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):** 01/01/2010 – 06/30/2011

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

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

5. **Grant SAP Number:** #4100050916

6. **Project Number and Title of Research Project:** *1: Modeling the Epigenetic Changes in Alternative Promoters of Cancer Genes*

7. **Start and End Date of Research Project:** 01/01/2010 – 06/30/2011

8. **Name of Principal Investigator for the Research Project:** Ramana V. Davuluri, PhD

9. **Research Project Expenses**.

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

   $ 456,729.00

   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	Position Title	% of Effort on Project	Cost
Davuluri	Principle Investigator	18%	65,179.17
Kim	Staff Scientist	88%	86,361.77
Bi	Postdoctoral Fellow	9%	4,549.28
Gupta	Postdoctoral Fellow	100%	63,031.63
Pal	Postdoctoral Fellow	69%	27,742.36

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	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:

American Cancer Society; $150,000 direct costs per year

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___X____ No_____

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:
Informatics platform for mammalian gene regulation at isoform-level (Pending IRG review)	☒NIH ☐ Other federal (specify:_____) ☐ Nonfederal source (specify:_)	June, 2011	$ 1,269,750	$ Pending
Molecular sub-typing of gliomas by isoform-level gene network modeling	☒NIH ☐ Other federal (specify:_____) ☐ Nonfederal source (specify:_)	June, 2011	$ 2,754,986	$ Pending

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:

Based on the promising results on isoform-level expression analysis of genes in melanoma and breast cancer cell lines, and validations in breast cancer tissue samples, we are planning to submit the following grant applications in the coming year.

a) RO1 application to NIH, "Molecular sub-typing of triple negative breast cancers by isoform-level gene network modeling"
b) RO1/R21 application to NIH, "Gene isoforms as diagnostic and therapeutic targets in melanoma"

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

We will continue the development of bioinformatics analysis methods for integrating transcriptome (RNA-seq and exon-array data) and epigenome (DNA methylation) data from public data repositories (e.g. TCGA consortium glioblastoma data) to find cancer-specific gene isoforms.

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_____ No____X_____

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

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.

CURE funding has increased our capability to conduct large-scale ChIP-seq and RNA-seq studies and integrative computational modeling for genome-wide profiling of the transcriptome and TF-DNA interactions in mammalian genomes, specifically in cancer genomes.

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:

a) New collaboration with Dr. Donald O'Rourk, Director of Penn Brain Tumor Tissue Bank, Neurosurgeon, University of Pennsylvania. This collaboration has led to submission of an R01 application to NIH.
b) New collaboration with Dr. Vivek Mittal, an internationally recognized cancer researcher and the Director of the Lehman Brothers Lung Cancer Laboratory at Weill Cornell Medical College, Cornell University, New York, NY.

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 application's strategic plan). Summarize the progress made in achieving these goals, objectives and aims <u>for the entire grant award period.</u> 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. <u>Provide detailed results of the project.</u> 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

Aim 1. Development of computational tools for annotating alternative promoters and their corresponding pre-mRNA isoforms that are differentially used in different human cancers.

Aim 2. Identify differentially regulated alternative promoters in cancer genomes, corresponding pre-mRNAs and associated local epigenetic modifications in various cancer genomes by integrative ChIP-seq and bioinformatics approaches.

Aim 3. Perform bioinformatics analysis to discover tissue- and cancer-specific promoters, and model the altered chromatin structures to define unique and common epigenetic signatures in different cancer cells by using various statistical data-mining approaches.

Results pertaining to each specific aim

Aim 1. Development of computational tools for annotating alternative promoters and their corresponding pre-mRNA isoforms that are differentially used in different human cancers. We have developed MPromDb (Mammalian Promoter Database) that strives to annotate gene promoters identified from ChIP-seq results with the goal of providing an integrated resource for mammalian transcriptional regulation and epigenetics.

