University of Pennsylvania

Annual Progress Report: 2011 Formula Grant

Reporting Period

July 1, 2013 – June 30, 2014

Formula Grant Overview

The University of Pennsylvania received $7,809,060 in formula funds for the grant award period January 1, 2012 through December 31, 2015. Accomplishments for the reporting period are described below.

<u>Research Project 1: Project Title and Purpose</u>

Cognitive Training for Nicotine Dependence – The purpose of this project is to evaluate cognitive enhancing interventions for smoking cessation. The focus will be on an alpha4beta2 nicotinic receptor partial agonist medication.

Duration of Project

4/1/2012 – 6/3/2014

Project Overview

The broad objective is to develop better treatments for nicotine dependence. To achieve this objective, we aim to develop novel approaches to treatment that target executive cognitive function (i.e., working memory, sustained attention, and behavioral inhibition), and evaluate the downstream effects on smoking behaviors.

<u>Aim 1:</u> To evaluate the effects of ABT-089 when administered as 40 mg oral once daily dose for 10 days, compared to placebo, on abstinence-induced cognitive deficits and smoking behavior in smokers. The following domains will be assessed: working memory (N-back Task; Word Recognition Task); sustained attention (Continuous Performance Task), and response inhibition (Stop Signal Reaction Time Task).
<u>H1:</u> Healthy adult smokers receiving daily 40 mg doses of ABT-089 for 10 days will show reductions in abstinence-induced cognitive deficits, and have a greater likelihood of maintaining 3 days abstinence from smoking.

<u>Aim 2:</u> To test the effects of ABT-089 when administered as 40 mg oral once daily dose for 10 days, compared to placebo, on: (a) subjective withdrawal symptoms as assessed by the Minnesota Nicotine Withdrawal Scale (MNWS) and the Brief Questionnaire on Smoking Urges,

respectively; (b) the subjective reward value of a cigarette following brief abstinence as measured by the Cigarette Evaluation Scale (CES); (c) positive and negative affect as measured using the Positive and Negative Affect Schedule (PANAS); and (d) inattention and hyperactive/impulsive symptoms as measured by the ADHD self-report checklist.
H2: Healthy adult smokers receiving daily 40 mg doses of ABT-089 for 10 days will show reductions in subjective withdrawal symptoms and smoking reward, compared to when receiving placebo.

Principal Investigator

Caryn Lerman, PhD
Professor of Psychiatry
University of Pennsylvania
Center for Interdisciplinary Research on Nicotine Addiction
3535 Market Street, Suite 4100
Philadelphia, PA 19104

Other Participating Researchers

Allison Gold, PhD; Rebecca Ashare, PhD – employed by the University of Pennsylvania

Expected Research Outcomes and Benefits

This research will determine the efficacy a novel pharmacologic treatment for nicotine dependence. If found to be efficacious, this treatment could be developed further in collaboration with industry for eventual FDA approval.

Summary of Research Completed

Recruitment was initiated in April, 2013 as scheduled. However, recruitment was discontinued on December 13, 2014 after the results of an interim futility analysis on the first 10 study completers did not produce any signals for efficacy.

We originally proposed to enroll up to 30 eligible smokers who met final eligibility at a rate of ~3-4 per month over a ~9-12 month period. The target was to have up to 25 subjects complete the study. As noted above, an interim futility analysis did not produce evidence for efficacy for any of the study endpoints. Therefore, recruitment was discontinued on December 13, 2014. Subjects already enrolled who had not completed were given the opportunity to complete the study. Last dose, last subject was on January 13, 2014. The expenses for completing this intensive study were higher than anticipated and the budget has been expended on the work completed thus far.

At the conclusion of the recruitment period, 53 subjects were enrolled (i.e., signed consent) and 17 of these subjects are included in the Intent-to-Treat (ITT) sample. A subject was considered ITT if they successfully completed the Baseline/Period 1 (P1) Medication Pick-Up Visit and

received their P1 Study Medication. Of the 17 ITT subjects, 13 completed the entire study.

Table 1 (below) includes the values for the outcome measures in the active medication and placebo periods. There was no effect of ABT-089 on number of days of self-reported abstinence during the monitoring period (1.2 days in each period out of 4 days; $\beta = -0.39$, 95% CI -2.1 to 1.4, $p = .66$). Nor were there any medication effects on subjective measures after 24 hours of mandatory abstinence that preceded the 4-day quit attempt: craving ($\beta = -0.84$, $p = .85$), withdrawal ($\beta = 0.23$, $p = .88$), and mood (positive affect: $\beta = -0.85$, $p = .66$; negative affect: $\beta = -0.38$, $p = .63$). There was no medication effect on working memory accuracy ($\beta = 5.4x10^{-17}$, $p = 1.0$) or response time ($\beta = -0.84$, $p = .74$) after 24 hours of mandatory abstinence, nor were there any medication by memory load interactions ($ps > .6$).

With respect to the lapse cigarette after the mandatory abstinence period, participants rated the lapse cigarette slightly less satisfying (CES satisfaction subscale) while taking ABT-089 compared to placebo ($\beta = -.58$, 95% CI -1.09 to -0.07, $p = .026$); however, this effect would not survive correction for multiple hypothesis testing. There was no effect on the psychological relief or toxicity subscales, or on the overall cigarette evaluation ($ps > .4$).

The overall incidence of self-reported adverse events was similar for ABT-089 versus placebo. Participants endorsed a mean of 7.9 events (SD 8.5) during ABT-089 treatment versus 9.3 (SD 9.1) during placebo ($\beta = -1.4$, $p = .57$). The most commonly endorsed adverse events (ABT-089 > placebo) were: agitation (61.5% vs. 53.8%); insomnia (23.1% vs. 15.4%); and diarrhea (15.4% vs. 7.7%). No participants discontinued treatment due to adverse events. Medication compliance did not differ by treatment period (90.9% for ABT-089, 89.5% for placebo, $p = .17$).

In conclusion, we found no signals for potential efficacy of ABT-089 on ability to maintain abstinence for 4 days, or on craving, withdrawal, mood or working memory performance during the prior mandatory abstinence period. Although we did detect a slight reduction in satisfaction ratings following the programmed lapse cigarette, this effect would not survive correction for multiple testing, and is difficult to interpret in light of the absence of effects on other clinical measures.

Table 1. Primary outcome measures

Measure	Testing day (ABT-089 period)	Testing day (placebo period)
Number of days abstinent during monitoring period	1.2 (1.7)	1.2 (1.8)
Withdrawal	6.8 (6.1)	6.6 (6.7)
Craving	37.4 (17.2)	38.3 (15.8)
Positive affect	25.4 (9.2)	26.2 (9.6)
Negative affect	13.2 (3.6)	13.5 (4.6)
Subjective effects of lapse cigarette	2.8 (0.78)	2.9 (0.76)
N-back number correct	47.5 (8.4)	47.5 (8.4)
N-back correct response time (ms)	646.5 (107.9)	638.9 (101.7)

Table 1 Legend: Mean (+/-SD) values for the primary outcome measures during each medication period.

Research Project 2: Project Title and Purpose

Mesothelin Chimeric Antigen Receptor Lentiviral Vector Production and IND Development – The overall hypothesis of this project is that immuno-gene therapy with chimeric antigen receptor (CAR) T cells with specificity for the tumor antigen mesothelin can be used successfully to treat cancers that overexpress mesothelin: most commonly, pancreatic cancer, ovarian cancer and mesothelioma. In order to advance to clinical trials to test the safety and feasibility of mesothelin CAR T cell therapy, a number of translational steps need to be met in the aspects of manufacturing and regulatory. This research project will enable the manufacturing of the mesothelin CAR lentiviral vector that will be used to transduce T cells, and facilitate development of an investigational new drug (IND) application for submission to the FDA.

Duration of Project

1/1/2012 – 6/30/2013

Summary of Research Completed

This project ended during a prior state fiscal year. For additional information, please refer to the Commonwealth Universal Research Enhancement C.U.R.E. Annual Reports on the Department's Tobacco Settlement/Act 77 web page at http://www.health.state.pa.us/cure.

Research Project 3: Project Title and Purpose

Identification of Moderate Penetrance Alleles in Young Women with Breast Cancer: A Multiplex Approach – The use of genetic testing to guide cancer risk management improves patient survival. However, most individuals and families in whom genetic susceptibility is suspected

have uninformative results using current approaches for clinical genetic testing. As one such example, 10% women diagnosed with breast cancer are under age 40. Whereas a genetic contribution to the development of such early onset cancer is suspected, at most 10% of such women will have a detectable mutation in *BRCA1* or *BRCA2*. However, large scale genetic studies of early onset breast cancer have not been done examining many of the recently identified moderate risk breast cancer susceptibility genes.

Anticipated Duration of Project

1/1/2012 – 12/31/2015

Project Overview

Genetic testing for high penetrance mutations, such as in BRCA1/2, and resulting changes in medical management, have improved morbidity and mortality in mutation carriers. However, we are not able to identify mutations in many high risk individuals and families. A specific example of such a lack is in women diagnosed with breast cancer under age 40, in whom currently we are only able to identify a heritable cause in 10%. Research efforts have revealed that mutations in moderate penetrance genes, such as *PalB2*, *CHEK2* and *PTEN*, also are associated with elevated risks of breast cancer of approximately 40%. While mutations in each gene are assumed to be infrequent (found in 1-2% of women), no study has been done to assess multiple breast cancer susceptibility genes simultaneously in women with early onset breast cancer to determine their overall frequency and contribution to breast cancer. We hypothesize that the genetic contribution to early onset breast cancer is underestimated, and that over 10% of women will have an identifiable mutation associated with cancer susceptibility other than *BRCA1/2.* Of even more critical importance, due to our lack of knowledge about their frequency and because the mutations themselves are incompletely penetrant, medical management guidelines for women carrying mutations in moderate risk genes have not yet been established. The development of a next generation sequencing approach, necessary to assess all of the risk genes concurrently, will lead to more accurate risk assessment and prediction in women with early onset breast cancer and has the potential for great impact on prevention and early detection.

Specific Aims:
Aim 1: To develop a targeted next generation sequencing array of moderate and high penetrance breast cancer susceptibility genes

Aim 2: To generate a robust set of data on controls and patients of Caucasian and African American ethnicity with early onset breast cancer

Aim 2A: To determine the rate of mutations in high and moderate risk breast cancer susceptibility genes in 250 women with early onset breast cancer

Aim 2B: To develop a data set and algorithms for mutation detection for potential use in the clinical molecular diagnostics laboratory

Principal Investigator

Katherine L. Nathanson, MD
Associate Professor
University of Pennsylvania
351 BRB 2/3
421 Curie Blvd.
Philadelphia, PA 19104

Other Participating Researchers

Jennifer Morrissette, PhD; Nandita Mitra, PhD; Susan M Domchek, MD – employed by the University of Pennsylvania

Expected Research Outcomes and Benefits

Gaining a more complete picture of the genetic component of early onset breast cancer, particularly in African-American women, will allow us to appropriately target screening and prevention strategies for these women and their relatives. In this proposal, at least 30 genes will be simultaneously assessed for deleterious mutations. Mutations in these genes have been associated with cancer risk, but each individually has been seen quite rarely. The strength of this proposal is to perform testing in parallel, not sequentially, with the use of a multiplex system to provide us with a large volume of information. Given the large number of women who will be screened, we will have a much better understanding than has been possible to date of the total contribution of moderate and high penetrance genes to early onset breast cancer. In addition, we will gain a better understanding of the other cancer risks associated with many of these genes. For example, mutations in *PalB2* have been associated with pancreatic cancer and *RAD51D* with ovarian cancer, but the studies demonstrating this have done in limited patient populations. In summary, this project will advance the field significantly by 1) demonstrating the ability to perform multiplex testing of a large number of cancer susceptibility genes, 2) provide an estimate of the contribution of mutations in these genes to early onset breast cancer and 3) advance the knowledge of the risks associated with these genes to better inform prevention and early detection.

Summary of Research Completed:

Aim 1 and 2:

<u>Patient identification:</u> All patients diagnosed under age 40 are referred for genetic counseling and testing, regardless of family history. Eligibility criteria for the study were: 1) diagnosis of breast cancer under age 40; 2) negative *BRCA1/2* sequencing in a CLIA-approved laboratory; and 3) negative personal or family history of ovarian cancer. Women were ascertained from academic and community hospital sites within the Penn Cancer Network (www.penncancer.org/patients/penn-cancer-network) and The Karmanos Cancer Institute at Wayne State University (details in Supplementary Methods). Targeted massively parallel

sequencing was completed in 278 women with early onset breast cancer. Analysis for *BRCA1/2* large genomic rearrangements was not required, although negative clinical testing was available for 28% of patients. The clinical characteristics of the women included in the study are reported in Table 1. In brief, the majority were white/Caucasian (70%) with black/African American patients comprising 24% of the overall group. Of the women, 70% had a family history of breast cancer. Most women (92%) had invasive breast cancer, usually ER+ (69%) with a wide variation in stage of disease.

