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):** 1/1/2009-6/30/2010

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 1, Computational and Experimental Approaches for Predicting miRNAs and their Targets

7. **Start and End Date of Research Project:** 1/1/2009-6/30/2010

8. **Name of Principal Investigator for the Research Project:** Veronica Hinman, 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:

   $ $31,570.06

   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
Kadri	Grad Student	100%	$23,559.05

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
Hinman	Principal Investigator	15

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

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_____ No____X_____

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

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:
The evolution and development of microRNAs	☐NIH X Other federal (specify: NSF) ☐ Nonfederal source (specify: _____)	Jan 2010	$720K	$500K
	☐NIH ☐ Other federal (specify:_____ _____) ☐ Nonfederal source (specify: _____)		$	$

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

Yes_____ No_X_____

If yes, please describe your plans:

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

This project will continue under the NSF funding described 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:

	Undergraduate	Masters	Pre-doc	Post-doc
Male				
Female			1	
Unknown				
Total			**1**	

	Undergraduate	Masters	Pre-doc	Post-doc
Hispanic				
Non-Hispanic			1	
Unknown				
Total			**1**	

	Undergraduate	Masters	Pre-doc	Post-doc
White				
Black				
Asian			1	
Other				
Unknown				
Total			**1**	

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.

Research performed provided preliminary data needed to obtain National Science Foundation (NSF) funding.

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:

We have established a collaboration with Takis Benos, PhD (University of Pittsburgh School of Medicine Department of Computational and Systems Biology)

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____X_____ No_____

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

We have initiated an outreach program to enhance molecular and computational biology training for local public high school students. This outreach program will receive funding from NSF.

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 (ß) should not print as boxes (□) and include the appropriate citation(s). DO NOT DELETE THESE INSTRUCTIONS.

Computational and experimental approaches for predicting miRNAs and their targets

I. Computational validation of miRNAs found using homology search in sea urchin embryos.

The first aim of this project was to develop computational tools to predict the presence of miRNA genes using genome sequence. Sea urchins, as deuterostomes, are closely related to vertebrate animals and yet have the advantages offered by an experimentally tractable invertebrate model system. Thus, we used the sea urchin as our model organism for these analyses so that the tools we develop will be specifically suited to identify vertebrate miRNAs. The sea urchin genome sequence and its reverse complement were divided into windows of 500 bases, with an overlap of 150 bases between consecutive windows. These windows were folded using the Vienna package (Hofacker 2003) to obtain the RNA minimum free energy secondary structure for each window. Single looped hairpins were then extracted from these windows, with minimum hairpin length 55 bases, maximum bulge size 25 bases, and having at least 10 base pairs in the stem (Kadri et al. 2009). We previously developed HHMMiR, a hierarchical hidden Markov model (HHMM)-based model to classify miRNA containing stem-loops from random hairpins in the genome. Using cross-validation with human sequences, HHMMiR was able to achieve a high average sensitivity of 84% and specificity of 88%. We obtained ~4.5 million single-loop hairpins in total. A homology search using miRBase version 12 showed that 16 of

these hairpins contained known miRNAs. Table 1 shows a list of these miRNAs in the sea urchin.

Of particular interest, we found a sea urchin ortholog of *miR-71*, a miRNA that has so far been identified only in protostomes and is unknown in vertebrates. We were able to verify this miRNA in sea urchin adult tissue (Fig.1) using the Stratagene High Specificity miRNA qRT-PCR kit. Total RNA was extracted from various embryonic stages and adult tissue of the *S. purpuratus* using the miRVana^{TM} kit from Ambion. We observed the presence of miR-71 transcripts in the sea urchin embryonic stages (0hrs, 24hrs, and 48hrs after fertilization) as well as in the unfertilized egg, confirming the presence of miR-71 in deuterostomes and indicating that, upon closer inspection, it may well be found within vertebrates.

