

RESEARCH ARTICLE

# Comparison of Nanostring nCounter<sup>®</sup> Data on FFPE Colon Cancer Samples and Affymetrix Microarray Data on Matched Frozen Tissues

Xi Chen<sup>1,8\*</sup>, Natasha G. Deane<sup>2,7</sup>, Keeli B. Lewis<sup>2</sup>, Jiang Li<sup>3</sup>, Jing Zhu<sup>2</sup>, M. Kay Washington<sup>6,7</sup>, R. Daniel Beauchamp<sup>2,4,5,7\*</sup>

**1** Division of Biostatistics, Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, Florida, United States of America, **2** Department of Surgery, Vanderbilt University, Nashville, Tennessee, United States of America, **3** Affymetrix Inc., Santa Clara, California, United States of America, **4** Department of Cell and Developmental Biology, Vanderbilt University, Nashville, Tennessee, United States of America, **5** Department of Cancer Biology, Vanderbilt University, Nashville, Tennessee, United States of America, **6** Department of Pathology, Vanderbilt University, Nashville, Tennessee, United States of America, **7** Vanderbilt Ingram Cancer Center, Vanderbilt University, Nashville, Tennessee, United States of America, **8** Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, Florida, United States of America

\* [xi.steven.chen@gmail.com](mailto:xi.steven.chen@gmail.com) (XC); [daniel.beauchamp@vanderbilt.edu](mailto:daniel.beauchamp@vanderbilt.edu) (RDB)



**OPEN ACCESS**

**Citation:** Chen X, Deane NG, Lewis KB, Li J, Zhu J, Washington MK, et al. (2016) Comparison of Nanostring nCounter<sup>®</sup> Data on FFPE Colon Cancer Samples and Affymetrix Microarray Data on Matched Frozen Tissues. PLoS ONE 11(5): e0153784. doi:10.1371/journal.pone.0153784

**Editor:** Xiaofeng Wang, Cleveland Clinic Lerner Research Institute, UNITED STATES

**Received:** February 12, 2016

**Accepted:** April 4, 2016

**Published:** May 13, 2016

**Copyright:** © 2016 Chen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are available in Gene Expression omnibus (GEO) database (GSE62932).

**Funding:** This research was supported by the National Institutes of Health grant CA158472 and CA095103 to XC, NGD, KBL, JZ, MKW, and RDB. JL received support in the form of a salary from Affymetrix. The specific role of this author is articulated in the Author Contributions section. These funders played no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

The prognosis of colorectal cancer (CRC) stage II and III patients remains a challenge due to the difficulties of finding robust biomarkers suitable for testing clinical samples. The majority of published gene signatures of CRC have been generated on fresh frozen colorectal tissues. Because collection of frozen tissue is not practical for routine surgical pathology practice, a clinical test that improves prognostic capabilities beyond standard pathological staging of colon cancer will need to be designed for formalin-fixed paraffin-embedded (FFPE) tissues. The NanoString nCounter<sup>®</sup> platform is a gene expression analysis tool developed for use with FFPE-derived samples. We designed a custom nCounter<sup>®</sup> codeset based on elements from multiple published fresh frozen tissue microarray-based prognostic gene signatures for colon cancer, and we used this platform to systematically compare gene expression data from FFPE with matched microarray array data from frozen tissues. Our results show moderate correlation of gene expression between two platforms and discovery of a small subset of genes as candidate biomarkers for colon cancer prognosis that are detectable and quantifiable in FFPE tissue sections.

## Introduction

The identification of colorectal cancer (CRC) patients who either more or less likely to benefit from adjuvant systemic chemotherapy after surgical resection poses a major unmet need in providing safe and effective care. The current practice may result in under-treatment of some

**Competing Interests:** The authors have declared that no competing interests exist.

high-risk stage II patients, and potential over-treatment of low-risk stage II and stage III patients [1]. The core obstacle is the lack of definitive diagnostic biomarkers to identify cancers with a high probability of metastasis and correspondingly poor clinical outcome that are more likely to benefit from systemic chemotherapy; and conversely, to identify those patients at very low risk (<10%) who are unlikely to derive significant benefit from chemotherapy [2–6]. Translation of microarray-based profiles into clinical diagnostics as biomarkers is complicated by the fact that the technology required to reproduce them has previously required fresh unfixed tissue samples, and there is a gap between an emerging body of genomic information and diagnostic application. The ability of gene expression signatures to predict recurrence and clinical outcome strongly argues that high- and low-risk phenotypes are molecularly encoded in primary tumors [7–9]. Nevertheless, a robust prognostic signature applicable in the clinical setting has yet to be developed for colorectal cancer due to variation in methods and procedures for preservation of surgical specimens, variation in the amount and quality of purified RNA available for analysis, tumor heterogeneity issues and the reproducibility and robustness of the assay platform. Because collection of fresh frozen tissues is not routine, a clinical test for improving staging of colon cancer will need to be designed for formalin-fixed paraffin-embedded (FFPE) tissues in order to be widely applicable.

The NanoString nCounter<sup>®</sup> system has been applied to quantify gene expression of various gene signatures for multiple tissue types including FFPE samples [10–13]. We have designed a custom nCounter codeset for quantitative assessment of expression of 414 gene elements. This assay consists of multiple published gene signatures for colon cancer prognosis plus several candidate gene elements derived from ongoing studies in intestinal stem cell biology and epithelial-to-mesenchymal transition (EMT). To determine which elements of prognostic signatures can be translated from fresh frozen tissue to archival FFPE derived patient RNA samples, we systematically compared the expression results for the 414 genes from the nCounter platform using FFPE-derived tissue and from a microarray array platform using matched frozen tissues.

## Materials and Methods

### Experiment design and sample description

Human tissues used for microarray analysis were collected and annotated according to established protocols and approved by the appropriate Institutional Review Boards (IRB) at Vanderbilt University (VUMC). All tissues were collected over the time period from 1999–2011. Tumor stage was assessed by American Joint Commission on Cancer guidelines. Written informed consent was obtained from all patients prior to inclusion in the studies. Quality assessment slides were obtained to verify the diagnosis and the amount of cellular material for each sample. De-identified human tumor tissues for immunohistochemistry were obtained with VUMC IRB approval. For microarray studies, representative sections of fresh tissue specimens were flash frozen in liquid nitrogen within 20 minutes of resection and stored at –80°C until the RNA isolation step. The median (minimum-maximum) frozen tissue storage time from date of resection to RNA extraction are 455 (35–2858) days. RNA was purified from tissue sections containing >80% epithelial tumor tissue using RNeasy (QIAGEN, Valencia, CA) according to manufacturer's instructions. Samples were hybridized to Affymetrix arrays Human Genome U133 Plus 2.0 GeneChip Expression Arrays, Santa Clara, CA), which included approximately 39,000 human genes. The samples included four healthy control patient tissues, 12 stage I, 17 stage II, 20 stage III and 15 stage IV CRC patient tissues. The detailed sample information including survival data can be found in [S1 Table](#).

