance of all genes in the pathway (**FIGURE 2**).

Our examination of the TCGA dataset also allowed for identification of genes with consistently low variance across many cancers. A subset of these genes was selected as "housekeepers" intended to aid in data normalization. These 40 genes were also selected based on their ability to provide coverage of the wide range of expression levels typically observed in experimental datasets.

Pathway-based Analyses of Gene Expression Data

The current understanding of cancer pathway architecture and the state of statistical methodologies for pathway-based analysis have progressed to the point where a pathway-based analysis can provide an exceptionally informative first look at a gene expression dataset. In order to demonstrate the power of pathway-based analysis, we examined a subset of publicly available gene expression information from 823 breast cancers and 105 normal breast samples available at TCGA. Breast cancers are known to cluster into four subtypes: Luminal A, Luminal B, HER2-enriched, and Basal-like[6,9]. We used the PAM50 algorithm[20] to estimate the "intrinsic subtype" of every sample in the dataset.

In order to gain initial insight into the biology of these breast cancer samples, we transformed our measurements of gene expression into measurements of pathway deregulation. A number of methods exist for using gene expression to score pathway activity or deregulation[12,21,22]. We found the Pathifier algorithm[12] to extract the greatest clarity from this dataset. Pathifier scores pathway deregulation by fitting a curve that captures the maximal variability of pathway gene expression and then projecting every observation onto that curve. An observation's deregulation score is its distance along the curve from the average normal sample.

In **FIGURE 3,** we display a heatmap of pathway deregulation scores in the breast cancer data, highlighting the tremendous clarity that examination of gene expression data at a pathway level affords. We immediately gain insight into the intrinsic biology of the four breast cancer subtypes. For example, one cluster of Luminal A samples has very low deregulation of all pathways, behaving almost like normal tissue. Other clusters of Luminal A and B samples exhibit much higher pathway deregulation in a set of eight pathways (Notch, STAT, TGF-β, Transcription Regulation, WNT, Hedgehog, RAS and MAPK). These observations suggest that the luminal breast cancer phenotype may result from multiple sources of pathway deregulation. Basal-like and HER2-enriched samples show very high deregulation of an almost entirely different set of pathways (Apoptosis, DNA Damage Control, Chromatin Modification, PI3K and Cell Cycle), suggesting these tumors rely on a fundamentally different mechanism to gain a selective advantage, which is consistent with the observations made by others[9]. The boxplots of pathway deregulation scores in **FIGURE 4** make these observations clear. Interestingly, our analysis shows that both cell fate and cell survival pathways are deregulated within the cancer samples analyzed. Two pathways, STAT and TGF-β, are deregulated in both Basal-like tumors and the non-normal-like Luminal tumors. This pathway-based analytical framework provides a clear means of visualizing the need for loss of regulatory control in both key cellular processes in order to support tumor proliferation. **FIGURE 5** examines the deregulation scores of the five cell fate pathways. Some pathways appear to behave in concert having highly correlated deregulation scores, such as the Wnt and Hedgehog pathways (**FIGURE 5A**). In contrast, while both the Wnt and Chromatin modification pathways are deregulated in cancer, the extent of their deregulation appears to negatively correlated (**FIGURE 5B**). A comparison of activity for the Notch and Wnt pathways in these samples highlight a third interesting mode of co-regulation: while all Luminal samples exhibit similar levels of Wnt deregulation, deregulation of the Notch pathway separates Luminal samples into a highly deregulated cluster and a minimally deregulated cluster (**FIGURE 5C**). Finally, **FIGURE 5D** shows consistent levels of deregulation of the Transcriptional Regulation pathway across most samples, with a small subset of samples of all subtypes experiencing extreme deregulation relative to normal.

