Fox Chase Cancer Center

Annual Progress Report: 2010 Formula Grant

Reporting Period

January 1, 2011 – June 30, 2011

Formula Grant Overview

The Fox Chase Cancer Center received $2,851,328 in formula funds for the grant award period January 1, 2011 through December 31, 2012. Accomplishments for the reporting period are described below.

Research Project 1: Project Title and Purpose

The Role of Histone-Complexes in Residue Specificity of Post-Translational Modifications - The purpose of this project is to develop the technology to determine the impact of histone presentation on the location(s) of post-translational modification of histones. *In vivo*, histones are predominately either bound to DNA (nucleosome) or bound to a histone chaperone. While there are clear correlations between changes in histone modifications and histone chaperones, their relationship is still unclear. Histone chaperones could alter post-translational modifications by interacting with the modification enzyme, protecting specific residues, or by increasing their apparent concentration by disrupting specific complexes. We aim to determine the effects of histone complexes on the location and identity of histone modifications by developing mass spectroscopy based assays to monitor multiple modifications at once.

Anticipated Duration of Project

1/1/2011 - 12/31/2012

Project Overview

The level of DNA compaction in the cell is equivalent to reducing the Empire State Building to the size of a nickel. The first level of this compaction is the nucleosome; the repeating structural unit in all eukaryotic chromatin. The nucleosome consists of an octamer of two copies each of the four histone proteins H2A, H2B, H3, and H4, around which 147 base pairs of DNA are wrapped. Therefore, histones are ultimately responsible for compacting the chromosomal DNA, and present a major obstacle for proteins needing to access the DNA.

Outside of the nucleosome, histones are primarily bound by histone chaperones. Histone chaperones function as histone transporters and to prevent non-nucleosomal histone-DNA interactions, thus facilitating nucleosome assembly. There is an emerging link between histone modification enzymes and histone chaperones and a multitude of histone post-translational modifications (PTM) are implicated in regulating DNA accessibility. Both the location and

identity of some histone PTMs have been shown to change in cancer. It is thought that histone-PTMs act by redistributing histones between nucleosomes and histone chaperones but histone-complexes can alter the locations of PTMs on a given histone. Determining how various histone complexes alter the location of PTMs is the one step in understanding the complex regulation of histone/chromatin homeostasis and its impact on gene expression.

The goals of this project are to understand: (1) how histone complexes influence which residues are post-translationally modified, and; (2) what factors lead to this specificity. To answer these questions we must develop and improve methods to simultaneously monitor the identities of the modification, and all the sites that are modified. These new methods will allow us to achieve a comprehensive and exhaustive analysis of histone modifications and will require the extensive use of a TSQ quantum access mass spectrometer to be purchased with this funding. We will use recombinant histones and histone chaperones as substrates for either recombinant enzymes (acetylases, methylases, kinases etc.) or cell lysate. This will provide a new look at the histone modifications by studying the enzymes that are the most active/abundant in the cell and the optimal substrate (free histone, histone-chaperone complex or nucleosomes).

Principal Investigators

Andrew J. Andrews, PhD
Assistant Professor
The Fox Chase Cancer Center
333 Cottman Avenue
Philadelphia, PA 19111

Other Participating Researchers

Yibai Chen, MD - employed by The Fox Chase Cancer Center

Expected Research Outcomes and Benefits

This project has the potential to produce a broad range of benefits for cancer research and treatment. The use of selective reaction monitoring (SRM) for histone modifications allows us to study which histone residue is being modified in a complex mixture. With this methodology, we can ask specific questions about what factors/conditions induce site specific modifications and how cancer drugs might alter the landscape of histone modifications and thus the epigenome of cancer cells.

There is a well established link between cancer and changes in global histone modification patterns. However, there are multiple modification enzymes that exist in multiple different complexes. Our methods will allow us to identify which complexes are responsible for these changes in specific modification patterns. This should lead to the development of screens to target these particular complexes.

Summary of Research Completed

In the first six months of the project we have accomplished the following: 1. Purchased the Waters Acuity H-class UPLC attached to a Quantum Access triple quadrupole mass spectrometer (QqQ); 2. Completed required lab renovations (custom bench, venting, power, etc); 3. Optimized protocols for chemical derivatization of histones; and 4. Determined the mass transitions for measuring acetylation and methylation for eleven of the twelve lysines in H3.