For nCounter studies, archived tumor FFPE samples matched to frozen tissues described above were identified and processed by the Vanderbilt Translational Pathology and Imaging Core facility. Tissues were fixed in 4% paraformaldehyde, embedded in paraffin and stored at room temperature until RNA extraction. The median (minimum-maximum) FFPE tissue storage time from date of resection to RNA extraction are 1021 (364–3483) days. The first 20µm of the tissue was discarded before cutting sections for RNA extraction. Tissue sections were mounted on uncharged glass slides and one of the tissue section slides was stained with H&E for quality control from each tissue block. RNA was purified from 5µm thick tissue sections containing greater than 80% tumor using High Pure FFPE RNA Micro Kit (Roche) according to manufacturer’s instructions. A minimum of 4 sections per sample were required. The mRNA hybridization, detection and scanning were followed the protocol and performed by NanoString Technologies.

### nCounter codeset development

We designed a custom nCounter<sup>®</sup> assay (NanoString Technologies, Seattle, WA) for quantitative assessment of expression of 414 gene elements. This multiplexed assay can detect expression of up to 800 transcripts at very low mRNA concentrations (0.1fM/1 copy per cell) [14] even in RNA samples that are significantly degraded, as long as more than 20% of the sample has RNA fragments of greater than 300 base pairs [13]. Table 1 lists all gene signatures involved in the nCounter Codeset. We included gene elements from our 34-gene signature [9], as well as elements from published signatures [7, 8, 15–24] in the nCounter assay, selecting those that were represented in ≥2 signatures. Fifty additional candidate genes of interest related to EMT and the study of metastatic behavior in cancer cells (eg, E-cadherin, Smad4, Zeb1, Snail, Slug,

**Table 1. Sources of Nanostring codeset of 414-gene list.**

Source	Genes included	Description
Eschrich 2005 [8]	21	Outcomes-based supervised analysis of 78 cases, stages I-IV
Barrier 2006 [7]	18	Outcomes-based supervised analysis of 50 cases, stage II only
Ki 2007 [19]	36	Differential expression between 25 primary colon tumors with and without synchronous metastasis
Lin 2007 [20]	24	Outcomes-based supervised analysis of 2 cohorts (n = 149, all stages; n = 55, stage I & II)
Wang 2004 [24]	23	Outcomes-based supervised analysis of 74 Duke’s B cases.
Grade 2007 [17]	121	Differential expression of 73 stage II & stage III cases based lymph node involvement
Paik 2004 [21]	9	Clinically approved classifier for recurrence in tamoxifen treated, node-negative breast cancer
Kanies 2008 [18]	5	Experimentally derived signature of Epithelial-to-Mesenchymal Transition
Fritzmman 2009 [15]	82	Differential expression of 95 stage matched primary tumors with and without synchronous metastasis
Smith 2010 [9]	34	Experimentally derived classifier based on immunocompetent mouse model of metastasis
Salzar 2011 (Coloprint) [22]	13	Supervised analysis of 188 cases based on metastasis-free survival
Garman 2008 [16]	50	Outcomes-based supervised analysis of 52 stages I and stage II cases.
Tripathi 2014 [23]	57	Systems-based classifier based on experimental mouse model of metastasis

doi:10.1371/journal.pone.0153784.t001

Twist, beta-catenin, etc.) were also added to the assay. The full list with gene symbols and the sources of the 414-genes can be found in [S2](#) and [S3](#) Tables.

## Bioinformatics and statistical data analysis

Affymetrix microarray data were normalized using the Robust MultiChip Averaging (RMA) algorithm as implemented in the Bioconductor package *Affy* [25]. The batch effects were adjusted by ComBat approach available in the Bioconductor package *SVA* [26]. NanoString nCounter data quality control and normalization were performed using R package *NanoString-Norm* [27]. For group comparisons, t-test in the *Limma* package in Bioconductor was used to identify differentially expressed probe sets between the tumor and normal tissues in both Affymetrix and nCounter platforms [28]. For survival analysis, the Cox proportional hazard was applied to test the association with overall survival outcomes for each candidate gene using R package *survival*.

In this study, in order to make direct comparison between Affymetrix HGU133plus 2 and nCounter gene expression data, we first found the best-matched Affymetrix microarray probes to NanoString nCounter probes. Toward this end, we used gene symbol or transcript ID to link NanoString nCounter IDs and Affymetrix Probe IDs. If the gene had only one unique probe, then we used that Probe ID as the match for NanoString. Otherwise, if the gene had multiple probes, then we picked the one with the largest Smith–Waterman alignment score, which was calculated by Smith–Waterman algorithm to identify homologous regions between sequences by searching for optimal local alignments, between NanoString nCounter and Affymetrix Probe sequences as the match for nCounter [29]. Both microarray and nCounter data are available in Gene Expression omnibus (GEO) database (GSE62932).