Massively parallel sequencing capture and DNA library preparation: For this project, we decided to use Agilent SureSelect Technology for the capture baits. The baits were designed to contain the entire gene plus 10kb on either side, for 3x coverage across the regions. We are using a custom designed targeted massively parallel sequencing panel covering all coding exons and the flanking 10 base pairs of the following 22 genes (20 study genes plus *BRCA1* and *BRCA2*): 1) high penetrance breast cancer susceptibility genes (*CDH1, PTEN*, *STK11, TP53*); 2) genes known to cause other cancer susceptibility (*CDKN2A, MLH1, MSH2*, *MSH6, PMS2*); 3) genes known or postulated to be moderate penetrance cancer susceptibility genes (*ATM*, *BARD1*, *BRIP1*, *CHEK2, FAM175A, MRE11A, NBN, RAD50, PALB2, RAD51C*); and 4) *MUTYH*, which leads to autosomal recessive polyposis.

For each patient, constitutional DNA was sheared to fragments of 150-200 base pairs, blunt ended and ligated with adaptors-embedded indexes. DNA quality, fragment size and concentration were measured with an Agilent 2100 Bioanalyzer. DNA libraries of sufficient quality were pooled pre-capture to 24-plex and hybridized to an Agilent SureSelect custom target library. The captured fragments underwent paired-end sequencing using read lengths of 100 base pairs on an Illumina HiSeq at the Next Generation Sequencing Core of the University of Pennsylvania. Samples were sequenced to a mean coverage of 224X. Three samples were removed from the analysis for having >10% of targets with 0% coverage or <50% of targets with >10X coverage.

In order to determine the efficiency and accuracy of our sequencing platform and bioinformatics pipeline, we analyzed samples with variants identified by clinical sequencing in *BRCA1*, *BRCA2*, *MSH2*, or *PALB2*; these included two nonsense mutations, four indels, two large genomic rearrangements, and 34 single nucleotide variants. 100% of the 42 known variants were identified.

Data analysis: Sequencing data were analyzed with a custom bioinformatics pipeline to identify all single nucleotide variants, small and medium sized insertion/deletions (indels) and large genomic rearrangements. Variants were called into a five-tiered classification scheme: Deleterious (D), Likely Deleterious (LD), Variants of Uncertain Significance (VUS), Likely Benign, and Benign. All D/LD variants were confirmed with Sanger sequencing using a separate stock aliquot of the patient's DNA sample from the aliquot used for MPS.

Statistical analysis: Statistical comparisons between data groups were made using a two-tailed Fisher's exact test or a two-tailed, type 2 Student's t-test. ESD (extreme studentized deviate) method was used to identify and remove outlier values from the data sets.

<u>Results:</u> Of the 278 patients, 169 (61%) had at least one variant found at <0.1% allele frequency in control populations (1000G and ESP6500). After variant classification, 86 patients (31%) were found to have at least one deleterious or likely deleterious (D/LD) variant or VUS (Figure 1). Thirty-one patients (11%) overall were identified to carry a total of 34 D/LD variants, 54 patients (19%) had 57 VUSs, and six patients (2.2%) had seven *MUTYH* variants.

Seven patients were identified to have D/LD variants in a high penetrance cancer susceptibility gene (Figure 1) which are clinically actionable. Four patients were found to carry a known pathogenic *TP53* mutation. Two patients, including one African American, were found to carry LD variants in *TP53*. One patient was identified to have a large genomic rearrangement deleting exon 5 of *MSH2*, leading to an in-frame deletion of 65 amino acids of the DNA interacting domain of MSH2. A patient with a history of both early-onset BC and sarcoma was found to carry a known pathogenic missense mutation in *CDKN2A*. Finally, one patient, with a personal history of early onset colon cancer and two primary breast cancers, was found to be a compound heterozygote for a known pathogenic mutation and an LD variant in *MUTYH*.

Twenty-four patients were found to have D/LD variants in genes in which mutations have been associated with a moderate risk of breast cancer. The majority of D/LD variants in moderate penetrance genes were found in *ATM* and *CHEK2*. Single D/LD variants were found in *ATM* in seven patients and in *CHEK2* in nine patients. One patient was found to carry D variants in both *ATM* and *CHEK2*; of note both variants also were found in her brother with early onset prostate cancer. In addition, one patient was found to carry two LD variants *in trans* in *CHEK2*. The remaining six patients had D variants in *MRE11A* (2), *BARD1* (1), *BRIP1* (1), *NBN* (1), and *RAD50* (1). Twenty-seven patients carried a VUS in a high penetrance cancer susceptibility gene, and three of those patients also had a D/LD variant. Nine patients were found to have a single VUS in *BRCA1* or *BRCA2*, three patients in *TP53* and 12 patients in *MLH1*, *MSH2*, *MSH6*, or *PMS2*; no VUSs were found in *CDH1*, *CDKN2A*, *STK11* or *PTEN*. Three additional patients each carried two VUSs in a high penetrance cancer susceptibility gene. Twenty-six patients were found to have VUSs in moderate penetrance cancer susceptibility genes, *ATM*, *BRIP1*, *CHEK2*, *FAM175A*, *MRE11A*, *NBN*, *PALB2*, *RAD50*, and *RAD51C*; no VUSs were found in *BARD1*. Finally, six patients carried a single D/LD variant or VUS in *MUTYH* (Figure 1). Three patients were heterozygous for the same known pathogenic *MUTYH* mutation and three were heterozygous for VUSs in *MUTYH*.

The proportion of patients identified to have a clinically reportable variant varied by race, such that 26% of self-reported white patients were found to have at least one reportable variant versus 37% of non-white patients (p=0.05). The proportion of patients with a D/LD variant did not vary significantly between white and non-white patients (13% versus 6%, p=0.36). The proportion of non-white patients found to carry a VUS was statistically significantly higher than the proportion of white patients, 31% versus 13% (p=0.003). Of the 66 African Americans, 7.5% carried a D/LD variant, which was not statistically significantly different than the proportion of white patients. Of the 27 Ashkenazi Jewish individuals, 22% were found to have a D/LD variant, compared with 9% of the 262 non-Ashkenazi Jewish individuals (p=0.08).

In comparison to D/LD variant negative patients, there was a statistically significant increase in

the rate of second primary malignancies (excluding non-melanoma skin cancers, Table 1, 19% vs 6%, p=0.02) in the D/LD variant positive patients. In addition, there was a trend towards a higher rate of a bilineal family history of breast cancer in D/LD variant positive versus negative patients (23% vs 11%, p=0.08). The Penn II *BRCA1/2* prior probability score were statistically significantly higher (27% vs 19%, p=0.04) in D/LD variant positive patients, as was the BOADICEA score (29% vs 14%, p=0.005).

Only three of the 22 patients with D/LD variants had ER- invasive breast cancer (Table 1, 14%), one had triple negative breast cancer (*BARD1* S551X) and two had ER- Her2+ breast cancer (*TP53* P151T and *CHEK2* c.444+1A>G). In contrast, 33% of the patients with no D/LD variant (+/- a VUS) had ER+ invasive breast cancer (p=0.09). Seven of the 20 patients (35%) with a D/LD variant had Her2+ breast cancer versus 26% of the patients with no D/LD variant (+/- a VUS, p=NS). Finally, D/LD variants were found in 13% of the 23 patients with DCIS, 11% of the 116 patients with node positive invasive cancer, and 11% of the 130 patients with node negative invasive breast cancer. The stage distribution was similar between D/LD variant positive versus negative patients.

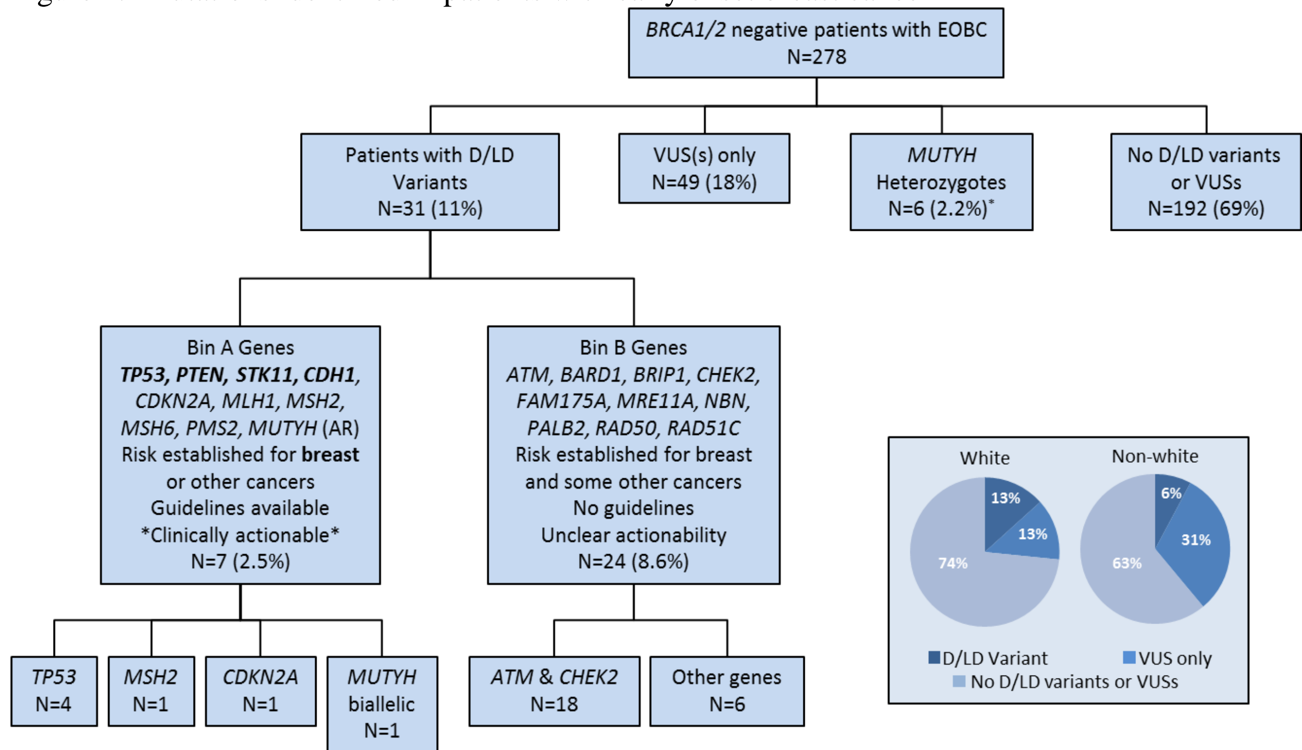
These data demonstrate that massively parallel sequencing identifies reportable variants in known cancer susceptibility genes in over 30% of patients with early onset breast cancer. However, only rare patients (2.5%) have definitively actionable mutations given current clinical management guidelines. Large-scale cooperative group studies are needed to determine the clinical utility of multiplex panel testing in patients with early onset breast cancer.

Development of data set and algorithms for mutation detection for potential use in the clinical molecular diagnostic laboratory: We are in the process of developing a Variant Classification Tool, which is a combined automated/user input Excel-based method of variant classification based on the American College of Medical Genetics (ACMG) guidelines. The tool uses multiple parameters to determine whether or not a missense variant (nonsynonymous) is pathogenic or not. The tool has several tabs including 1) Fields tab describing the variables used for ACMG classification; 2) GeneInfo with the characteristics of genes for which variants will be classified; 3) VarClass, the working variant classification tab; 4) CodesTable, a lookup table for the ACMG code calls in the VarClass tab and 5) FamTable, a lookup table for the samples in the study. The user determines the gene level characteristics and fills out the GeneInfo tab. ANNOVAR (http://www.openbioinformatics.org/annovar/) , an efficient software tool to utilize update-to-date information to functionally annotate genetic variants detected from diverse genomes (including human genome hg18, hg19, as well as mouse, worm, fly, yeast and many others), is used to fill out the VarClass tab. This input allows the incorporation of multiple pathogenicity software callers. Additional user input is required for information such as co-segregation in the family, whether the variants are *de novo*, *in trans* or *in cis* with each other. This tool is still in the process of development and will be validated on known variants.

Table 1. Patient Characteristics

Characteristic	Study population (n=278)	Mutation positive (n=31)	VUS Positive (n=55)	Mutation/ VUS negative (n=192)	Mutation positive vs rest of population
Clinical characteristics					***p-value***
Average age of onset of BC	34 (20-39)	34 (23-39)	34 (24-39)	34 (20-39)	NS
Self-reported race/ethnicity					
White/Caucasian	190 (69%)	25 (81%)	29 (53%)	136 (71%)	NS
African American/Black	66 (24%)	5 (13%)	19 (34%)	42 (22%)	NS
Other	12 (4.0%)	0	5 (9%)	7 (4%)	NS
Ashkenazi Jewish	27 (9.7%)	6 (19%)	3 (5.4%)	18 (9%)	NS
Personal cancer history					
Contralateral Breast Cancer	36 (13%)	6 (19%)	5 (11%)	25 (13%)	NS
Second primary malignancy[4]	21 (8%)	6 (19%)	3 (5%)	12 (6%)	0.02
Family cancer history					
Breast cancer	188 (70%)	25 (81%)	34 (62%)	129 (67%)	NS
Breast cancer age<40	76 (27%)	9 (29%)	52 (27%)	15 (27%)	NS
Bilineal breast cancer	34 (12%)	7 (23%)	5 (9%)	22 (11%)	0.08
BRCA1/2 Prediction Models					
Penn II prior probability	21%	27%	20%	19%	0.04
BOADICEA	15%	29%	13%	14%	0.005
Pathological data	***N (%)***	***N (%)***	***N (%)***	***N (%)***	***p-value***
Ductal carcinoma in situ (DCIS)	23/278 (8%)	4/31 (13%)	7/55 (13%)	12/192 (6%)	NS
ER+ invasive BC	147/214 (69%)	19/22 (86%)	29/44 (66%)	99/148 (67%)	0.09
Her2+ invasive BC	49/175 (28%)	7/20 (35%)	10/30 (33%)	32/125 (26%)	NS
Stage I	67/208 (32%)	6/21 (29%)	14/41 (34%)	47/146 (32%)	NS
Stage IIA/B	96/208 (46%)	9/21 (43%)	16/41 (39%)	71/146 (49%)	NS
Stage IIIA/B/C	43/208 (21%)	6/21 (29%)	11/41 (27%)	26/146 (18%)	NS
Stage IV	2/208 (1.0%)	0/21	0/41	2/146 (1.4%)	n/a

Figure 1. Mutations identified in patients with early onset breast cancer



Research Project 4: Project Title and Purpose

Mechanisms of Cancer Progression – The natural history of human cancer is characterized by the progressive selection and outgrowth of cells that possess increasingly aggressive properties. Among these properties, the propensity of cancers to recur following surgery and adjuvant therapy is the most important determinant of clinical outcome, since recurrent cancer is often incurable. This project will purchase the equipment and fund the operator to analysis the experiments elucidating the pathways that contribute to cancer recurrence. The studies proposed in this application will advance the therapeutic goals of preventing cancer recurrence. This knowledge has the potential to facilitate the development of more effective therapeutic approaches for treatments.

