Response Form for the Final Performance Review Report*

1. Name of Grantee: The Wistar Institute

2. Year of Grant: 2009 Formula Grant

A. For the overall grant, briefly describe your grant oversight process. How will you ensure that future health research grants and projects are completed and required reports (Annual Reports, Final Progress Reports, Audit Reports, etc.) are submitted to the Department in accordance with Grant Agreements? If any of the research projects contained in the grant received an "unfavorable" rating, please describe how you will ensure the Principal Investigator is more closely monitored (or not funded) when conducting future formula funded health research.

In order to ensure an appropriate and timely response to all requests for reports from the Department of Health, The Wistar Institute has designated a single person, Ms. Joyce Macauley, to serve as the sole liaison with the Department. Ms. Macauley gathers all documents, communicates with all researchers, assures appropriate certifications and assembles all grants, progress and final reports. She has had this responsibility since the inception of the Tobacco Formula Funding program. Her experience and familiarity with personnel in the Department of Health assures the Institute's immediate response to all requests.

* Please note that for grants ending on or after July 1, 2007, grantees' Final Performance Review Reports, Response Forms, and Final Progress Reports ***will be made publicly available on the CURE Program's Web site.***

Project Number: 0991601
Project Title: Modeling the Epigenetic Changes in Alternative Promoters of Cancer Genes
Investigator: Davuluri, Ramana V.

B. Briefly describe your plans to address <u>each</u> specific weakness and recommendation in <u>Section B</u> of the Final Performance Summary Report using the following format. As you prepare your response please be aware that the Final Performance Review Summary Report, this Response Form, and the Final Progress Report will be made publicly available on the CURE Program's Web site.

Reviewer Comment on Specific Weakness and Recommendation (*Copy and paste from the report the reviewers' comments listed under Section B - Specific Weaknesses and Recommendations)*:

Response (*Describe your plan to address each specific weakness and recommendation to ensure the feedback provided is utilized to improve ongoing or future research efforts*):

Section B. Recommendations

SPECIFIC WEAKNESSES AND RECOMMENDATIONS

<u>Reviewer 1:</u>

1. It would be beneficial to see more in-depth analysis on the methods behind CancerPromDb. The ability to accurately analyze Chip-Seq and particularly RNA-Seq is a rapidly developing area of research that would require more in-depth descriptions. Furthermore, understanding the technical comparisons between exon array and RNA-Seq would be beneficial for interpreting the results.

   <u>Response:</u>
   The Methods behind CancerPromDb are same as MPromDb (Mammalian Promoter Database – http://mpromdb.wistar.upenn.edu). Based on feedback from the users, we have integrated both databases, and the data is now maintained in a single database to minimize the maintenance tasks. The methods behind these databases are published in the following papers.
   a) Bhattacharjee, M., Gupta, R. and **<u>Davuluri, R.V.</u>** (2012) Estimation of gene expression at isoform level from mRNA-Seq data by Bayesian hierarchical modeling, ***Frontiers in Genetics***, **3**, Article 239. doi: 10.3389/fgene.2012.00239.
   b) Kim, H., Bi, Y., Pal, S., Gupta, R., and **<u>Davuluri, R.V.</u>** (2011) IsoformEx: Isoform level expression estimation using weighted non-negetive least squares from mRNA-seq data, ***BMC Bioinformatics***, **12(1):305.**
   c) Sun, H., Wu, J., Wickramasinghe, P., Pal, S., Gupta, R., Bhattacharyya, A., Agosto-Perez, F.J., Showe, L.C., Huang, T.H. and **<u>Davuluri, R.V.</u>** (2011) Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq, ***Nucleic Acids Res***, **39**, 190-201.

d) Gupta, R., Bhattacharyya, A., Agosto-Perez, F.J., Wickramasinghe, P. and **Davuluri, R.V.** (2011) MPromDb update 2010: an integrated resource for annotation and visualization of mammalian gene promoters and ChIP-seq experimental data, ***Nucleic Acids Res***, **39**, D92-97.
e) Gupta, R., Wikramasinghe, P., Bhattacharyya, A., Perez, F.A., Pal, S. and **Davuluri, R.V.** (2010) Annotation of gene promoters by integrative data-mining of ChIP-seq Pol-II enrichment data, ***BMC Bioinformatics***, **11 Suppl 1**, S65.