Pathway Deregulation Values Across All Cancer Pathways

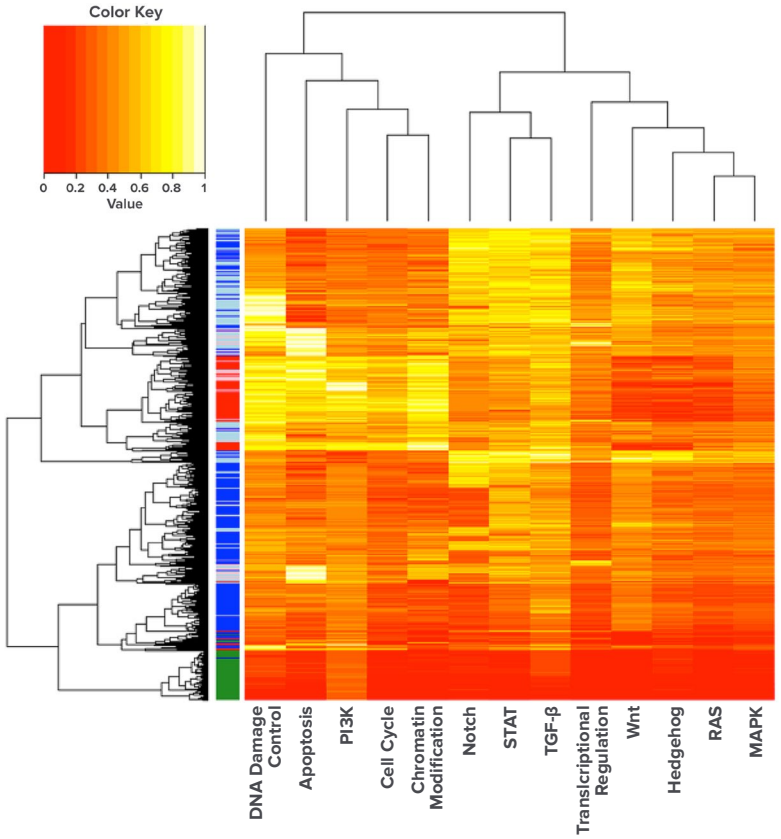


FIGURE 3: Pathway deregulation scores in breast cancer samples. Pathifier was used to calculate deregulation scores for each pathway (x-axis) in each sample (y-axis). Samples are colored according to intrinsic subtype, with Basal-like (red), HER2-enriched (pink), Luminal A (dark blue), Luminal B (light blue), and normal samples (green). Deregulation scores were generated relative to expression in normal breast tissue and are shown on a continuum from no deregulation (red) to highly deregulated (yellow).

