Final Progress Report for Research Projects Funded by Health Research Grants

Instructions: Please complete all of the items as instructed. Do not delete instructions. Do not leave any items blank; responses must be provided for all items. If your response to an item is "None", please specify "None" as your response. "Not applicable" is not an acceptable response for any of the items. There is no limit to the length of your response to any question. Responses should be single-spaced, no smaller than 12-point type. The report **must be completed using MS Word**. Submitted reports must be Word documents; they should not be converted to pdf format. Questions? Contact Health Research Program staff at 717-783-2548.

1. **Grantee Institution:** The Wistar Institute

2. **Reporting Period (start and end date of grant award period):** 01/01/2013 – 06/30/2014

3. **Grant Contact Person (First Name, M.I., Last Name, Degrees):** Russel E. Kaufman, M.D.

4. **Grant Contact Person's Telephone Number:** 215-898-3926

5. **Grant SAP Number:** 4100062226

6. **Project Number and Title of Research Project:** 1: Informatics Solutions for NextGen Sequence Data Analysis

7. **Start and End Date of Research Project:** 01/01/2013 – 06/30/2014

8. **Name of Principal Investigator for the Research Project:** Ramana Davuluri, Ph.D.

9. **Research Project Expenses**.

   9(A) Please provide the total amount of health research grant funds spent on this project for the entire duration of the grant, including indirect costs and any interest earned that was spent:

   $313,831.10

   9(B) Provide the last names (include first initial if multiple individuals with the same last name are listed) of **all** persons who worked on this research project and were supported with health research funds. Include position titles (Principal Investigator, Graduate Assistant, Post-doctoral Fellow, etc.), percent of effort on project and total health research funds expended for the position. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

Last Name, First Name	Position Title	% of Effort on Project	Cost
Davuluri	Principal Investigator	33%	$69,625
Bi	Postdoctoral Fellow	63%	$49,605
Jung	Postdoctoral Fellow	55%	$5,575
Pal	Staff Scientist	100%	$104,927

9(C) Provide the names of **all** persons who worked on this research project, but who *were not* supported with health research funds. Include position titles (Research Assistant, Administrative Assistant, etc.) and percent of effort on project. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

Last Name, First Name	Position Title	% of Effort on Project
None		

9(D) Provide a list of **all** scientific equipment purchased as part of this research grant, a short description of the value (benefit) derived by the institution from this equipment, and the cost of the equipment.

Type of Scientific Equipment	Value Derived	Cost
None		

10. Co-funding of Research Project during Health Research Grant Award Period. Did this research project receive funding from any other source during the project period when it was supported by the health research grant?

Yes____X_____ No_____

If yes, please indicate the source and amount of other funds:

R01 LM011297 (Dr. R.V. Davuluri)
Title: Informatics platform for mammalian gene regulation at isoform-level
Source of Funding: NIH/NLM (Project Period: 05/02/2013 – 04/30/2016)
Total Funding Amount: $387,000/Year (for 3 years)

11. Leveraging of Additional Funds

11(A) As a result of the health research funds provided for this research project, were you able to apply for and/or obtain funding from other sources to continue or expand the research?

Yes_____ No____X_____

If yes, please list the applications submitted (column A), the funding agency (National

Institutes of Health—NIH, or other source in column B), the month and year when the application was submitted (column C), and the amount of funds requested (column D). If you have received a notice that the grant will be funded, please indicate the amount of funds to be awarded (column E). If the grant was not funded, insert "not funded" in column E.

Do not include funding from your own institution or from CURE (tobacco settlement funds). Do not include grants submitted prior to the start date of the grant as shown in Question 2. If you list grants submitted within 1-6 months of the start date of this grant, add a statement below the table indicating how the data/results from this project were used to secure that grant.

A. Title of research project on grant application	B. Funding agency (check those that apply)	C. Month and Year Submitted	D. Amount of funds requested:	E. Amount of funds to be awarded:
None	☐NIH ☐ Other federal (specify:_____) ☐ Nonfederal source (specify:_)		$	$

11(B) Are you planning to apply for additional funding in the future to continue or expand the research?