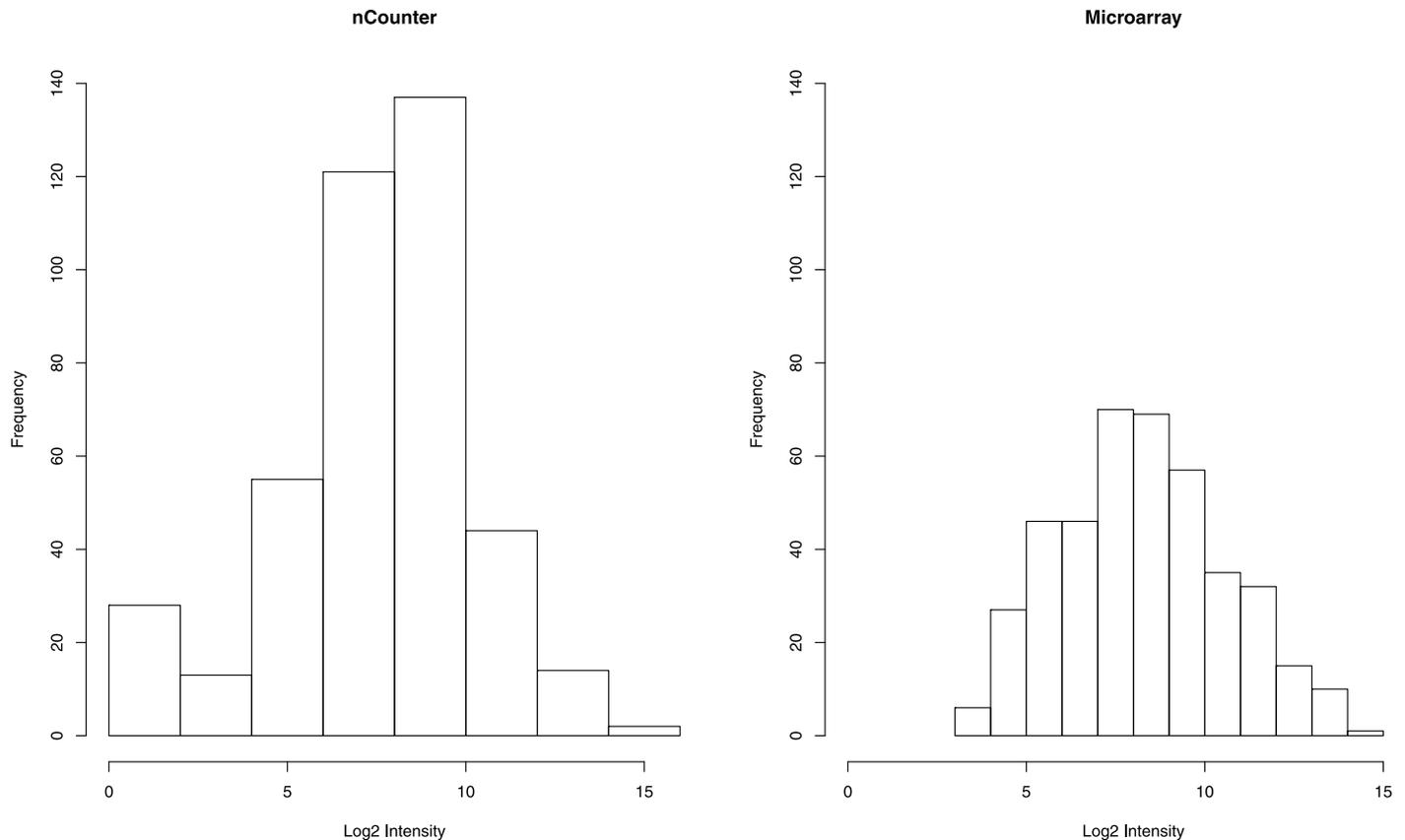
## Results

### Quality of extracted RNA

RNA Integrity Numbers (RIN) values were obtained using an Agilent Bioanalyzer instrument for both frozen and FFPE derived RNA samples. The instrument software generates a RIN score based on its entire electropherogram. RIN values range from 1–10, with 1 being totally degraded RNA and 10 being high quality (intact) RNA. For microarray studies, a cut-off of RIN = 7.0 was used. RIN values ranged from 7–10, with a median of 8.5 and mean of 8.4. For nCounter studies, RIN values ranged from 0–8, with a median of 2.4 and mean of 2.5. In addition, to be acceptable for analysis for nCounter,  $\geq 20\%$  of the RNA sample had to contain fragments of at least 300 base pairs in length. The details of the results were included in [S4 Table](#).

### Data quality of nCounter and microarray

After pre-processing and normalization, log<sub>2</sub> transformed data from 414 genes of both microarray and nCounter were compared. The distributions of signal intensities for all samples are displayed with boxplots in [S1 Fig](#). To draw the gene expression distribution, each of the 414 genes was represented by the median value for 68 samples. [Fig 1](#) compares gene expression distributions based on nCounter and microarrays. nCounter data showed a bi-modal distribution that is likely due to the RNA degradation in FFPE samples. In more detail, we found that  $\sim 70\%$  of genes had median normalized expression values in the 5–10 arbitrary unit range on both platforms. In contrast, about 15% of the genes showed very low normalized expression values ( $\leq 5$  arbitrary units) on the nCounter platform while only 8% of the genes performed as poorly on the microarray platform. Conversely, only 14% of the genes showed expression robust median expression values ( $> 10$  arbitrary units) on the nCounter platform while closer to 23%



**Fig 1. Histograms for nCounter and microarray.** The frequency of gene probes (414 elements total) is graphed against median normalized expression values across the x-axis.

doi:10.1371/journal.pone.0153784.g001

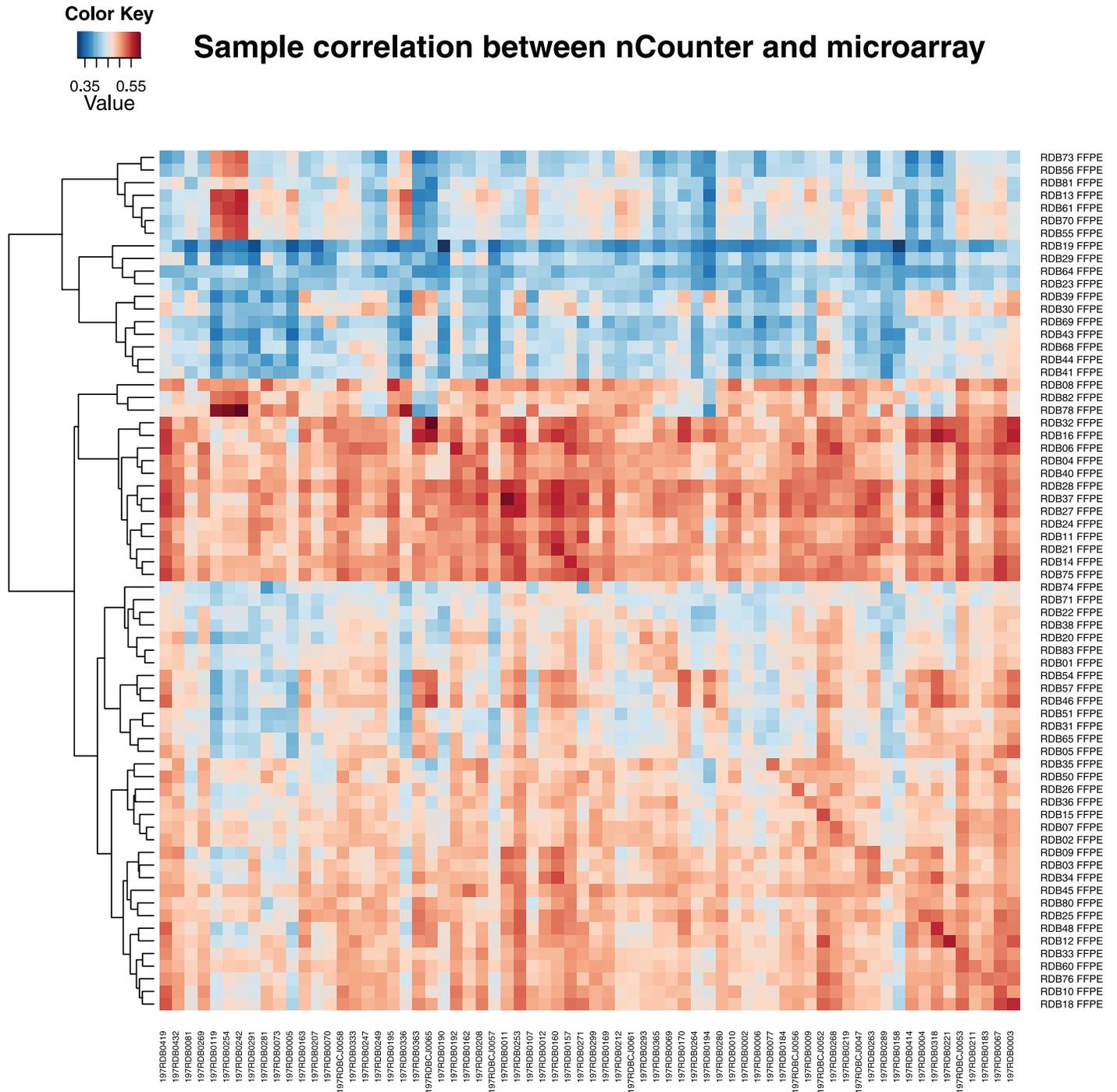
of the genes gave robust signals on the microarray platform. In summary, the amount of useable data achieved using the nCounter platform was low relative to that achieved on the microarray platform.

