State of the Hub 2017

Adam Labadorf, PhD
labadorf@bu.edu

Director, BU Bioinformatics Hub
http://bubhub.bumc.bu.edu/

November 22, 2017

Executive Summary

The mission of the Hub from its inception has been to increase the bioinformatics profile at BU in four ways: through collaboration, education and outreach, mentorship, and as a research conduit. In its first 18 months, the Hub has achieved specific accomplishments in each of these areas. Productive collaborations between the Hub and investigators across multiple academic departments have been established. These collaborations were funded in large part by both internal and external grants that would likely not otherwise have been awarded without the expertise provided by the Hub. Seven master's students have trained with the Hub, one of whom has now graduated and been hired by a biotechnology company in San Diego that was looking for someone with exactly the skillset provided by the Hub. Over half of the master's students have continued to work on their projects past the initial 400 hour minimum, or have been invited to do so by the collaborating PI. The Hub was asked to aid in the mentorship of four PhD students from within and without the Bioinformatics Program by serving as either an official mentor or on thesis advisory committees. The Hub led the development and deployment effort for a set of computational skills workshops for first year PhD students, with plans to expand the materials and offer them to a greater audience in the future. When the Hub could not provide help to researchers due to capacity or skillset, connections were made to groups around BU like the Bioinformatics Applications group at IS&T and the Biostatistics consulting group. Overall, the activities of the Hub have led to direct, demonstrable impact on the BU biological community in all of its stated domains.

At least four current Hub projects have already produced significant scientific findings that will soon be submitted for publication. Gene fusion events were discovered by our analysis of osteosarcoma cell lines with Rachel Flynn, and a very interesting set of differentially expressed genes was discovered and is being evaluated for mechanistic and biomarker potential. Important insights into processes related to neurodegeneration in humans and epitranscriptomics in zebrafish have been discovered. Novel algorithms for analyzing small and large RNA-Seq datasets are currently under development, contributing both methodological and biological innovation in multiple contexts. The Hub team is creating a software package that combines many common operations for working with and manipulating counts matrices for use in pipelines and custom software, for which no other dedicated packages are available, that will be published in the form of an application note when it becomes fully mature. Every Hub project employs a strategy for ensuring replicability and transparency that we plan to describe in a short publication. In a way that is emblematic of the interdisciplinary nature of bioinformatics, the activities of the Hub span a broad range of areas from scientific software engineering tools and approaches to algorithm development to biological discovery.

The Hub has played a critical role in 10 new grant applications, 6 of which were funded and the remaining are under review. The Hub's participation in these grants varied, but almost all include some amount of salary support, most often for Hub students. Adam was listed as key personnel, co-PI, or PI on all of the grants listed for providing his expertise in high throughput sequence analysis. Newly awarded grants have or will provide 10% of Adam's support and $\$14,800$ for Hub student salary, with more than $\$16,800$ additional support provided directly from existing collaborators sources. Overall, the grants awarded with collaboration of the Hub amount to a total of $\$3.2M$ in combined direct costs. Current Hub projects leading to publication are certain to generate new NIH grant proposals both with collaborators and for internal Hub projects.

Despite this strong showing of success, challenges remain. The Hub was immediately brought to maximum capacity with projects and students, where many new potential collaborators had to be turned away due to lack of bandwidth. Without additional PhD level scientists, the Hub cannot expand to meet the clear need for its work. Identifying consistent sources of funding is the primary challenge to address this, and an appropriate financial model for the Hub has been challenging to devise. In addition to the financial challenges, the lack of faculty appointments for active Hub scientists presents an unfortunate credibility issue, in spite of the clear track record of success thus far. These critical issues must be addressed before the Hub can be established as a viable entity in pursuit of its mission.

Accomplishments

- November 2017 - Five posters accepted to GSI Research Symposium; Eve Byington won a top poster presentation prize for her work "Discovery of Topologically Associated Domains using RNA-Seq Datasets in Human Brain".