We applied in-house-developed bioinformatics approaches on 507 million uniquely aligned Pol-II ChIP-seq reads from 26 different data sets that include six human cell-types (Jurkat, K562, NB4, CD4+TCells, lymphoblastoid, Hela) and nine distinct mouse cell/tissues (MEF, kidney, brain, liver, lung, spleen, 3T3L1, ES, Bone marrow). The Pol-II ChIP-seq datasets include ChIP-seq data generated in our laboratory and publicly available datasets available at NCBI repository. Our method identified 34,172 and 38,850 promoters bound by Pol-II in human and mouse, respectively. We also identified more than 9,000 novel promoters in human and mouse. The predicted and annotated promoters can be accessed from our recent version of mammalian promoter database (MPromDb: http://mpromdb.wistar.upenn.edu/).

*Further Development of Cancer Gene Promoter Database***.** We integrated the cancer gene information into MPromDb, where users can search the database based on Enterz gene id/symbol, or by tissue/cell (normal or cancer) specific activity and filter results based on any combination of tissue/cell specificity. To date, we have analyzed 507 million uniquely aligned RNA Pol-II ChIP-seq reads from 26 different datasets. We plan to integrate RNA-seq data in the near future into the dataset.

Aim 2. Identify differentially regulated alternative promoters in cancer genomes, corresponding pre-mRNAs and associated local epigenetic modifications in various cancer genomes by integrative ChIP-seq and bioinformatics approaches.

Integrative analysis of exon-array data-sets revealed that isoform-level expression profiles provide better cancer signatures than gene-level expression profiles. We analyzed Affymetrix exon-array datasets comprising 160 cell lines of normal or tumor tissue origins. We obtained the transcript-level and gene-level expression estimates for the whole-transcript arrays based on the Ensemble database (version 56), which contains a total of 114,930 different transcript annotations that correspond to 35,612 different gene models. The expression estimates across all samples were normalized using the locally weighted scatter plot smoothing algorithm. We used unsupervised and supervised clustering algorithms implemented in R to study the profile similarity between the samples at both gene and isoform levels. The isoform-level expression estimates were validated by performing qRT-PCR on TPM4 isoforms in four normal breast tissues and four breast cancer patient tissue samples.

Hierarchical clustering, based on isoform-level expressions, effectively grouped the normal and oncogenic cell lines with a virtually perfect homogeneity grouping rate (97.5%) regardless of the tissue origins of the cell lines. However, at the gene-level, this rate was much lower and was 75% at best. Statistical analyses at the isoform-level between groups of cancer and normal samples revealed the existence of numerous genes having differentially expressed isoforms, which otherwise were not significant at the gene-level. The opposite expression pattern of two TPM4 isoforms observed in the cell lines was confirmed in breast cancer patient's tumor tissues over normal neighboring breast tissues. Finally, pathway-based enrichment analysis revealed "protein ubiquitination," "purine metabolism," and "breast cancer regulation by stathmin1" as significantly enriched canonical pathways with the gene set that showed differential expression at the isoform-level but not at the gene-level.

In summary, we have identified a common, isoform-level signature that can be used to effectively discriminate cancer and non-cancer cell lines and provide potentially important biological implications for a better understanding of common defining characteristics of tumors of varied types.

We will continue this type of analysis using RNA-seq and ChIP-seq datasets to predict alternative gene promoters and corresponding transcript variants specifically altered in one or more cancers.

Aim 3. Perform bioinformatics analysis to discover tissue- and cancer-specific promoters, and model the altered chromatin structures to define unique and common epigenetic signatures in different cancer cells by using various statistical data-mining approaches. We performed integrative analysis of data generated in our laboratory with the data generated through the Encyclopedia of DNA elements (ENCODE) and The Cancer Genome Atlas (TCGA) consortia.