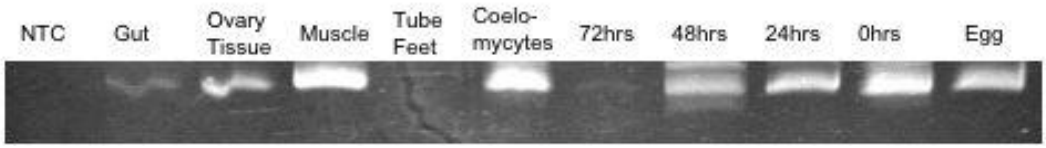

Figure 1: Gel showing presence of *miR-71* in adult tissue and embryos of the sea urchin *S.purpuratus*

We sought to further validate that these predicted miRNAs might be expressed in sea urchin embryos by analyzing transcriptome data. An analysis of expression of sea urchin transcripts in embryos up to late gastrula stage using high density arrays discovered ~55,000 regions, called "enigmatic regions" because they could not be annotated. We used these "enigmatic" non-coding transcripts and extracted hairpins as described above. A Basic Local Alignment Search Tool (BLAST) search was performed with the version 12.0 of the miRNA registry. We got 26 hits for these hairpins, with similarity between the sequences being greater than 75%. The top hits with identity greater than 90% are shown in Table 2.

Table 1: Results of 100% homology search on entire genome.

miRNA	Scaffold	Sequence	Energy of hairpin kcal/mol	Hairpin
spu-miR-7	Scaffold15487	UGGAAGACUAGUGAUUUUGUUGU	-39.1	ACCUCUGGCCCUUUGGAAGACUAGUGAUUUUGUUGUUGUGAAUGUCAAAGAACUAACAAAAAAUCACUGUCGGCCACUGGGGACAGAGGUCCA
spu-miR-9	Scaffold32321	UCUUUGGUUAUCUAGCUGUAUGA	-29.7	GCUUUUCUCUUUGGUUAUCUAGCUGUAUGAGUGAUGUCAAUGCAUCAUAAAGCUAGGUUACCAAAGAUAAGUGC
spu-miR-10a	Scaffold44497	ACCCUGUAGAUCCGAAUUUGU	-45.71	CAUGGAUUCGGCUUUCUUCACUUUUUGGGCUUACAUCAACCCUGUAGAUCCGAAUUUGUGUCCGAGUUUCUCGCCUCUGGUAGUCACAGAUUCGUAUCUCUGGGUAACUGUAUCCCAGGACUUACAACUGCUGGAUGACGAUG
spu-miR-71	**Scaffold1649**	**UGAAAGACAUGGGUAGUGA**	**-32.1**	**GUGUUCCCUGAGUGAAAGACAUGGGUAGUGAGAUUUGACUAUCACAAAACAAUCUCACUAUUCUGUUUUUCCCCUGGUGGAUACUUAUUACAUCUACUA**
spu-miR-153	Scaffold36366	UUGCAUAGUCACAAAAGUGA	-37.7	CCCGAUCAUUUUUGUGUUUAUGCAACUUUUGUAUUUUCACCGAUCAGAGUAGUUGCAUAGUCACAAAAGUGAUUGGGA
spu-miR-92	Scaffold40011	UAUUGCACUUGUCCCGGCCUG	-44.4	UGGUGAUUGAUGGUCAAGUCGGACCGAGCGCAAUGUUGUUCCUCUAUUGAGGUUUUUCGAAUAUUGCACUUGUCCCGGCCUGCUCAUCAACACCAAAA
spu-miR-1	Scaffold4906	UGGAAUGUAAAGAAGUAUGUA	-31.1	UCCCAUGCCAUAAACAUACUUCUUUAGAAUUCCAUACUGAAUCUCCUCAACUCUAUGGAAUGUAAAGAAGUAUGUAUUGAAACUGGGACCUCACAUCACA
spu-miR-31/72	Scaffold56613	AGGCAAGAUGUUGGCAUAGCUG	-38	CUGGUACAAACCAGUGGAGAAAAGGCAAGAUGUUGGCAUAGCUGUGAUUUAAAUAUUAACCCAGCCUGUGUCUUCAUACUGCCAUUUAUUCACUUGGUUGUUCGCCCAGAAUCUU
spu-miR-137	**Scaffold67887**	**UUAUUGCUUGAGAAUACACGUA**	**-36.2**	**GACCUCAUGAUUACAGGUAUUCUCGAGUGAAUAAUACAGAAAAGCCUGAUGUUAUUGCUUGAGAAUACACGUAGUAUACGGGUCACUC**
spu-miR-184	Scaffold74947	UGGACGGAGAACUGAUAAGGGC	-32.93	UAUUUGUCUUUUAUCUUUCCUUAUCAUUCUCUUGCCCGGCCGAAUACUAUGUUAUUGAAAUUGCGCUGGACGGAGAACUGAUAAGGGCUUAUAUAAGACACGUGGCACC
spu-miR-124	Scaffold6417	UAAGGCACGCGGUGAAUGCC	-21	UCUUAGUGUUCACUGUGAUCCUUGGAUUUAUAUUCUAAAACAAUUAAGGCACGCGGUGAAUGCCAAGGCUCGA
spu-miR-125	Scaffold77741	UCCCUGAGACCCUAACUUGUGA	-32.04	CCAAAUCCCUGAGACCCUAACUUGUGAUGUGCUUUUAUCAAAUCACACAGGUUGGUAUCUCAGGAAUUUGGU
spu-miR-466f	**Scaffold87820**	**CAUACACACACACAUACACAC**	**-6.21**	**GUCAAAUCACUAAAGUUGGUUGCAUAAGUGGACACCCCCCCCCCCAUACACACACACAUACACACGCAAAGGAGGACUAUGAGAAG**