- October 2017 - Awarded Alzheimer's Disease Center Pilot Grant (via National Institute of Aging) to study Chronic Traumatic Encephalopathy post mortem human brains, in collaboration with Thor Stein of the ADC. Includes support for a Hub student.
- October 2017 - Valentina Perissi received an 8th percentile score on an R01 that lists Adam as key personnel with salary support for mentorship of her PhD student Joey Orofino in the design and execution of ChIP-Seq and RNA-Seq analysis for their project *Regulation of mitochondrial homeostasis through retrograde signaling*
- August 2017 to present - Led effort for development and delivery of a Computational Skills Workshop series to first year Bioinformatics PhD students in collaboration with Gary Benson and current Bioinformatics PhD students. http://foundations-in-computational-skills.readthedocs.io
- June 2017 to August 2017 - *2'-O-methylation epitranscriptomics in zebrafish development* with Daniel Cifuentes, Biochemistry Dept. Project completed by Kylie Shen.
- June 2017 - Invited to speak about the Hub at the Biochemistry Department Retreat
- June 2017 - With Daniel Cifuentes, awarded GSI Seed Grant entitled *Decoding the epitranscriptomic role of 2'-O-methylation during vertebrate embryogenesis* with support for a Hub student.
- May 2017 - Recruited Eve Byington, Kylie Shen, and Lingyu Zhou to work on Hub projects with Daniel Cifuentes, Sabine Schnyder, and two internal Hub projects
- April 2017 - Organized Bioinformatics master's Program Alumni Panel. Invited Chet Birger, David Jiang, and Jalini Rajapakse to come discuss their experiences post master's and answer questions from current students.
- February 2017 - *Epi-transcriptome m6A Detection by 2OMe-Seq in Zebrafish Development.* Completed pilot project with Daniel Cifuentes analyzing small RNA-Seq data.
- January 2017 - Recruited Filisia Agus, Diego Crespo, Angel Dai, and Ryan Ingram to work on Hub projects with Rick Myers and Rachel Flynn.
- January 2017 - With Rachel Flynn, awarded CTSI Pilot Grant entitled *Using Gene Signatures to Predict ALT Status in Human Cancer* with support for a Hub student.
- July 2016 to November 2016 - Developed projects in collaboration with multiple investigators in several different departments.
- July 2016 to present - Developed software infrastructure and methodological strategies for replicable analysis.

Works In Progress

Ongoing Projects

- *Alternative Lengthening of Telomeres (ALT) in osteosarcoma* with Rachel Flynn, Pharmacology Dept. Project work performed by Angel Dai. We have identified novel gene fusion events in a common osteosarcoma cell line. Differential expression analysis identified a set of consistently perturbed genes across all cell lines that implicate novel genes and processes that consistent with known factors thought to contribute to the ALT pathway. Analysis of patient derived xenograft osteosarcoma tumors grown up in a murine host is currently underway to investigate whether the DE signature from the cell lines has predictive power in human tumors.
- *A novel RNA-Seq analysis method for comparing genetically and clinically distinct S. pneumoniae strains* with Sabine Schnyder, Pediatric Infectious Disease, Department of Medicine. Project work performed by Lingyu Zhou. Novel methodological development is underway to better understand how to relate RNA-Seq abundance measurements between SP strains with related but distinct genomic sequences. There are no publications currently available treating this problem, and the bioinformatics and statistical aspects of analysis are challenging.
- *IsomiR Analysis in Parkinson's Disease Brain and Cerebrospinal Fluid (CSF)* with Rick Myers. Project work performed by Ryan Ingram. We are developing a methodology for comparing the IsomiR profile detected in small RNA-Seq datasets between conditions. There is remarkable diversity in the number of IsomiRs across known miRNAs. We have determined that the IsomiR profile abundance is very consistent between brain and CSF, suggesting measuring small RNAs in the CSF offers insight into what is happening in the brain. This consistency is irrespective of disease or control status.