Transcript variants of numerous genes were differentially expressed between normal and cancer cell lines. We evaluated differential gene expression, both at gene and isoform-level, between (i) all normal and all cancer cell lines, (ii) normal breast (HMEC) and breast cancer (MCF7), and (iii) normal melanocytes and melanoma cell lines. After performing the three comparisons independently, we overlapped the identified gene sets to identify those genes/gene isoforms that were consistently Up/Down regulated in cancer in comparison to normal cell lines. We denoted the genes that were found to be significantly different between normal and cancer groups in all three comparisons as core set of genes/gene isoforms. Interestingly, we found numerous genes that were significantly differentially expressed at isoform-level but not at gene-level. A gene was declared as differentially expressed at isoform-level if at least one of its isoforms showed significantly differential expression between normal and cancer groups. For example, 29 and 13 genes were found to be significantly up- and down-regulated, respectively, at isoform-level but not at gene-level in all three comparisons. Overall, we found a total of 260 different transcript variants or gene isoforms of 182 genes that showed statistically significant gene expression differences either at isoform or gene-level.

In each pair of comparisons, of the total genes identified to be significant at isoform level, at least 30% (range from 30% to 55%) were found to be significant at isoform-level only. In other words, more than one isoform of these genes displayed differential expression between normal and cancer samples, but the gene level expression differences got canceled out by the combined effect of various isoforms of the same gene. These genes display alternate splicing between normal and cancer cell lines. For example, MITF (micro-ophthalmia transcription factor) gene uses at least nine different promoters and first exons to generate a remarkably diverse set of mRNAs and protein isoforms that differ at N-terminus. Affymetrix GeneChip Human Exon 1.0 ST Array platform has probe-sets corresponding to 16 different transcript variants of this gene, according to the ENSEMBL gene annotations. The alternative promoters of MITF reflect the tissue specificity of its isoforms, which are selectively expressed in melanocytes, macrophages, osteoclasts, heart muscle, or retinal pigmented epithelium. MITF, generally believed to play a primary role in melanocyte stem cell proliferation and expression of a set of pigment-related genes, has been shown to be amplified in a small percentage (10 to 20%) of melanomas and

appears to confer a poor prognosis when over-expressed. In the comparative analysis between normal melanocytes and melanoma cell lines, MITF was found not differentially expressed at gene-level analysis. However, its isoform ENST00000352241 was found to be significantly over expressed in melanoma over melanocytes (FC=3.4), while the transcript variants ENST00000433517 (FC=-5.6) and ENST00000472437 (FC=-3.4) were seen to be under expressed in melanoma (Figure 1A). Similarly, TPM4 gene was seen to be weakly differentially expressed in MCF7 in comparison to HMEC cell-line samples. However, while one of its isoform ENST0000030933 was found to be highly over expressed (FC=5.47), another isoform, ENST00000344824, was found to be significantly under expressed in MCF7 samples (FC=-7.75) (Figure 1B). These two isoforms thus canceled off with each other, resulting in the overall gene expression to be less or not significantly different between the normal and oncogenic cell lines. TPM4 has been shown to be differentially expressed in breast cancer in the literature. Our analysis suggests that, while gene-level expression estimates of TPM4 and MITF contribute little to the discrimination of cancer cell lines from normal cells, expression estimates specific to one or more isoforms of these genes have better discriminating power. Interestingly, we found a total of 294 isoforms, corresponding to 110 genes in melanoma, and 75 transcript isoforms, corresponding to 16 genes in breast cancer that showed opposing expression patterns at isoform-level.

To validate the opposing expression patterns of gene isoforms in patient samples, between normal and cancer tissues (Figure 1A & 1B), we selected TPM4 gene in breast cancer. The opposite expression pattern of the TPM4 gene isoforms was confirmed in the ER+ and TNBC sample. Although the Her2 sample didn't show the opposing pattern of expression, one isoform showed significant and highest fold-change among all the samples (Figure 1C). In all samples, the simple studentized t test between the averaged fold expressions of the two isoforms were all significant ($p<0.001$). These results strongly support our hypothesis that the isoforms of multi-transcript genes can function in opposing roles in cancer.