The difficulty in studying post-translational modification of histones is due in large part to the complexity arising from enzymes modifying more than one location on a single histone. Modification of these distinct locations can result in large differences in gene expression. For example, K14 acetylation is directly involved in gene activation while acetylation of K9 blocks methylation of this residue, preventing gene suppression. To ascertain both the location and extent of acetylation within histones, we developed a triple quadrupole mass spectrometer (QqQ) based assay. Briefly, in a QqQ, the first quadrupole is set to filter a parent ion of interest (i.e. our acetylated histone peptide), the selected ion is fragmented in the second quadrupole (collision cell), and the third quadrupole is set to filter for a specific fragment ion. This type of analysis is termed "selective reaction monitoring" (SRM). The dual filtering approach results unmatched sensitivity and selectivity, and the non-scanning nature of the analysis extends the quantitative dynamic range over five orders of magnitude (compared to 2-3 orders of magnitude for conventional full-scan analysis). SRM experiments can be designed to monitor multiple sites of acetylation in a high through-put manner, and can generate relative or absolute quantitation when combined with an isotopically labeled standard. This technique is widely used for the study of protein modifications *in vivo* due to its high specificity for an analyte of interest within a complex sample.

Our basic experiment begins with a standard histone modification assay. After the samples are quenched they are propinylated by propionoic anhydride and digested with trypsin. The propinylated histone fragments have a longer retention time than acetylated or unmodified peptides. This allows for greater resolution because we have both the retention time and the SRM to confirm acetylation. We designed 74 SRM transitions that identify every acetylated or propionylated lysine residue in H3, to determine the modification state (another 74 methylated/proinoylated) of each lysine. Currently, we observe the expected transitions for all peptides but the one containing K115, this is likely due to larger m/z and methionine oxidation. We have purchased acetylated and non-acetylated peptides to validate the assays; we will also treat these peptides with isotopically labeled propinyl anhydride to allow absolute quantification.

The synthetic peptides (all but K115) are in good agreement with both the predicted parent and product ions as well as the observed retention times on the C18 column. Our current detection level is <1% of total protein concentration, providing the sensitivity level of radiochemical assays. This level of sensitivity is sufficient to begin our initial set of enzyme assays using recombinant Gcn5 as proof of principle. Using the QqQ based assay to measure modification of Gcn5, we observe acetylation at K14, K9, K23, K18, K37, K36, and K27 (in order of modification). We are currently working on optimizing the assay to increase both sensitivity and throughput.

In summary, we have developed a powerful label free histone modification assay that can be used for both *in vitro* enzyme kinetics assays and to monitor changes from histones taken from *in vivo* samples. This allows us to develop detailed models of the specificity of histone modifications and test these models *in vivo* using the same assay. In addition to the project described here, this one assay will be able to be used for both drug screens and to evaluate patient response.

Research Project 2: Project Title and Purpose

DUSP6 Regulates the Response to EGFR Inhibitors in Cancer - Much effort has been devoted to the development of drugs targeted at signaling proteins thought to be essential for tumor cell viability. A central disappointment has been the fact that many of these inhibitors have limited efficacy in cancer patients in spite of potently inhibiting their drug targets. DUSP6 is a cytosolic member of the family of dual specificity phosphatases, a group of proteins that regulate ERK1/2 and p38. We have found that depletion of DUSP6 significantly increases the apoptosis caused by Epidermal Growth Factor Receptor (EGFR) blockade. Although DUSP6 expression is associated with poor prognosis in cancers, the mechanism for DUSP6 action is not well understood. The purpose of this project is to discover the mechanism by which DUSP6 regulates tumors responsiveness to clinically important EGFR antagonists.

Anticipated Duration of Project

1/1/2011 - 12/31/2012

Project Overview

Our immediate objective is to characterize the mechanism by which DUSP6 regulates tumor responsiveness to EGFR-targeting drugs. DUSP6 is known to be a cytosolic phosphatase for mitogen-activated protein kinases ERK1/2 and p38. Although changes in expression of DUSP6 have been reported as associated with specific prognoses in cancers, no clear mechanism for DUSP6 action has been established. Using stably integrated shRNA, we recently found that silencing DUSP6 markedly increased the cytotoxicity of the EGFR antagonists cetuximab, erlotinib and dasatinib, but not of inhibitors of RAF, MEK, PI3K, and PKC. Subsequent phosphoproteomic analysis of 46 candidate DUSP6 dephosphorylation targets in A431 cells treated with erlotinib revealed markedly increased phospho-p38 and p-CHK2 (activated) proteins in DUSP6-depleted cells. We thus hypothesized that DUSP6 regulates apoptosis commitment in EGFR-dependent cancers via regulation of the G2/M checkpoint. We will continue to investigate this hypothesis in the following Specific Aims:

Aim 1. Investigate the role of DUSP6 in regulation of cell growth and the G2/M checkpoint. We will probe whether DUSP6-regulated activation p38 regulates the cell cycle via the ATM-Chk2 pathway, and if DUSP6-mediated anti-EGFR sensitivity is dependent on p53 status in the tumor cell. We have created all necessary tools, including cell lines stably expressing inducible DUSP6-targeting shRNA, and overexpression plasmids for wild type, truncated and catalytically inactive mutants of DUSP6, to directly study this mechanism.

