**SCIENTIFIC REVIEW OFFICER'S NOTES**

**RESUME AND SUMMARY OF DISCUSSION:** The application is for support for a Data Management Resource and Repository for the Extracellular RNA (exRNA) Communication Program, to provide access to exRNA data, protocols and technologies for the consortium. exRNAs are emerging as a novel and significant class of RNAs that may play a role in extracellular communication; thus the development of a central resource and repository for the exRNA community would be highly significant. The investigators are highly qualified, with excellent prior experience in leading large scale consortia and good expertise in exRNA and vesicles. There were minor concerns on the distance between the investigators and lack of a strong track record in working together, and the roles for some of the co-investigators were not well specified. There is some innovation in adapting existing tools and approaches such as Genboree to the exRNA field and the application of BioGPS is also innovative. The proposed analysis pipeline in the DIAC is comprehensive and focused on the unique aspects of exRNA data; however, reviewers noted that plans for analysis by the groups who are generating the data were not well considered. The efforts within the SOC on protocols and community annotation efforts in general are well described and complement the analysis pipeline, although metadata standards and details about ontologies were not sufficiently addressed; however, the investigators have good expertise in this area, which alleviates the concern. In addition, the description of the outreach, especially for the upstream part of the pipeline, was somewhat generic. However after discussion, the strong track record and success of the investigators in other large consortia with similar challenges and the overall strong and detailed plans for the resource mostly overcame the concerns, making overall enthusiasm for the application very high.

**DESCRIPTION (provided by applicant):** Extra-cellular RNAs (exRNAs) are emitted into the human bloodstream and other body fluids by different types of cells in the human body and may be uptaken by other cells. The exRNAs may also originate from edible plants and from microbes that inhabit the human body. The Extracellular RNA Communication Program (ERCP) will explore this newly discovered mechanism of communication in healthy individuals and in pathological conditions such as cancer. The Data Management Resource and Repository for the exRNA Atlas (DMRR) will integrate the efforts of the FRCP and serve as a community-wide resource for the development of the exRNA Atlas database. DMRR will consist of three components and an Administrative Core. The Data Coordination Component (DCC) will develop data and metadata standards, establish data flow into the exRNA Atlas database; develop tools for download, visualization and analysis of exRNA data; and integrate exRNA Atlas database with other relevant resources. The Scientific Outreach Component (SOC) will develop the exRNA Atlas Web Portal to disseminate and provide for visualization of the exRNA Atlas data; ensure accessibility of ERCP-generated resources; and initiate community engagement in exRNA biology using leading biological Wiki sites. In close coordination with the DCC, the SOC will engage the community through knowledge curation jamborees, scientific workshops and symposia. The Data Integration and Analysis Component (DIAC) will provide large-scale integrative and analytic support; evaluate tools and build pipelines to be hosted by DCC and used to populate the exRNA Atlas; build tools to be deployed and distributed by the DCC for use by other consortium participants and the wider scientific community for exRNA data; and lead consortium-wide advanced integrative analyses. Through these coordinated efforts of its DCC, SOC, and DIAC components, the DMRR will help organize the ERCP consortium and open opportunities for rapid progress in the nascent field of exRNA biology.

**PUBLIC HEALTH RELEVANCE:** Cells in different parts of the human body communicate over long distances through the nervous and endocrine systems. There is accumulating evidence about yet another such system of long-range intercellular communication involving RNA molecules. This project will examine this newly discovered system of communication and its role in human diseases. We will focus on the construction of an exRNA Atlas, a chart of extracellular RNA communication in the human body.

**CRITIQUE 1:**

Significance: 3

Investigator(s): 2

Innovation: 4

Approach: 4

Environment: 3

**Overall Impact:** This is very nice albeit largely generic RNA expression-oriented proposal to administer a large program devoted to data collection, management, and analysis of the universe of extracellular RNAs. There is as would be the case from any such generic effort, excellent consideration of meta-information, data provenance, and the issues of many cooks, sources of biological materials, and data consumers. In general all of that is well done and looks to be adequate to the task. However, despite the experience of the SOC leader Dr. Galas with the generation and analysis of exRNAs of different types and from different sources of biological samples, there is a rather striking lack of planning for addressing the technical and experimental challenges associated with sample generation, primary data generation, and data analysis and interpretation approaches that would be critical to move the field forward. This lack of consideration of what it will take to create a useful community resource diminishes enthusiasm for the proposed organizational structure.