Deregulation of Cell Fate and Cell Survival Pathways

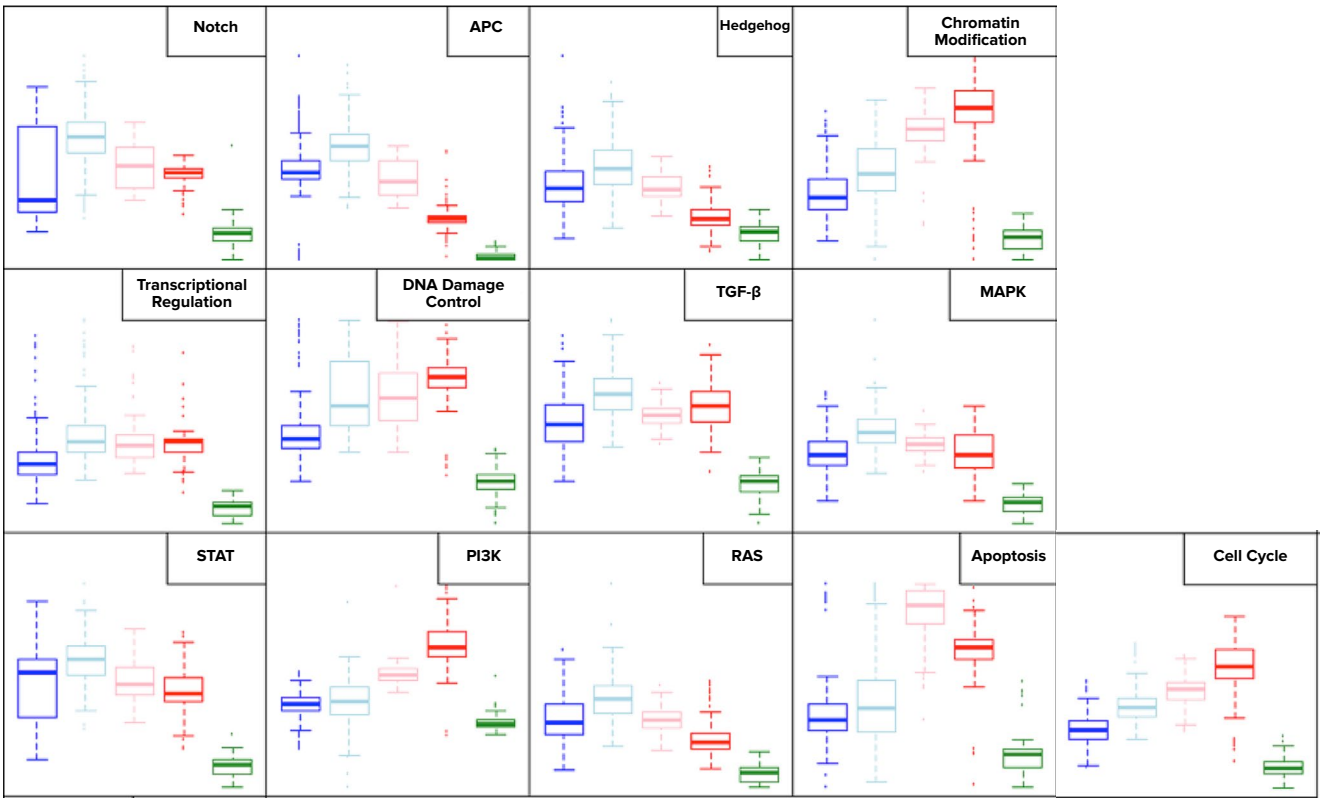


FIGURE 4: Boxplots of pathway deregulation scores by intrinsic subtype. The distribution of Pathifier deregulation scores of each pathway is plotted for each intrinsic subtype. Samples are colored according to intrinsic subtype, with Basal-like (red), HER2-enriched (pink), Luminal A (dark blue), Luminal B (light blue), and normal samples (green). The top and bottom of the box delineate the upper and lower quartiles, with the thick line within each box representative of the median. Whiskers extend to capture all data within two standard deviations of the mean.