Also, we are investigating various statistical methods for comparisons between exon array and RNA-Seq platforms, based on analysis of TCGA Glioblastoma datasets, and validations using RT-qPCR assays on independent tumor tissue samples from University of Pennsylvania Brain Tumor Registry.

2. What are the technical limitations of evaluating 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? What noise is introduced by integrating these data sets? And how do the researchers rule out that the consistently up-/down-regulated genes are not artifacts in the data? What about tumor specific alterations?

Response:
The technical limitations of estimating transcript/isoform-level expression from RNA-seq are discussed, and improved methods are presented, in the following publications.

a) Bhattacharjee, M., Gupta, R. and **Davuluri, R.V.** (2012) Estimation of gene expression at isoform level from mRNA-Seq data by Bayesian hierarchical modeling, ***Frontiers in Genetics***, **3**, Article 239. doi: 10.3389/fgene.2012.00239.
b) Kim, H., Bi, Y., Pal, S., Gupta, R., and **Davuluri, R.V.** (2011) IsoformEx: Isoform level expression estimation using weighted non-negetive least squares from mRNA-seq data, ***BMC Bioinformatics***, **12(1):305.**

We have recently developed a new method, called NPEBseq (Nonparametric Empirical Bayesian-based Procedure for Differential Expression Analysis of RNA-seq Data). The manuscript describing these results is currently in review.

a) Bi, Y. and **Davuluri, R.V.** (2013) NPEBseq: Nonparametric Empirical Bayesian based Procedure for Differential Expression Analysis from RNA-seq Data. *(Submitted)*

3. The researchers describe analyzing 160 cell lines of tumor tissue origin. However, when they conduct the pathway analysis they identify a breast-cancer regulation by stathmin1 as a significantly enriched pathway. What is the interpretation of this finding? Perhaps more work could be placed on determining, via in silico and experimental methods, the accuracy of such an approach. It is unclear whether this is relevant or a false positive.

Response:
Thank you very much for the suggestion. We have discussed these findings in the published paper (Zhang et al. 2013, ***Genome Medicine***, Apr 17;**5(4)**:33).

An isoform or gene was selected if both its fold change was greater than a cut-off value of 2, and the false discovery rate (FDR)-adjusted P value was smaller than a cut-off value of 0.01 for all comparisons between the two conditions. Ingenuity Pathway Analysis (IPA) was used to associate the identified gene sets with biological functions, canonical pathways, and networks. To identify pathway differences arising from gene sets identified at either isoform or gene level, we used the counting method on the P values of pathways from the IPA analysis; the P values were used as an indicator of association strength between the gene sets and pathways. For the three pairwise oncogenic/non-oncogenic comparisons (all oncogenic cell lines versus non-oncogenic cell lines, melanocyte versus melanoma, HMEC versus MCF7), a pathway was selected and reported if it was significantly associated with the gene sets identified at isoform-level in all three pairs of comparisons, but was not significantly associated with the gene sets identified at gene-level in all three pairs of comparisons, or vice versa. The significance level was set at $P < 0.05$ for all comparisons.

4. The annotation resource will be valuable to the community, but the inability to access the web site is concerning.

   Response:
   The database is now available at http://mpromdb.wistar.upenn.edu.

5. It would be beneficial to see more of the progress made looking at specific cancer cohorts such as TCGA to assess the significance of this approach in a cancer context.

   Response:
   We have applied our methods on TCGA cohorts, currently on glioblastoma multiforme (GBM), genetically heterogeneous and most aggressive primary brain tumors.