- *Discovery of Topologically Associated Domains using RNA-Seq Datasets in Human Brain*, internal Hub project. Project work performed by Eve Byington. Our novel methodology examines windowed correlation between abundance measurements the GTEx brain RNA-Seq datasets using a combination of bioinformatic, mathematical, and statistical techniques. We have determined that TAD structure can indeed be predicted using correlation structure across genomic bins using RNA-Seq data by examining overlap with experimentally determined TAD regions. Current work is in building a size-invariant predictive model to identify correlated regions that are consistent with the true positive TAD regions matching this profile.

- *Asymptomatic caudate nucleus gene expression resembles prefrontal cortex in Huntington's Disease*, with Rick Myers. Project work performed by Filisia Agus and Diego Crespo. This study seeks to assess whether the brains of asymptomatic HD gene positive individuals show similar gene expression activity in the caudate nucleus (CAU), the primarily affected brain region in HD, as is observed in symptomatic HD cortex (BA9), when compared with neurologically normal controls. We have encouraging evidence that the gene expression profile in CAU does indeed reflect what is observed in BA9. Current work is in comparing the CAU and BA9 of asymptomatic HD individuals directly.

- *Regulation of mitochondrial homeostasis through retrograde signaling* The Hub is serving a mentorship role of Joey Orofino, a PhD in Valentina Perissi's lab. Joey graduated from the Bioinformatics master's program prior to becoming a PhD student, and has included significant bioinformatics techniques including RNA-Seq and ChIP-Seq experiments into his thesis. During his thesis, Adam has agreed to serve as a mentor to Joey in guiding these analyses. This mentorship was established when Valentina included 10% of Adam's salary support on a recent RO1 proposal which, though funding has not been confirmed, was scored 8th percentile in October 2017. Mentorship of Joey is ongoing.

- *Graph-based ab initio small RNA reference using miRNA-Seq datasets*, internal Hub project. Project work performed by Adam. Through our IsomiR work, it is clear that the currently accepted annotation (miRBase) does not capture the full diversity of miRNA-related transcripts observed in humans. It is known that miRNAs may undergo a number of non-templated sequence alterations by established enzymatic mechanisms, particularly on their 3' ends, resulting in miRNA sequences that are functional but do not match any reference genome exactly. Current miRNA-Seq analysis pipelines rely on alignment against a reference genome and often involve filtering on exact annotation coordinates from miRBase. We have observed that many reads do not align exactly to known annotated miRNA sites, and that there are IsomiRs with higher abundance than those mapping to annotated coordinates exactly. This work explores the idea of using sequence graph methodologies to build an ab initio small RNA reference that can be used to better understand and identify important miRNA species without the biases introduced by using a single reference sequence. Current work is in developing the sequence graph methodology, and in devising a graph-mapping strategy for aligning reads to these graphs.

- `de_toolkit` - a software tool that consolidates many mechanical and diagnostic operations on counts matrices from, e.g. RNA-Seq data, used in all current analysis projects. Application note publication planned.
  http://de-toolkit.readthedocs.io

Hub Organization

- *Establishment of the Hub as an official organization* Met with the heads of groups with potential 'customers' around BU: David Harris (Biochemistry), David Center (CTSI), Katya Ravid (Medicine), Darrell Kotton and George Murphy (CReM), Kim McCall (Biology). The goal was to identify the Bioinformatics needs of these departments as well as generate enthusiasm for the Hub model. Several of these contacts have offered to write letters of support for establishing the Hub as part of the proposal to BU. Current work is meeting with more department heads Avi Spira (CBM), David Farb (Pharmacology), and Lindsay Ferrer (Genetics), to qualify and quantify the current and projected Bioinformatics needs of their organizations.