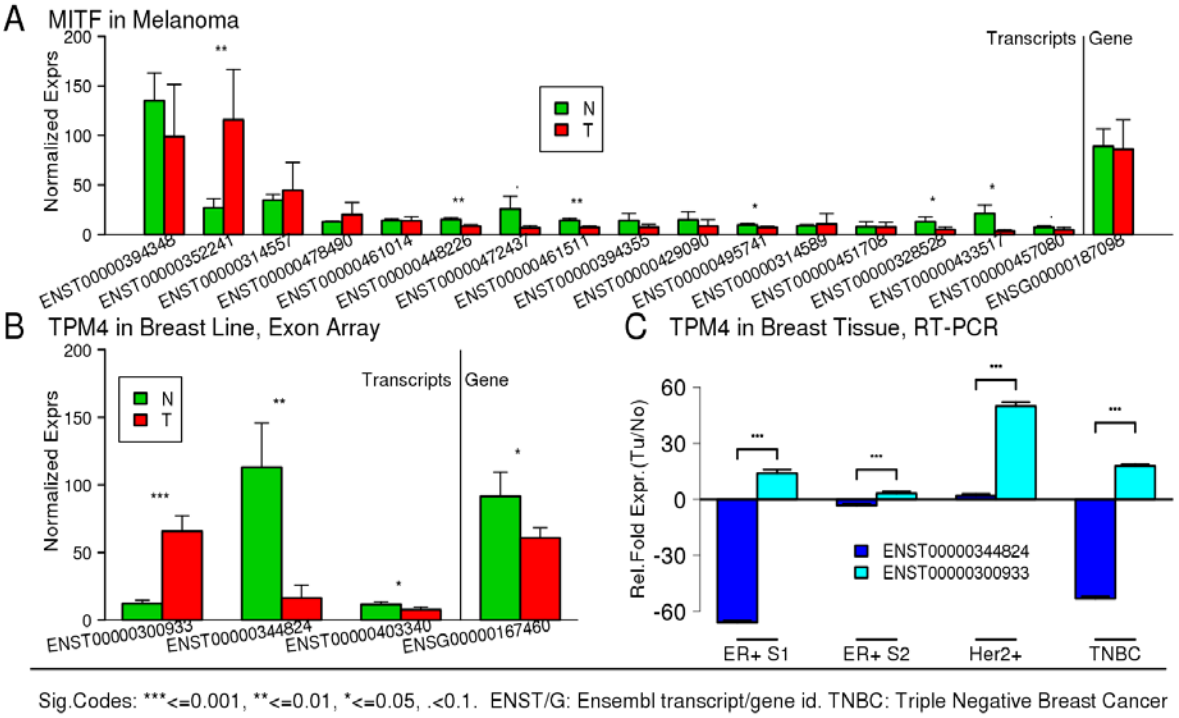

Figure 1: Bar chart of the mean normalized expression estimates of (A) MITF and its transcript variants in melanocytes (N) and melanoma (T) cell lines, (B) TPM4 and its transcript variants in HMEC (N) and MCF7 (T) cell lines. Notice that, although the gene-level expression of MITF is not significantly different between melanoma and melanocytes, three of its isoforms showed significant difference in their expression. While ENST00000352241 is over-expressed, ENST00000433517 and ENST00000472437 were under-expressed, in melanoma. Similarly, two transcript variants of TPM4 showed opposing expression patterns between HMEC normal and MCF7 breast cancer cell lines. (C) Validation results via RT-PCR experiment showing the relative fold expressions of TMP4's two major transcript variants (ENST00000300933 and ENST00000344824) in various human breast cancer tissue subtypes compared to the surrounding matched normal breast tissues. Here the subtypes are ER+: estrogen receptor positive, Her2+: Her2 gene positive, TNBC: Triple negative breast cancer. MCF7 is ER+.