To examine the reproducibility of the nCounter platform, six FFPE samples were selected to generate two more technical replicates for each sample. [S2 Fig](#) shows the scatterplots of the log<sub>2</sub>-transformed counts between all pairs of replicates. The average correlation coefficient is 0.983, which indicates that the nCounter platform is reproducible.

### Sample and gene correlations between nCounter and microarray

For each pair of FFPE sample quantified by nCounter and fresh frozen sample quantified by microarray, the Pearson correlation was calculated based on normalized, log<sub>2</sub> transformed gene expression values of the 414 genes. The mean correlation was 0.501 with 95% bootstrap confidence interval (0.412, 0.589), and the minimum and maximum correlation were 0.337 and 0.591 respectively. The heatmap in [Fig 2](#) displays all pairwise correlations between FFPE and fresh frozen samples. The analysis based on Spearman correlation showed similar results, which suggested the correlation pattern between samples was not affected by different normalization procedures since the normalization methods don't change the gene rankings within each sample.

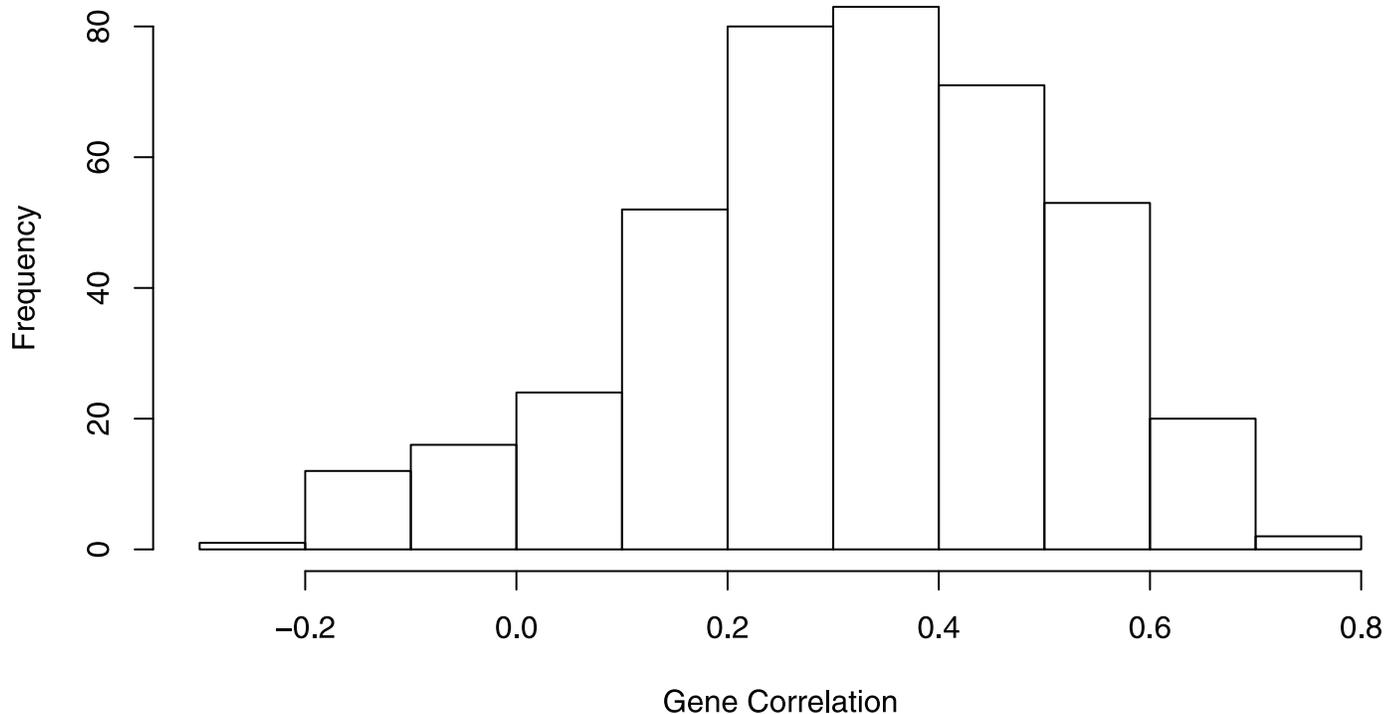
For each of 414 genes, the Pearson correlation was also calculated between two tissue types. We plotted the histogram of gene-wise correlation coefficients in [Fig 3](#). The mean correlation



**Fig 2. The heatmap of pairwise correlations between individual matched FFPE and fresh frozen samples.** X-axis represents the fresh frozen samples and y-axis represents matched FFPE samples.

doi:10.1371/journal.pone.0153784.g002

of the 414 genes was 0.315 with 95% bootstrap confidence interval (0.248, 0.385), and the minimum and maximum correlation were -0.250 and 0.784 respectively. Gene-wise correlation represents the concordance of gene expression values from two expression platforms between two matched tissue type. Due to nCounter’s single probe design, some genes did not match well