Anticipated Duration of Project

1/1/2012 – 12/31/2015

Project Overview

Understanding the biology of residual cancer ½ cells and elucidating the molecular pathways that contribute to their recurrence is a critical priority in cancer research. We have generated conditional transgenic mouse models for breast cancer that leave behind residual cancer cells following tumor regression induced oncogene down-regulation, and these residual cancer cells spontaneously recur in a manner similar to that observed in women with breast cancer. In this

project, we will use these models to define the pathways that contribute to breast cancer recurrence in order to develop improved therapeutic approaches to human cancers. To date, these models have been used to identify a panel of 300 genes that are consistently and concordantly either up- or down-regulated in recurrent tumors from multiple independent mouse models, and that predict recurrence-free survival in women with breast cancer in the direction predicted based upon findings in mouse models. In this project, we intend to identify molecular pathways that are responsible for tumor dormancy and recurrence.

Specifically, this proposal will leverage federal support from three funded R01 grants: NIH R01 CA143296 ("Minimal residual disease and mechanisms of breast cancer recurrence"), NIH R01 CA98371 ("In vivo oncogene-induced tumorigenesis and escape"), and NIH R01 CA148774 ("Survival and recurrence of dormant cancer cells"). Each of these grants is focused on investigating the pathways that contribute to breast cancer recurrence. These projects use tissue culture and mouse models to systematically evaluate the impact of specific genes and pathways on the cellular processes that drive tumor recurrence: cellular dormancy, cellular proliferation and cell survival. A rate-limiting problem that has emerged in this work is the highly time-consuming nature of the studies aimed at quantifying the effects of up-regulating or down-regulating genes of interest on cellular proliferation and apoptosis evaluated in tissue culture systems as well as mouse tissues. We now propose to purchase a slide scanning system in order to greatly accelerate the pace with which these quantification studies can be performed.

Principal Investigator

Lewis A. Chodosh, MD, PhD
Professor and Chair, Department of Cancer Biology
Perelman School of Medicine at the University of Pennsylvania
612 BRB II/III
421 Curie Blvd.
Philadelphia, PA 191904-6160

Other Participating Researchers

Hans Seidel – employed by the University of Pennsylvania

Expected Research Outcomes and Benefits

Breast cancer is typically treated by resection of the primary tumor followed by treatment with a combination of radiation, chemotherapy, and adjuvant hormonal therapy. Depending upon a spectrum of factors, patients will either remain disease-free following this course of treatment, or will develop recurrent breast cancer. While the pathways leading to the initial growth of primary breast cancers have been extensively studied, far less is known about the genes and processes that govern the ability of tumor cells to survive and recur following therapy.

The strategy outlined in this proposal should help to illuminate the genes and pathways that ultimately give rise to tumor recurrence. This project will improve our understanding of breast

cancer in several important ways. First, understanding the mechanisms by which tumors recur should aid in the diagnosis and clinical management of breast cancer. Furthermore, the tumor markers and pathways identified will represent potential targets for therapeutic intervention. Drugs targeting pathways that promote survival of residual tumor cells are most likely to extend the survival of patients with dormant, residual disease.

In summary, we have developed a novel set of genetically engineered mouse models to study tumor progression and have coupled these with a unique approach to study the biology of tumor cells that are important for recurrence. This research strategy should provide important insight into the biology of breast cancer recurrence, a process that receives scant attention and remains poorly understood, but that is critically important for the survival of patients with breast cancer.

Summary of Research Completed

There is no activity to report for this reporting period and no expenses were incurred.

Research Project 5: Project Title and Purpose

High Resolution RNA Functional Genomics – The purpose of this project is to utilize the state-of-art RNA sequencing technology, molecular imaging, and computational analysis to address problems in cellular fate decisions and responses to drugs. In the first part, we will use marker-assisted RNA sequencing measurements to test the hypothesis that a dynamical sequence of regulatory protein expression governs cell fate decisions. In the second part, we will use RNA sequencing and digital gene expression assay to address the hypothesis that dynamic variability of model genomes to pharmaco-kinetic treatments will identify components of the Non-Steroidal Anti-Inflammatory Drug response network.

Anticipated Duration of Project

1/1/2012 – 12/31/2015

Project Overview

The broad research objectives of this study is to use high-resolution dynamical RNA assays at the whole-genome level to uncover cellular decision processes involved in cell fate determination and to use targeted drug perturbation and comparative data to identify novel components of a critical molecular pathway responding to Non-Steroidal Anti-Inflammatory Drugs (NSAIDs).

Specific Aims:
Aim 1: To use a genome-wide cellular resolution map of embryonic gene expression in C. elegans to develop models for cell fate decisions.
Aim 2: To use the dynamics and variability of the NSAID response network across species and tissues to identify components of the response network.

The following two hypotheses are addressed in this project:

Hypothesis 1: High resolution of dynamic pattern of gene expression in a model organism will reveal the key decision network for cell differentiation and organismal development.

Hypothesis 2: Phamaco-kinetic perturbations coupled to comparative RNA profiling will reveal novel components of NSAID pathway.

These hypotheses will be tested by developing novel methods for deconvolving multiplexed RNA measurements into high-resolution single cell inferences and also by using comparative perturbation data from multiple species to identify common network components. The resulting work will develop new methods for high-resolution RNA profiling and help uncover novel components of a critical drug response pathway.

Principal Investigator

Junhyong Kim, PhD
Co-Director, Penn Genome Frontiers Institute
University of Pennsylvania
302 Lynch Laboratories
433 S. University Avenue
Philadelphia PA 19104-6018

Other Participating Researchers

John Hogenesch; John Murray, PhD; Garret FitzGerald, PhD; James Eberwine, PhD – employed by the University of Pennsylvania

Expected Research Outcomes and Benefits

The results of this project will uncover genomic regulatory pathways related to cell development and to drug responses. One of the frontiers of modern medicine is cell-based therapeutics including regenerative medicine, cell-based tissue remodeling, and gene-therapy. While there has been considerable applied work in cell-based therapeutics, the approach fundamentally requires knowledge of how a genome, through development, specifies cell fate and the functional phenotype of each cell type. The results from the Aim 1 of this project will help develop a new technique for constructing genome-scale model of cell fate decisions, which through homology modeling will be applicable to cell-based therapeutics in human systems. In addition, the techniques and algorithms developed in this aim will be directly applicable to human cell-based systems. In Aim 2, we will use comparative functional genomics data to identify components of molecular pathways responsive to Non-Steroidal Anti-Inflammatory Drugs (NSAIDs). NSAIDs are one of the most widely used pharmacological agents. Despite their efficacy they are also associated with gastrointestinal bleeding; while at the same time, scattered evidence suggests that NSAIDs may have other side benefits such as suppression of cancer pathways. Therefore, the uncovering novel components of this pathway will help understand the basic biological effects of

NSAIDs use and open the pathway to personalized medicine by identifying potential genomic factors that modulate its action and side effect risk.

Summary of Research Completed

During this reporting period, we made progress on both aims of this project and we discuss each in turn.

Specific Aim 1: To use a genome-wide cellular resolution map of embryonic gene expression in C. elegans to develop models for cell fate decisions.

We improved our earlier lineage-tracing methods, which were limited to the 350-cell stage, when one round of cell divisions still remains for most cells. Coupling new software with the increased speed and image quality of a Leica SP5 resonance-scanning confocal microscope has allowed us to optimize the lineage tracing procedure; we now reliably lineage embryos to the 600-cell stage, when essentially all embryonic cell divisions have occurred. In addition, one person can image and curate up to eight embryos per day through the 350-cell stage, four times faster than with previous methods. Our work and previous work with lower-throughput lineage tracing methods have demonstrated the power of lineage tracing at these later stages for phenotyping. In particular, this stage has more phenotypic information because there are more cell movements and asymmetries in cell deaths and cell cycle length. We have now traced the lineages of 36 wild-type embryos to the 600-cell stage and used these data to compute a detailed statistical model of normal patterns and variability of cell position, cell division orientation and cell cycle length (Richards et al., 2013). Since all of these features are highly reproducible between WT embryos, they provide a sensitive background for identification of defects in mutants. We furthermore examined variability in WT development changes under temperature stress. Surprisingly, we found that moderate (22C) and high (25C) temperatures have essentially identical levels of variability, while an additional 1-degree increase to 26C causes a step function-like dramatic increase in variability concomitant with a decrease in viability. This occurs in part because of homeotic transformations in lineages that are normally specified by the Notch signaling pathway.

We have tested the potential for using lineage tracing to define cellular defects in lineage-specific TF mutants. We analyzed three lethal TF mutants, including two early ABpxp lineage-specific TFs (*ceh-36/Otx* and *nob-1/Hox9-13*) and one ABpxp TF with later expression (*mls-2/Hmx*). For *mls-2* and *ceh-36*, previous studies had identified defects in only a few specific cell types, with no information about these factors' importance for the vast majority of expressing cells. We developed algorithms to identify cells whose lineage, migration path and neighbors differ from the wild-type reference set. Applying these to collections of deep lineages for each mutant strain allowed us to identify defects in each mutant. For example, previous work had identified four cells with cell fate defects in *ceh-36* mutants, three of which require *ceh-36* in later-stage elongating embryos or larvae. Our analysis identified the one early embryonic cell known to be defective along with 28 additional cells whose defects ranged from mis-migration, early or late division, and failure to undergo programmed cell death. All of these defects occurred in cells that normally express *ceh-36*, suggesting most are likely cell-autonomous.

Some of these lineage phenotypes are correlated to the phenotypes of arrested larvae; some *ceh-36* mutant larvae appear to have intestinal blockage, and among the defective cells were two visceral muscle cells important for defecation. Similarly analysis of *nob-1* and *mls-2* mutants identified defects both in known cells and in numerous other cells that express each TF. These results are consistent with our hypothesis that lineage-specific factors are broadly important for fate regulation across their expression domain, and indicate that lineage tracing of mutants is a powerful approach to identify defective cells.

Specific Aim 2: To use the dynamics and variability of the NSAID response network across species and tissues to identify components of the response network.

For this aim, in this reporting period we investigated the role of microsomal prostaglandin synthesis 1 in inflammatory pain. Nonsteroidal anti-inflammatory drugs (NSAIDs) alleviate inflammatory pain by inhibiting prostanoid formation. One of the major prostanoids involved in pain signaling is prostaglandin (PG) E_2 generated by microsomal prostaglandin E synthase-1 (mPGES-1). Reduced inflammation and pain hypersensitivity has been reported in mPGES-1 null mice, however, its involvement in sensitization at specific (peripheral or central) sites has not been investigated either. We explored the relative contribution of peripheral and central mPGES-1 in inflammatory pain as a prerequisite to understanding which sources of variability may contribute to differences in the NSAID response between individuals. We have generated mice that are deficient in mPGES-1 in myeloid cells by crossing floxed mPGES-1 fx/fx mice with LysM-Cre +/+ mice (designated myeloid mPGES-1 KOs). We observed that inflammatory responses and the pain response to inflammation are markedly reduced in myeloid mPGES-1 KOs. Macrophages are the major source of PGE_2 in inflammation and we observed a reduced chemotaxis phenotype in macrophages isolated from myeloid mPGES-1 KOs. Health Research Funds were used to explore in an unbiased fashion whether removal of this enzyme affects the base line phenotype of macrophages and their response to an inflammatory challenge. Peritoneal inflammatory macrophages were isolated from male LysM-Cre +/+ mPGES-1 fx/fx mice and wildtype littermates following intraperitoneal stimulation with the sterile inflammatory agent thioglycollate. Macrophages were allowed to adhere to culture dishes and stimulated with 10ng/ml lipopolysaccharide for 0, 6 and 24 hours prior to total RNA isolation. Libraries were prepared using the Illumina Truseq method and barcoding. Two samples were multiplexed per lane pooling equimolar concentrations. Multiplexing at concentrations that result in similar numbers of clusters remains a technically challenging problem, because the “clusterable” concentration of a library can be quite different from its measured concentration. Here we resolved this problem as follows: If the number of reads for a given multiplexed sample was less than 80 mio (paired end reads), we adjusted the concentration based on the actually obtained read numbers and sequenced the sample again. 8 samples (out of 24) were repeated based on these criteria. A total of 24 data sets were generated and aligned using RUM. Statistical expression analysis was performed at the exon level using the Patterns from Gene Expression (PaDE) algorithm.