Education and Training

- *Development of Bioinformatics master's Curriculum* As part of the overall Hub mission to train master's students in bioinformatics techniques, Adam has recruited five Bioinformatics PhD students to develop a new course, BF528 Applications in Translational Bioinformatics, that focuses on training students in the methods and skills needed to perform analysis of high throughput genomic datasets in real-world contexts. The course follows the proven case-study model, where teams of students with a spectrum of backgrounds work together to implement and replicate analyses from published studies that made these data publicly available. The course is targeted to master's students but is suitable for any student with an interest in and some exposure to computational skills. The course is currently under development and will be offered in SP18.

- *RNA-Seq Pipeline Tutorial Materials* A set of tutorials that guide Hub students, and eventually others, through the technologies and tools needed for RNA-Seq analysis is currently under development. Current and past Hub students have contributed their knowledge to the materials incrementally, as they become facile with the technologies through their experiences in projects. The development of these materials is still in-progress. https://bitbucket.org/bubioinformaticshub/bubhub_mrnaseq_primer

- *Computational Skills Workshops* The computational skills workshops we developed and presented to the first year Bioinformatics PhD students were very well received. The materials cover many of the fundamental concepts required to effectively implement computational projects, but more advanced topics remain to be covered. In particular, there is a growing recognition that replicability and transparency of analysis is a critical aspect of biological studies that has any computational component. One final workshop is currently under development that covers computational workflow software like Snakemake and Nextflow. The ultimate aim of developing these materials is to offer fee-based workshop sessions. The current materials can be found below. http://foundations-in-computational-skills.readthedocs.io

Publications

These manuscripts are in preparation:

- Filisia Agus*, Diego Crespo*, Richard H. Myers, Adam Labadorf. *Asymptomatic caudate nucleus gene expression resembles prefrontal cortex in Huntington's Disease*.

- Ryan Ingram, Jeanne C. Latourelle, Richard H Myers, Adam Labadorf. *miRNA Profiling and IsomiR Analysis in Parkinson's Disease*

- Eve Byington, Adam Labadorf. *Discovery of Topologically Associated Domains using RNA-Seq Datasets in Human Brain*

Grant Opportunities and Support

All of the grants below listed Dr. Labadorf as either PI, co-investigator, or key personnel.

Awarded

Regulation of mitochondrial homeostasis through retrograde signaling
NIH R01
Award Number: PA-16-160
PI: Dr. Valentina Perissi, Dept. of Biochemistry, BUSM
Award amount: $3,117,108
Hub Support: 10% salary support for Dr. Labadorf
Scored 8th percentile, funding to be confirmed in January 2018

Chronic Traumatic Encephalopathy and the Pan-Neurodegenerative Disease Phenotype
BU-ADC Pilot Grant Program, with review and concurrence by the NIA
PI: Dr. Adam Labadorf, Dept. of Neurology, BUSM
Award amount: $25,000
Hub Support: $6,400 salary for Hub intern

Using Gene Signatures to Predict ALT Status in Human Cancer
BU-CTSI Integrated Pilot Grant Program
NIH Award Number: 1UL1TR001430
Award amount: $20,000
PI: Dr. Rachel Flynn, Dept. of Pharmacology, BUSM
Hub Support: $6,400 salary for Hub intern Angel Dai

Regulation of mitochondrial biogenesis during muscle differentiation
BU-CTSI Integrated Pilot Grant Program
NIH Award Number: 1UL1TR001430

Award amount: $20,000
PI: Dr. Valentina Perissi, Dept. of Biochemistry, BUSM
Hub Support: Mentorship of Joey Orofino

Crosstalk Between Histone Ubiquintination and Methylation for the Regulation of Mitochondrial Gene Expression
Genome Science Institute Pilot Grant Program
Award amount: $20,000
PI: Dr. Valentina Perissi, Dept. of Biochemistry, BUSM
Hub Support: Mentorship of Joey Orofino

Decoding the epitranscriptomic role of 2'-O-methylation during vertebrate embryogenesis
Genome Science Institute Pilot Grant Program
Award amount: $20,000
PI: Dr. Daniel Cifuentes, Dept. of Biochemistry, BUSM
Hub Support: $2,000 salary for Hub intern Kylie Shen