Pathway Deregulation Scores in Selected Cell Fate Pathways

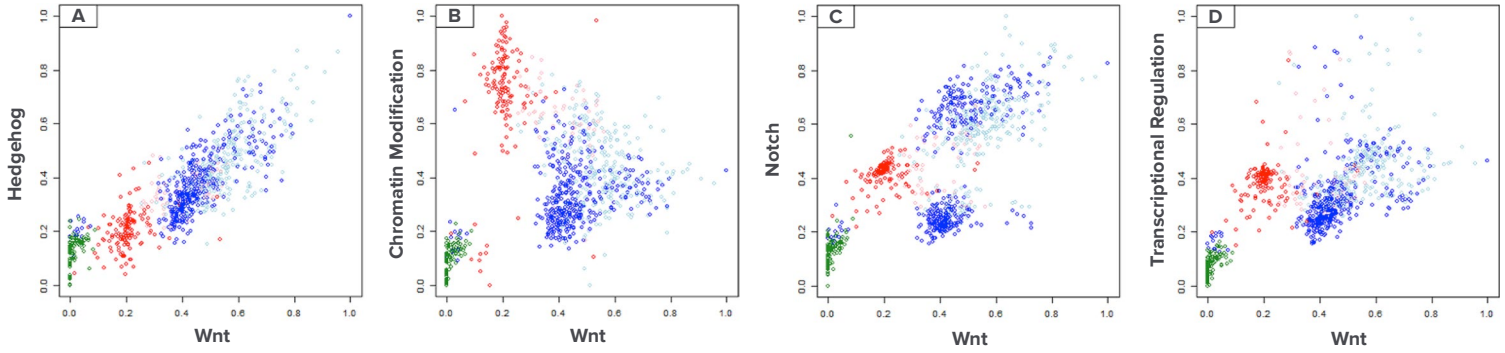
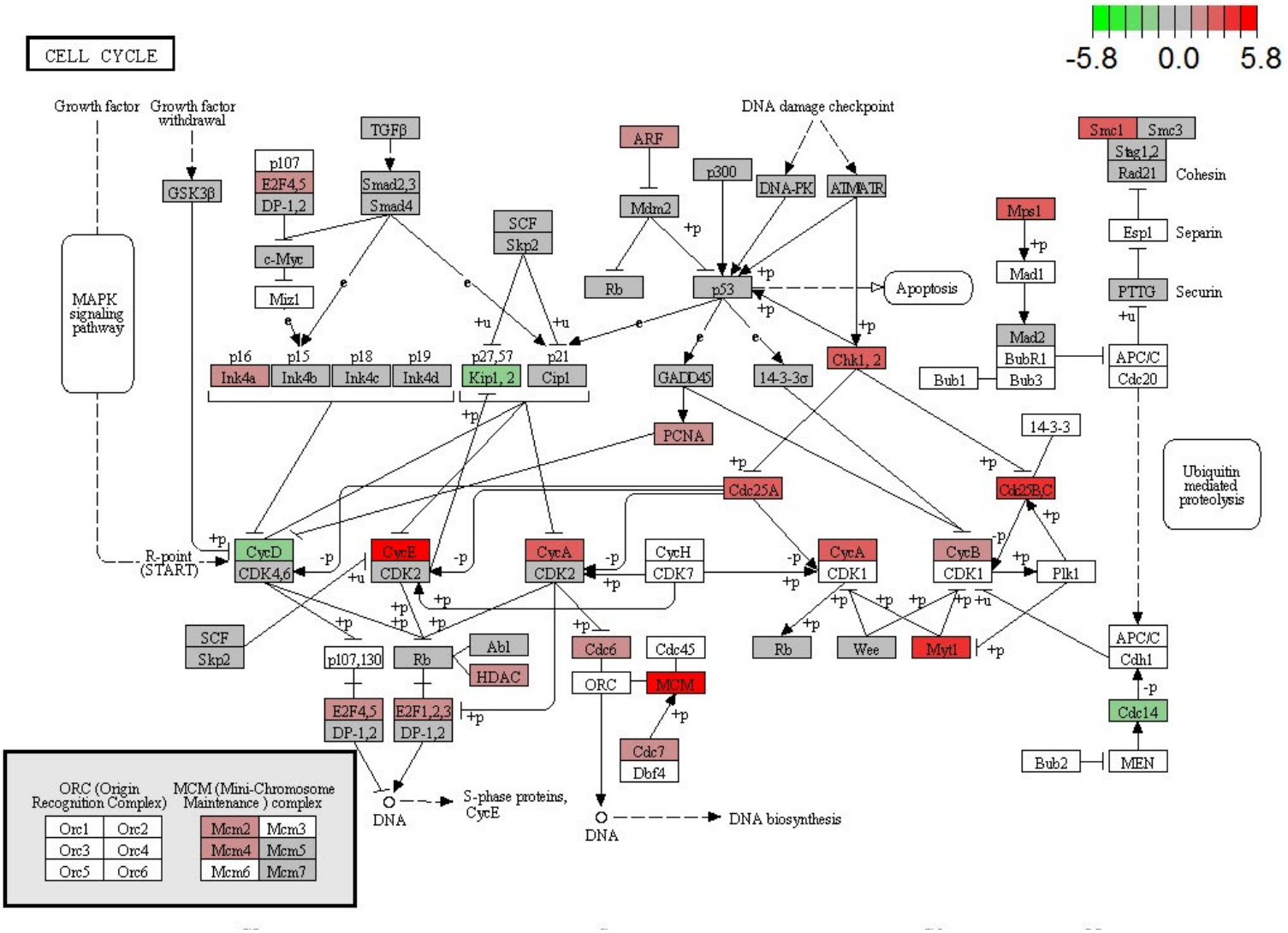