The experimental aim was to explore the candidate molecular mechanisms for the reduced pain and inflammation in myeloid mPGES-1 KO mice. We found that exons in 3913 genes were differentially expressed between wild type and knockout at all three time points at the q=0.1

FDR level (figure 1). In two independent RT-PCR validation experiments we focused on genes that are known to be associated with inflammatory processes and those that are considered markers of macrophage phenotypes. Generally, inflammatory genes (e.g. COX-2 IL1-β, TNFα, iNOS, IL-6, IL-18, IL12α, IL12β) were suppressed and anti-inflammatory genes (e.g. IL-10, Arg1) increased suggesting that mPGES-1 deficient macrophages had adapted to a less inflammatory phenotype. Interestingly, the classical segregation into M1 and M2 did not capture fully the phenotypic specification displayed by knockout vs. wild type macrophages.
While the gene expression profile suggested that knockout macrophages might be less responsive to chemo-attraction, and Boyden chamber experiment indeed identified a small, but reproducible migration delay in response to chemokines, this was not reflected by a reduction of macrophage migration into inflamed tissues in vivo. Visualization of macrophage infiltration in the inflamed paw by injection of fluorescently labeled HDL particles, revealed no differences in the macrophage content of the inflamed paw between wild type and myeloid mPGES-1 KO mice (figure 2)

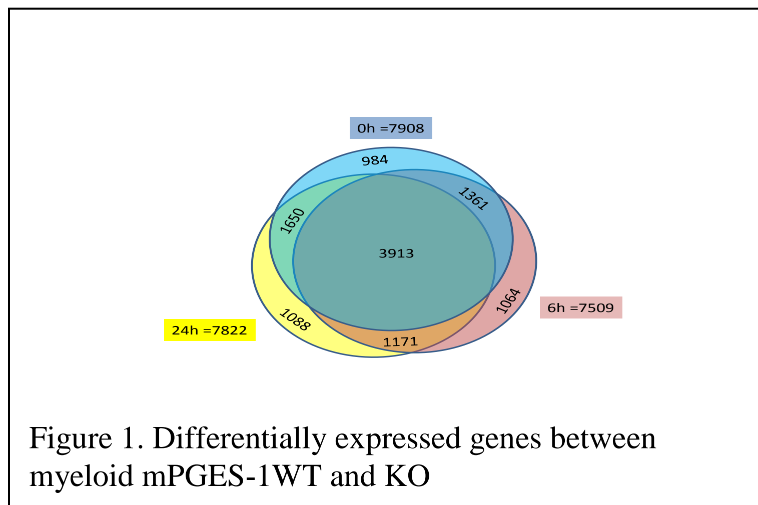


Figure 1. Differentially expressed genes between myeloid mPGES-1WT and KO

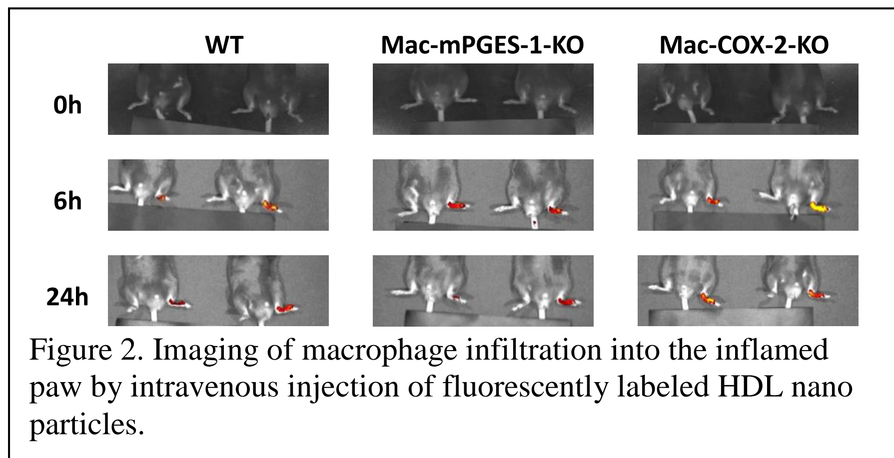


Figure 2. Imaging of macrophage infiltration into the inflamed paw by intravenous injection of fluorescently labeled HDL nano particles.

Richards JL, Zacharias AL, Walton T, Burdick JT, Murray JI. A quantitative model of normal Caenorhabditis elegans embryogenesis and its disruption after stress. Developmental biology. 2013;374(1):12-23.

Research Project 6: Project Title and Purpose

Enhancement of Systems and Computational Neuroscience Space - Research Infrastructure Project – The purpose of this research infrastructure project is to renovate the vivarium in Stemmler Hall to support systems and computational neuroscience. The project will provide a total of 15,059 square feet of renovated space and infrastructural improvements. This includes the consolidation and improvement of research functions such as animal holding, testing and procedure space from different locations of outmoded areas in the Richards Building. In addition, the renovated space will support investigators across multiple departments, including Otorhinolaryngology, Neuroscience and Psychology, whose closely related research programs address critical needs common to Penn, the NIH road map and the Commonwealth of Pennsylvania.

Anticipated Duration of Project

1/1/2012 – 8/31/2014

Project Overview

The renovation of the Stemmler Hall vivarium is the first phase of a multi-phase project to relocate our animal programs from the Richards Building. This first phase will upgrade and enhance space that has remained largely in its original state since Stemmler Hall was constructed in 1978. Although the infrastructure has been upgraded, the devices specific to this area are original to the fit-out and are nearing the end of their expected life span. Renovations to this area will allow for the consolidation of the research of several investigators, thereby increasing research efficiencies and creating opportunities for collaboration. Furthermore, relocating this group to Stemmler Hall will enable space to be cleared for infrastructure improvements in the Richards Building (the current location of the group), which will ultimately support University-wide initiatives.

The total area of modifications is approximately 15,059 square feet, and the scope of this project includes the following spaces:

1. Four testing labs
2. Surgical suite
3. Three special-purpose animal holding rooms
4. Five multi-purpose animal holding rooms
5. Two bathrooms and a touch-down/locker room
6. University Laboratory Animal Resources (ULAR) break room and office
7. Replacement of existing cage wash equipment with state-of-the-art rack washer

Principal Investigator

Glen N. Gaulton, PhD
Executive Vice Dean and Chief Scientific Officer
University of Pennsylvania, Perelman School of Medicine
240 John Morgan Building
3620 Hamilton Walk
Philadelphia PA 19104-6160

Other Participating Researchers

None

Expected Research Outcomes and Benefits

The renovations to Stemmler Hall are being completed to allow for increased programmatic efficiencies in support of systems and computational neuroscience research. The research that will be conducted in the renovated space provides novel insights into fundamental issues of brain organization and functions, and all the studies are designed to translate into clinical advances. The resulting renovated vivarium space will include behavioral testing laboratory and testing laboratory preparation space, animal holding rooms, holding room anteroom and a support / administrative area. The renovations will provide capacity for investigators from multiple disciplines, including the departments of Otorhinolaryngology, Neuroscience, Neurosurgery and Psychology. The space is also ideally positioned in the midst of the broader neurosciences research communities on other floors of Stemmler Hall and the adjoining Johnson Pavilion.

Evaluation of this project will be based on two sets of criteria: (1) the successful construction of research space in support of the mission and goals of the systems and computational neuroscience program; and (2) the number of extramural sponsored projects awarded to the investigators directly supported in this space.

Summary of Research Completed

Construction of the animal holding rooms were completed in summer 2013, and animals were relocated to the new facility shortly thereafter. Although unforeseeable challenges arose due to low range noise transmissions in several behavioral testing booths, the project team partnered with the University facilities office to implement solutions. This enabled research to commence in the new facility on schedule. The facility is now functioning successfully within the sound isolation range for which it was designed.

As of this date, we have successfully completed 90% of the renovation and improvements for the testing laboratories, surgical spaces, additional administrative areas and the installation of new cage wash equipment. We are working on finalizing the renovations and making preparations for scientists from the departments of Otorhinolaryngology, Neuroscience, Neurosurgery and Psychology to occupy the space by the end of the summer.

Research Project 7: Project Title and Purpose

Research Infrastructure: Renovation for Laboratory Space for BioPhysical Chemistry – The purpose of this project is to upgrade and renovate 2300 square feet of laboratory space to standards appropriate for a 21st century laboratory performing research at the interface between physical chemistry and biology. The renovations include outfitting the laboratory space with modern fume hoods, electric, and casework as well as the creation of 3 rooms controlled to +-.2 deg C to accommodate highly sensitive laser-based evaluation of biological samples. The renovated space will be occupied by researchers from the Chemistry Department whose research includes work studying the biophysical properties and functions of biological membranes.

Anticipated Duration of Project

11/16/2012 – 3/31/2014

Project Overview

A total of 2300 square feet will be dedicated to the work of a recently tenured investigator and will consist of spaces devoted to light, temperature, and vibration sensitive laser laboratories, light, temperature, and vibration sensitive microscopy, student spaces, and biological sample preparation and evaluation. Renovations will include updates to the HVAC control systems and diffusers to improve airflow, temperature control and energy efficiency in the laboratories, new wall finishes and floor surfaces, replacement of casework, laboratory benches and shelving, replacement of fume hoods, and upgrades to the electrical infrastructure in the rooms to improve power sources for the equipment to be installed.

Principal Investigator

Richard M Schultz, PhD
Associate Dean School of Arts and Sciences
University of Pennsylvania
116 College Hall
University of Pennsylvania
Philadelphia, PA 19104

Other Participating Researchers

None

Expected Research Outcomes and Benefits

The principal outcome of the present infrastructure project will be to improve the workspace quality of the faculty performing research at the scientific frontier where biology and medicine intersect the chemical and physical sciences. The improved laboratory space will enable the Department of Chemistry to retain critical faculty members and will indirectly expand the

amount, continuity and quality of space available to other junior faculty associated with Biochemistry in the Department. Such faculty would not consider working at the University of Pennsylvania without the provision of adequate laboratory space that meets the highest quality standard.

Summary of Research Completed

Design and construction of the laboratories was completed and occupants have been moved into the space. During the reporting period, we completed construction, punch-list and move-in. All electrical, HVAC, plumbing, fume hood and casework installation is complete. Active research has recently started in all renovated spaces in this laboratory.

<u>Research Project 8: Project Title and Purpose</u>

The Dynamic Microbiome of Chronic Diabetic Foot Ulcers – Bacterial colonization and infection significantly impair healing of the diabetic foot ulcer (DFU). This study will utilize less-biased genomic methods of surveying microbial diversity and functionality to more precisely track and characterize temporal microbiome dynamics of the DFU. Analyzing how the dynamic microbiome correlates with clinical phenotypes and wound outcomes provides a vital foundation for the development of improved biomarkers and therapeutics that leverage microbial community dynamics.

Anticipated Duration of Project

8/7/2013 – 12/31/2015

Project Overview

Aim 1: Longitudinal characterization of the dynamic microbial communities colonizing diabetic foot ulcers (DFUs). This aim will test the hypothesis that the dynamic microbiome of the DFU is associated with clinical phenotypes and has predictive and prognostic value for clinical outcome.

- Sub-aim 1.1: Sequence 16S rRNA genes to identify microbial richness and diversity in 40 time series of DFU, at 2-week intervals.
- Sub-aim 1.2: Determine if temporal dynamics of microbial colonization are associated with wound outcome.
- Sub-aim 1.3: Determine if clinical variables are associated with temporal patterns of microbial colonization.

Aim 2: Metagenomic analysis of the diabetic foot ulcer. This aim will test the hypothesis that whole metagenomic shotgun sequencing can identify the functional complement of microbial genes present in a DFU. This aim will set the groundwork for future studies aimed at associating the functional features of the metagenome with clinical outcomes.

- Sub-aim 2.1: Establish and optimize methods for metagenomic sequencing of the DFU.
- Sub-aim 2.2: Perform comparative metagenomic sequencing on 10 DFU time-series.

Principal Investigator

Elizabeth A. Grice, PhD
Assistant Professor
University of Pennsylvania
415 Curie Blvd.
242A Clinical Research Building
Philadelphia, PA 19104

Other Participating Researchers

Sue E. Gardner, RN, PhD – employed by the University of Iowa
Brendan Hodkinson, PhD – employed by the University of Pennsylvania

Expected Research Outcomes and Benefits

By the end of the project, we will have a much greater understanding of how the microbiome of the diabetic foot ulcer (DFU) fluctuates throughout injury, infection, and healing. This data will enable us to then determine if the DFU microbiome has prognostic or predictive value. Furthermore, we will reveal how clinical parameters, such as blood glucose control and tissue oxygenation, relate to the dynamic DFU microbiome. This study will be the first longitudinal analysis of DFU microbiota using genomic, culture-independent methods to characterize microbial diversity. This study will also set the groundwork for metagenomic analysis of DFU to identify functional traits of the microbiome associated with wound outcomes and clinical phenotypes. Lastly, this work provides a critical foundation towards improved biomarkers and therapeutics based on the dynamic microbial community.

Summary of Research Completed

Sub-aim 1.1: Sub-aim 1.1 was completed, for which we sequenced 16S rRNA genes from 100 patients with DFUs, exceeding our initial goal of 40 patients. Amplification and sequencing of the 16S-V4 region was performed using the Illumina MiSeq platform with 150 bp paired-end 'V2' chemistry. Samples were multiplexed with one another and samples from other studies such that 192 samples were sequenced per run.