   Based on the preliminary results generated through this project, we hypothesize that gene expression signatures at isoform-level would allow us to identify novel molecular markers and generate a more robust and clinically translatable assay for GBM patient stratification. We are currently investigating the molecular heterogeneity of GBMs at gene-isoform level rather than at gene-level, and aim to identify gene-isoforms as novel diagnostic and prognostic markers for GBM patient stratification using a multi-disciplinary approach. We showed that isoform level expression profiles provide better cancer signatures than gene-level expression profiles (Zhang et al. 2013, Genome Medicine). We recently developed a tumor classification assay based on isoform-level expression signature, which demonstrated improved prognostic stratification of GBM patients over gene-level expression signature (Pal et al. 2013, submitted). In this study, we designed a novel classification model based on isoform-level gene expression profiles and translated the gene-panel to a high-throughput RT-qPCR validation assay. Using isoform-level expression analyses of The Cancer Genome Atlas (TCGA) samples, we identified four GBM subgroups with significant ($p=0.0103$) survival differences. A four-class classifier, built with 121 transcript-variants, assigns GBM patients' molecular subtype with 92% accuracy. The classifier was translated to an RT-qPCR-based assay and validated on an independent cohort of 206 glioblastoma samples. We

found that proneural patients have the worst prognosis except for younger patients (<40 years), while a better prognosis is observed in older (≥40 years) neural group of patients.

We will continue these investigations on TCGA datasets of ovarian cancer and low-grade gliomas during the next one year.

Reviewer 2:

1. One manuscript was submitted to Genome Medicine in August, 2011. I was unable to find evidence that this paper has been published. This is a relatively low level of productivity in terms of peer-reviewed publication. No other activities involving licenses, patents, or commercial development were reported. The researchers should do a better job of promptly publishing the results of their funded research.

   Response:
   The lag in publishing the results was due to the departure of lead authors of the work to better positions in industry and academia. The following articles, funded by this work, are now published.
   a) Zhang, Z., Pal, S., Tchou, J. and **Davuluri, R.V.** (2013) Isoform-level expression profiles provide better cancer signatures than gene-level expression profiles, ***Genome Medicine***, Apr 17;**5(4)**:33.
   b) Bhattacharjee, M., Gupta, R. and **Davuluri, R.V.** (2012) Estimation of gene expression at isoform level from mRNA-Seq data by Bayesian hierarchical modeling, ***Frontiers in Genetics***, **3**, Article 239. doi: 10.3389/fgene.2012.00239.
   c) Pal, S., Gupta, R. and **Davuluri, R.V.** (2012) Alternative transcription and alternative splicing in cancer. ***Pharmacology & Therapeutics***, **136**(3):283-94.

2. The investigators should either carry out all the proposed research or explain why some of the planned research was not completed. Specifically in question is the work on finding cancer-associated epigenetic modifications and modeling altered chromatin structures to define unique and common epigenetic signatures in different cancer cells.

   Response:
   While we were successful in conducting ChIP-seq experiments using fresh mouse tissues (Pal et al. 2011, ***Genome Research*, 21(8)**: 1260-1272), we failed to get reproducible ChIP-seq data using frozen cancer specimens. Moreover, we could not extract enough cells/DNA from the tissue specimens. We, therefore, decided to focus more on the analysis of publicly available datasets, such as TCGA. For example, we diverted our attention to design novel classification systems based on isoform-level gene expression profiles – an avenue unexplored for diagnostic and prognostic use. Using isoform-level expression clustering of The Cancer Genome Atlas (TCGA) samples, we identified four GBM subgroups with significant (p=0.0103) survival differences. A four-class classifier, built with 121 transcript-variants, assigns GBM patients' molecular subtype with 92% accuracy. The classifier was translated to a high-throughput RT-qPCR assay and validated on an independent cohort of 206 glioblastoma samples. In patients ≥40 years old, we found that proneural patients have the worst prognosis, while a better prognosis is observed in neural patients. The isoform-level classifier, transformed from a high-dimensional platform to a low-dimensional RT-qPCR platform, provides a quantitative and reproducible stratification of GBM patients with

prognostic significance, a requirement towards personalized medicine. The resulting diagnostic assay from the classifier has immediate clinical implications towards prioritizing patients for standard care versus aggressive therapy regimen. The following manuscript is currently in review.

a) Pal, S., Bi, Y., Macyszyn, L., Showe, L.C, O'Rourke, D.M., and **Davuluri, R.V.** (2013) An Isoform-level gene signature improves prognostic stratification and accurately classifies glioblastoma subtypes (*Submitted).*

Also, the following patent application is pending.

a) United States Patent Application (Pending) WST141P1**: Methods and compositions for diagnosis of glioblastoma or a sub-type thereof**. (Inventors – Sharmistha Pal, Yingtao Bi, Luke Macyszyn, Louise C. Showe, Donald M. O'Rourke, and Ramana V. Davuluri).