Yes___X_____ No_____

If yes, please describe your plans:

I moved my research laboratory to Northwestern University – Feinberg School of Medicine in February 2014. Currently I am developing a research proposal in response to the NIH program announcement on Big Data (PA-14-155: Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science (R01)).

12. Future of Research Project. What are the future plans for this research project?

A major innovation of this proposal is development of novel algorithms for analysis of NextGen sequencing data by investigating various machine learning methods and nonparametric approaches. The overarching and long-term goal of this proposal is to develop an informatics platform to shift the fundamental paradigm from "one gene –> one protein –> one functional pathway" to "one gene –> multiple protein isoforms –> multiple functional pathways". We will continue to develop novel algorithms and data-discretization methods for pre-processing isoform-level gene expression data from different profiling platforms;

1. novel methods for indexing shared datasets and curate minimal metadata to support discovery across different data resources (e.g., exon-array, RNA-Seq); and a
2. novel database and website that can provide information at all levels (both human and

machine-readable interfaces).

13. **New Investigator Training and Development**. Did students participate in project supported internships or graduate or post-graduate training for at least one semester or one summer?

    Yes_____ No____X_____

    If yes, how many students? Please specify in the tables below:

	Undergraduate	Masters	Pre-doc	Post-doc
Male				
Female				
Unknown				
Total				

	Undergraduate	Masters	Pre-doc	Post-doc
Hispanic				
Non-Hispanic				
Unknown				
Total				

	Undergraduate	Masters	Pre-doc	Post-doc
White				
Black				
Asian				
Other				
Unknown				
Total				

14. **Recruitment of Out-of–State Researchers**. Did you bring researchers into Pennsylvania to carry out this research project?

    Yes____X_____ No_____

    If yes, please list the name and degree of each researcher and his/her previous affiliation:

    Dr. Segun Jung, Ph.D. (Computational Biology), New York University, New York.

15. **Impact on Research Capacity and Quality**. Did the health research project enhance the quality and/or capacity of research at your institution?

    Yes____X_____ No_____

If yes, describe how improvements in infrastructure, the addition of new investigators, and other resources have led to more and better research.

The research project contributed to the expansion of bioinformatics research at The Wistar Institute.

16. Collaboration, business and community involvement.

16(A) Did the health research funds lead to collaboration with research partners outside of your institution (e.g., entire university, entire hospital system)?

Yes____X_____ No_____

If yes, please describe the collaborations:

Based on observations in the Pal paper on isoform-level expression in brain and publications by others, showing that a majority of genes produce multiple transcript and protein isoforms, which could be involved in multiple functional pathways, we hypothesized that the isoform-level expression profiling could generate better classification biomarkers to identify the molecular sub-groups of cancer. We first undertook a project to identify biomarker panels that could accurately classify glioblastoma subtypes using The Cancer Genome Atlas (TCGA) data. Using results from this *in silico* study, since better molecular classification of cancer is a requirement for the quest of personalized therapy, we established a new collaboration with Dr. Donald O'Rourke, Associate Professor and Neurosurgeon, Department of Neurosurgery, University of Pennsylvania. The identified markers are validated using new tissue samples provided by Dr. O'Rourke. The results were published in the following publication:

Pal S, Bi Y, Macyszyn L, Showe LC, O'Rourke DM and Davuluri RV. (2014) Isoform-level gene signature improves prognostic stratification and accurately classifies glioblastoma subtypes. *Nucleic Acids Res.* Epub 2014/02/08. doi: 10.1093/nar/gku121. PubMed PMID: 24503249

16(B) Did the research project result in commercial development of any research products?

Yes_____ No____X_____

If yes, please describe commercial development activities that resulted from the research project:

16(C) Did the research lead to new involvement with the community?