One hundred subjects were enrolled and had at least one sample that produced microbial sequence data. The mean age of the subjects was 54 years. Seventy-eight were male and 91 were white. Eighty-seven had type 2 diabetes, whereas the remainder had type 1 diabetes. The mean ulcer duration was 31 weeks, and the mean wound tissue oxygen was 46.9 mmHg. Seventy-three of the DFUs were located on the plantar forefoot, while the remaining DFUs were located on the plantar mid foot or hind foot. The mean white blood cell count was 7951, the mean hemoglobin A1c level was 8.2%, and the mean serum C-reactive protein level was 2.2. Of the 92 patients remaining in the study at week 4, the average surface area percent change of the wound was 75.6. Thirty-one of the patients in the study experienced complication (osteomyelitis, amputation, wound deterioration).

The most abundant genus was *Lactococcus* (Streptococcaceae), which was present in 374 of the 377 samples, with an average relative abundance of 27.03%. The second, third, and fourth most abundant genera were *Staphylococcus* (19.92% average relative abundance; present in all samples), *Corynebacterium* (7.54% average relative abundance; present in all samples), and *Streptococcus* (5.69% average relative abundance; present in 375 of the 377 samples), respectively. All other genera represented <5% of the total diversity in the sequence set. An example of the longitudinal output of these analyses is presented in **Figure 1**. For each patient, we measured various alpha-diversity measures (Shannon index, Faith's Phylogenetic Diversity, number of observed species, and Chao1 index), bacterial load (as measured by 16S rRNA qPCR), and bacterial relative abundance.

Sub-Aim 1.2: Outcomes were defined by the following criteria: 1) binary outcome of whether the wound was healed or not healed by week 12; and 2) binary outcome of development of complications (as defined by amputation, osteomyelitis, or severe infection) or no development of complication. We first analyzed baseline samples from DFUs from the first time the patient presented and before treatment was started. Using a regression model (LOGISTIC procedure in SAS), we found that the number of observed species-level OTUs in the DFU was significantly associated with development of complications. However, when covariates were included in the regression model, the association was no longer significant. Covariates included in the analyses were: ulcer duration, ulcer size, ulcer depth, mean tissue oxygen tension, ulcer location, white blood cell count, hemoglobin A1c, and C reactive protein level.

Next, we analyzed 16S rRNA gene data from the longitudinal DFU samples that were collected every two weeks. A mixed model was then used to determine if incorporating longitudinal data on bioburden was more informative than baseline data alone for predicting the outcomes defined above. A general trend we observed was that lower microbiome diversity was associated with the development of complications. This was true of several measures we employed to measure diversity, including 1) the Shannon diversity index, illustrated in **Figure 2**, which takes into account evenness and abundance of OTUs (P=0.0031, odds ratio (OR)=0.676; 95% Wald confidence limits (CL)=0.522-0.876); 2) the raw number of species-level OTUs observed in each DFU (P=0.0036, OR=0.996, CL=0.993-0.999) and 3) Faith's phylogenetic diversity index, which takes into account phylogenetic diversity in addition to OTU diversity (P=0.0019, OR=0.924, CL=0.879-0.971). There were no detected associations between 16S rRNA-based measures and outcome as defined by the binary measure of healed at 12 weeks. We also examined the microbiome at the timepoint directly before healing, to determine if there were features that were consistent in the microbiome, and detected no significant features.

Sub-aim 1.3: Work is in progress to complete sub-aim 1.3. All data is collected and available and analyses are currently underway.

Aim 2: Now that Aim 1 is near completion, we are beginning to optimize procedures for metagenomic sequencing of DFU. A major obstacle we have encountered is the high amount of host DNA present in the DFU samples. This DNA overwhelms the microbial DNA in the sample, and thus, we obtain >90% human reads when sequencing is performed. We are working to optimize and increase the amount of microbial DNA recovered.

In the meantime, we have optimized the metagenomic analysis pipeline that we will employ for the analyses following metagenomic sequence generation. Following sequence generation, our generalized bioinformatics preprocessing pipeline is to first remove human contaminating sequences using DeconSeq. Low quality sequences are then filtered and barcodes and adaptors removed using FASTX-toolkit, using the FASTQ Quality Filter and FASTQ Trimmer, respectively. Reads are then assembled into contigs. To assemble metagenomic reads requires reconstructing genomes without referencing a sequenced genome, or de novo assembly, as reference databases of microbial genomes are far from complete. We employ Ray Meta, a de novo assembler based on Eulerian tour of de Bruijn graphs, but also allows distribution of the assembly job across multiple compute nodes. In our experience, a dataset of the size proposed could take up to 3 months to assemble on a single server. Utilizing the Penn Medicine Academic Computing Service (PMACS) High Performance Cluster, we are now able to distribute the job over multiple nodes using Ray Meta and assembly will complete in <1 week. Ray Meta incorporates a scaffolding function, utilizing paired end information to assemble contigs into larger genome scaffolds.

Once experimental techniques are optimized to recover higher amounts of microbial DNA, we anticipate sequence analysis to proceed smoothly.

Figures:

Figure 1: Example of a longitudinal profile showing changes in the microbiome as the DFU heals.

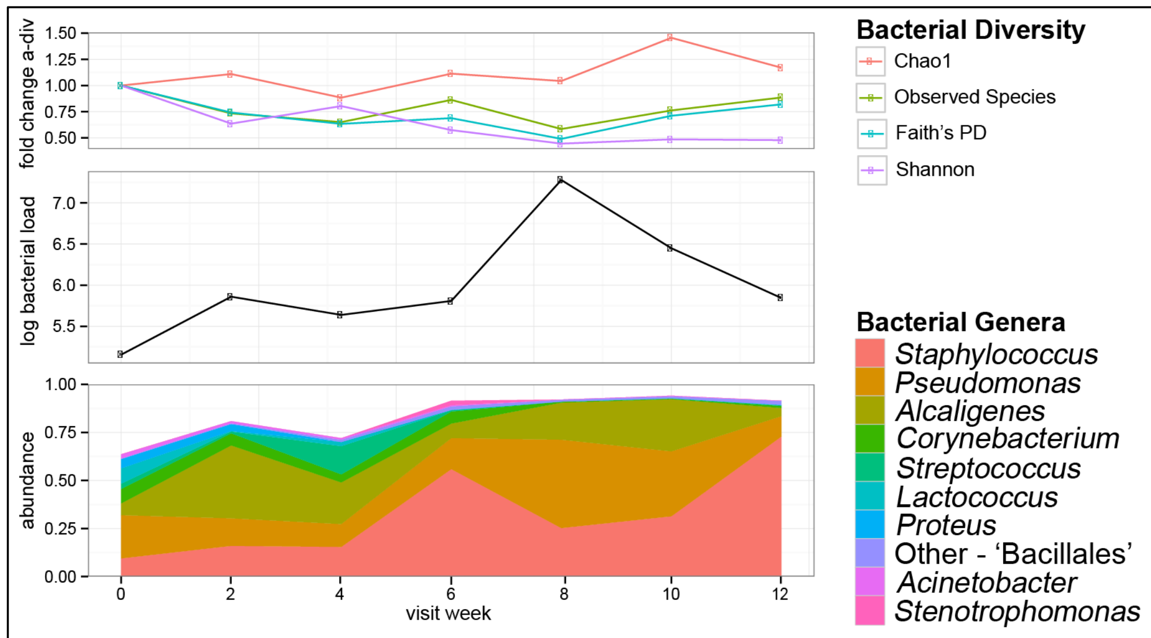
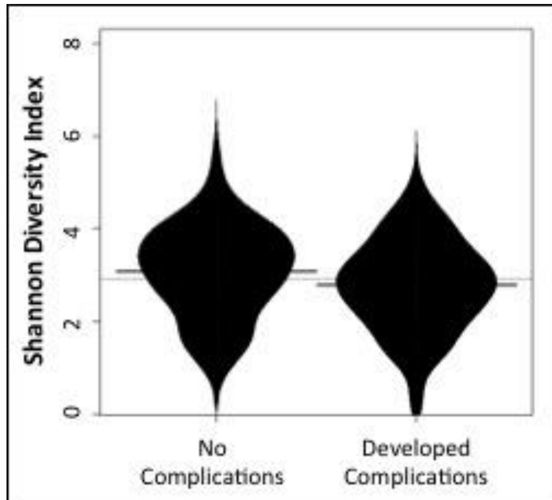


Figure 2: Violin plot depicting probability density of Shannon Diversity Index values in subjects who developed complications and those that did not. The median is represented with the line through the middle of the plot.



Research Project 9: Project Title and Purpose

Direct Reprogramming of Somatic Cells to Spermatogonial Stem Cells – We propose to directly convert fibroblasts using defined factors to spermatogonial stem cells, those stem cells in males which give rise to sperm. This research will provide a new mechanism by which men rendered infertile without the opportunity for sperm-banking will have the opportunity to repopulate their testes through the generation of patient-specific stem cells.

Anticipated Duration of Project

8/7/2013 – 12/31/2015

Project Overview

The reprogramming of cells either to a pluripotent, embryonic-like state or directly from one fate to another demonstrates the plasticity of cells and is one of the most electrifying recent advancements in biomedical research. The technologies developed for reprogramming cells enable the generation of patient-specific treatments for a variety of diseases. They also provide a new strategy to study normal cell development, disease progression, and a system to generate new pharmacological screening methods. We are interested in studying spermatogonial stem cells (SSCs), those stem cells found in the testes of males that give rise to sperm. Studies from the past few decades have demonstrated we can isolate SSCs from several mammalian species, culture them in vitro, and in the mouse transplant them to an infertile testis where they are capable of producing viable sperm. This isolation and transplantation method has enabled study of the molecular mechanisms controlling the establishment and differentiation of SSCs within the testes and further provides an exciting new way to potentially preserve the fertility of patients

rendered infertile without the opportunity to bank their sperm. In this proposal, we intend to reprogram embryonic and adult mouse fibroblasts directly to SSCs using defined factors and to transplant these reprogrammed cells to a germ cell depleted testis. The research proposed here will provide a novel source of patient-specific SSCs for men rendered infertile due to radiation and chemotherapy from cancer treatments and will offer a novel mechanism to study SSC establishment, maintenance, and differentiation.

SPECIFIC AIMS:
AIM 1: Reprogram mouse embryonic fibroblasts to spermatogonial stem cells using defined factors.
AIM 2: Reprogram mouse adult fibroblasts to spermatogonial stem cells.
AIM 3: Determine in vivo function of directly reprogrammed mouse spermatogonial stem cells using transplantation assays.
AIM 4: Reprogram human neonatal foreskin fibroblasts to spermatogonial stem cells.

Principal Investigator

John D. Gearhart, PhD
Director, Institute for Regenerative Medicine
University of Pennsylvania
Smilow Center for Translational Research
3400 Civic Center Blvd., Bldg. 421
Philadelphia, PA 19104

Other Participating Researchers

Ralph L. Brinster, VMD, PhD; Lori D. Kellam, PhD – employed by the University of Pennsylvania

Expected Research Outcomes and Benefits

We expect that spermatogonial stem cells produced from direct reprogramming of fibroblasts will provide a novel source of stem cells for the treatment of male infertility. This approach bypasses the need for embryonic stem cells as a source for cell therapy.

Summary of Research Completed

AIM 1: Reprogram mouse embryonic fibroblasts to spermatogonial stem cells using defined factors.

We generated doxycycline inducible lentiviral vectors to ectopically express cDNAs representing transcription factors known to be critical for the establishment or maintenance of spermatogonial stem cells in mouse embryonic fibroblasts. These factors include GDNF responsive genes *Bcl6b*, *Egr2*, *Egr3*, *Etv5*, *Lhx1*, and *Oct6,* the pluripotency related genes *Oct4, Klf4, Lin28b*, *Myc*, and *Sox2* as well as factors implicated in maintaining SSCs in an undifferentiated state in

vivo including, *Ngn3*, *Plzf*, and *Sohlh2*. Following forced expression in and conversion of somatic cells to spermatogonial stem cells we can test the stability of reprogrammed cell phenotype by withdrawal of doxycycline, shutting off the exogenous transcription factors. We transduced primary mouse embryonic fibroblasts (MEFs) derived from male B6.129S7-Gtrosa26 (ROSA26-B6) mice, a mouse model that constitutively expresses the *Escherichia coli* LacZ (*E.Coli* lacZ) gene and can be detected by staining with the B-galactosidase substrate 5-bromo-4-chloro-3-indolyl b-D-galactoside (X-gal). After transduction of many combinations of the listed factors, ROSA26-B6 MEFs were placed in media to support the initial stages of reprogramming and after 9 days, the cells were assessed for colony formation. If colonies arose from transduced cells, individual clones were selected and plated onto SNL 76/7 STO (STO) feeder layers in media containing growth factors known to induce germ cell development from mouse embryonic stem cells (mESCs), including bone morphogenic protein 4 (BMP4), stem cell factor (SCF), epidermal growth factor (EGF), and leukemia inhibitory factor (LIF), for 6 days. Following this treatment, clones were transitioned to a defined culture media supporting long-term culture of SSCs, including glial derived neurotropic factor (GDNF) and basic fibroblast growth factor (bFGF). After testing a variety of combinations of transcription factors, one specific 4-factor combination induced morphological changes in MEFs consistent with SSCs in culture - with cells growing as indicative 'clump-forming' colonies – and the induced spermatogonial stem cells (iSSCs) were supported specifically by the defined SSC culture media over at least 15 passages. Analysis of putative iSSC clones cells for germ cell specific and pluripotency proteins, including mouse vasa homolog (MVH), promyelocytic leukaemia zinc finger protein (PLZF), and octamer-binding transcription factor 4 (OCT4) by immunocytochemistry (FIG 1) and gene expression analysis by quantitative RT-PCR (FIG 2) indicates these cells have significantly up-regulated germ cell specific markers (MVH, PLZF) as well as markers of pluripotency (OCT4). Morphological and gene expression data described support the notion that we have generated a germ cell phenotype similar to SSCs found in vivo. Following withdrawal of doxycycline the iSSC clones maintain SSC-like morphology as well as expression of germ cell specific genes and proteins.