Reviewer 3:
The productivity of this research group in terms of publication is relatively weak. With this relatively novel research area and the amount of funding available to this group, one would expect better and more publications, which would allow reviewers to better evaluate the success of this group. Currently, this review has to take the investigator's words for it and relies more on the potential of the project rather than the tangible outcome for the rating.

Response: Please see the responses to reviewers 1 and 2. Also, we have successfully competed for the following NIH R01.

1R01 LM011297 (PI: DAVULURI) 5/2/13–4/30/16
NLM/NIH Grant amount: $378,000 per year
Title: Informatics platform for mammalian gene regulation at isoform-level
The major goal of the proposed research is to develop novel algorithms and an integrated database for understanding and modeling gene regulation at isoform-level between normal and disease conditions by integrating heterogeneous datasets from high throughput and next generation sequencing experiments including ChIP-seq and RNA-seq.

C. If the research project received an "unfavorable" rating, please indicate the steps that you intend to take to address the criteria that the project failed to meet and to modify research project oversight so that future projects will not receive "unfavorable" ratings.

Response: None

D. Additional comments in response to the Final Performance Review Report (OPTIONAL):

Response:
By using the results from this funding, we have also applied for the following multi-PI R01:

1R01CA185784-01 (MPI: DAVULURI (Contact); O'ROURKE) NIH/NCI .
Title: Molecular Stratification of Glioblastomas Using Isoform-level Gene Signatures
This proposal addresses the validation of our recently developed tumor classification assay based on isoform-level expression signature. We seek to validate the classification assay in a larger GBM cohort and investigate intra- and inter-tumor heterogeneity using independent GBM samples from University of Pennsylvania Brain Tumor Registry.

Project Number: 0991602
Project Title: Characterizing Mechanisms of Transcriptional Activation Using Live Cell Imaging
Investigator: Janicki, Susan M.

B. Briefly describe your plans to address <u>each</u> specific weakness and recommendation in <u>Section B</u> of the Final Performance Summary Report using the following format. As you prepare your response please be aware that the Final Performance Review Summary Report, this Response Form, and the Final Progress Report will be made publicly available on the CURE Program's Web site.

Reviewer Comment on Specific Weakness and Recommendation (*Copy and paste from the report the reviewers' comments listed under Section B - Specific Weaknesses and Recommendations)*:

Response (*Describe your plan to address each specific weakness and recommendation to ensure the feedback provided is utilized to improve ongoing or future research efforts*):

Section B. Recommendations

SPECIFIC WEAKNESSES AND RECOMMENDATIONS

<u>Reviewer 1:</u>

1. I believe that Aim2 is still salvageable with a better fluorescence detector/camera. I recommend that the PI be given additional funds to support the purchase of a new detector for the microscopy system she has in the laboratory. She should seek advice from experts in light microscopy or fluorescence detectors (James Pawley, University of Wisconsin-Madison comes to mind) before purchasing a new detector.

   <u>Response:</u>
   We agree with Reviewer #1 that results are always limited by the technology available for the experiment. It is likely that improved detection capabilities could increase the information and, ultimately, the insight that could be gained from these experiments.

2. I believe the gene silencing work is very exciting with great potential for cancer research. This work should be given additional funds for development.

   <u>Response:</u>
   We plan to continue to pursue this course of investigation.

<u>Reviewer 2:</u>
There are several weaknesses in the execution of the assays. First, the dynamic changes of the fluorescent intensity should be translated into actual rates (preferably initial rates) to tease out subtle differences in biological processes. Second, the investigators missed an opportunity to study the turnover rates of these proteins on the complexes. While the apparent molecular

recruitment may not show appreciable distinction in dynamics, the actual turnover rate and therefore the dissociation constants may differ, and this can be teased out easily using simple techniques such as fluorescence recovery after photobleaching (FRAP).

