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: MPC Corporation

2. Reporting Period (start and end date of grant award period): 1/1/09-6/30/10

3. Grant Contact Person (First Name, M.I., Last Name, Degrees): Shannon M. Barnes, MS

4. Grant Contact Person’s Telephone Number: 412-648-9676

5. Grant SAP Number: 4100047641

6. Project Number and Title of Research Project: Project 3, Computational Modeling of miRNA Involvement in Pulmonary Gene Networks

7. Start and End Date of Research Project: 1/1/09-6/30/10

8. Name of Principal Investigator for the Research Project: Panagiotis Benos, PhD


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:

$29,273.11

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).
<table>
<thead>
<tr>
<th>Last Name</th>
<th>Position Title</th>
<th>% of Effort on Project</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang, Tzu-Wei (Grace)</td>
<td>Graduate Student Researcher</td>
<td>4.17%, Year 1</td>
<td>$1,541.86</td>
</tr>
<tr>
<td>Huang, Tzu-Wei (Grace)</td>
<td>Graduate Student Researcher</td>
<td>75%, Year 2</td>
<td>$27,731.25</td>
</tr>
</tbody>
</table>

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).

<table>
<thead>
<tr>
<th>Last Name</th>
<th>Position Title</th>
<th>% of Effort on Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benos</td>
<td>Principal Investigator</td>
<td>10</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Type of Scientific Equipment</th>
<th>Value Derived</th>
<th>Cost</th>
</tr>
</thead>
</table>

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:

NIH-RO1-Modeling InVivo DNA Year 1-$11,092.50
NIH-RO1-Modeling InVivo DNA Year 2-$9,243.75

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.

<table>
<thead>
<tr>
<th>A. Title of research project on grant application</th>
<th>B. Funding agency (check those that apply)</th>
<th>C. Month and Year Submitted</th>
<th>D. Amount of funds requested:</th>
<th>E. Amount of funds to be awarded:</th>
</tr>
</thead>
<tbody>
<tr>
<td>□NIH</td>
<td>□ Other federal (specify:________)</td>
<td>$</td>
<td>$</td>
<td></td>
</tr>
<tr>
<td>□ Nonfederal source (specify: ____________)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□NIH</td>
<td>□ Other federal (specify:________)</td>
<td>$</td>
<td>$</td>
<td></td>
</tr>
<tr>
<td>□ Nonfederal source (specify: ____________)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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:

During the duration of the project, Ms. Huang developed a novel method to combine sequence (static) and gene expression (dynamic) data. The novelty of the method lies in the way it combines different lines of evidence in a unified framework. We are currently testing this method on publicly available data and data from our collaborators’ labs on idiopathic pulmonary fibrosis (IPF). If the method is proven to be useful in identifying IPF-related gene networks, we will apply for a joint NIH grant to: (1) further develop the method and (2) collect and analyze data to elucidate the IPF-related gene regulatory pathways.

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

See above.

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 X No

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

<table>
<thead>
<tr>
<th></th>
<th>Undergraduate</th>
<th>Masters</th>
<th>Pre-doc</th>
<th>Post-doc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Undergraduate</th>
<th>Masters</th>
<th>Pre-doc</th>
<th>Post-doc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Undergraduate</th>
<th>Masters</th>
<th>Pre-doc</th>
<th>Post-doc</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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_________  No____ X____

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


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_________  No____ X____

If yes, please describe the collaborations:

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 for the entire grant award period. 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 (β) should not print as boxes (□) and include the appropriate citation(s). DO NOT DELETE THESE INSTRUCTIONS.**

The two aims of this project were: (1) to evaluate existing methods and develop new methods for identifying critical regulatory network modules and (2) to analyze high-throughput data related to idiopathic pulmonary fibrosis (IPF) in order to predict critical regulatory modules associated with it. In the last 12 months (funding period), we accomplished Aim 1 and we continue to work toward Aim 2.