AIM 2: Reprogram mouse adult fibroblasts to spermatogonial stem cells.

We have initiated experiments to reprogram mouse adult tail tip fibroblasts with the same specific four factor combination that successfully reprogrammed mouse embryonic fibroblasts and have generated a single clone that morphologically resembles the clones generated from embryonic fibroblasts in addition to endogenous SSCs explanted to culture (FIG 4). We have yet to further analyze these clones to determine germ cell specific gene and protein expression.

AIM 3: Determine in vivo function of directly reprogrammed mouse spermatogonial stem cells using transplantation assays.

Selected iSSC clones expressing germ cell specific genes and proteins were transplanted by injection into individual seminiferous tubules of the testis of adult 129xB6 F1 male mice treated with busulfan, an agent that selectively ablates the spermatogonial stem cell population. One caveat of these first experiments is that the initial iSSC clones we generated were doxycycline (dox) dependent; therefore, the mice were also given doxycycline in the water to maintain the

iSSC phenotype after transplant. Transplants with dox-dependent clones resulted in tumor formation in the testis that subsequently stained positive for X-gal indicating the dox-dependent iSSCs contributed to the tumor formation. Examination of the tumors revealed they were made up of a homogenous cell type indicative of other germ cell tumor types, supporting the notion the iSSCs may be germ cell-like but continued expression of the exogenous transcription factors may have resulted in over-proliferation of the cells within the testis. As indicated in AIM 1, we have since generated iSSC lines that are dox-independent and have initiated transplants with these iSSC clones into the busulfan treated 129xB6 F1 male mice as described above.

AIM 4: Reprogram human neonatal foreskin fibroblasts to spermatogonial stem cells.

We have not yet attempted reprogramming human fibroblasts to induced spermatogonial stem cells.

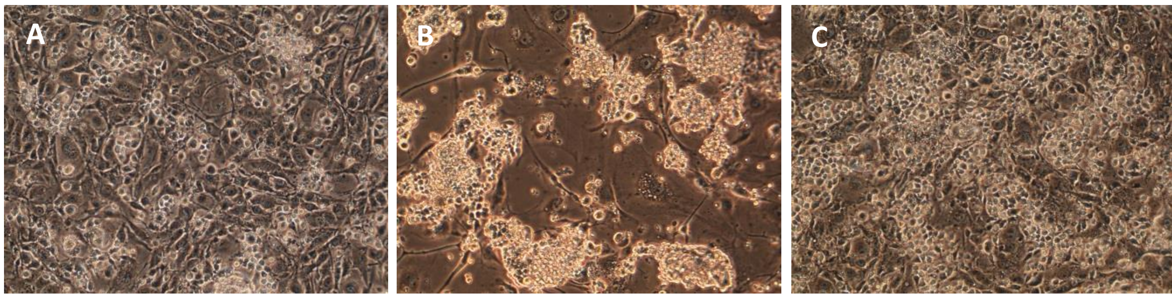


FIG 1. MEFs transduced with 4 factor combination exhibit clump-forming morphology at passage 7. Representative iSSC clones 25B2 (*A*), 14B4 (*B*), and 24A2 (*C*).

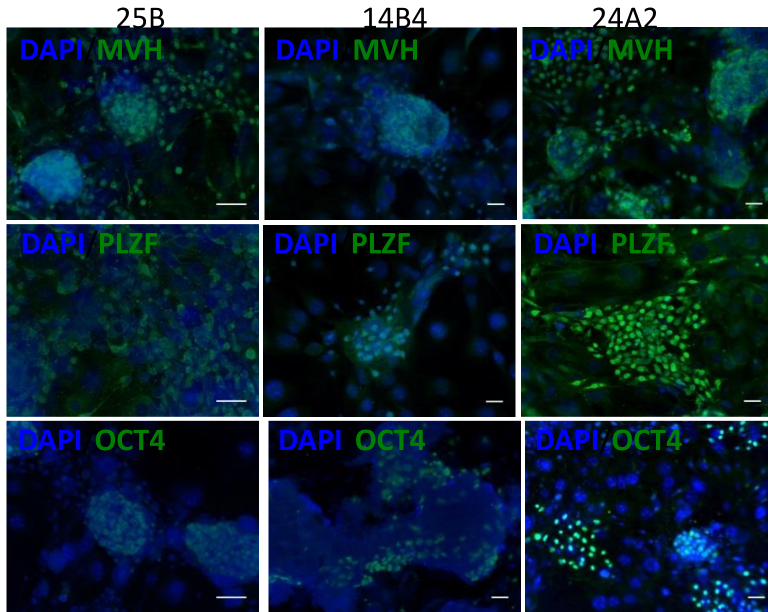


FIG 2. SSC marker expression in 4 Factor reprogrammed cells. Expression (*green*) of MVH, PLZF, and OCT4 in three representative clones (25B2, 14B4, and 24A2) of 4 factors transduced pMEFs reprogrammed to resemble SSCs morphologically. DAPI labeling nuclei (*blue).* Scale bar: 50um.

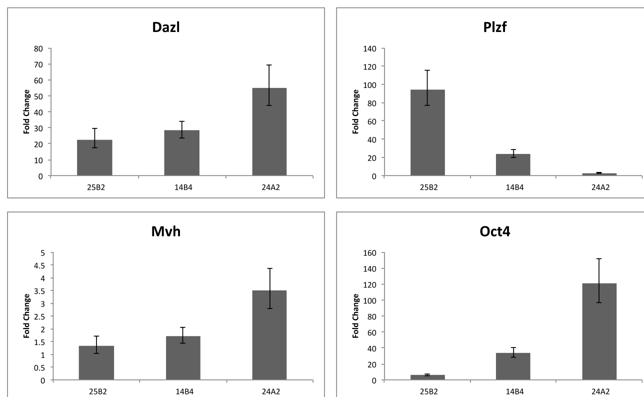


FIG 3. Quantitative RT-PCR for iSSC marker expression in 4 Factor reprogrammed cells. Fold change for germ cell specific genes, *Dazl*, *Plzf,* and *Mvh*, as well as the pluripotency gene *Oct4* in 4 factors transduced pMEFs reprogrammed to iSSCs. Relative expression to starting pMEF population is shown for three representative clones (25B2, 14B4, 24A2).

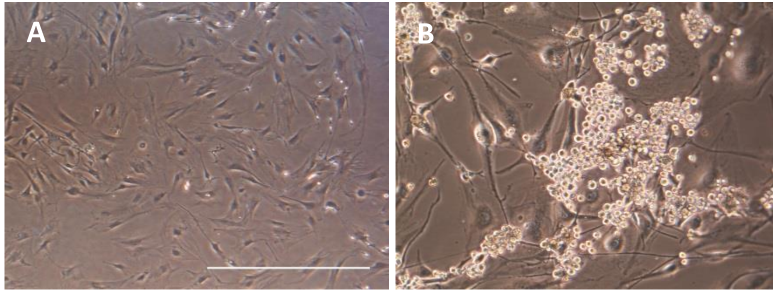


FIG 4. TTFs transduced with 4 factor combination exhibit clump-forming morphology at passage 2. Images of starting TTF population (*A*) TTF:iSSC clone at passage 2 (*B*).

Research Project 10: Project Title and Purpose

Causes and Consequences of Altered Dosage Compensation in Humans – Female reproductive cancers exhibit epigenetic instability of X-Chromosome Inactivation (XCI): loss of XIST RNA expression, loss of heterochromatic markers of the inactive X (Xi), partial reactivation of Xi, and acquisition of additional copies of active Xs. Consequently, these cancers suffer from improper dosage of X-linked genes, yet it is unknown how this misregulated expression contributes to the initiation and pathogenesis of disease. Human female pluripotent stem cells (hPSCs) also exhibit XCI instability (similar to reproductive cancers), and are a powerful model system to investigate the causes and consequences of altered dosage of X-linked genes.

Anticipated Duration of Project

8/7/2013 – 12/31/2015

Project Overview

Regenerative medicine holds great promise for treating genetic and trauma-induced disease. Pluripotent stem cells have the potential to differentiate into all cell types, underscoring their importance for therapeutic applications. However, human pluripotent stem cells (hPSCs) are susceptible to genetic and epigenetic instability, which raises questions about their utility and safety in clinical settings. Female hPSCs display remarkable variability of X-Chromosome Inactivation (XCI), a hallmark example of epigenetic gene regulation, and can be a powerful model system for investigating mechanisms of epigenetic instability. These cells frequently undergo epigenetic transformations and silence the XIST gene, the long non-coding RNA required for XCI and proper dosage of X-linked genes. Loss of XIST RNA is frequently observed in aggressive female reproductive cancers, yet mechanistic information addressing the role of XIST in cancer initiation or progression is unknown. Recently, I found that female hPSCs lacking XIST RNA exhibit cancer-related phenotypes: increased cell growth rates, reduced in vivo cellular differentiation potential, and altered coding and long noncoding RNA expression patterns. The objective of this study is to use female hPSCs as a model system to determine the causes and consequences of altered X-gene dosage. We are particularly interested in examining roles for known and novel long noncoding RNAs in these pathways. Long

ncRNAs are frequently misregulated in cancer yet lack important functional characterization. Our Specific Aims will be: 1) to determine a role for XIST RNA in preventing epigenetic instability and partial X-reactivation; 2) to identify how cancer/testes antigen genes (MAGEA2, A6) become reactivated in female hPSCs and the biological consequences of increased expression; and 3) the discovery of novel long noncoding RNAs that become misregulated when dosage compensation is compromised. The long-term goal of this project is to identify ways to prevent or correct alterations in dosage compensation that could be used for both regenerative medicine (to stabilize female hPSCs for clinical applications) and female reproductive cancer research.

Principal Investigator

John D. Gearhart, PhD
Director, Institute for Regenerative Medicine
University of Pennsylvania
Smilow Center for Translational Research
3400 Civic Center Blvd., Bldg. 421
Philadelphia, PA 19104

Other Participating Researchers

Montserrat C Anguera, PhD – employed by the University of Pennsylvania

Expected Research Outcomes and Benefits

Aim 1: This Aim will determine whether XIST RNA loss is the cause or consequence of epigenetic instability in female hPSCs, an important question in the human stem cell field. We expect that XIST RNA deletion will result in X-reactivation, and we will learn which regions of the X (and particular genes) are that prone to X-reactivation. We also expect heterochromatic markers of the inactive X to be affected, revealing novel regulatory mechanisms for XIST RNA specific for humans. We will use genetic and chemical approaches to reactivate XIST expression in female hPSCs to 'repair' low quality lines.

Aim 2: This Aim will examine the consequences of MAGEA gene reactivation, which are frequently overexpressed in a variety of cancers, aggressive breast cancers, and many XIST-negative female hPSCs. These experiments will determine whether the cancer phenotypes of these female hPSCs are attributable to MAGEA expression, and the epigenetic mechanisms responsible for their reactivation. The biological function of these genes is unknown, and we will also investigate putative roles in lineage commitment using hPSCs.

Aim 3: There are thousands of long ncRNAs yet only 15-20 have been functionally characterized. Long ncRNAs are poorly conserved between mouse and human, so we will identify human-specific long ncRNA expression signatures for XIST-positive (high quality) and XIST-negative (low quality) female hPSCs. We expect that some of these long ncRNAs will be robust markers for assessing X-dosage status and quality of hPSC lines. We will characterize

long ncRNAs misregulated in XIST-negative female hPSCs and investigate their expression and regulation in female cancer cells lines that also lack XIST RNA.

Summary of Research Completed

Since the initiation of funding of the project on August 1, 2013, we have made significant progress on the specific aims proposed in the strategic plan.

Aim 1:
Specific Aim 1 proposed to examine the contribution of XIST RNA expression from the inactive X in regulating gene expression (specifically silencing) of X-linked genes in human female pluripotent stem cells. Our approach is to disrupt XIST RNA expression using shRNAs in human female cells, then determine the fidelity of X-silencing in various cell contexts (cancer cells; pluripotent stem cells). We designed 2 different shRNA constructs for reducing XIST RNA levels: a. 2 distinct lentiviral shRNAs (cloned into both the psico-Puro-R and psico-eGFP constructs) designed against the highly conserved Repeat A region; b. the same two shRNA sequences cloned into a targeting vector for constitutive expression from the AAVS.1 locus on chromosome 19. We tested these constructs in 293T cells, an anueploid female cancer line with 2 inactive X-chromosomes coated with XIST RNA (Figure 1B). The reason we chose this line was because these cells are simple to transfect (often 80-100% transfection efficiency), and because these cells have multiple inactive X-chromosomes (increasing the probability of XIST reduction). We observed efficient XIST RNA knockdown using the lentiviral shRNA constructs using RNA FISH, where 2/3 clones exhibited dramatic reduction of XIST RNA clouds compared to uninfected cells (Figure 1A; clones 2 and 3 compared to clone 1 and uninfected 293). However, we found that the AAVS-targeted 293 cells did not disrupt XIST RNA as efficiently as the lentiviral infected 293 clones. We are currently using the XIST shRNA lentivirus to disrupt XIST localization (& reduce XIST expression) in human female cell lines with just 1 inactive X, including human female hiPSCs.