Table 2: Results of BLAST on sea urchin transcriptome.

miRNA	Scaffold in genome	Sequence	Identity Percentage	E-value
miR-153	34878	UUGCAUAGUCACAAAAGUGA	100	5.00E-05
miR-292-5p	40147	ACUCAAACUGGGGGCUCUUUUG	91	4.00E-03
miR-14	8155	UCAGUCUUUUUCUCUCUCCUA	90.5	1.30E-02
miR-297	4729	AUGUAUGUGUGCAUGUGCAUG	95	1.10E-02
miR-31/72	37510	AGGCAAGAUGUUGGCAUAGCUGA	91.3	6.40E-02
miR-466b-5p	31259	GAUGUGUGUGUACAUGUACAUG	91	6.00E-05
miR-466c-5p	35028	GAUGUGUGUGUGCAUGUACAUA	91	6.00E-05
miR-466d-5p	31259	UGUGUGUGCGUACAUGUACAUG	95.5	1.00E-03
miR-466h	16346	UGUGUGCAUGUGCUUGUGUGUA	91	4.00E-03
miR-466k	31259	UGUGUGUGUACAUGUACAUGUGA	91.3	2.00E-05
miR-467g	47995	UAUACAUACACACACAUAUAU	95.2	3.00E-03
miR-574	1491	CACGCUCAUGCACACACCCACA	91	4.00E-03
miR-598	50743	UACGUCAUCGUUGUCAUCGUCA	91	6.00E-05
miR-1007	18473	UAAGCUCAAUUAACUGUUUGCA	91	0.004

II. Computational Extension to HHMMiR

HHMMiR acts as a classifier to distinguish miRNA-containing hairpins from random hairpins in the genome. The biggest weakness of conventional HMM is the exponential duration density of a state in a Markov chain. The probability of d consecutive observations in state S_i can be modeled as:

$$p_i(d) = (a_{ii})^{d-1}(1 - a_{ii})$$

where a_{ii} is the self-transition coefficient.

Thus, HHMMiR is unable to decode the miRNA part of the hairpin. We have developed an extension to the Hierarchical Hidden Markov model (HHMM) for miRNA prediction by adding an explicit state duration density to each state of the HMM. This model is called the Hierarchical Explicit State Duration HMM (HESD-HMM). This model allows for efficient decoding of the hairpin into its constituent regions, namely loop, extension, miRNA, and pri-extension. We have developed modified algorithms for this model and tested its performance on limited human data sets. We observed a drastic increase in specificity (~90%) but a less drastic decrease in sensitivity (~60%). The datasets used for these calculations were used in Kadri et al (2009). We also found that introducing explicit duration densities for production states (leaf nodes in the hierarchy of states) reduces accuracy and, thus, are added only for the four internal states representing the four distinct regions described above.