Yes_____ No___X_____

If yes, please describe involvement with community groups that resulted from the research project:

17. Progress in Achieving Research Goals, Objectives and Aims.

List the project goals, objectives and specific aims (as contained in the grant agreement). Summarize the progress made in achieving these goals, objectives and aims for the period that the project was funded (i.e., from project start date through end date). Indicate whether or not each goal/objective/aim was achieved; if something was not achieved, note the reasons why. Describe the methods used. If changes were made to the research goals/objectives/aims, methods, design or timeline since the original grant application was submitted, please describe the changes. Provide detailed results of the project. Include evidence of the data that was generated and analyzed, and provide tables, graphs, and figures of the data. List published abstracts, poster presentations and scientific meeting presentations at the end of the summary of progress; peer-reviewed publications should be listed under item 20.

This response should be a DETAILED report of the methods and findings. It is not sufficient to state that the work was completed. Insufficient information may result in an unfavorable performance review, which may jeopardize future funding. If research findings are pending publication you must still include enough detail for the expert peer reviewers to evaluate the progress during the course of the project.

Health research grants funded under the Tobacco Settlement Act will be evaluated via a performance review by an expert panel of researchers and clinicians who will assess project work using this Final Progress Report, all project Annual Reports and the project's strategic plan. After the final performance review of each project is complete, approximately 12-16 months after the end of the grant, this Final Progress Report, as well as the Final Performance Review Report containing the comments of the expert review panel, and the grantee's written response to the Final Performance Review Report, will be posted on the CURE Web site.

There is no limit to the length of your response. Responses must be single-spaced below, no smaller than 12-point type. If you cut and paste text from a publication, be sure symbols print properly, e.g., the Greek symbol for alpha (α) and beta (ß) should not print as boxes (□) and include the appropriate citation(s). DO NOT DELETE THESE INSTRUCTIONS.

Specific Aims:

The central hypothesis of the project is that the isoform-level gene products – "transcript variants" and "protein isoforms" – are the basic functional units in a mammalian cell; and, accordingly, the informatics platforms for managing and analyzing gene regulation data in both normal and diseased cells should adopt "gene isoform centric" rather than "gene centric" approaches. The recent advances in Next-Generation Sequencing (NGS) brought the field closer to the goal of studying gene expression and regulation at isoform-level by facilitating massive scale sequencing at much lower costs. The advent of NGS-based molecular technologies, such as Chromatin Immunoprecipitation Sequencing (ChIP-seq) and

Ribonucleic Acid Sequencing (RNA-seq), is enabling genome-wide identification of gene isoforms and the target genes of transcription factors (TFs) in different cells, tissues and disease conditions. However, there are a number of new informatics challenges and difficulties that must be investigated to improve the current state and fulfill the promise of studying gene regulation and expression at gene isoform level. We propose to develop novel algorithms and build an informatics platform for understanding gene regulation at isoform level (alternative promoter or alternative transcript level) by developing statistically rigorous bioinformatics applications for processing NGS data.

Specific Aim #1: Develop statistically rigorous novel algorithms and bioinformatics pipelines to identify the orthologous promoters, corresponding transcript variants and protein isoforms that are conserved between human and mouse.

Specific Aim #2: Develop novel algorithms and informatics pipelines for integrative analysis of NGS datasets to predict the activity of both known and novel promoters (from ChIP-seq) and transcript variants (from RNA-seq) the promoters transcribe, and annotate their expression in both normal and diseased cells/tissues.

Studies and Results

1. NPEBseq – Nonparametric Empirical Bayesian-based procedure for differential expression analysis of RNA-seq data: We developed a novel nonparametric empirical Bayesian-based approach (NPEBseq) to model the RNA-seq data. The prior distribution of the Bayesian model is empirically estimated from the data without any parametric assumption, and hence the method is "nonparametric" in nature. Based on this model, we proposed a method for detecting differentially expressed genes across different conditions. We also extended this method to detect differential usage of exons from RNA-seq data. The evaluation of NPEBseq on both simulated and publicly available RNA-seq datasets and comparison with three popular methods showed improved results for experiments with or without biological replicates. NPEBseq can successfully detect differential expression between different conditions not only at gene level but also at exon level from RNA-seq datasets. In addition, NPEBSeq performs significantly better than current methods and can be applied to genome-wide RNA-seq datasets. Sample datasets and R package are available at http://bioinformatics.wistar.upenn.edu/NPEBseq.