Under Review

Transcriptomic and Phosphorpoteomic Analysis of Synaptotoxicity in Alzheimer's and Prion Diseases NIH R21
PI: Dr. David Harris
Award Amount: TBD
Submitted 10/16/2017
Hub Support: Dr. Labadorf salary support % TBD

Molecular-anatomic mapping of AD-related changes in white matter neuronal populations NIH R21
PI: Dr. Kathleen Rockland
Award Amount: TBD
Submitted 10/16/2017
Hub Support: Dr. Labadorf salary support 5%

The neuropathology of mild traumatic brain injury in Alzheimer's disease.
VHA/VA Merit Award Clinical Science Research and Development Program (CSR&D)
PI: Dr. Thor Stein
10/01/2018-9/30/2022
Award amount: $649,911
The goal of this project is to understand and characterize how mild traumatic brain injury (mTBI) influences the pathologies of Alzheimer's disease (AD) and Chronic Traumatic Encephalopathy (CTE). Hub Support: 1.2 Calendar Months for Dr. Labadorf, $6400+ for Hub intern

Regulation of the hippocampal epigenetic DNA methylation clock by choline nutrition in a mouse model of Alzheimer's disease
NIA NIH R21
Award amount: $275,000
PI: Dr. Jan Krzystof Blusztajn, Dept. of Pathology & Laboratory Medicine, BUSM
Hub Support: 1.2 calendar month support for Dr. Labadorf, 2 x $6,400 salary for Hub intern

Challenges

Project Timeline Projections and Load The primary challenge in the operation of the Hub in its first 18 months was managing the project load. As anticipated, the demand for collaborations immediately exceeded the Hub's capacity, even before the Hub was widely advertised. The six projects initiated by the Hub in the first year required substantial time and effort from a management stand point, both in the mentoring of Hub students performing the work and in managing relationships with the collaborating PI. The Hub projects were scoped appropriately for 400 hours of work each, but variability in the number of hours each student could work per week made the projects extend in calendar time further than anticipated. Consequently, projects that were anticipated to have wrapped up by the end of the summer did not do so and extended into the fall semester, making the

combined load of managing projects and teaching burdensome. The Hub began using new collaboration tools (basecamp.com) to help make communication and project management more efficient, which was immensely helpful in this regard. Even so, mentoring seven students concurrently, spread across as many projects, in addition to all of the other Hub activities is too much to manage at once. Reducing the number of projects completed in a year threatens the financial viability of the Hub model.

CRC and BUMC Have Different Needs In conversations with labs on CRC, it was clear that the bioinformatics needs for the two campuses differ. BUMC labs tend to use high throughput sequencing data in more canonical ways, e.g. differential expression with RNA-Seq, or transcriptional regulatory patterns with ChIP-Seq. CRC labs, on the other hand, tend to have more 'boutique' analysis needs at present, e.g. designing custom probes on protein binding arrays, or analyzing and interpreting poorly annotated fungal genomes. While these are clearly bioinformatics problems, the custom nature of the analysis needs in these investigations poses challenges to the Hub model, requiring more time and custom code than their BUMC counterparts. Nonetheless, the availability of bioinformatics expertise on CRC is just as limited as on BUMC, and some labs do not have the persistent need for bioinformatics to justify hiring a full time grad student or bioinformatician. Thus, the same fundamental need remains on both campuses. While the Hub is motivated to address these needs on CRC as well, the present capacity and skillset of the Hub is not specialized to these kinds of projects. Additional diversified expertise is needed to expand the Hub to operate on CRC. The Hub was contacted by Juan Fuxman-Bass in the Biology Department to provide mentorship on a project involving genomics data. This is perhaps a sign that the needs of CRC are changing, but at present this is a major area for growth and development for the Hub.