As part of Aim 1, we designed a novel system for initiating X-Chromosome Inactivation in human cells. All human female cell lines have undergone X-inactivation, and cannot be used to determine how the human active X becomes heterochromatic and transcriptionally silent. Our system has doxycycline inducible control of the endogeneous XIST gene, which is the first event in X-silencing and heterochromatin formation. Introduction of Xist transgenes or an inducible promoter upstream of Xist can initiate X-chromosome silencing in male mouse embryonic stem cells, which do not undergo X-chromosome Inactivation. Use of XIST transgenes can silence autosomes in human cell lines, suggesting that induction of XIST expression should be capable of silencing the human active X in male cells. We designed a donor construct for inserting a tetracycline response element with a CMV promoter upstream of the human XIST gene (deleting 800 bp of the XIST promoter) (Figure 2). We also generated 2 sets of TALE nucleases and 3 CRISPR constructs, designed to introduce double strand breaks in the deleted region, which would enhance homologous recombination at the XIST locus. Inducible expression from the tetO promoter requires expression of rtTA. Thus, for our inducible XIST expression system, we generated a series of human male & female cell lines (cancer cell lines and pluripotent stem cell lines) expressing rtTA from the AAVS.1 locus (using TALENs specific for the AAVS.1 intron)

(Figure 2B). First, we transfected the 293 cell line with XIST and rtTA constructs. This female cell line contains multiple active and inactive X-chromosomes, and we did not know which X would be preferentially targeted (or perhaps both). We obtained correctly targeted clones, and addition of doxycycline for 24hrs dramatically increased XIST RNA expression and the size of the XIST clouds by RNA FISH (Figure 2C, 2D). We also observed that we targeted the inactive X (which expresses XIST), because the number of XIST clouds did not change before/after transfection. Next we successfully targeted the active X in human male cells, using a placental cancer line (JEG-3) containing 1 X-chromosome (data not shown). We are currently targeting male and female human pluripotent stem cells with our system, to generate a developmentally relevant line where XCI takes place, and we anticipate obtaining correctly targeted clones in the near future. This system will allow us to investigate the molecular mechanisms of heterochromatin formation on the human X, and how silencing is altered during distinct developmental stages (pluripotent and after germ lineage formation).

Aim 2:
No progress was achieved for this aim during this reporting period.

Aim 3:
For the third Aim, we examined the transcriptional profile of human female pluripotent stem cells, comparing XIST-expressing lines with those that have irreversibly silenced the XIST gene. We used allele-specific high throughput RNA sequencing (RNA-Seq) to examine 5 isogenic hiPS lines (XIST+ or XIST-) and the HUES-9 XIST-expressing human embryonic stem cell line. We obtained roughly 50 million reads for each sample using the Ilumina HiSeq instrument at the BGI CHOP facility. We have finished aligning these reads to the Human Ref Seq database, working in collaboration with Dr. Kyoung Jae Won (UPenn, Genetics), an expert in bioinformatics analyses. First, we compared expression between XIST-expressing and XIST-lacking cell lines, and observed that the greatest difference with expression occurred at the X-chromosome. XIST-lacking female hiPSCs had greater numbers of reads mapping to the X, suggesting that the loss of XIST RNA expression resulted in reactivation of the silent X-chromosome (Figure 3A). We also observed a reduction in the number of reads mapping to chromosome 6 in cells lacking XIST RNA expression, and future experiments are investigating how XIST gene silencing could correlate with altered expression from this autosome. Next we developed a bioinformatics pipeline to identify novel long noncoding RNAs expressed in human female pluripotent stem cells. Unlike many of the recent publications profiling long noncoding RNA expression in human pluripotent stem cells, we did not exclude transcripts overlapping coding genes. We identified over 350 novel transcripts (not annotated in ENSMBL nor Ref Seq databases) expressed throughout the genome, with the largest chromosomes containing more novel transcripts (Figure 3B). Next, we examined the average length for the group of novel long noncoding RNAs, and found no correlation with chromosome size. Interestingly, we observed that the X and chromosomes 6 and 13 had the largest gene sizes for these novel long noncoding RNAs (Figure 3C). Finally, we sought to determine the most biologically relevant novel long noncoding RNAs in female pluripotent stem cells. Previously, we found that XIST RNA expression is a robust marker of stem cell quality, where cells lacking XIST expression are low quality and have cancer-related characteristics. We assembled a list of the top 10 candidate long noncoding RNAs (for future investigation) that were exclusively (or abundantly) expressed in

high quality, XIST-expressing lines (Figure 3D). We are now validating the expression of these transcripts using qRT-PCR, and will characterize their function in human stem cells in the future.

Figure 1

A.

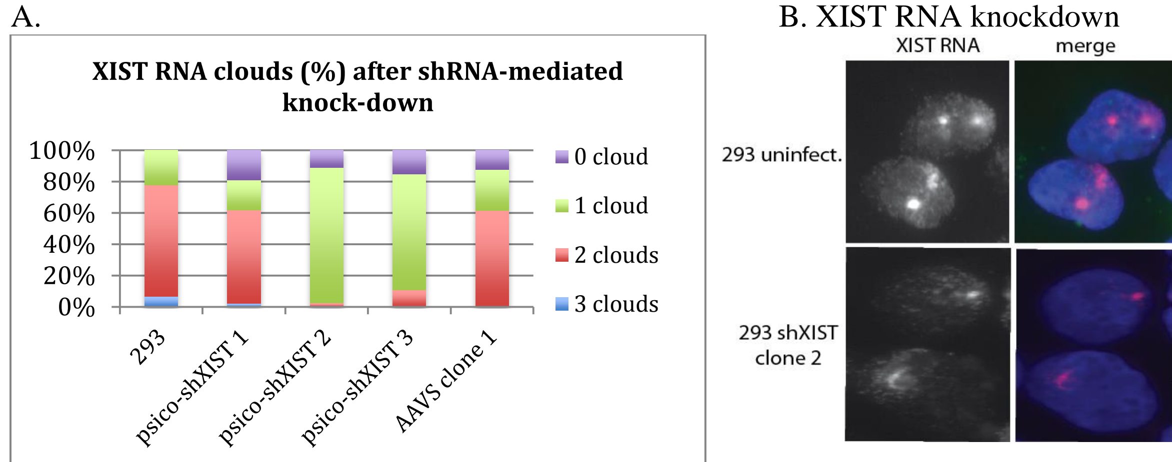
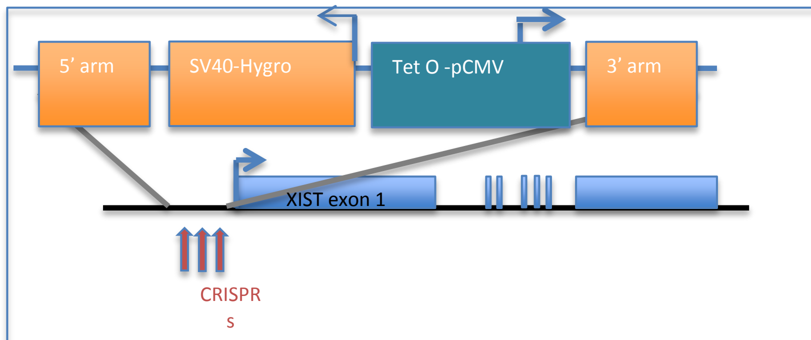
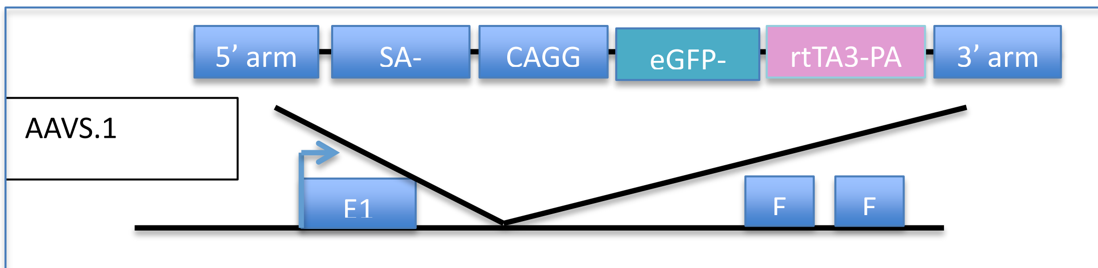


Figure 2

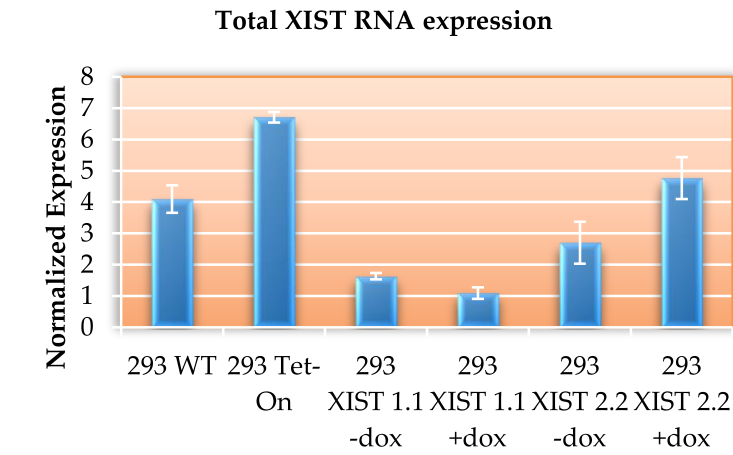
A. Targeting strategy for inducible promoter upstream of XIST gene

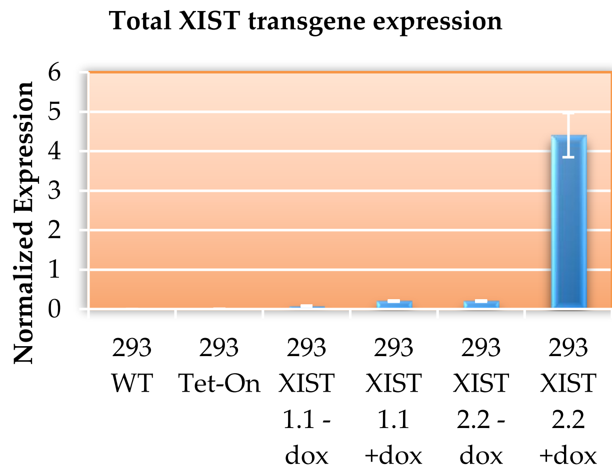


B. Targeting strategy for rtTA expression from AAVS1 locus



C. XIST RNA upregulation with doxycycline treatment in 293 cells





D. Doxycycline induced XIST RNA upregulation generates larger XIST clouds.

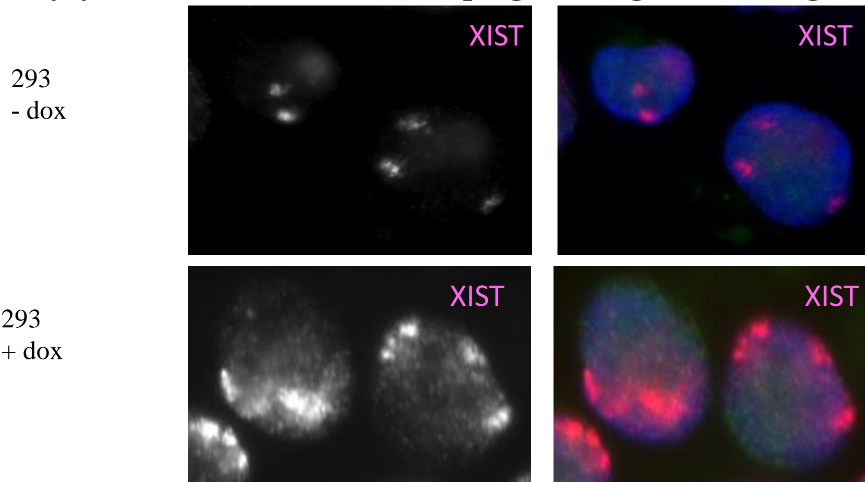
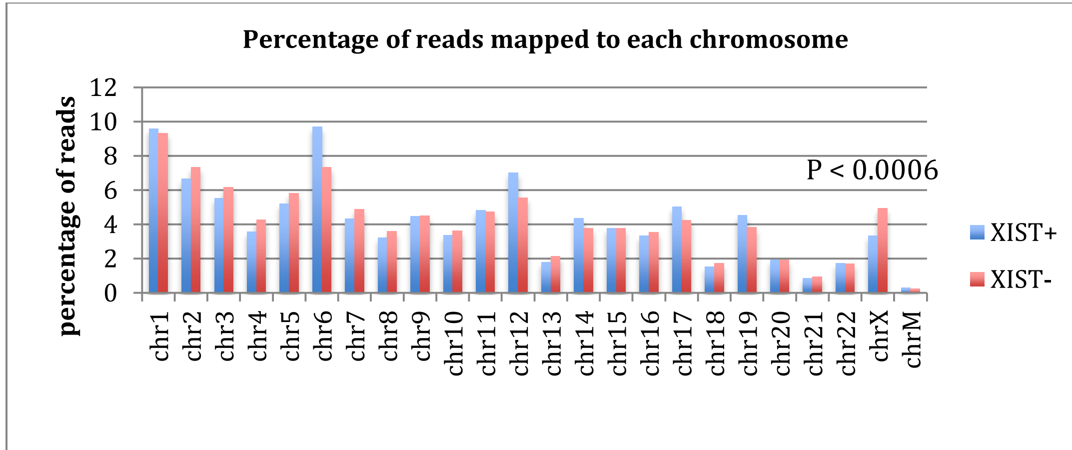
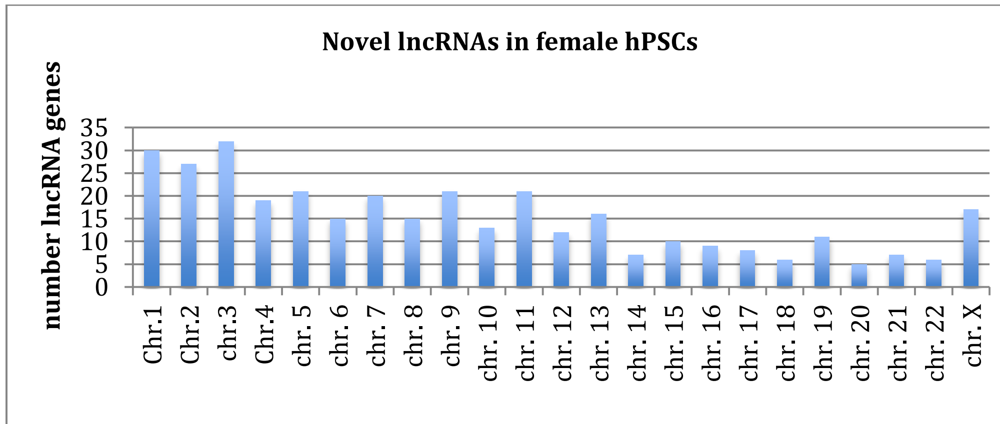