2. Short-read based annotation pipeline for mapping and analyzing ChIP-seq data: Mapping genome-wide data to human subtelomeres has been problematic due to the incomplete assembly and challenges of low-copy repetitive DNA elements. In collaboration with Dr. Riethman's laboratory at The Wistar Institute, we developed a novel bioinformatics pipeline incorporating multi-read mapping for annotation of the updated assemblies using short-read data sets.

3. *Evaluation of algorithms for isoform-level expression estimation from RNA-seq data:* Both hybridization and sequencing-based approaches have been developed for transcriptome studies. High density exon-array has been successfully used to study alterative splicing events. With the advent of NGS technologies, RNA-seq has also emerged as a powerful tool for transcriptome analysis. The analysis of RNA-seq data is a challenging task. In the last few years, many tools have been developed to estimate the abundance of gene expression at transcript level. We

performed a comparative study between different exon-array and RNA-seq tools using The Cancer Genome Atlas (TCGA) Glioblastoma Multiforme (GBM) data. We designed a Real-Time Polymerase Chain Reaction (RT-qPCR) assay as a gold standard to benchmark the gene expression values by RNA-seq and Exon-array. Previous studies indicate strong correlation between gene-level expression estimates from RNA-seq and Exon-array data. However, we found poor correlation between RT-qPCR and Exon-array/RNA-seq for isoform/transcript-level expression estimates and better correlation for fold change values. We found that, while the transcript-level fold-change estimates (tumor over normal) obtained by Multi-Mapping Bayesian Gene eXpression (MMBGX) from exon-array data correlated well with RT-qPCR values. For RNA-seq, RNA-Seq by Expectation-Maximization (RSEM) outperformed the other algorithms (Table 1). We expect to submit a manuscript within two to three months.

Table 1: Correlation between RT-qPCR based values and isoform-level expression estimates by MMBGX (Exon-array) and RNA-seq algorithms. The Pearson correlation coefficients were computed using 159 transcripts measured by all platforms.

Algorithm	Correlation coefficient based on	
	Expression	Fold-change
Cufflinks (http://cufflinks.cbcb.umd.edu/)	0.156	0.684
RSEM (http://deweylab.biostat.wisc.edu/rsem/)	0.338	0.743
eXpress (http://bio.math.berkeley.edu/eXpress/index.html)	0.321	0.587
IsoformEx (http://bioinformatics.wistar.upenn.edu/isoformex)	0.286	0.660
MMBGX (Exon-array) http://www.bgx.org.uk/software/mmbgx.html	0.43	0.825

18. Extent of Clinical Activities Initiated and Completed. Items 18(A) and 18(B) should be completed for all research projects. If the project was restricted to secondary analysis of clinical data or data analysis of clinical research, then responses to 18(A) and 18(B) should be "No."

18(A) Did you initiate a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
___X__No

18(B) Did you complete a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
__X___No

If "Yes" to either 18(A) or 18(B), items 18(C) – (F) must also be completed. (Do NOT complete 18(C-F) if 18(A) and 18(B) are both "No.")

18(C) How many hospital and health care professionals were involved in the research project?

_____Number of hospital and health care professionals involved in the research project

18(D) How many subjects were included in the study compared to targeted goals?

_____Number of subjects originally targeted to be included in the study
_____Number of subjects enrolled in the study

Note: Studies that fall dramatically short on recruitment are encouraged to provide the details of their recruitment efforts in Item 17, Progress in Achieving Research Goals, Objectives and Aims. For example, the number of eligible subjects approached, the number that refused to participate and the reasons for refusal. Without this information it is difficult to discern whether eligibility criteria were too restrictive or the study simply did not appeal to subjects.

18(E) How many subjects were enrolled in the study by gender, ethnicity and race?

Gender:
_____Males
_____Females
_____Unknown

Ethnicity:
_____Latinos or Hispanics
_____Not Latinos or Hispanics
_____Unknown

Race:
_____American Indian or Alaska Native
_____Asian
_____Blacks or African American
_____Native Hawaiian or Other Pacific Islander
_____White
_____Other, specify:_____
_____Unknown

18(F) Where was the research study conducted? (List the county where the research study was conducted. If the treatment, prevention and diagnostic tests were offered in more than one county, list all of the counties where the research study was conducted.)