**Results of Aim 1.** The student funded by this project (Grace Huang) developed a new method that combines sequence (static) data and gene expression (dynamic) data in order to infer relations among transcription factors (TFs), microRNA (miRNA), and other genes. Sequence data are used to score possible TF → gene interactions (“gene” here indicates miRNA gene, TF, or another protein coding gene) by examining various features of the corresponding gene’s promoter. Currently the features include the score for TF binding on
the promoter, as well as the accessibility of the promoter as it is evidenced by various chromatin immunoprecipitation (ChIP)-chip or ChIP-seq experiments (H4 histone acetylation, CCCTC-binding factor (CTCF) binding, etc). We plan to expand the feature set as more features become available. The miRNA → gene interactions (here “gene” indicates a TF or other protein coding gene) are inferred by combining predictions of five algorithms (including the popular TargetScan, PicTar, and PITA). Currently, the combined score is calculated empirically, but we plan to extend the method to combine the target prediction scores in a unified framework using a support vector machine (SVM) or another general machine learning algorithm.

Gene expression data of protein coding and miRNA genes are analyzed together, and pairwise expression correlations are identified between various genes (TFs, miRNAs, other genes) using standard statistical methods. The p-values for these associations are combined with the p-values from the static data to generate a confidence score. A new web server has been developed by our programmer (Harry Athanassiou; funded by other sources) to graphically present the output network of interactions. The pipeline of the algorithm is presented in Figure 1.

**Future directions of Aim 1.** We plan to include more features in the determination of active versus non-active promoters, we will experiment with existing methods or develop new methods for combining miRNA target prediction scores, and we will finish the development of the analysis pipeline web server by including tools for clique finding and gene ontology (GO) analysis. We will also extend our static network calculations to mouse and other species (currently available only for humans). We anticipate that this work will result in one publication describing the server and—depending on the results of Aim 2—one or two more publications describing the method and presenting some interesting biological findings.

**Results of Aim 2.** In Aim 2 we proposed to use our tool to identify gene regulatory modules related to IPF. In particular, we are interested in identifying significant feed-forward-loops (FFLs) (Figure 2). As a test case, we analyzed a dataset consisting of 17 IPF samples and 7 control samples (matched mRNA + miRNA arrays), provided by our collaborator, Dr. Naftali Kaminski. Although our algorithm can work with protein coding gene expression data only,
we wanted to test its efficiency on combined gene expression datasets. The mRNA and miRNA profiles were “matched,” that is, they were collected from the same tissue from the same patient. The preliminary analysis identified a number of potentially interesting FFLs (Figure 3). For example, we predict that miR-34a and miR-10b participate in an FFL by downregulating signal transducer and activator of transcription 3 (STAT3) and its target additional sex comb-like 1 (ASXL1).

Future directions of Aim 2. In the future, we plan to extend the analysis of the IPF and control cases using new data from the Lung Genomics Research Consortium (LGRC). This NIH-funded consortium generates large-scale clinical, gene expression (microarray and deep sequencing), and methylation data. Dr. Kaminski is a key participant in this project. We anticipate at least one publication from the analyses of these datasets.

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

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.

<table>
<thead>
<tr>
<th>Title of Journal Article:</th>
<th>Authors:</th>
<th>Name of Peer-reviewed Publication:</th>
<th>Month and Year Submitted:</th>
<th>Publication Status (check appropriate box below):</th>
</tr>
</thead>
</table>
| 1.                       |          |                                   |                          | Submitted
| 2.                       |          |                                   |                          | Accepted
| 3.                       |          |                                   |                          | 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 anticipate that this project will generate 2-3 publications. One article will describe the web server that student Grace Huang (funded by this project) and programmer Harry Athanassiou (funded from other sources) are putting together. This paper will most likely be submitted to *Nucleic Acids Research* web server issue. Another publication will most likely result from the description of the method itself. We plan to submit to a journal like *Genome Research* or *PLoS Comput Biology*, and it will include a limited analysis of some of the data. A third publication will result from the analysis of the full-scale LGRC dataset (see above).