**1. Significance:**

**Strengths**

ExRNA is an exciting and potentially high impact direction of research.

There is a strong need for a central source that could provide useful data integration and authoritative material that could help the community advance the field and improve data collection, understanding of the factors associated with technical failures, weakness, and strength, as well as the development of robust analysis approaches that will maximize capabilities of investigators to interpret, hypothesize and validate from the underlying data.

There is expertise among the included groups that will promote the inclusion of strong administrative, statistical, and biological expertise.

**Weaknesses**

The proposal lacks a framework that would facilitate addressing the specific problems associated with exRNAs at both a technical level and at the level of a framework for an atlas. Should there be different atlas types, how could they be combined, how can they be mined as a function of sample types, technological analysis methods, and biological area? How can strong pilot studies be facilitated where these questions can be addressed?

Approaches to categorizing sample and data generating technology/methodologies here are so generic that the end product is not likely to have significance for the field as it should.

**2. Investigator(s):**

**Strengths**

The investigators have broad experience in post-hoc genomic analyses and as a group are stellar.

**Weaknesses**

There is a lack of technical expertise among the administration and analysis groups pertaining to variables associated with optimizing sample collection, factors that affect technical data quality, and different approaches to building highly informative atlases.

There is certainly great expertise if the problems were dealing with a well organized atlas data generation project. However, this does not make up for a lack of expertise surrounding the most critical problems that this field faces.

Much of the investigator support in the overall proposal is designed for doing post-hoc analyses for which the biomedical research community has plenty of available expertise.

**3. Innovation:**

**Strengths**

**-**

**Weaknesses**

Nothing particularly innovative, but the general scheme of efforts is en par with current good practices in large scale genomic project organization.

**4. Approach:**

**Strengths**

Proposed approaches to the problems of the SOC and DCC related to sample and technology-specific profiling analyses, the data provenance of all of that, and the effects of this on downstream data analyses are reasonably well considered and do represent a formidable challenge that must be solved.

The re-use of technology that was developed for conventional RNA profiling studies is good.

Proposed approaches to providing users with access to quality data sets with full annotation of where it came from and how it was generated looks good.

**Weaknesses**

The DCC has taken a good first look into both the data categories, technical approach categories, and current datasets. This is fine, but it has not as convincingly shaped the proposed Center’s ability to focus systems, tools, and user interfaces around this so as to convince one that a highly significant work product will emerge. Looks like there will be a lot of extraneous work related to biological network analyses, when in fact the real questions will be revolving around the issue of whether the data is worth anything. Thus the balance seems way off and this diminishes enthusiasm.

The large proposed efforts here encompass many different directions with respect to data display and follow-through, but without the focus on technical analysis that is necessary to make these kinds of analyses and data from the atlas valuable for the community. Namely, support for rapid, comparative, and integrative evaluation of different approaches to sample and data generation that would allow meaningful atlases to be created that could in turn instruct the field(s).

**5. Environment:**

**Strengths**

Lots of great bioinformatics expertise and tools are available between Baylor, Yale, San Diego, and San Francisco.

**Weaknesses**

**-**

**Budget and Period of Support:**

Recommend as requested

**CRITIQUE 2:**

Significance: 2

Investigator(s): 1

Innovation: 2

Approach: 1

Environment: 1

**Overall Impact:** The proposal draws on successful experiences of the PIs in other large-scale projects to build an informatics resource for exRNA research. If all of this were to come to pass, and with the efficacy that is desired in the program, it would be a highly impactful suite of tools and approaches for exRNA collaborative research.

**1. Significance:**

**Strengths**

This application addresses needs for data management, access, analysis, community interactions, and other aspects of collaborative research into exRNAs, a significant new research area.

**Weaknesses**

None noted.

**2. Investigator(s):**

**Strengths**

This is an outstanding team. Drs. Milosavljevic and Gerstein have led other consortium type informatics efforts for the Roadmap epigenomics and ENCODE projects, respectively. Dr. Galas has been a leader in a variety of genomics activities from the start of the Human Genome Project.

**Weaknesses**

None noted.

**3. Innovation:**

**Strengths**

A large number of tools and approaches that are being developed or are in use in other projects are being brought to bear on the extracellular RNA program. This is innovative in the sense that many of these approaches from other data-intensive fields have not been applied to genetic data so far. In other cases, where there is a genetic research connection, this would be the first application to research that is heavily sequence based.