19. Human Embryonic Stem Cell Research. Item 19(A) should be completed for all research projects. If the research project involved human embryonic stem cells, items 19(B) and 19(C) must also be completed.

19(A) Did this project involve, in any capacity, human embryonic stem cells?
_____Yes
__X____ No

19(B) Were these stem cell lines NIH-approved lines that were derived outside of Pennsylvania?
_____Yes
_____ No

19(C) Please describe how this project involved human embryonic stem cells:

20. Articles Submitted to Peer-Reviewed Publications.

20(A) Identify all publications that resulted from the research performed during the funding period and that have been submitted to peer-reviewed publications. Do not list journal abstracts or presentations at professional meetings; abstract and meeting presentations should be listed at the end of item 17. **Include only those publications that acknowledge the Pennsylvania Department of Health as a funding source** (as required in the grant agreement). List the title of the journal article, the authors, the name of the peer-reviewed publication, the month and year when it was submitted, and the status of publication (submitted for publication, accepted for publication or published.). Submit an electronic copy of each publication or paper submitted for publication, listed in the table, in a PDF version 5.0.5 (or greater) format, 1,200 dpi. Filenames for each publication should include the number of the research project, the last name of the PI, and an abbreviated title of the publication. For example, if you submit two publications for Smith (PI for Project 01), one publication for Zhang (PI for Project 03), and one publication for Bates (PI for Project 04), the filenames would be:

Project 01 – Smith – Three cases of isolated
Project 01 – Smith – Investigation of NEB1 deletions
Project 03 – Zhang – Molecular profiling of aromatase
Project 04 – Bates – Neonatal intensive care

If the publication is not available electronically, provide 5 paper copies of the publication.

Note: The grant agreement requires that recipients acknowledge the Pennsylvania Department of Health funding in all publications. Please ensure that all publications listed acknowledge the Department of Health funding. If a publication does not acknowledge the funding from the Commonwealth, do not list the publication.

Title of Journal Article:	Authors:	Name of Peer-reviewed Publication:	Month and Year Submitted:	Publication Status (check appropriate box below):
1. Subtelomeric CTCF and cohesin binding site organization using improved subtelomere assemblies and a novel annotation pipeline.	Stong N, Deng Z, Gupta R, Hu S, Paul S, Weiner AK, Eichler EE, Graves T, Fronick CC, Courtney L, Wilson RK, Lieberman P, Davuluri RV, Riethman H.	*Genome Research*	September, 2013	☐Submitted ☐Accepted ☒Published
2. NPEBseq: nonparametric empirical bayesian-based procedure for differential expression analysis of RNA-seq data.	Bi Y, Davuluri RV.	*BMC Bio informatics*	February, 2013	☐Submitted ☐Accepted ☒Published

20(B) Based on this project, are you planning to submit articles to peer-reviewed publications in the future?

Yes____X_____ No_____

If yes, please describe your plans:

We are currently working on the following research manuscripts:

1. Evaluation of algorithms for isoform-level expression estimation from RNA-seq data
2. Evaluation of Data Discretization Methods for Cross Platform Transfer of Gene-expression based Tumor Subtyping Classifier

21. Changes in Outcome, Impact and Effectiveness Attributable to the Research Project. Describe the outcome, impact, and effectiveness of the research project by summarizing its impact on the incidence of disease, death from disease, stage of disease at time of diagnosis, or other relevant measures of outcome, impact or effectiveness of the research project. If there were no changes, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

None

22. Major Discoveries, New Drugs, and New Approaches for Prevention Diagnosis and Treatment. Describe major discoveries, new drugs, and new approaches for prevention,

diagnosis and treatment that are attributable to the completed research project. If there were no major discoveries, drugs or approaches, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

None

23. Inventions, Patents and Commercial Development Opportunities.

23(A) Were any inventions, which may be patentable or otherwise protectable under Title 35 of the United States Code, conceived or first actually reduced to practice in the performance of work under this health research grant? Yes_____ No____X_____