FIGURE 5: Deregulation scores of selected cell fate pathways in TCGA breast cancer data. Pathifier-derived deregulation scores from five pathways related to cell fate are plotted against each other in order to highlight patterns of coexpression within instrinsic subtypes. Samples are colored according to intrinsic subtype, with Basal-like (red), HER2-enriched (pink), Luminal A (dark blue), Luminal B (light blue), and normal samples (green). **(A)** Hedgehog and Wnt pathway regulation is consistent for all breast cancer subtypes. **(B)** Chromatin Modification and Wnt pathway deregulation and discordant in the majority of samples from each intrinsic subtype. **(C)** Notch and Wnt pathway regulation is consistent except for a subgroup of Luminal tumors. **(D)** Transcriptional Regulation and Wnt pathway regulation is consistent in Basal and HER2-enriched and discordant in Luminal A and B tumors.

Data Analysis

Pathways also provide an excellent lens through which to examine genes' differential regulation between tumors and controls. This type of analysis can be done at the pathway level with gene set enrichment analysis (GSEA)[23], while tools like Pathview[24] provide for visualization of differential expression of individual genes in the context of KEGG pathways. A Pathview plot of differential expression between Basal-like and normal samples in the Cell Cycle pathway highlights genes that are up- and down-regulated in these tumors (**FIGURE 6**). Overall, there is broad up-regulation of many cell cycle genes in Basal-like samples consistent with the dysregulation of this pathway in these samples. Interestingly, a Pathview plot of differential expression between Basal-like and normal samples in the Apoptosis pathway (**FIGURE 7**) highlights that a relatively small number of genes likely contribute to the significant dysregulation identified by Pathifier analysis in **FIGURES 3–5**.