**Weaknesses**

The use of existing approaches and tools is not a weakness per se, but it is not purely innovative but rather a re-use of existing tools. For this reason the score is slightly reduced.

**4. Approach:**

**Strengths**

The project draws on the successful experiences from the Epigenomics and ENCODE projects, although some tools were also part of the early TCGA project. Other existing resources such as Gene Wiki, WikiPathways, and Cytoscape are also prominently featured. Thus there is a well-vetted framework to build on, which makes for a strong approach.

**Weaknesses**

None noted

**5. Environment:**

**Strengths**

The environments are excellent as they are already supporting a number of such projects as noted above.

**Weaknesses**

None noted

**Protections for Human Subjects:**

Not Applicable (No Human Subjects)

**Inclusion of Women, Minorities and Children:**

G4A - Gender Unknown, Acceptable

M4A - Minority Representation Unknown, Acceptable

C4A - Children Representation Unknown, Acceptable

**Vertebrate Animals:**

Not Applicable (No Vertebrate Animals)

**Biohazards:**

Not Applicable (No Biohazards)

**Applications from Foreign Organizations:**

Justified

**Resource Sharing Plans:**

Acceptable

**Budget and Period of Support:**

Recommend as requested

**CRITIQUE 3:**

Significance: 3

Investigator(s): 3

Innovation: 2

Approach: 2

Environment: 1

**Overall Impact:** Milosavljevic and colleagues propose a multi-institute led data management and resource repository for the exRNA atlas. PIs are Milosavljevic (Baylor), Galas (PNRI), Gerstein (Yale), Mathivanan (LaTrobe), Pico (Gladstone), Su (Scripps). The large number of PIs and participating institutes lead to a significant spreading of resources. The DIAS shows a clear expertise and understanding of the unique aspects of exRNA and clearly anticipates several challenges that will need to be addressed with exRNA. It is clear that simply placing RNA-seq or 1000G tools on top of an exRNA seq data will not lead to unexpected advances by default. Rather a concerted understanding of knowledge is used as a prior to designing the data resource and management tools.

**1. Significance:**

**Strengths**

This proposal is significant in the focus on exRNA specific issues and shows clear consideration of issues that are of key concerns to exRNA community members. It is clear that simply placing RNA-seq or 1000G tools on top of exRNA seq data will not lead to unexpected advances by default. Rather a concerted understanding of knowledge is used as a prior to designing the data resource and management tools

**Weaknesses**

**-**

**2. Investigator(s):**

**Strengths**

The team of PIs has experience related to both large-scale consortium style projects where coordination is needed across a set of teams, as well as specific experience in the area of epigenetics and exRNA. The team is recognized across their fields for their leadership.

Key PIs have a strong history of working together

**Weaknesses**

The large number of PIs and participating institutes lead to a significant spreading of resources, and some of the PIs lack significant commitment of effort or effort of their teams creating some concern of their ability to commit to the project.

**3. Innovation:**

**Strengths**

The project builds from architecture (Genboree system) that is conceivably capable of carrying out many goals needed by the exRNA atlas.

The project is focused on the unique aspects of exRNAs particularly their frequently small size creating unique analysis opportunities.

The investigators pay particular attention to allelic diversity and polymorphisms, and specifically address how they impact their analysis.

**Weaknesses**

**-**

**4. Approach:**

**Strengths**

The DIAC seems very focused on the unique aspects of analysis of exRNA data, recognizing their differences from traditional RNA-seq analysis.

The DCC is focused on standards and implementation within the Genboree system. The experience with the Epigenomics projects will likely be highly relevant.

Significant effort is placed on databases with APIs that will be accessible to other core members.

**Weaknesses**

**-**

**5. Environment:**

**Strengths**

The individual teams environments are largely some of the strongest available. In particular the Baylor and Yale infrastructures and support system are remarkable, lending one to believe that they will be beneficial and not a hindrance towards the success of the projects.