If "Yes" to 23(A), complete items a – g below for each invention. (Do NOT complete items a - g if 23(A) is "No.")

a. Title of Invention:

b. Name of Inventor(s):

c. Technical Description of Invention (describe nature, purpose, operation and physical, chemical, biological or electrical characteristics of the invention):

d. Was a patent filed for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes_____ No____

If yes, indicate date patent was filed:

e. Was a patent issued for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes_____ No____
If yes, indicate number of patent, title and date issued:
Patent number:
Title of patent:
Date issued:

f. Were any licenses granted for the patent obtained as a result of work performed under this health research grant? Yes_____ No____

If yes, how many licenses were granted?_____

g. Were any commercial development activities taken to develop the invention into a commercial product or service for manufacture or sale? Yes___ No____

If yes, describe the commercial development activities:

23(B) Based on the results of this project, are you planning to file for any licenses or patents, or undertake any commercial development opportunities in the future?

Yes_____ No____X_____

If yes, please describe your plans:

24. Key Investigator Qualifications. Briefly describe the education, research interests and experience and professional commitments of the Principal Investigator and all other key investigators. In place of narrative you may insert the NIH biosketch form here; however, please limit each biosketch to 1-2 pages.

BIOGRAPHICAL SKETCH

NAME **DAVULURI, Ramana V.**	POSITION TITLE **Professor of Preventive Medicine** **Director of Cancer Informatics**
eRA COMMONS USER NAME **DAVULURI10**	

INSTITUTION AND LOCATION	DEGREE	YEAR(s)	FIELD OF STUDY
Nagarjuna University, India	B.S.	1985-1988	Mathemathics
Indian Agricultural Statistics Research Institute, IARI, New Delhi, India	M.S.	1989-1991	Statistics & Computer Applications
Indian Agricultural Statistics Research Institute, IARI, New Delhi, India	Ph.D.	1991-1996	Statistics & Computer Applications

A. Personal Statement

Dr. Davuluri will provide bioinformatics, statistics and data-mining expertize as a PI on this project. He has worked in the areas of bioinformatics and computational genomics since 1998, starting with the postdoctoral training in Dr. Michael Zhang's laboratory at CSHL, NY. Coming from a background in quantitative sciences, he gained synergistic expertize in statistical modeling, genomics and bioinformatics, while leading several projects over the past 12 years. During his tenure at OSU, he was a Co-PI of an NSF's Biological Infrastructure grant; and Project Leader (Project #3) and Core Director (Data Management & Computational Modeling Core) of a P50 grant supported by NCI's Integrative Cancer Biology Program (OSU center for Cancer Epigenetics). He has been also awarded three NIH R01s and American Cancer Society Research Scholar Grant to study the epigenetic modifications across alternative promoters and associated isoform-level gene signatures in normal and cancer genomes. His research program is interdisciplinary in nature with a complement of experimental investigation. His lab currently focuses on developing data-mining algorithms and informatics solutions for problems in biology and medicine. The overarching goal of the lab is to translate data from high dimensional (-omic) platforms (e.g., NextGen sequencing) to derive experimentally interpretable and testable discovery models towards genomics-based clinical decision support systems. His group is developing bioassays that can rapidly identify biomarkers from human tissue and blood samples. Towards these goals, his group applies a combination of state-of-the-art statistically rigorous data-mining methods and high-throughput experimental procedures in a systems biology setting.

B. Positions and Honors

Professional Experience

1993-1998	Assistant Professor of Statistics, ANGR Agricultural University, Hyderabad, INDIA
1998-1999	Special Project Scientist (Bioinformatics), International Crops Research Institute for the Semi Arid Tropics, (ICRISAT), Patancheru, AP, INDIA
2001	Research Investigator, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2001-2007	Assistant Professor, Bioinformatics & Computational Biology, Human Cancer Genetics (HCG) Program, Department of Mol. Virology, Immunology and Medical Genetics, OSU, Columbus, OH
2001-2007	Adjunct Assistant Professor, Department of Biomedical Informatics, OSU, Columbus, OH
2002-2008	Head of Bioinformatics Unit, HCG, Comprehensive Cancer Center, OSU,