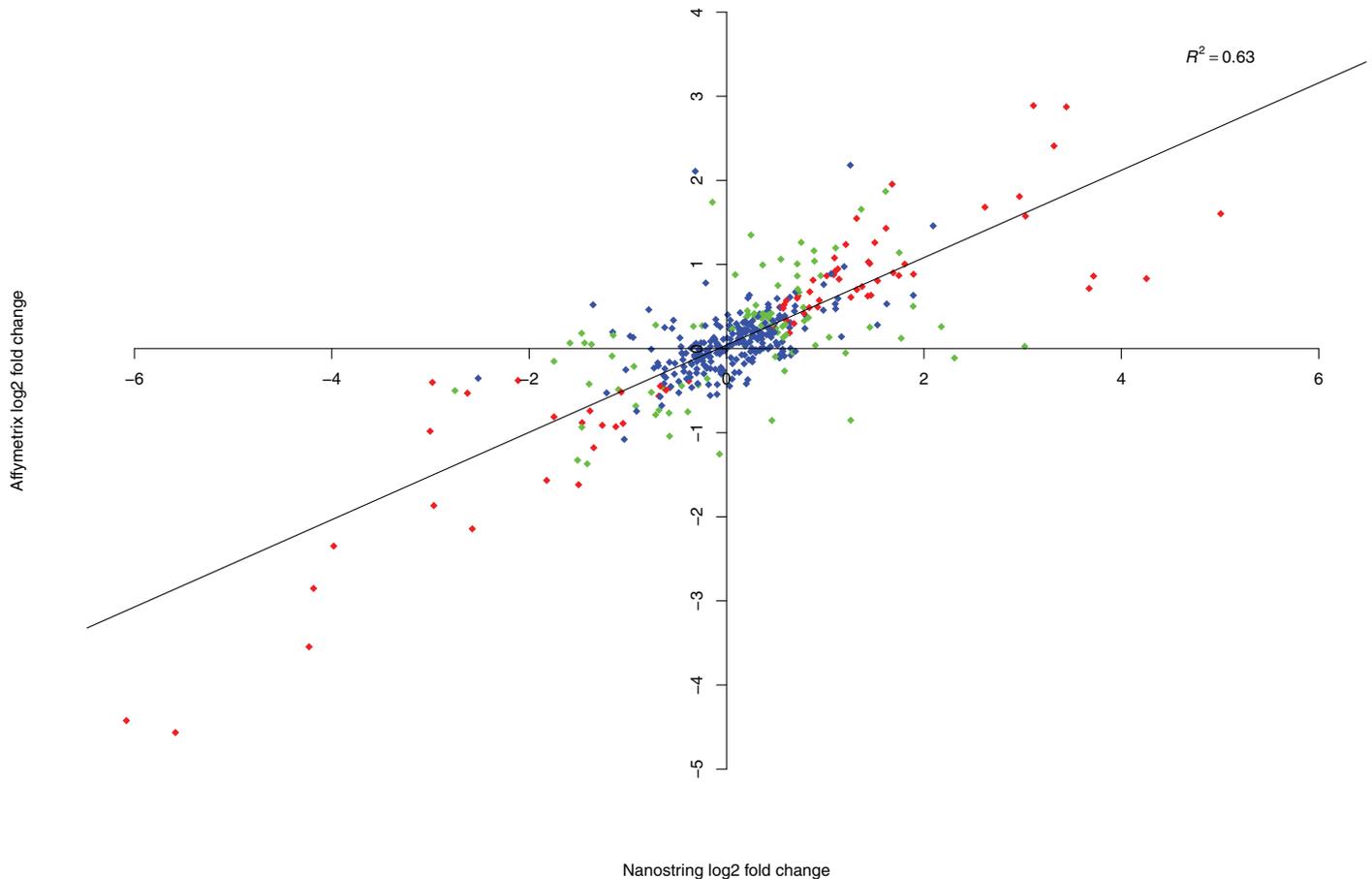
**Fig 3. The histogram of gene-wise correlations between FFPE and fresh frozen samples.** X-axis represents Pearson correlation of each of 414 genes between fresh frozen samples and FFPE samples and y-axis represents the frequency.

doi:10.1371/journal.pone.0153784.g003

between two platforms, which may have contributed to the weak or negative correlations. It is reasonable to observe this moderate level of concordance because the variations in gene expressions were from two different platforms and tissue types. The medium level of concordance between genes also led to the ambiguous pattern between fresh frozen and FFPE samples displayed in Fig 2.

### Comparison of gene expression between nCounter and microarray

After preprocessing and normalization, to detect differentially expressed genes between 64 tumor and 4 normal samples in nCounter or microarray platforms separately, we used regularized t-test as implemented in the *Bioconductor* package *LIMMA*. Using nominal p-value 0.05 as the cutoff, 76 genes were identified as significantly differentially expressed genes in both platforms. Among the 76 genes, 40 and 26 genes were overexpressed and downregulated in cancer samples. There were 30 and 60 significant genes discovered in only one platform by nCounter and microarray respectively, and 248 genes were not significant by either platform. The complete gene testing lists including coefficient, nominal p-value and false discovery rate can be found in S5 and S6 Tables for nCounter and microarray, respectively. Fig 4 shows the comparison of log<sub>2</sub> fold change in tumor to normal signal intensity between nCounter and microarray for each of 414 genes that were selected for the nCounter codeset. The red, green and blue dots represent 76, 90 and 248 genes significantly differentially expressed in both, either, and neither platforms respectively. If we use FDR 0.1 as the cutoff, there were 48, 72 and 294 genes significantly differentially expressed in both, either, and neither platforms respectively.