**Weaknesses**

**-**

**Protections for Human Subjects:**

Acceptable Risks and/or Adequate Protections

**Inclusion of Women, Minorities and Children:**

G4A - Gender Unknown, Acceptable

M4A - Minority Representation Unknown, Acceptable

C4A - Children Representation Unknown, Acceptable

**Vertebrate Animals:**

Not Applicable (No Vertebrate Animals)

**Biohazards:**

Not Applicable (No Biohazards)

**Applications from Foreign Organizations:**

Not Applicable (No Foreign Organizations)

**Select Agents:**

Not Applicable (No Select Agents)

**Resource Sharing Plans:**

Acceptable

**Budget and Period of Support:**

Recommend as Requested

**CRITIQUE 4:**

Significance: 4

Investigator(s): 3

Innovation: 2

Approach: 4

Environment: 3

**Overall Impact:** A major strength of the proposal is the combined expertise and experience of the investigators relevant to the goals of the DMRR. The integrative approach to make use of established Wikipedia like resources covering RNA and vesicles is innovative. Using these as a basis for an Atlas and a partnership with the journal, Gene, are strengths. Also use of the established Genboree system is appropriate. Although metadata standards were identified as critical not much detail was provided on their usage. The community annotation (SOC) and comprehensive RNA analysis pipeline (DIAC) plans are impressive but getting ERCP investigators to invest their data and participate in a timely manner could be challenging. Having leadership roles of exRNA and vesicle experts in the SOC should significantly help.

**1. Significance:**

**Strengths**

The focus of the SOC on RNA annotation and protocols is a strength. The involvement of Dr. Galas (as SOC PI) and Dr. Mathivanan in the SOC is a strength as having their exRNA and vesicle expertise will be very useful for understanding ERCP and community needs and building community involvement.

**Weaknesses**

Although the investigators have documented experience with data standards, they have not really addressed metadata standards. If successful, much information will be collected and distributed through the Genboree and Wiki systems but details of how ontologies will be selected and applied were not provided.

**2. Investigator(s):**

**Strengths**

The assembled group of investigators brings an impressive array of expertise and established resources to bear on creating the DMRR.

**Weaknesses**

Although individually strong, the investigators do not have a track record of working together. There is a statement in the Administrative Core of “three experienced PIs who have collaborated previously” but a PubMED search did not reveal joint papers aside from a many author paper on ChIP-seq guidelines with Gerstein and Milosavljevic.

**3. Innovation:**

**Strengths**

The application of BioGPS developed as a Gene Atlas integrating different resources to being an extracellular RNA Atlas is a strength. It is likely to be challenging to get community involvement in annotation but Dr. Su and his group are best positioned to do it.

**Weaknesses**

**-**

**4. Approach:**

**Strengths**

The distinction between endogenous and exogenous exRNAs is addressed well in the proposal as well as mechanisms to capture, analyze, and visualize these types of RNA.

Use of structured spreadsheet systems such as ISA to collect metadata is a strength.

The incorporation of linked data and semantic web technologies to access knowledgebases such as miRBase is a strength.

The DIAC pipeline is very powerful and comprehensive for analyzing RNA-seq data. It serves as a nice complement to the community annotation approach for populating the exRNA Atlas. Getting investigators to submit their data for this purpose before they do their own analyses could be challenging so some mechanism of authorship should be considered.

**Weaknesses**

While the crowd sourcing aspect for providing annotation through the wikis is innovative for the exRNA Atlas, details about how the editors will be recruited are missing aside from inviting domain experts.

It is not clear that ERCP investigators will have the expertise to make use of SPARQL endpoints but getting other relevant RNA biology resources to do so (as described for internal data exchange with WikiPathways) would be a major plus.

Although the importance of proteins and lipids in vesicles is acknowledged there wasn’t much about how data on those molecules if generated by the ERCP would be handled. May not be an issue with the focus on RNA but could be a problem.

**5. Environment:**

**Strengths**

Individually the resources at the different institutions are excellent

**Weaknesses**

Ability to coordinate and communicate over so many different locations and time zones a concern

**Protections for Human Subjects:**

Acceptable Risks and/or Adequate Protections

**Inclusion of Women, Minorities and Children:**

G4A - Gender Unknown, Acceptable

M4A - Minority Representation Unknown, Acceptable

C4A - Children Representation Unknown, Acceptable

**Vertebrate Animals:**

Not Applicable (No Vertebrate Animals)

**Biohazards:**

Not Applicable (No Biohazards)

**Applications from Foreign Organizations:**

Justified

The group from LaTrobe, Autralia brings important expertise and resources on extracellular vesicles highly relevant to the DMRR

**Select Agents:**

Not Applicable (No Select Agents)

**Resource Sharing Plans:**

Acceptable

**Budget and Period of Support:**

Recommend as Requested