Response:
We agree with Reviewer #2 that FRAP experiments should be undertaken in this system in order to examine the turnover rate of factors on the chromatin. Important new insights could be gained by examining the rapid turnover of factors at the transcription site.

Reviewer 3:

1. The proposed research is innovative and exciting. However, the PI encountered several technical hurdles that prevented the PI from pursuing the original goals. This may be related to the short funding period. One strategy is to allow the PI to have a no-cost extension to solve the technical challenges. For example, the established transgene array did not permit the proposed research. If time permits, the PI could modify the transgene array or establish different arrays.

   Response:
   We are in the process of modifying the transgene arrays in order to solve the technical difficulties. We agree that changes to the underlying DNA sequence could advance the scientific aims.

2. The funding appears mainly to provide salary support to the PI (74% effort). Is the funding designed to relieve the PI from other duties (teaching, service, etc.)? Also the effort for the PI seems to be too high.

   Response:
   The majority of the PI's time was spent on this project. The focused effort resulted in the production of data that resulted in the funding of a grant application by NIH.

C. If the research project received an "unfavorable" rating, please indicate the steps that you intend to take to address the criteria that the project failed to meet and to modify research project oversight so that future projects will not receive "unfavorable" ratings.

Response:

D. Additional comments in response to the Final Performance Review Report (OPTIONAL):

Response: None

Project Number: 0991603
Project Title: Elucidation of the Integrator Composition and Function
Investigator: Shiekhattar, Ramin

B. Briefly describe your plans to address <u>each</u> specific weakness and recommendation in <u>Section B</u> of the Final Performance Summary Report using the following format. As you prepare your response please be aware that the Final Performance Review Summary Report, this Response Form, and the Final Progress Report will be made publicly available on the CURE Program's Web site.

Reviewer Comment on Specific Weakness and Recommendation (*Copy and paste from the report the reviewers' comments listed under Section B - Specific Weaknesses and Recommendations)*:

Response (*Describe your plan to address each specific weakness and recommendation to ensure the feedback provided is utilized to improve ongoing or future research efforts*):

Section B. Recommendations

SPECIFIC WEAKNESSES AND RECOMMENDATIONS

<u>Reviewer 1:</u>

1. The novel roles for the integrator complex that were identified are important and are the major strength of the project. The project made acceptable progress, but failed to meet most of its stated objectives. The reasons for not pursuing the stated objectives should be clearly outlined. The replacement research should be well-rationalized and outlined, as well.

   <u>Response:</u>
   We thank the reviewer for pointing this out and below is a further explanation of the rationale for pursuing the role of Integrator complex in regulation of protein-coding genes. During the execution of the project using genome-wide approaches, we realized that a large amount of the Integrator complex resides at protein-coding genes. This was surprising to us as our initial functional data had suggested that Integrator is critical for processing of small nuclear RNAs, which are a group of RNA polymerase II genes that are not subject to polyadenylation. These are a small set of non-protein coding transcripts (about 50) of RNA polymerase II genes, and it was not clear why such an elaborate multiprotein complex (Integrator contains a core of 12 subunits) could be dedicated to a small subset of genes. To really understand the scope of Integrator function, it was paramount to uncover its function at protein-coding genes. During the course of these experiments, we uncovered a role for Integrator in regulation of transcriptional elongation by recruiting the super elongation complex (SEC) containing the CDK9 module. Our additional, functional analysis using global run on (GRO-seq) experiments have also identified an important function for Integrator in processing of a large number of other non-protein coding genes including the RNA component of the telomerase complex known as TERC. Therefore, the major impetus

was to delineate a greater role for Integrator in transcriptional regulation and RNA processing.

2. Efforts should be made to recruit new scientists and to be involved in trainee education. Note: Though the application cites support for post-doctoral trainees, Question 13 was answered, "No." The four post-doctoral trainees should be included in the table, and the question should have been answered, "Yes."

   Response:
   This was an administrative oversight when completing the final report. Question 13 should have shown four postdoctoral trainees.

3. When the work is published, the Health Research CURE Program should be appropriately cited.

   Response:
   We will indicate the contribution of Health Research CURE Program to the publication of Integrator manuscripts.