Funding Model While the Hub has attracted substantial support to fund the students who perform the work, this amount is far less than is sufficient to support a scientist level salary. While mentoring master's students in the completion of analysis projects has many upsides, there are significant complementary downsides to the approach; namely, projects completed by master's students require more mentorship time, more analysis time as they are learning, and they generate less income for the Hub. The alternative is to instead have the scientists themselves perform the majority of project work, which would enable more projects to be completed within the calendar year and would have direct salary contribution. Since mentorship is a primary mission of the Hub, this poses a challenge to the viability of the Hub model and is a significant hurdle to making the Hub self-sustaining.

Appointment Status The long term viability of the Hub requires that the scientists have faculty appointments. As a scientific entity that makes intellectual contributions in its own right, without such an appointment, the Hub does not have sufficient academic credibility from the perspective of many collaborating PIs or the institution as a whole. This obviously poses a problem for the entire Hub model, as faculty members hired into departments in the traditional way are expected to bring their own specific research program with them. Requiring Hub scientists to both collaborate on a large number of projects as well as maintain full independence with their own research amounts to nothing more than a traditional PI position, which does not meet the fundamental need that exists, and indeed is why the Hub was created in the first place. In such an environment, Hub scientists must either compromise the collaborative mission of the Hub in pursuit of their own independence, or have reliable support from some other mechanism. If this challenge cannot be solved, the Hub as a concept and as an organization will fail.

Next Objectives

The first and critical next objective is to identify a way to organize and fund the Hub as a sustaining organization. This includes identifying sources of funding and finding a solution to the problem of appointment status. If this can be accomplished, the next objectives will follow in parallel:

- Establish a presence on CRC through new collaborations
- Devise and implement a plan in collaboration with the Bioinformatics Program to organize computational workshops for the BU community and beyond, with the goal of fulfilling both the Hub mission of outreach as well as generating revenue from outside the institution
- Continue development of methodologies for the analysis of small RNA-Seq data, TAD discovery with RNA-Seq data, and bacterial RNA-Seq expression analysis to provide the basis for writing grants to further develop these ideas as internal Hub projects
- Publish software tools and communicate the Hub strategies for replicable analysis via publication to increase Hub visibility and scientific footprint

Personnel

Director: Dr. Adam Labadorf, PhD

Master's Students:

- Filisia Agus - Asymptomatic HD, with Rick Myers (1/2017-present)
- Eve Byington - TAD Detection with RNA-Seq Data, internal Hub project (5/2017-present)
- Diego Crespo - Asymptomatic HD, with Rick Myers(1/2017-present)
- Angel Dai - Alternative Lengthening of Telomeres, with Rachel Flynn (5/2017-present)
- Ryan Ingram - miRNA-Seq analysis in Parkinson's Disease, with Rick Myers (1/2017-present)
- Kylie Shen - 2-0-methylation in zebrafish, with Daniel Cifuentes (5/2017-10/2017)
- Lingyu Zhou - mRNA-Seq analysis in closely related Strep Pnemo strains, with Sabine Schnyder (5/2017-present)

PhD Students (advisement/mentorship):

- Joey Orofino - Mechanisms of GPS2 in adipose and muscle tissue, student of Valentina Perissi (Biochemistry)
- Jason Lui - Transcriptional and regulatory mechanisms of YAP/TAZ in skin cancer, student of Deborah Lang (Dermatology)

Project Summaries

A novel RNA-Seq analysis method for comparing genetically and clinically distinct S. pneumoniae strains
Lingyu Zhou, Sabine Schnyder, Stephen Pelton, Adam Labadorf