We will continue our bioinformatics analysis by integrating transcriptome (RNA-seq and exon-array data) and epigenome (DNA methylation) data from TCGA consortium (e.g. glioblastoma data).

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, listed in the table, in a PDF version 5.0.5 format, 1,200 dpi. Filenames for each publication should include the number of the research project, the last name of the PI, the number of the publication and an abbreviated research project title. For example, if you submit two publications for PI Smith for the "Cognition and MRI in Older Adults" research project (Project 1), and two publications for PI Zhang for the "Lung Cancer" research project (Project 3), the filenames should be:

Project 1 – Smith – Publication 1 – Cognition and MRI
Project 1 – Smith – Publication 2 – Cognition and MRI
Project 3 – Zhang – Publication 1 – Lung Cancer
Project 3 – Zhang – Publication 2 – Lung Cancer

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. Isoform-level expression profiles provide better cancer signatures than gene-level expression profiles	Zhang, Z., Pal, S., Tchou, J. and Davuluri, R.V.	*Genome Medicine*	August 2011	☒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 planning to submit the following manuscript: "Novel molecular classification of gliomas based on isoform-level gene expression profiles" (In preparation).

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.

Novel method to classify cancers (e.g. gliomas) based on isoform-level gene expression signatures is currently under development.

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. *For Nonformula grants only – include information for only those key investigators whose biosketches were not included in the original grant application.*

BIOGRAPHICAL SKETCH

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME	POSITION TITLE
DAVULURI, Ramana V.	**Associate Professor**
eRA COMMONS USER NAME **DAVULURI10**	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

INSTITUTION AND LOCATION	DEGREE	YEAR(s)	FIELD OF STUDY
Nagarjuna University, India	B.S.	1985-1988	Mathematics
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
Dept. of Plant Genetics, VIB, University of Ghent, Gent, BELGIUM.	Post Doc. Fellow	1998	Bioinformatics
Cold Spring Harbor Laboratory (CSHL), Cold Spring Harbor, NY 11724, USA	Post Doc. Fellow	1999-2001	Computational Biology

A. Personal Statement

Dr. Davuluri 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. His research focuses on two closely related fundamental aspects of mammalian genomics and cancer: **alternative promoters** and **isoform-level gene regulatory networks**. His research program is interdisciplinary in nature with a complement of experimental investigation. Coming from a background in quantitative sciences, Davuluri gained synergistic expertise in statistical modeling, genomics and bioinformatics, while leading several projects over the past 10 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. He was also awarded an NIH R01 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.

B. Positions and Honors

Professional Experience

1993-1998 Asst. 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-Present Associate Professor; Director of Computational Biology, Center for Systems and Computational Biology, The Wistar Institute, Philadelphia, PA
2008-Present Scientific Director, Bioinformatics Facility, The Wistar Institute Cancer Center, Philadelphia, PA
2008-Present Wistar Associate Professor, Dept of Genetics, University of Pennsylvania, Philadelphia, PA

Honors and Professional Service

1989-1993 IASRI (Indian Council of Agricultural Research) junior & senior research fellowships
1997 Young Scientist Award (Statistics): 1996-97 by *Indian Science Congress Association*, 84th annual session, Delhi, India. [A prestigious award presented by the prime minister of India]

2003-2005	V Scholar Award, The V foundation for Cancer Research, USA
2006-2009	Ad Hoc Reviewer, National Science Foundation, USA; National Institutes of Health, USA.
2007	Program Committee Member, *Interface 2007: the 39th Symposium on the interface of statistics, computing science, and applications*, May 23-27 Philadelphia, USA
2007-08	Program Committee Member, *PSB 2008, 2009: (Pacific Symposium of Biocomputing)*
2007-present	Editorial board member, *Cancer Research* Official journal of AACR; *Advances in Bioinformatics*
2008-present	Editorial board member, *Journal of Nucleic Acids Investigation*
2007-present	Ad hoc member of various NIH Special Emphasis Panel Study Sections,
2011	Ad hoc member, Tumor Biology & Genomics Peer Review Committee, American Cancer Society
2008-present	Philadelphia Healthcare Trust Endowed Chair Professor, Centre for Systems and Computational Biology, The Wistar Institute, Philadelphia.