III. Expression of miRNAs within embryos of sea urchin and sea stars

We used Northern blots to verify the presence of mature miRNAs in sea urchin. Total RNA was extracted from embryos and adult tissues using the miRVana kit (Ambion). Standard northern blot protocols were performed using 10-15μg of total RNA and an antisense *miR-31* Starfire™ (IDT) α-P^{32} oligonucleotide-labeled probe. In sea urchin, positive bands of the predicted size for *miR-31* were found in RNA from 12h, 24h, 36h, and 48h embryos (Fig. 2).

Figure 2: Northern blot showing the expression of *miR-31* in sea urchin embryos

III. Validation of computational predictions of miRNAs in sea urchin embryos using high throughput Illumina sequencing

We previously designed a computational prediction algorithm, HHMMiR, for discovery of miRNA from genome sequence (Kadri et al. 2009). This algorithm was used in addition to homology searches (above) with the miRBase registry (Griffiths-Jones 2006) to make genomic predictions for miRNAs in the sea urchin, as noted previously. To verify the predictions made

and discover new miRNAs, if any, we used Illumina sequencing of small RNA libraries in sea urchin embryos.

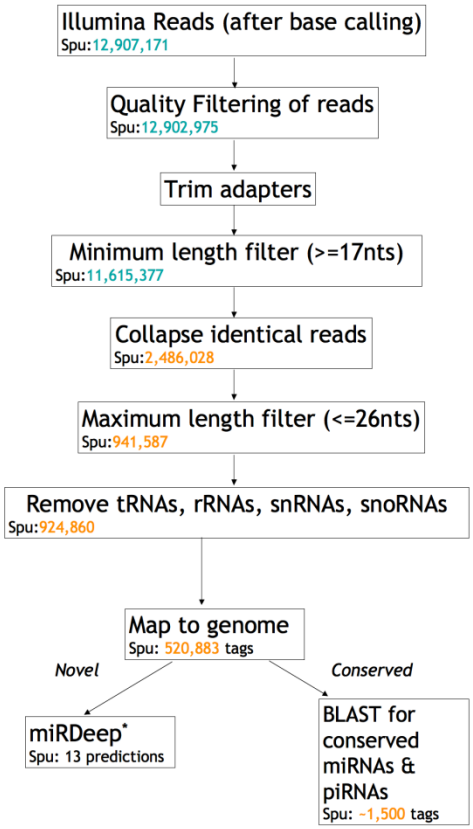

Figure 3: Computational pipeline (Kadri et al., *in preparation*) to search for miRNAs in deep sequencing data

Mixed RNA populations from multiple embryonic stages (1, 2, 3 days post-fertilization) in sea urchin were outsourced to the Wistar Institute, Philadelphia, for small RNA library preparation and sequencing. We developed a computational pipeline (Fig. 3) to analyze the sequenced data and search for miRNAs (Kadri et al., *in preparation*). The sea urchin library returned ~13 million reads, which were then filtered and preprocessed to remove low quality 3' ends and linker sequences. The pre-processed reads were collapsed into "tags" based on sequence identity, and reads mapping to other non-coding RNAs (such as tRNAs, rRNAs, etc.) were removed. We found 44 conserved miRNAs, of which 11 were predicted previously with 100% homology search in the genome (Table 3). The precursor sequences of the five most abundant miRNAs in the sample are shown in Fig. 4. miRDeep (Friedländer et al. 2008) was used to make novel miRNA predictions in these samples. We found 13 potentially novel miRNAs using this method (Fig. 5).