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.

Currently there is no cure for IPF, largely because of our limited knowledge about the mechanisms that contribute to its onset. This project develops tools that will facilitate analyses of mRNA and miRNA gene expression data, thus contributing to the better understanding of the mechanism of the disease.

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.

At this phase, our research focused on the development of an efficient algorithm for the identification of critical regulatory network modules. We expect that the major discoveries will arise from analyzing the high-throughput data generated by LGRC and other sources (see above).

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_______ X____ No____
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:

At this phase, our research is focused on the development of an efficient algorithm for the identification of critical regulatory network modules. If the anticipated analysis of the high-throughput data results in interesting findings, all appropriate steps will be taken to ensure that the knowledge generated by this project will ultimately help patients.
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.*

Name: Panagiotis Benos, PhD  
Current position: Associate Professor (tenured), University of Pittsburgh  
Citizenship: USA  
Business Address: 3501 Fifth Avenue, 3078 Biomedical Science Tower 3, Pittsburgh, PA 15260  
E-Mail Address: benos@pitt.edu  
Business Phone: 412-648-3315  

**EDUCATION and TRAINING**  
**UNDERGRADUATE:**  
1984-1990 University of Crete, BS Mathematics, Heraklion, Crete, Greece  

**GRADUATE:**  
1991-1993 University of Crete, MS equivalent Molecular Biology, Heraklion, Crete, Greece  
1993-1997 University of Crete, PhD Molecular Biology, Heraklion, Crete, Greece  

**POSTGRADUATE:**  
1999-2002 Washington University in St. Louis Prof. Gary Stormo, School of Medicine, Department of Genetics, St. Louis, MO  

**APPOINTMENTS and POSITIONS**  
**ACADEMIC:**  
1992-1996 Teaching Assistant, University of Crete School of Arts and Sciences, Department of Biology, Heraklion, Crete, Greece  
1997-1999 Research Fellow, EMBL- European Bioinformatics Institute, Hinxton, Cambridge, UK  
1999-2002 Research Associate, Washington University in St. Louis School of Medicine, Department of Genetics, St. Louis, MO  
2002-2005 Assistant Professor, University of Pittsburgh Department of Human Genetics until August 2002 and Visiting Assistant Professor, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA  
2005-2008 Assistant Professor, University of Pittsburgh School of Medicine, Department of Computational Biology, Pittsburgh, PA  
2008-present Associate Professor, University of Pittsburgh School of Medicine, Department of Computational Biology, Pittsburgh, PA
OTHER AFFILIATIONS
- Co-director, Carnegie Mellon University-University of Pittsburgh (CMU-Pitt) Joint PhD Program in Computational Biology (2009-present).
- Associate Professor (secondary appointment), University of Pittsburgh School of Medicine, Department of Biomedical Informatics, 2007-present.
- Member, CMU-Pitt Joint PhD Program in Computational Biology, 2004-present
- Member, Training Program Core Faculty, University of Pittsburgh School of Medicine, Department of Biomedical Informatics, 2003-present
- Member, Molecular and Cellular Biology Program, University of Pittsburgh Cancer Institute, 2002-present.
- Member, Molecular Virology Program, University of Pittsburgh Cancer Institute, 2004-present.

HONORS AND ACHIEVEMENTS
Aug. 1993: Ranked first in graduation class at post-graduate educational program (M.Sc. equivalent).
1995-1998: Elected member and Secretary of the Education and Training Project Committee, European Molecular Biology Network (EMBnet)
2005-2006: Elected Vice Chair of the Recruitment Committee, Interdisciplinary Biomedical Graduate Program, School of Medicine, University of Pittsburgh.
2007-current: Elected member of the Faculty of 1000 (http://www.f1000biology.com/home).

RECENT PUBLICATIONS