Streptococcus pneumoniae (SP) is a potentially deadly prokaryotic human pathogen. The virulence of individual strains, evaluated using antibody binding affinity, shows large variability within the same multi-locus sequence type (MLST), but the specific causes of this variability can be difficult to ascertain. The SP genome is highly dynamic due to its high capacity for horizontal gene transfer and genome rearrangement, resulting in variation of homologous gene positions and sequences that make comparative genomics challenging. This work describes a framework that clusters homologous genes from different genomes together for comparative genomics and RNAseq analysis. The framework is applied to a dataset of paired genomic sequence and RNA-Seq generated using twelve clinical SP isolates from six strains that show differentiated binding affinity with the host complement cascade system, a critical aspect of the adaptive host immune response. Despite having the same MLST and capsule type, three of the strains exhibit high complement binding affinity, thus resulting in less severe disease, while the other three strains exhibit relatively lower binding. Using our novel method, we identified genes with RNA abundance measurements that are associated with complement binding status. These genes may be implicated in differential complement binding and are therefore potentially clinically relevant. To our knowledge, the analytical framework presented here is the first to directly compare RNA-Seq data from closely related but genetically distinct strains, and is sufficiently general that it can be employed in other experiments with similar data.

Comparative analysis of prefrontal cortex and caudate nucleus in post-mortem Huntington's Disease brains by RNA-Seq
Filisia Agus*, Diego Crespo*, Richard H Myers, Adam Labadorf

Huntington's Disease (HD) is a devastating neurodegenerative disease caused by an expanded trinucleotide CAG repeat in the HTT gene. The caudate nucleus (CAU) is the primarily affected brain region in HD, which is massively degenerated in late stage disease, but other brain regions including Brodmann Area 9 (BA9) are relatively unaffected. While studying CAU directly in post-mortem human brains is desirable, the lack of neurons in highly degenerated tissue makes comparisons with unaffected brains problematic. CAU samples from post-mortem human brains of asymptomatic HD gene positive (HDpos) individuals, who died before any significant degeneration has occurred, avoid this difficulty but unfortunately are exceedingly rare. The Myers lab has obtained both BA9 and CAU tissues from a small number of HDpos brains. Previously, the Myers lab has performed whole transcriptome analysis with high throughput sequencing (RNASeq) in BA9 of HD and control individuals, and identified key pathways associated with disease. The HDpos samples provide an opportunity to evaluate whether the changes detected in BA9 are reflective of the disease process in CAU. This study presents an RNASeq analysis of BA9 and CAU tissues from 2 HDpos individuals in comparison with a larger set of HD BA9 samples previously studied. A custom differential expression analysis pipeline that includes a stringent gene filtering process to attain the highest confidence comparison was developed to interrogate the concordance between HDpos and HD BA9. In parallel, BA9 and CAU RNASeq from healthy individuals in the GTEx dataset were filtered and processed using a pipeline similar to the original HDpos and HD BA9 set. This formed the basis of comparison between these two brain regions to identify genes that were consistently perturbed between these brain regions within the same HDpos individuals. The results show

encouraging concordance in gene expression differences in HDpos BA9 and CAU when compared to HD BA9, suggesting the study of BA9 is valuable in understanding the disease process in the most affected region of the brain. Differentially expressed genes in the GTEx dataset have also proven useful when analyzing the differentially expressed genes in the HDpos set.

miRNA Profiles and IsomiR Analysis in Parkinson's Disease

Ryan Ingram, Jeanne Latourelle, Richard H Myers, Adam Labadorf

Recent studies in whole-genome miRNA expression (miRNA-Seq) in Parkinson's disease (PD) human post-mortem brain tissue have shown coherent differences between PD and unaffected brain. However, brain tissue is not a viable tissue source for developing biomarkers in living subjects, and no predictive small RNA-based biomarkers have been identified in peripheral fluid to date. Cerebrospinal fluid (CSF) presents an intriguing option for the collection of miRNA samples as it can be performed in an outpatient procedure. The goal of this study is to compare the miRNA profiles of prefrontal cortex tissue (BA9) with CSF to test whether miRNA abundance differences in CSF accurately mirror those observed in the brain. Small RNA sequencing of paired BA9 and CSF tissue in 32 PD and 9 control (C) individuals was performed and analyzed with a custom miRNA-Seq pipeline to identify differentially expressed miRNAs between PD and C, as well as assess the concordance between BA9 and CSF. Results suggest low concordance of differential expression, with only miR-133a-5p detected as differentially expressed in both BA9 and CSF (p-values: 0.011 and 0.014, respectively). In addition to the differential expression analysis, the sequencing data was also analyzed to identify the presence of templated miRNA isoforms (isomiRs) using a novel bioinformatic quantification method. This analysis sought to identify miRNA reads that do not map exactly to known miRBase annotations, with the hypothesis that there are some non-canonical miRNA species that are more abundant than the annotated form and that may also be predictive of disease status. Preliminary results show that many miRNAs exhibit evidence of isomiRs within PD BA9 and CSF, with some miRNAs showing a larger abundance of alternative isomiRs than the precise miRBase annotations. These results suggest that isomiRs may provide a new source of diagnostic biomarkers in PD.