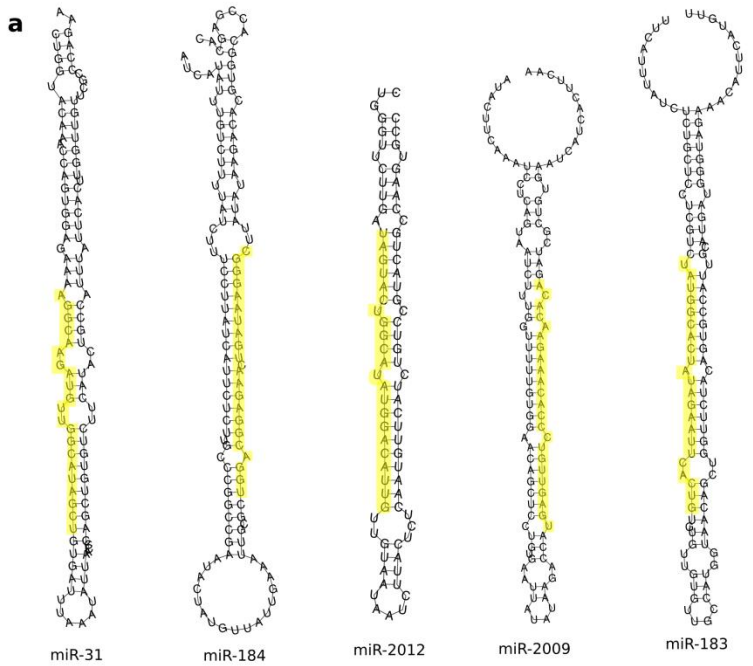

Figure 4: Stem loop structures of precursors of the most abundant sea urchin miRNAs (from the registry)

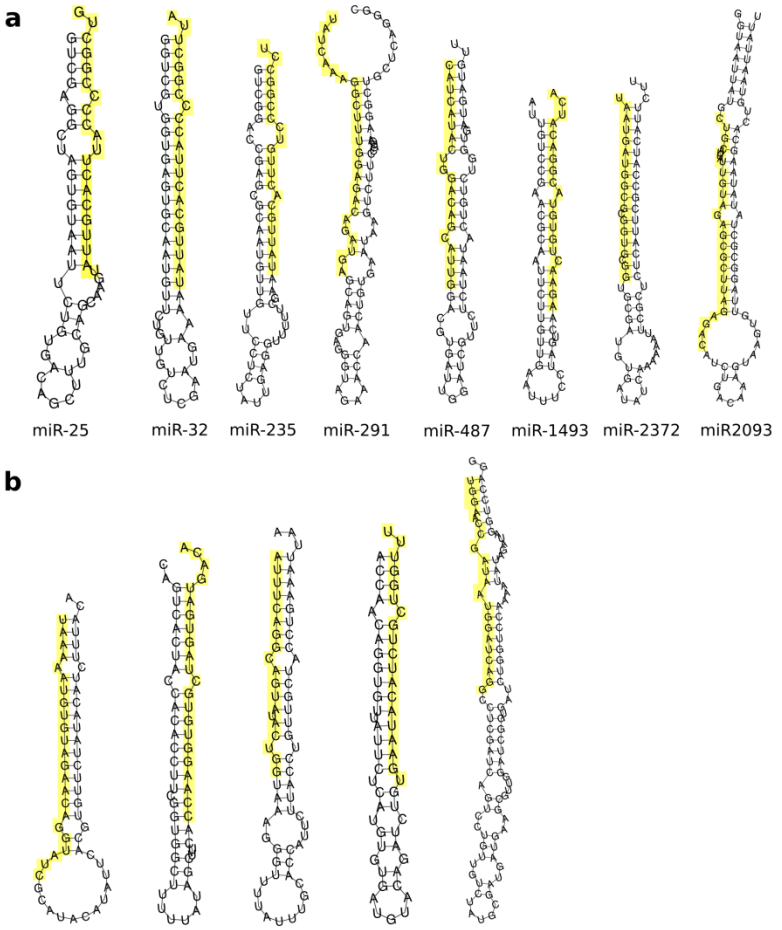

Figure 5: Stem loop structures for 13 novel miRDeep predictions. (a) The eight predictions with seed regions (2-8nts) identical to known miRNAs (b) five completely potentially novel miRNAs.