4. This was a strong basic science research project that yielded unexpected but interesting data. Similar research should continue to be supported, but training and recruitment should be fostered as part of the projects.

   Response:*:*
   Indeed, training and recruitment will be part of the projects.

Reviewer 2:

1. Publication of the data is an important goal. Leveraging the results for future research outcomes is crucial, and future goals are not clearly outlined.

   Response:
   The revised manuscript regarding our analysis of Integrator function to regulate protein-coding genes is under consideration. Thus far we have uncovered a role for Integrator in responsiveness of immediate early genes to epidermal growth factor (EGF). Depletion of Integrator subunits substantially diminished the activation of transcription following EGF stimulation. We detailed the molecular mechanism to include a defect in transcriptional elongation following the activation process. Using genome-wide analysis, Integrator was deemed critical for recruitment of elongation factors during transcriptional activation. Additionally, we uncovered a role for Integrator in processing of the 3'-end of other non-coding RNAs whose transcripts are not polyadenylated. Therefore, the major trust of the future work will be to: 1) understand whether the contribution of Integrator to transcriptional activation is specific for immediate early genes and the EGF signaling or whether Integrator has a larger role in transcriptional activation through other signaling pathways, and 2) delineate the impact of Integrator RNA processing activity on other species of non-poyadenylated RNA in mammalian cells.

2. A clearer explanation of the progress on Aim 1, including how and why the project direction has apparently changed, would have been helpful in evaluation of the overall progress in the project.

   Response:
   During the experiments examining the occupancy and the role of Integrator on small nuclear RNAs (snRNAs; Aim 1), we realized a much larger scope of function for this complex beyond that of ~50 snRNA genes. Both our analysis of the occupancy of Integrator and its functional impact in transcriptional regulation compelled us to look beyond small nuclear RNA genes and to examine the possible role for Integrator in regulation of protein-coding genes and other forms of non-coding RNAs. Therefore, the extension of experiments proposed in Aim 1 to analyze the role of Integrator in transcription and RNA processing of protein-coding genes was the logical expansion of the original aims with the intent to uncover the scope of the function for this multiprotein complex.

3. There should be a better explanation of the roles of individual investigators. In addition, how the funds were used to improve the research capacity of the institution should be included.

   Response:
   Dr. Baillat performed experiments using recombinant subunits of the Integrator complex in order to establish *in vitro* RNA processing assays such that we could assess the contribution of individual subunits of Integrator to RNA processing. Drs. Carre and Beringer performed Genome-wide chromatin immunoprecipitation (ChIP-seq) and high throughput RNA sequencing (RNA-seq) experiments. Dr. Gardini provided guidance to Ms. Saviolaki to analyze the role of Integrator in processing of other noncoding RNA in mammalian cells that used Integrator in their 3'-end processing. Ms. Acevedo provided experimental support in tissue culture and general preparation of the reagents for these experiments. This project provided support for the genomic facilities at The Wistar institute and was important for developing a bioinformatics infrastructure in the institute.

Reviewer 3:

1. The execution of the project veered away in part from the original proposal. This is not a substantial weakness because that is how science works, and the end results were no less interesting or important.

   Response:
   I would like to thank the reviewer for his/her understanding of the extension to the original aims. This extension to the original aims was required to gain a better understanding of the scope of Integrator function.

The value for money and benefits seem to be out of line with the investment, if the figure of $907K in 18 months is correctly understood by this reviewer.

Response:
Gaining a fundamental understanding of the Integrator function required growing large amounts of mammalian cells using cell culture media as well as performing a number of genome-wide experiments including ChIP-seq, RNA-seq and Gro-Seq. We also had to use a large amount of bioinformatics power to analyze these results. Therefore, the funds were used to support personnel and supplies, as well as core facility expenditures to accomplish our goals. It should also be noted that the direct costs received during this time period totaled $542,141.74 and the indirect costs totaled $364,861.39 ($907,003.13).

C. If the research project received an "unfavorable" rating, please indicate the steps that you intend to take to address the criteria that the project failed to meet and to modify research project oversight so that future projects will not receive "unfavorable" ratings.

Response:

D. Additional comments in response to the Final Performance Review Report (OPTIONAL):

Response: None