C. Selected Peer-reviewed Publications (Selected from 66 peer-reviewed publications)

1. **Davuluri RV**, Suzuki Y, Sugano S, Zhang MQ. (2000) CART classification of human 5'UTR sequences, ***Genome Res.*** 10: 1817-1827.
2. **Davuluri VR**, Grosse, I. Zhang, MQ. (2001) Computational identification of first exons and promoters in the human genome. ***Nature Genetics,*** 29: 412-417.
3. Jin V, Leu YW, Liyanarachchi S, Sun H, Fan M, Nephew KP, Huang TH, **Davuluri RV**. (2004) Identifying Estrogen Receptor α Target Genes Using Integrated Computational Genomics and Chromatin Immunoprecipitation Microarray. ***Nucleic Acids Res.*** 32*:* 6627-6635.
4. Cheng ASL, Jin VX, Fan M, Smith LT, Liyanarachchi S, Yan PS, Leu YW, Chan MWY, Plass C, Nephew KP, **Davuluri RV**, Huang TH. (2006) Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor α- responsive promoters. ***Molecular Cell*** 21:393-404.
5. **Davuluri RV**, Suzuki Y, Sugano S, Plass C, Huang TH (2008) The functional consequences of alternative promoter use in mammalian genomes. ***Trends Genet.*** **24**:167-177.
6. Gupta,R, Wikramasinghe P, Bhattacharyya A, Agosto-Perez FJ, Pal S and **Davuluri RV** (2010) Annotation of gene promoters by integrative data-mining of ChIP-seq Pol-II enrichment data. ***BMC Bioinformatics***, 2010, 11 Suppl 1:S65.
7. Gupta R, Bhattacharyya A, Agosto-Perez FJ, Wickramasinghe P, **Davuluri RV** (2011) MPromDb update 2010: an integrated resource for annotation and visualization of mammalian gene promoters and ChIP-seq experimental data. ***Nucleic Acids Res***, **39** (Database issue):D92-97.
8. Sun H, Wu J, Wickramasinghe P, Pal S, Gupta R, Bhattacharyya A, Agosto-Perez FJ, Showe LC, Huang TH, **Davuluri RV** (2011) Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq. ***Nucleic Acids Res***, **39**(1):190-201.
9. Kim H, Bi Y, Pal S, Gupta R, **Davuluri RV** (2011) IsoforrmEx: Isoform level gene expression estimation using weighted non-negative least squares from mRNA-Seq data. ***BMC Bioinformatics***. **12(1)**:305.
10. Pal S, Gupta R, Kim H, Wickramasinghe P, Baubet V, Showe LC, Dahmane N and **Davuluri RV** (2011) Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development, ***Genome Research,*** **21(8)**: 1260-1272**.**

D. Research Support

Ongoing Research Support

R01 CA135444 (Drs. G. Calin, R. Davuluri, S. Hamilton, E. Kopetz, Co-PIs) 7/17/09 – 6/30/11
NCI/NIH (Title – MicroRNAs and other non-coding RNAs in Colorectal Metastasis)

P30 CA 010815 (Dr. D. Altieri) 3/1/09 – 2/28/14
NCI/NIH
Consolidated Basic Cancer Research Program

R01 CA 132098 (Dr. L. Showe) 7/1/09 – 5/31/14
NIH (Title – HDAC Inhibitors and CTCL)

1 R01 DA 028775-01A1 (Dr. L. Montaner, Co-PI Dr. D. Speicher) 3/15/10-1/31/15
NIH (Title – NK Cell Activation and Function in HIV-1 Exposed Uninfected IV Drug Users)