Table 3: The miRNAs found in small RNA library of sea urchin, with conserved sequences found in the sea urchin registry. High abundance miRNAs are highlighted in yellow.

let-7	miR-1810	miR-202	miR-2011	miR-252a	miR-34	miR-92c
miR-1	miR-182	miR-2003	miR-2012	miR-252b	miR-7	miR-96
miR-10	miR-183	miR-2004	miR-2013	miR-278	miR-71	
miR-125	miR-184	miR-2007	miR-210	miR-29	miR-79	
miR-1280	miR-1937	miR-2008	miR-219	miR-31	miR-9	
miR-137	miR-200	miR-2009	miR-22	miR-31b	miR-92a	
miR-153	miR-2001	miR-2010	miR-242	miR-33	miR-92b	

IV. Computational prediction tool for miRNA target prediction

The ultimate goal of this project is to make better predictions of miRNA-target interactions involved in regulatory gene networks. A number of methods have been developed for predicting targets of miRNAs, including TargetScan (Lewis et al. 2005), RNAhybrid (Rehmsmeier 2004), miRanda (John et al. 2004), and Pita (Kertesz et al. 2007). While miRanda and TargetScan use sequence complementarity and free energies of the RNA:RNA duplex, RNAhybrid makes predictions based entirely on thermodynamics, and PITA uses site accessibility as the main prediction criteria. All of these methods suffer from high false positive rates (Fig. 6). The underlying hypothesis for our improved prediction tool is to study whether the given UTR sequence of a gene is under regulation by miRNA(s). We hoped to combine the prediction power of these weak classifiers into a single strong classifier. We used the supervised learning method, boosting to achieve this goal.

For testing purposes, we used immunoprecipitation (IP) data from *C. elegans* (Zhang et al. 2007). Two proteins from the GW182 family, AIN1 and AIN2, were used for the IP studies. We used 721 mRNA transcripts (558 had available untranslated [UTR] sequence), enriched in both IP datasets, as the positive dataset. Similarly, we used the 51 miRNAs overlapping in both datasets as the miRNAs to be considered for this study. Six hundred-fifty-two mRNAs that were not enriched in both samples (474 of which had UTR sequence available) were used as the negative dataset. On 10-fold cross validation, the sensitivity and false positive rates of prediction were 92.4% and 10.6%, with a standard deviation of 16% and 22%, respectively (Fig. 7). To eliminate background noise from the datasets, we filtered the positive dataset by removing all UTRs that were not predicted as positive by any algorithm (67 were removed). UTRs were removed from the negative dataset if three or more algorithms predicted the UTR as positive (118 were removed). This process reduced the standard deviation of the accuracy statistics drastically. The sensitivity and false positive rates were 90.4 ± 4.8% and 7.7 ± 5%, respectively (Fig. 7).

We hope to perform a more detailed testing of this method and apply it to the sea urchin gene set.

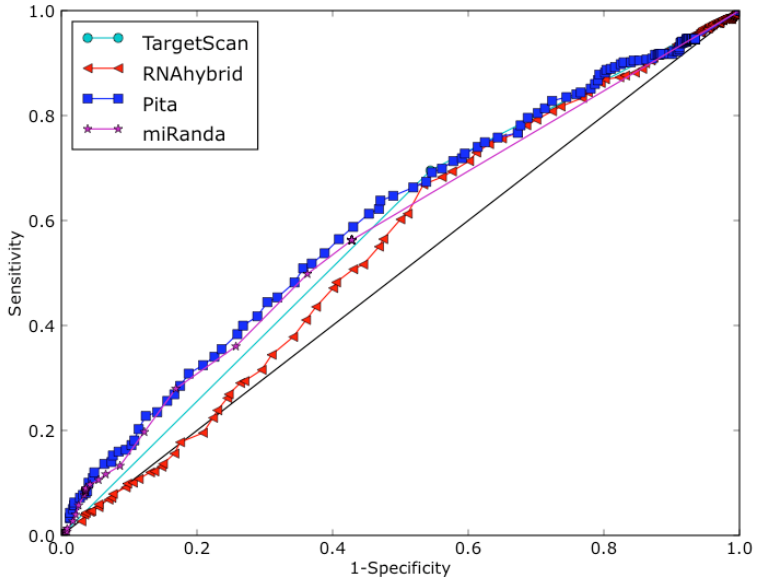

Figure 6: ROC curves for the four known miRNA target prediction algorithms using the *C. elegans* IP dataset