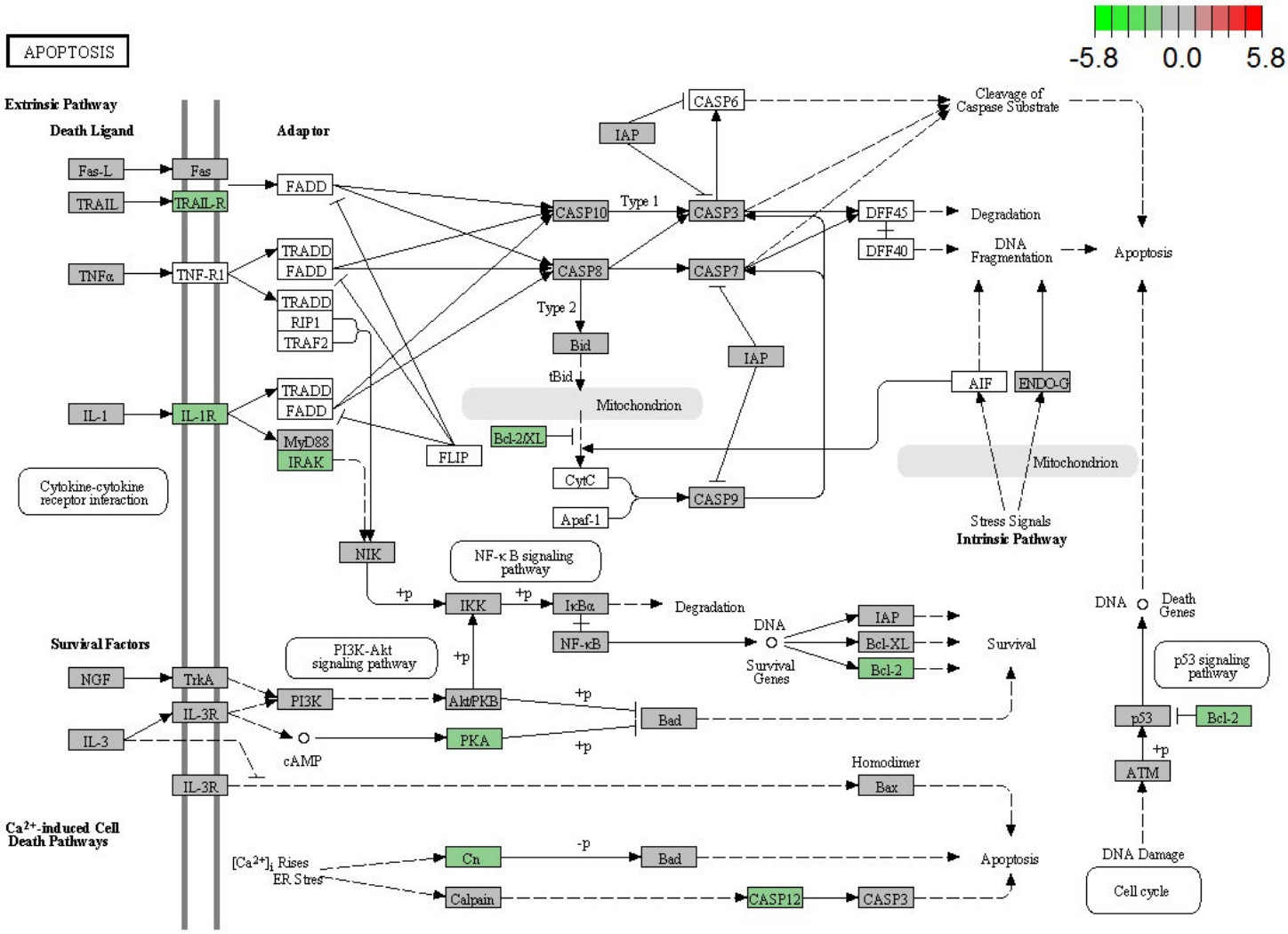
Differential Gene Expression Mapped to Proteins in the Cell Cycle Pathway



Data on KEGG graph
Rendered by Pathview

FIGURE 6: Differential expression in Basal-like tumors relative to normal tissue within the cell cycle pathway. Differential expression results comparing expression of individual cell cycle genes between normal and Basal-like samples are mapped to a KEGG representation of the pathway using Pathview. Proteins whose corresponding genes are up-regulated in Basal-like samples are colored red; proteins with down-regulated genes are colored green.

Differential Gene Expression Mapped to Proteins in the Apoptosis Pathway



Data on KEGG graph
Rendered by Pathview

FIGURE 7: Differential expression in Basal-like tumors relative to normal tissue within the apoptosis pathway. Differential expression results comparing expression of individual apoptosis genes between normal and Basal-like samples are mapped to a KEGG representation of the pathway using Pathview. Proteins whose corresponding genes are up-regulated in Basal-like samples are colored red; proteins with down-regulated genes are colored green.