	Columbus, OH
2003-2008	Mentor in Biosciences, Mathematical Biosciences Institute, OSU, Columbus, OH
2007-2008	Associate Professor (tenured), Bioinformatics & Computational Biology, HCG Program, Dept. of Mol. Virology, Immunology and Medical Genetics, OSU, Columbus, OH
2008-2014	Associate Professor; Director of Computational Biology, Center for Systems and Computational Biology, The Wistar Institute, Philadelphia, PA
2008-2011	Scientific Director, Bioinformatics Facility, The Wistar Institute Cancer Center, Philadelphia, PA
2008-2014	Wistar Associate Professor, Dept of Genetics, University of Pennsylvania, Philadelphia, PA
2012-2014	Scientific Investigator, Penn Brain Tumor Center, University of Pennsylvania, Philadelphia, PA
2014-Present	Professor and Director of Cancer Informatics; Division of Health and Biomedical Informatics, Department of Preventive Medicine, Professor of Neurological Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL.

C. Selected Peer-reviewed Publications (Selected from 82 peer-reviewed publications)

1. Jung S, Bi Y, **Davuluri RV**. Comparison of Data Discretization Methods for Cross Platform Transfer of Gene-expression based Tumor Subtyping Classifier. *Proceedings of 4th IEEE Intl Conference on Computational Advances in Bio and Medical Sciences (ICCABS).* 2014.
2. Stong N, Deng Z, Gupta R, Hu S, Paul S, Weiner AK, Eichler EE, Graves T, Fronick CC, Courtney L, Wilson RK, Lieberman PM, **Davuluri RV**, Riethman H. Subtelomeric CTCF and cohesin binding site organization using improved subtelomere assemblies and a novel annotation pipeline. ***Genome Res.*** 2014. PubMed PMID: 24676094. PMCID: In Process
3. Pal S, Bi Y, Macyszyn L, Showe LC, O'Rourke DM, **Davuluri RV**. Isoform-level gene signature improves prognostic stratification and accurately classifies glioblastoma subtypes. ***Nucleic Acids Res.*** 2014;42(8):e64. PubMed PMID: 24503249; PubMed Central PMCID: PMC4005667
4. Ota H*, Sakurai M*, Gupta R*, Valente L, Wulff B-E, Ariyoshi K, Iizasa H, **Davuluri RV** and Nishikura K. ADAR1 complexes with Dicer and plays a role in microRNA processing and RNA-induced gene silencing mechanisms. ***Cell.*** 2013; 153(3):575-89. Epub 2013/04/30. doi: 10.1016/j.cell.2013.03.024. PubMed PMID: 23622242; PubMed Central PMCID: PMC3651894*.* (**Equal contribution)*
5. Bi Y, **Davuluri RV**. NPEBseq: nonparametric empirical bayesian-based procedure for differential expression analysis of RNA-seq data. ***BMC Bioinformatics***. 2013;14:262. PubMed PMID: 23981227; PubMed Central PMCID: PMC3765716
6. Zhang Z, Pal S, Bi Y, Tchou J, **Davuluri RV**. Isoform level expression profiles provide better cancer signatures than gene level expression profiles. ***Genome medicine***. 2013;5(4):33. PubMed PMID: 23594586; PubMed Central PMCID: PMC3706752
7. Bhattacharjee M, Gupta R, **Davuluri RV**. Estimation of Gene Expression at Isoform Level from mRNA-Seq Data by Bayesian Hierarchical Modeling. ***Front Genet.*** 2012;3:239. PubMed PMID: 23293650; PubMed Central PMCID: PMC3536024
8. Pal S, Gupta R, Kim H, Wickramasinghe P, Baubet V, Showe LC, Dahmane N and **Davuluri RV.** Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development, ***Genome Res.*** 2011;21(8):1260-72**.** PMCID: PMC3149493
9. Kim H, Bi Y, Pal S, Gupta R, and **Davuluri RV**. IsoformEx: Isoform level expression estimation using weighted non-negetive least squares from mRNA-seq data, ***BMC Bioinformatics***, 2011;12:305. PMCID: PMC3180389