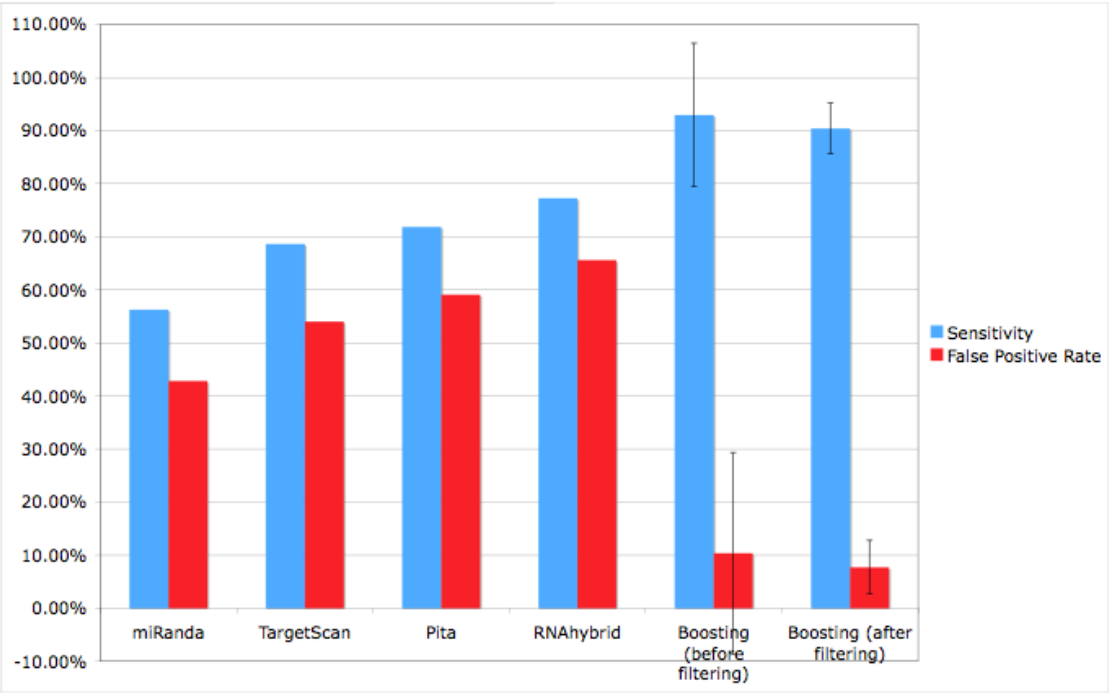

Figure 7: Comparison of performance among the four known prediction algorithms and our boosting tool, which combines the prediction power of all four methods

References:

Friedländer, M.R. et al., 2008. Discovering microRNAs from deep sequencing data using miRDeep. *Nature Biotechnology*, 26(4), 407-415.

Griffiths-Jones, S., 2006. miRBase: the microRNA sequence database. *Methods in Molecular Biology (Clifton, N.J.)*, 342, 129-138.

Hofacker, I.L., 2003. Vienna RNA secondary structure server. *Nucleic Acids Research*, 31(13), 3429-3431.

John, B. et al., 2004. Human MicroRNA Targets. *PLoS Biology*, 2(11), e363.

Kadri, S., Hinman, V. & Benos, P.V., 2009. HHMMiR: efficient de novo prediction of microRNAs using hierarchical hidden Markov models. *BMC Bioinformatics*, 10(Suppl 1), S35.

Kertesz, M. et al., 2007. The role of site accessibility in microRNA target recognition. *Nature Genetics*, 39(10), 1278-1284.

Lewis, B.P., Burge, C.B. & Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1), 15-20.

Rehmsmeier, M., 2004. Fast and effective prediction of microRNA/target duplexes. *RNA*, 10(10), 1507-1517.

Zhang, L. et al., 2007. Systematic identification of C. elegans miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Molecular Cell*, 28(4), 598-613.

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.

Title of Journal Article:	Authors:	Name of Peer-reviewed Publication:	Month and Year Submitted:	Publication Status (check appropriate box below):
1.				☐Submitted ☐Accepted ☐Published
2.				☐Submitted ☐Accepted ☐Published
3.				☐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 plan to submit the following manuscript within the next few weeks: Revealing microRNA populations during development of two evolutionarily divergent echinoderms. Authors: Kadri S, Hinman VF, Benos PV.