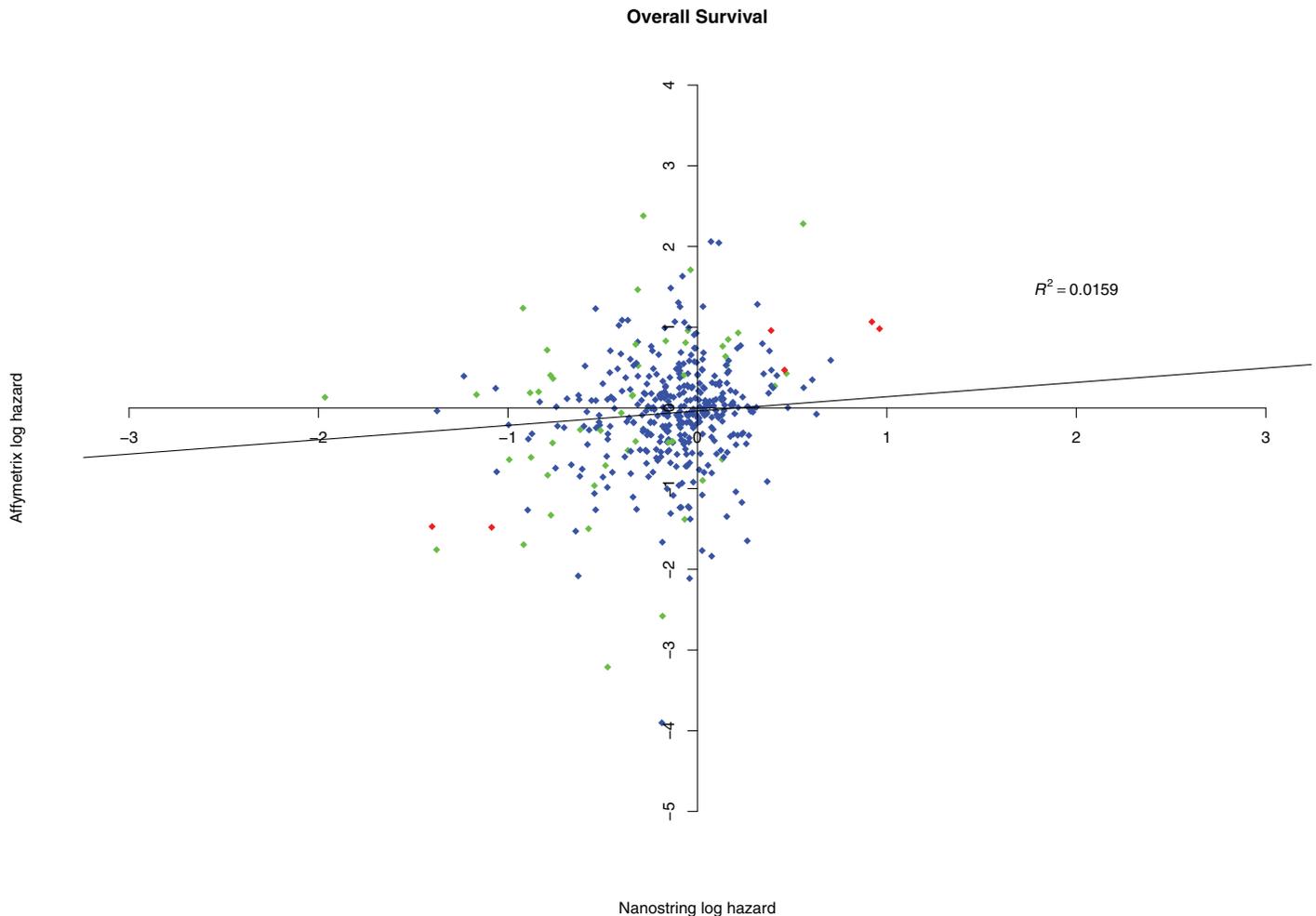


**Fig 4. Scatterplot of log<sub>2</sub> fold change.** Log<sub>2</sub> fold change in tumor to normal signal intensity for 414 genes comparing results from nCounter (x-axis) and microarray (y-axis). The red, green and blue dots represent 76, 90 and 248 genes significantly differentially expressed in both, either, and neither platforms respectively.

doi:10.1371/journal.pone.0153784.g004

### Association of nCounter and microarray gene expression with survival outcome

We also examined the association of gene expression with overall survival outcomes of 64 patients using a Cox proportional hazard model. Using a nominal p-value 0.05 as the threshold, there were 6 significantly associated genes including PPL, CCND1, EXT2, S100A11, USP9X, and PHLDA1 by both platforms, 25 significant genes by nCounter, and 24 significant genes by microarray. PPL, CCND1, S100A11, and PHLDA1 were associated with higher hazard, and EXT2 and USP9X were associated with better prognosis. The majority of the genes (359 genes) were not significantly associated with overall survival outcomes by either platform due to relatively small sample size. The complete gene testing lists including coefficient, nominal p-value and false discovery rate of this survival data analysis are in [S7](#) and [S8](#) Tables for nCounter and microarray, respectively. [Fig 5](#) shows the log hazard ratios for overall survival outcomes of the 414 genes compared between nCounter and microarray data, with the red, green and blue dots indicating the 6, 49 and 359 genes having significant survival association in the two, one and neither platforms. For the six genes identified by both platforms, previous studies showed that CCND1 and S100A1 can be used as biomarkers for colon cancer prognosis



**Fig 5. Scatterplot of log hazard ratio.** Log hazard ratio for tumor to normal signal intensity for 414 genes vs. overall survival outcomes comparing results from nCounter (x-axis) and microarray (y-axis). The red, green and blue dots represent 6, 49 and 359 genes significantly associated with survival outcomes in both, either, and neither platforms respectively.

doi:10.1371/journal.pone.0153784.g005

[30, 31], and the inhibition of USP9X might increase the tumor cell sensitivity to some chemotherapeutic agents [32, 33].

## Discussion

The quality of mRNA transcript quantification in FFPE samples is crucial for translational cancer research. In this study, our comparisons were based on the measurements of NanoString's nCounter platform using FFPE-derived samples and Affymetrix's Human Genome U133 Plus 2.0 GeneChip Expression Arrays on patient matched fresh frozen tissue. Ideally, mRNA quantification using the same platform on matched FFPE and frozen tissues would make a more direct comparison. However, in order to determine whether gene signatures developed using fresh frozen tissues are applicable for clinical samples, comparisons between the two platforms using matched, but differentially archived tissues, is required.

The nCounter-based mRNA profiling using FFPE technical replicates showed high reproducibility of the platform similar to the previous study [10]. However, the average correlation coefficients were 0.501 and 0.315 for paired FFPE/fresh-frozen sample correlations and

matched gene correlations respectively. These results indicated moderate correlations between nCounter (FFPE) and microarray (fresh-frozen) data.

For tumor and normal group comparisons, 166 of the 414 genes were identified as significantly differentially expressed genes by either platform and that of these 166 genes, 76 genes (46%) were expressed concordantly between the two platforms. Demonstrating an association with survival outcomes is more challenging, only 11% genes (6 genes) among significantly associated genes (55 genes) by either platform were concordant.

Multiple factors can affect these experimental results. The nCounter platform, by its very nature, represents a custom build of candidate gene elements as a biomarker, thus the selection of genes in an nCounter codeset inherently limits *in silico* comparisons with other published studies. The degree of RNA degradation is another challenging issue facing transcriptome profiling strategies using FFPE-derived samples. For example, a recent study on hepatocellular carcinoma demonstrated that mRNA transcript measurements achieved using the nCounter system for archived sectioned FFPE samples compared poorly with freshly cut FFPE tissues [11]. This issue relates both to the problem of variation in sample preparation and to the issue of inherent tumor heterogeneity reflected in measurable molecular differences between adjacent tissue sections. In this study, NanoString Technologies provided the single probe design based upon transcript frequency for our gene panel, but using the multiple probe design might be a strategy to improve the nCounter data quality.

Archived FFPE preserved tumor tissue remains the most available tissue specimen source for large-scale development and validation of prognostic and predictive gene signatures for colon cancer. Identification of subgroups of archival tumor tissues producing higher quality RNA samples for expression profiling is one approach to improving biomarker development from these samples. Improvements in technology such as new FFPE RNA extraction protocols, new FFPE RNA amplification and labeling methods, and modified data processing workflows can also help to side-step issues of RNA degradation and enhance the transcriptome data quality from RNA-Seq or microarray [34–36]. In the current study, we showed that the gene signature developed on microarray gene expression data using fresh frozen samples may not be directly applicable to nCounter data based on FFPE samples. However, we demonstrated that a small subset of genes had stable expression patterns in both the nCounter and microarray platforms, and independent of whether the lysates were derived from fresh frozen or FFPE-derived samples. Follow-up studies on these genes may help lead to more clinically useful biomarkers for colon cancer.

## Supporting Information

**S1 Fig. The distributions of signal intensities of 68 samples (individual box plots) by nCounter and microarray.**

(PDF)

**S2 Fig. Scatterplot of the normalized counts from 414 individual gene elements using a split sample from six FFPE preserved tumor extractions (blinded technical replicates) on the nCounter platform.**

(PDF)

**S1 Table. The clinical information of 68 samples.**

(CSV)

**S2 Table. The full list with gene symbols and the sources of the 414-genes for nCounter codeset.**

(CSV)

**S3 Table. The list of matched Affymetrix probe IDs of the 414-genes for nCounter codeset.**  
(CSV)

**S4 Table. The RNA extraction RIN values**  
(CSV)

**S5 Table. The differentially expressed gene testing results for nCounter.**  
(CSV)

**S6 Table. The differentially expressed gene testing results for microarray.**  
(CSV)

**S7 Table. The testing results of association with survival outcomes for nCounter.**  
(CSV)

**S8 Table. The testing results of association with survival outcomes for microarray.**  
(CSV)

## Author Contributions

Conceived and designed the experiments: XC NGD RDB. Performed the experiments: XC NGD KBL JL JZ MKW. Analyzed the data: XC JL JZ. Wrote the paper: XC NGD KBL JL JZ MKW RDB.