RNA-Seq analysis identifies Alternative Lengthening of Telomeres pathway gene expression signature

Angel Dai, Emily Mason-Osann, Adam Labadorf, Rachel Flynn

The Alternative Lengthening of Telomere (ALT) pathway is a non-telomerase dependent mechanism for telomere elongation, which is required for cellular immortality and a hallmark of cancer cells. The ALT pathway is estimated to be active in 10-15% of all human cancers, and is substantially more common in certain cancer types, including osteosarcoma. The specific mechanisms underlying the ALT pathway are incompletely understood. The Flynn lab has in its possession the most complete compendium of osteosarcoma cell lines available, including 8 verified to be ALT positive (ALT+) and 5 as ALT negative (ALT-). To better characterize the ALT gene expression profile, high throughput RNA sequencing datasets (RNA-Seq) were generated from these cell lines to identify a differentially expressed gene signature that can be used to discriminate between ALT+ and ALT- cell types. A state of the art RNA-Seq analysis pipeline was used to find differentially expressed genes using Firth's logistic regression. Out of 17626 consistently abundant genes, 1017 genes were significantly differentially expressed at FDR ¡ 0.01 between the two cell type conditions. Gene set enrichment analysis using the MsigDB gene set database identified pathways that strongly implicate Histone Cluster Family Member genes as consistently perturbed between cell types. In particular, HIST1H3C is down regulated in the ALT+ cell lines, suggesting that the incorporation of histones into the telomere ends may be compromised in cells lengthening the telomeres via the ALT pathway. In addition to the cell lines, the Flynn lab also has also performed RNA-Seq on a set of patient derived xenograft (PDX) osteosarcoma tumor samples with unknown ALT status. The differential expression profile discovered in the cell line analysis will be used to examine the PDX samples for evidence of ALT activity.

Discovery of Topologically Associating Domains using RNA-Seq datasets in human brain

Eve Byington, Adam Labadorf

Topologically associating domains (TADs) are a recently discovered feature of chromatin architecture. They consist of mega-base scale linear segments of DNA that form loops, essentially "chromosomal neighborhoods" of genetic expression. Disruption of TAD boundaries have been shown to cause disease and physiological malformations — thus, they play a critical role in human health and development. TADs are typically discovered using Hi-C, a chromatin conformation capture technique which is an expensive and laborious process. Here, we present a methodology that can identify TADs using only mRNA-Seq data by employing a genome-binning and windowed correlation strategy across a group of samples. Using a previously published set of TADs established by Hi-C in a human developmental model of neurons and the RNA-Seq datasets from human brain available from GTEx, some areas of highly correlated RNA expression across the genome were concordant with known TAD boundaries. A TAD width-invariant quadratic curve fitting methodology followed by k-means clustering successfully discovered TAD-like patterns genome wide. These correlated regions were used to build a model for identifying putative TADs using RNA-Seq data alone. When applied to a set of RNA-Seq datasets generated from Huntington's Disease

brains compared with neurologically normal controls, the methodology identified entire chromosomal regions with perturbed correlation structure, suggesting large scale disruption in chromatin architecture may be implicated in Huntington's Disease.