Figure 3

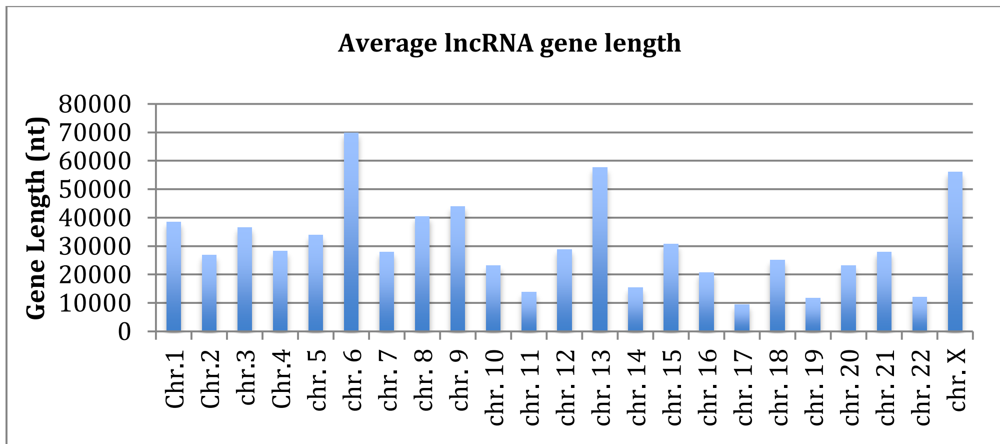
A.



B.



C.



D. Novel long noncoding RNAs (top hits) unique to high quality female hESCs/hiPSCs

Symbol	Chr.	hiPS-2 XIST+(FKPM)	hiPS-2 XIST- (FKPM)	HUES-9 XIST+ (FKPM)
SCARNA22	4	423.316	0	730.433
AP000459.4	21	54.689	0	166.787
AC004696.2	19	30.5662	0.0115084	199.591
CTD-2554C21.2	19	31.9814	0	15.5157
CTC-228N24.1	5	9.03407	0.716773	12.257
RP11-219G10.3	4	9.71603	1.23972	29.8745
CTD-2554C21.3	19	16.4063	0	7.90883
RP4-755D9.1	X	6.24138	0.970557	10.5533

Research Project 11: Project Title and Purpose

Optimizing Production of Cardiac Myocytes from Stem Cells and by Direct Reprograming – The purpose of this study is to enhance understanding of the production of cardiac muscle to replace cardiac muscle lost in the setting of myocardial infarction (heart attack). Recent work in mouse models indicates that defined combinations of transcription factors can convert connective tissue and fibroblasts into cardiac muscle cells. We propose to extend this work to human tissue and to explore the role of small molecules in enhancing this process. We will also extend to humans preliminary findings in mice that a gene called *Hopx* specifically identifies cardiac muscle precursor cells. This could allow for the isolation and purification of human cardiac precursors from stem cells to be used for cell-based therapy of myocardial damage.

Anticipated Duration of Project

8/19/2013 – 12/31/2014

Project Overview

The overall goal of this project is to enhance the ability to provide effective regenerative therapies for humans with heart failure and myocardial infarction. Recent evidence suggests that a low level of myocardial turnover and generation of new cardiac myocytes occurs in humans throughout life, and enhancement of this natural regeneration process might be an effective therapeutic strategy for heart diseases. Exciting recent work indicates that fibroblasts can be directly converted into functional cardiac muscle in mouse models by the use of a combination of several cardiac-specific transcription factors. Small molecules and micro-RNAs may also enhance this process. However, the precise factors that have been used successfully in mice are not sufficient to convert human cells. Additionally, work from our own laboratory indicates that,

in mice, the *Hopx* gene identifies cardiac precursor cells that are committed to the cardiac myocyte lineage. Thus, *Hopx* serves as a very early and specific marker of cardiac myoblasts that are restricted to the cardiac muscle lineage and do not produce other cell types. Such a marker would be very useful for human cell therapy approaches for cardiac disease in order to reduce the risk of tumor or teratoma formation resulting from inappropriate growth and differentiation of transplanted cells. Specific aims of this proposal include:

1. Adapt and optimize direct reprogramming protocols to convert fibroblasts to cardiac myoctyes from mouse to human.
    a. Using inducible lentiviral vectors, test combinations of candidate cardiac transcription factors for ability to convert adult human ventricular cardiac fibroblasts to cardiac muscle cells capable of cardiac troponin and myosin expression, contraction, and calcium oscillations.
    b. Optimize the use of a recently developed cardiac-specific GCaMP calcium reporter to detect and quantify conversion of human fibroblasts to cardiomyocytes.
    c. Purify and characterize soluble factors implicated by preliminary studies that enhance efficiency of reprogramming.
2. Determine if *HOPX* identifies human precursor cells that are committed to the cardiac muscle lineage using NIH approved human ES cell lines in which the *HOPX* locus will be tagged with a fluorescent marker.
    a. Utilize a TALEN-based gene targeting approach in human ES cells (the NIH approved MEL1 cell line) to insert the fluorescent tomato reporter in place of the coding region (ATG).
    b. Using standard hES differentiation protocols, determine if hES derivatives that express tomato from the *HOPX* locus are committed to the cardiomyocyte lineage. Compare results to the available NKX2-5GFP hESC cell line (MEL1) in which GFP-expressing cells are multipotential.

Principal Investigator

John D. Gearhart, PhD
Director, Institute for Regenerative Medicine
University of Pennsylvania
Smilow Center for Translational Research
3400 Civic Center Blvd., Bldg. 421
Philadelphia, PA 19104

Other Participating Researchers

Jonathan A. Epstein, MD; Russell Addis, PhD; Jamie Ifkovits, PhD; Rajan Jain, MD; Deqiang Li, PhD – employed by the University of Pennsylvania

Expected Research Outcomes and Benefits

Specific outcomes of this project include determination of whether human fibroblasts can be

directly converted into functional contractile cardiac myocytes, and whether the *HOPX* gene can be used as a marker of stem cells that have been induced to differentiate into cells that are committed to the cardiac muscle lineage. Each of these outcomes, if successful, would benefit and enhance efforts to develop effective regenerative approaches for heart disease and myocardial infarction. Animal studies suggest that it may be possible to use genes and small molecules to transform fibroblasts that comprise the scar tissue formed as a result of a heart attack into regenerated functional cardiac muscle. In order to translate these findings to humans, the type of work presented here is critical. Other data suggests that human-derived stem cells might be used to generate new heart muscle cells that could be injected into damaged hearts (cell-based therapy) but a major risk of this approach is the possible formation of tumors or teratomas derived from inappropriate growth or differentiation of the injected cells. The identification of a human gene or marker that identifies cells committed to become only heart muscle would provide mechanisms to minimize this risk.

Summary of Research Completed

Specific Aim 1:

The goal of these aims were to optimize the direct reprogramming protocols with respect to mouse and human cells by testing a variety of combinations of transcription factors and to improve upon the reporter that is used to identify the converted cells.

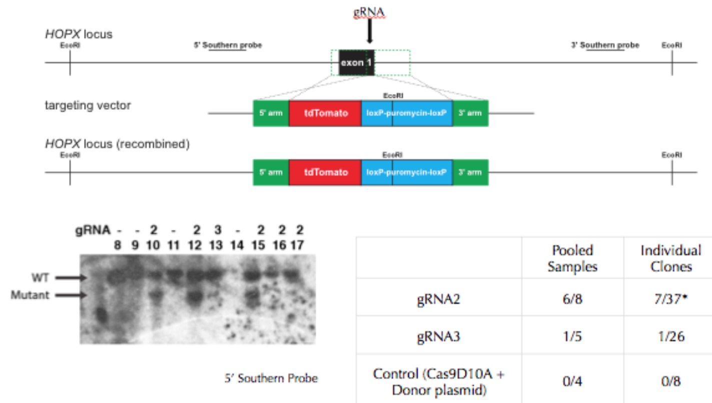
As a result of tests employing a large number of different combinations none proved more efficient than Hand2, Nkx2.5, Gata4, Mef2c, Tbx18 and Msx2 for the conversion of the mouse cells. The human cell conversions continue to be problematic. Human cardiac fibroblasts are more difficult to grow in culture, but most importantly, the various transcription factors (and their combinations) just are not effective in converting the human cells. This result is proving typical for all other investigators in this field and we must turn our attention to identifying 'new' factors to try for the human cells. It is not surprising that there would be species-specific factors involved in cell differentiation.

Specific Aim 2:

The Epstein lab has made significant progress with this aim. Given the rapid advances in the field of genome engineering, our approach has changed slightly from what was originally proposed. We initially proposed using TALENs to target the human *HOPX* locus, but our preliminary experiments, which compared TALENs to the more recent CRISPR approach suggested much higher targeting frequency at this locus with CRISPRs, and we have therefore pursued this approach. The goals of the experiments are unchanged even though the techniques for targeting the genomic locus have been modified.

The strategy that we have used is shown in Figure 1. In collaboration with Nathan Palpant and Charles Murry at the University of Washington, Seattle, we have successfully obtained human ES cells with tdTomato inserted in the *HOPX* locus. We tested 2 different guide RNAs and both yielded correctly targeted clones, though at different frequencies (Figure 1). One clone was

expanded, and TAT-Cre protein (which is readily taken up into cells) was used to induce Cre-mediated recombination and removal of the loxP-flanked puromycin resistance cassette. This was necessary because we did not detect robust tdTomato expression when this antibiotic resistance cassette remained in place, as has been reported by others. However, after removal of this cassette, targeted ES cells that were induced to differentiate into cardiac myocytes robustly expressed tdTomato (Figure 2). Thus, the major goals of Aim 2 have been achieved, although further validation is underway.



	Pooled Samples	Individual Clones
gRNA2	6/8	7/37*
gRNA3	1/5	1/26
Control (Cas9D10A + Donor plasmid)	0/4	0/8

Figure 1: Targeting strategy for CRISPR-mediated insertion of tdTomato into the human *HOPX* locus. A Southern blot is shown bottom left revealing multiple correctly targeted clones. Two different guide RNAs were used, and the targeting frequencies are shown in the table bottom right.

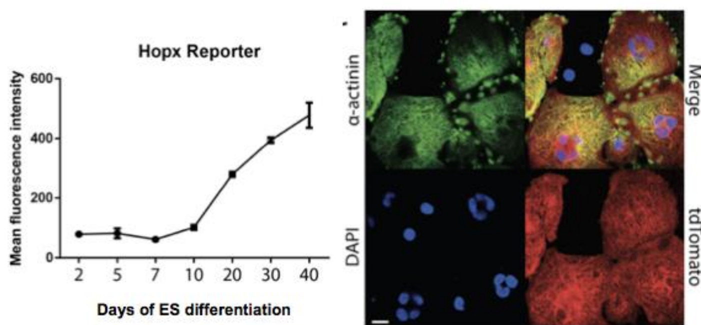


Figure 2: Targeted hES cells express tdTomato from the *HOPX* locus upon cardiac myocyte differentiation. At approximately 10 days of culture under differentiation conditions, tdTomato activity is activated and continues to increase through day 40. Differentiated cells express both alpha actinin (right panels) and tdTomato, indicating co-expression in cardiac myocytes. Cells that do not express alpha actinin do not express tdTomato.