## References

1. Ragnhammar P, Hafstrom L, Nygren P, Glimelius B, Care SB-gSCoTAiH. A systematic overview of chemotherapy effects in colorectal cancer. *Acta oncologica*. 2001; 40(2–3):282–308. PMID: [11441937](#).
2. Benson AB 3rd, Schrag D, Somerfield MR, Cohen AM, Figueredo AT, Flynn PJ, et al. American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2004; 22(16):3408–19. doi: [10.1200/JCO.2004.05.063](#) PMID: [15199089](#).
3. Figueredo A, Charette ML, Maroun J, Brouwers MC, Zuraw L. Adjuvant therapy for stage II colon cancer: a systematic review from the Cancer Care Ontario Program in evidence-based care's gastrointestinal cancer disease site group. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2004; 22(16):3395–407. doi: [10.1200/JCO.2004.03.087](#) PMID: [15199087](#).
4. Figueredo A, Coombes ME, Mukherjee S. Adjuvant therapy for completely resected stage II colon cancer. *The Cochrane database of systematic reviews*. 2008;(3):CD005390. doi: [10.1002/14651858.CD005390.pub2](#) PMID: [18646127](#).
5. Gill S, Loprinzi CL, Sargent DJ, Thome SD, Alberts SR, Haller DG, et al. Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2004; 22(10):1797–806. doi: [10.1200/JCO.2004.09.059](#) PMID: [15067028](#).
6. Mamounas E, Wieand S, Wolmark N, Bear HD, Atkins JN, Song K, et al. Comparative efficacy of adjuvant chemotherapy in patients with Dukes' B versus Dukes' C colon cancer: results from four National Surgical Adjuvant Breast and Bowel Project adjuvant studies (C-01, C-02, C-03, and C-04). *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 1999; 17(5):1349–55. PMID: [10334518](#).
7. Barrier A, Boelle PY, Roser F, Gregg J, Tse C, Brault D, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. *Journal of Clinical Oncology*. 2006; 24(29):4685–91. doi: [10.1200/Jco.2005.05.0229](#) WOS:000241191900004. PMID: [16966692](#)
8. Eschrich S, Yang I, Bloom G, Kwong KY, Boulware D, Cantor A, et al. Molecular staging for survival prediction of colorectal cancer patients. *Journal of Clinical Oncology*. 2005; 23(15):3526–35. doi: [10.1200/Jco.2005.00.695](#) WOS:000229352400033. PMID: [15908663](#)
9. Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang AX, et al. Experimentally Derived Metastasis Gene Expression Profile Predicts Recurrence and Death in Patients With Colon Cancer. *Gastroenterology*. 2010; 138(3):958–68. doi: [10.1053/J.Gastro.2009.11.005](#) WOS:000275109900027. PMID: [19914252](#)

10. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol.* 2008; 26(3):317–25. doi: [10.1038/nbt1385](https://doi.org/10.1038/nbt1385) PMID: [18278033](https://pubmed.ncbi.nlm.nih.gov/18278033/).
11. Kojima K, April C, Canasto-Chibuque C, Chen XT, Deshmukh M, Venkatesh A, et al. Transcriptome Profiling of Archived Sectioned Formalin-Fixed Paraffin-Embedded (AS-FFPE) Tissue for Disease Classification. *PLoS one.* 2014; 9(1). ARTN e86961 doi: [10.1371/journal.pone.0086961](https://doi.org/10.1371/journal.pone.0086961) WOS:000330617100042.
12. Payton JE, Grieselhuber NR, Chang LW, Murakami M, Geiss GK, Link DC, et al. High throughput digital quantification of mRNA abundance in primary human acute myeloid leukemia samples. *Journal of Clinical Investigation.* 2009; 119(6):1714–26. doi: [10.1172/Jci38248](https://doi.org/10.1172/Jci38248) WOS:000266601000036. PMID: [19451695](https://pubmed.ncbi.nlm.nih.gov/19451695/)
13. Reis PP, Waldron L, Goswami RS, Xu W, Xuan YL, Perez-Ordóñez B, et al. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *Bmc Biotechnology.* 2011; 11. ArtN 46 doi: [10.1186/1472-6750-11-46](https://doi.org/10.1186/1472-6750-11-46) WOS:000291021000001.
14. Fortina P, Surrey S. Digital mRNA profiling. *Nat Biotechnol.* 2008; 26(3):293–4. doi: [10.1038/nbt0308-293](https://doi.org/10.1038/nbt0308-293) PMID: [18327237](https://pubmed.ncbi.nlm.nih.gov/18327237/).
15. Fritzmann J, Morkel M, Besser D, Budczies J, Kosel F, Brembeck FH, et al. A Colorectal Cancer Expression Profile That Includes Transforming Growth Factor beta Inhibitor BAMBI Predicts Metastatic Potential. *Gastroenterology.* 2009; 137(1):165–75. doi: [10.1053/J.Gastro.2009.03.041](https://doi.org/10.1053/J.Gastro.2009.03.041) WOS:000267410100027. PMID: [19328798](https://pubmed.ncbi.nlm.nih.gov/19328798/)
16. Garman KS, Acharya CR, Edelman E, Grade M, Gaedcke J, Sud S, et al. A genomic approach to colon cancer risk stratification yields biologic insights into therapeutic opportunities (Retracted Article. See vol 106, pg 6878, 2009). *P Natl Acad Sci USA.* 2008; 105(49):19432–7. doi: [10.1073/Pnas.0806674105](https://doi.org/10.1073/Pnas.0806674105) WOS:000261706600073.
17. Grade M, Hormann P, Becker S, Hummon AB, Wangsa D, Varma S, et al. Gene expression profiling reveals a massive, aneuploidy-dependent transcriptional deregulation and distinct differences between lymph node-negative and lymph node-positive colon carcinomas (vol 67, pg 41, 2007). *Cancer Res.* 2007; 67(4):1877–. WOS:000244289200059.
18. Kanies CL, Smith JJ, Kis C, Schmidt C, Levy S, Khabar KSA, et al. Oncogenic Ras and transforming growth factor-beta synergistically regulate AU-rich element-containing mRNAs during epithelial to mesenchymal transition. *Molecular Cancer Research.* 2008; 6(7):1124–36. doi: [10.1158/1541-7786.Mcr-07-2095](https://doi.org/10.1158/1541-7786.Mcr-07-2095) WOS:000257895900005. PMID: [18644977](https://pubmed.ncbi.nlm.nih.gov/18644977/)
19. Ki DH, Jeung HC, Park CH, Kang SH, Lee GY, Lee WS, et al. Whole genome analysis for liver metastasis gene signatures in colorectal cancer. *International Journal of Cancer.* 2007; 121(9):2005–12. doi: [10.1002/ijc.22975](https://doi.org/10.1002/ijc.22975) WOS:000249718100016. PMID: [17640062](https://pubmed.ncbi.nlm.nih.gov/17640062/)
20. Lin YH, Friederichs J, Black MA, Mages J, Rosenberg R, Guilford PJ, et al. Multiple gene expression classifiers from different array platforms predict poor prognosis of colorectal cancer. *Clin Cancer Res.* 2007; 13(2 Pt 1):498–507. doi: [10.1158/1078-0432.CCR-05-2734](https://doi.org/10.1158/1078-0432.CCR-05-2734) PMID: [17255271](https://pubmed.ncbi.nlm.nih.gov/17255271/).
21. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *The New England journal of medicine.* 2004; 351(27):2817–26. doi: [10.1056/NEJMoa041588](https://doi.org/10.1056/NEJMoa041588) PMID: [15591335](https://pubmed.ncbi.nlm.nih.gov/15591335/).
22. Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, et al. Gene Expression Signature to Improve Prognosis Prediction of Stage II and III Colorectal Cancer. *Journal of Clinical Oncology.* 2011; 29(1):17–24. doi: [10.1200/Jco.2010.30.1077](https://doi.org/10.1200/Jco.2010.30.1077) WOS:000285712400015. PMID: [21098318](https://pubmed.ncbi.nlm.nih.gov/21098318/)
23. Tripathi MK, Deane NG, Zhu J, An HB, Mima S, Wang XJ, et al. Nuclear Factor of Activated T-cell Activity Is Associated with Metastatic Capacity in Colon Cancer. *Cancer Res.* 2014; 74(23):6947–57. doi: [10.1158/0008-5472.Can-14-1592](https://doi.org/10.1158/0008-5472.Can-14-1592) WOS:000346362400020. PMID: [25320007](https://pubmed.ncbi.nlm.nih.gov/25320007/)
24. Wang Y, Jatko T, Zhang Y, Mutch MG, Talantov D, Jiang J, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology.* 2004; 22(9):1564–71. doi: [10.1200/JCO.2004.08.186](https://doi.org/10.1200/JCO.2004.08.186) PMID: [15051756](https://pubmed.ncbi.nlm.nih.gov/15051756/).
25. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of affymetrix GeneChip probe level data. *Nucleic Acids Research.* 2003; 31(4). ArtN E15 doi: [10.1093/Nar/Gng015](https://doi.org/10.1093/Nar/Gng015) WOS:000181115100004.
26. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007; 8(1):118–27. doi: [10.1093/biostatistics/kxj037](https://doi.org/10.1093/biostatistics/kxj037) PMID: [16632515](https://pubmed.ncbi.nlm.nih.gov/16632515/).
27. Waggott D, Chu K, Yin SM, Wouters BG, Liu FF, Boutros PC. NanoStringNorm: an extensible R package for the pre-processing of NanoString mRNA and miRNA data. *Bioinformatics.* 2012; 28(11):1546–8. doi: [10.1093/Bioinformatics/Bts188](https://doi.org/10.1093/Bioinformatics/Bts188) WOS:000304537000024. PMID: [22513995](https://pubmed.ncbi.nlm.nih.gov/22513995/)

28. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004; 3:Article3. doi: [10.2202/1544-6115.1027](https://doi.org/10.2202/1544-6115.1027) PMID: [16646809](https://pubmed.ncbi.nlm.nih.gov/16646809/).
29. Smith TF, Waterman MS. Identification of Common Molecular Subsequences. *Journal of Molecular Biology*. 1981; 147(1):195–7. doi: [10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5) WOS:A1981LJ17700013. PMID: [7265238](https://pubmed.ncbi.nlm.nih.gov/7265238/)
30. Li Y, Wei J, Xu C, Zhao Z, You T. Prognostic significance of cyclin D1 expression in colorectal cancer: a meta-analysis of observational studies. *PloS one*. 2014; 9(4):e94508. doi: [10.1371/journal.pone.0094508](https://doi.org/10.1371/journal.pone.0094508) PMID: [24728073](https://pubmed.ncbi.nlm.nih.gov/24728073/); PubMed Central PMCID: PMC3984178.
31. Meding S, Balluff B, Elsner M, Schone C, Rauser S, Nitsche U, et al. Tissue-based proteomics reveals FXD3, S100A11 and GSTM3 as novel markers for regional lymph node metastasis in colon cancer. *The Journal of pathology*. 2012; 228(4):459–70. doi: [10.1002/path.4021](https://doi.org/10.1002/path.4021) PMID: [22430872](https://pubmed.ncbi.nlm.nih.gov/22430872/).
32. Harris DR, Mims A, Bunz F. Genetic disruption of USP9X sensitizes colorectal cancer cells to 5-fluorouracil. *Cancer Biol Ther*. 2012; 13(13):1319–24. doi: [10.4161/cbt.21792](https://doi.org/10.4161/cbt.21792) PMID: [22895071](https://pubmed.ncbi.nlm.nih.gov/22895071/); PubMed Central PMCID: PMC3493440.
33. Peddaboina C, Jupiter D, Fletcher S, Yap JL, Rai A, Tobin RP, et al. The downregulation of Mcl-1 via USP9X inhibition sensitizes solid tumors to Bcl-xl inhibition. *BMC Cancer*. 2012; 12:541. doi: [10.1186/1471-2407-12-541](https://doi.org/10.1186/1471-2407-12-541) PMID: [23171055](https://pubmed.ncbi.nlm.nih.gov/23171055/); PubMed Central PMCID: PMC3543233.
34. Hedegaard J, Thorsen K, Lund MK, Hein AM, Hamilton-Dutoit SJ, Vang S, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PloS one*. 2014; 9(5):e98187. doi: [10.1371/journal.pone.0098187](https://doi.org/10.1371/journal.pone.0098187) PMID: [24878701](https://pubmed.ncbi.nlm.nih.gov/24878701/); PubMed Central PMCID: PMC4039489.
35. Callari M, Lembo A, Bianchini G, Musella V, Cappelletti V, Gianni L, et al. Accurate data processing improves the reliability of Affymetrix gene expression profiles from FFPE samples. *PloS one*. 2014; 9(1):e86511. doi: [10.1371/journal.pone.0086511](https://doi.org/10.1371/journal.pone.0086511) PMID: [24489733](https://pubmed.ncbi.nlm.nih.gov/24489733/); PubMed Central PMCID: PMC3906036.
36. Williams PM, Li R, Johnson NA, Wright G, Heath JD, Gascoyne RD. A Novel Method of Amplification of FFPET-Derived RNA Enables Accurate Disease Classification with Microarrays. *J Mol Diagn*. 2010; 12(5):680–6. doi: [10.2353/Jmoldx.2010.090164](https://doi.org/10.2353/Jmoldx.2010.090164) WOS:000281690800017. PMID: [20688907](https://pubmed.ncbi.nlm.nih.gov/20688907/)