Conclusion

Gene expression signatures can be used to better understand basic cancer biology and to identify patterns of pathway deregulation in cancers. In this paper, we have highlighted the ability of the genes within the nCounter PanCancer Pathways panel to provide a valuable lens through which to examine cancer biology. Clustering tumors based on pathway signatures may aid in defining prognosis in patient populations based on the intrinsic biology and outcome of specific cancers[11], as well as help predict the sensitivity of these tumors to therapeutic agents[25–27]. Profiling of the driver genes included in our panel provides valuable information for those interested in exploring the relationship between driver genes and their impact on pathway activity. As such, tools like the nCounter PanCancer Pathways Panel, that provide a way to look at both pathway genes and driver genes, enables analysis of cancer for the translational research community and permits the further the development of molecular diagnostics and targeted therapeutics.

References

1. Weinberg RA, Hanahan D (2000) The Hallmarks of Cancer. *Cell* 100:57–70.
2. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144:646–674.
3. Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. *Nat Med* 10:789–799.
4. Vogelstein B *et al*. (2013) Cancer genome landscapes. *Science* 339:1546–58.
5. Khatri P, Sirota M, Butte AJ (2012) Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. *PLoS Comput Biol* 8:e1002375.
6. Perou CM *et al*. (2000) Molecular portraits of human breast tumours. *Nature* 406:747–752.
7. Sørlie T *et al*. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98:10869–10874.
8. Bhattacharjee A *et al*. (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 98:13790–13795.
9. Cancer T, Atlas G (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490:61–70.
10. Edelman EJ, Guinney J, Chi J-T, Febbo PG, Mukherjee S (2008) Modeling cancer progression via pathway dependencies. *PLoS Comput Biol* 4:e28.
11. Gatza ML *et al*. (2010) A pathway-based classification of human breast cancer. *Proc Natl Acad Sci USA* 107:6994–6999.
12. Drier Y, Sheffer M, Domany E (2013) Pathway-based personalized analysis of cancer. *Proc Natl Acad Sci USA* 110:6388–93.
13. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM (2002) Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* 62:4427–4433.
14. Efroni S, Schaefer CF, Buetow KH (2007) Identification of key processes underlying cancer phenotypes using biologic pathway analysis. *PLoS One* 2:e425.
15. Bild AH *et al*. (2009) An integration of complementary strategies for gene-expression analysis to reveal novel therapeutic opportunities for breast cancer. *Breast Cancer Res* 11:R55.
16. Ogata H *et al*. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 27:29–34.
17. Croft D *et al*. (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res* 42:D472–7.
18. Milacic M *et al*. (2012) Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers (Basel)* 4:1180–211.
19. Ashburner M *et al*. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–29.
20. Parker JS *et al*. (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27:1160–1167.
21. Lee E, Chuang H-Y, Kim J-W, Ideker T, Lee D (2008) Inferring pathway activity toward precise disease classification. *PLoS Comput Biol* 4:e1000217.
22. Vaske CJ *et al*. (2010) Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* 26:i237–i245.
23. Subramanian A *et al*. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102:15545–50.
24. Luo W, Brouwer C (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* 29:1830–1.
25. Bild AH *et al*. (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439:353–357.
26. Yauch RL, Settleman J (2012) Recent advances in pathway-targeted cancer drug therapies emerging from cancer genome analysis. *Curr Opin Genet Dev* 22:45–49.
27. Heiser LM *et al*. (2012) Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc Natl Acad Sci* 109:2724–2729.

NanoString Technologies, Inc.

530 Fairview Ave N
Suite 2000
Seattle, WA 98109 USA

Toll-free: **1 888 358 6266** | Fax: 206 378 6288
www.nanostring.com | info@nanostring.com

LEARN MORE

To learn more about the nCounter® PanCancer Pathways Panel, visit us at **www.nanostring.com/pancancer** or contact your local NanoString Technologies representative.

SALES CONTACT

United States: **us.sales@nanostring.com**
Europe: **europe.sales@nanostring.com**
Other Regions: **info@nanostring.com**

KEGG is a trademark of Kanehisa Laboratories, Kyoto University, and the University of Tokyo.

Pathview was created by: Luo, W. and Brouwer C., Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics, 2013, 29(14): 1830-1831

WP-PC010-01 March 2014