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

None

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

We characterized the expression of microRNAs in two organisms for which they had not been previously described. This work led to the discovery of some microRNAs previously unknown from any organism.

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

NAME Veronica Hinman	POSITION TITLE Assistant Professor of Biological Sciences
eRA COMMONS USER NAME veronicah	

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

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	MM/YY	FIELD OF STUDY
University of Queensland, Australia	BEng	08/1990	Mechanical Engineering
University of Queensland, Australia	BS	08/1995	Zoology
University of Queensland, Australia	PhD	08/2000	Zoology
Caltech, Pasadena. CA.	Postdoc	08/2006	Gene regulation, developmental GRNs

A. Positions and Honors

Positions

09/2006-present Assistant Professor, Carnegie Mellon University Department of Biology, Pittsburgh, PA.

09/2007–present Affiliated Faculty, Member of the Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA.

Honors

University of Queensland Medal (1994). University of Queensland. Australia. Awarded for outstanding scholarship throughout undergraduate degree.

Dean's commendation list for outstanding PhD theses (2000). University of Queensland. Australia.

B. Selected Peer-Reviewed Publications

Most relevant

McCauley B.S., Weideman, E. **Hinman V.F.** (**2009**). A conserved gene regulatory network subcircuit drives different developmental fates in the vegetal pole of highly divergent echinoderm embryos. Developmental Biology.

Hinman, V. F., Yankura KA, McCauley BS.**(2009)**. Evolution of gene regulatory network architectures: Examples of subcircuit conservation and plasticity between classes of echinoderms. Biochim Biophys Acta.1789(4):326-32.

Hinman, V.F., and Davidson E.H. **(2007)**. Evolutionary plasticity of developmental gene regulatory network architecture. Proceedings National Academy of Sciences (USA). 104(49):19404-9.

Hinman, V. F., Nguyen A., Davidson E.H. (**2007**). Caught in the evolutionary act: precise *cis*-regulatory basis of difference in organization of gene networks of sea stars and sea urchins. Developmental Biology. 312(2):584-95

Hinman, V. F., Nguyen, A. T., Cameron, R. A., and Davidson, E. H. **(2003)**. Developmental gene regulatory network architecture across 500 MY of echinoderm evolution. *Proceedings of the National Academy of Sciences of the United States of America,* **100**, 13356-13361.

Additional recent publications of importance

Kadri S, **Hinman V**, Benos PV. HHMMiR (**2009**). Efficient de novo prediction of microRNAs using hierarchical hidden Markov models. BMC Bioinformatics. 2009 Jan 30;10 Suppl 1:S35.

Hinman, V. F., and Davidson, E. H. **(2004)**. System-level Properties Revealed by a Gene Regulatory Network Analysis of Pregastrular Specification in Sea Urchins. **In** '*Gastrulation'*. Ed. C. Stern. Cold Spring Harbor Laboratory Press. Cold Spring Habor.

Hinman, V. F., Nguyen, A. T., and Davidson, E. H. **(2003)**. Expression and function of a starfish Otx ortholog, AmOtx: a conserved role for Otx proteins in endoderm development that predates divergence of the eleutherozoa. *Mechanisms of Development* **120,** 1165-1176.

Hinman, V. F., O'Brien, E. K., Richards, G. S., and Degnan, B. M. **(2003).** Expression of anterior hox genes during larval development of the gastropod *Haliotis asinina*. *Evolution and Development* **5,** 508-521.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., **Hinman, V**., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z. J., Schilstra, M. J., Clarke, P. J. C., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L., and Bolouri, H. **(2002)**. A genomic regulatory network for development. *Science* **295,** 1669-1678.

C. Research Support

Ongoing Research Support

National Science Foundation:
PI: Hinman. $500,000. Jan 2009-Jan 2013.
Evolution of Gene Regulatory Networks for Development of Novel Structures in Echinoderm Embryos

National Science Foundation.
PI: Hinman, Co-PI Benos. $500K Aug 2010-July 2014
The role of microRNA regulation in the evolution of development