Aim 2. Model the DUSP6-p38-CHK2 interaction network. We will perform systematic analyses of the interactions between DUSP6 and the MAPK signaling network. We hypothesize that genetic and physical interactions will be conserved between the yeast and human orthologs at the level of functional modules. We will define and evaluate DUSP6-associated functional modules for the effects on cell cycle checkpoints, growth and viability in DUSP6-silenced or DUSP6-overexpressing cells.

Aim 3. Investigate the impact of DUSP6 expression on outcomes in head and neck squamous cell carcinoma (HNSCC). We wish to determine the relevance of our results to the clinic. Elevated DUSP6 levels have been associated with poor outcomes in some tumor types. We will establish whether EGFR-targeted therapies, and use of radiotherapy, which induce cell cycle checkpoints correlate with DUSP6 expression in HNSCC, with the goal of generating a biomarker for therapy selection.

Principal Investigator

Igor Astsaturov, MD, PhD
Assistant Professor
The Fox Chase Cancer Center
333 Cottman Ave
Philadelphia, PA 19111-2434

Other Participating Researchers

Tetyana Bagnyukova, PhD, Illya Serebriiskii, PhD - employed by The Fox Chase Cancer Center

Expected Research Outcomes and Benefits

This research will establish DUSP6 as a novel regulator of cancer cell transition through the G2/M phase of the cell cycle, and will link this mechanism to the sensitivity of tumors to inhibitors of EGFR. We predict this mechanism is of particular relevance to head and neck squamous cell carcinomas harboring mutant p53 (associated with treatment refractoriness and metastatic behavior), in the context of cell treatment with EGFR-targeting inhibitors. Such a possibility is supported by our preliminary results showing significantly increased p38 and ATM/CHK22 activity, impaired growth kinetics and delayed G2/M transition following DUSP6 depletion. Conversely, we expect that increased expression of DUSP6 will confer resistance to EGFR targeting or chemotherapy agents with specific activity in G2/M phase of the cell cycle (inhibitors of topoisomerase I and II, cisplatin, doxorubicin, or ionizing radiation). These agents are commonly used to treat squamous cell carcinomas of the head and neck, and other cancers. By delineating the mechanism of treatment-sensitizing activity of DUSP6 inhibition, we ultimately hope to establish DUSP6 as a prognostic or predictive biomarker that will allow the more effective stratification of patients for treatment with different therapeutic regimens. Success in this goal would both improve cancer survival and reduce treatment costs based on more appropriate alignment of patients with the most appropriate therapeutic choices.

Summary of Research Completed

1. *Long-term DUSP6 depletion sensitizes A431 cells to EGFR inhibitors and DNA-damaging drugs.*

To address the question on the mechanisms underlying altered sensitivity of siRNA-depleted cells to EGFR inhibitors and/or chemotherapy, we made stable cell lines with depleted DUSP6 and DUSP7. We confirmed gene silencing by our shRNA constructs at the level of mRNA and protein (Fig. 1A). Long-term depletion of DUSP6, but not DUSP7 increased sensitivity of the A431 cells to EGRF inhibitors erlotinib and cetuximab (Fig. 1B), and caused enhanced apoptosis (Fig. 1C) as measured by surface labeling with Annexin V. Such effect of DUSP6 silencing may be associated with deregulation of the DUSP6 downstream targets, or predispose cells to apoptosis via a non-EGFR specific mechanism. To examine whether this effect is EGRF-specific, we tested sensitivity of shRNA-transduced A431 cells to a panel of signaling inhibitors and a DNA-damaging agent, CPT11, which inhibits topoisomerase I. We found that DUSP6 depletion sensitized A431 cells to dasatinib, a SRC/ABL/EGFR inhibitor, and to inhibitors of Aurora A and B kinases (Fig. 1D). Synergistic action of simultaneous inhibition of EGFR and Aurora kinases has been shown earlier and can be partly explained by transcriptional regulation of AURKA by Erk2. DUSP6-depleted cells also were sensitized to CPT11 suggesting DUSP6 activity on mechanisms not necessarily linked to the EGFR signaling. This is in keeping with the previously published observations of the general pro-survival function of several dual specificity phosphatases in cancer cells. To search for changes in different signaling pathways and possible targets for DUSP6, we performed phosphoproteomic analysis of 46 proteins involved in regulation of cell proliferation, apoptosis or survival.

2. *DUSP6 depletion is associated with activated p38 and the checkpoint kinase Chk2.*

To this point, the mechanism for pro-survival function of dual specificity phosphatases has been unclear. We argued that silencing DUSP6 with shRNA may increase levels of phosphatase substrates or activate pathways regulated by DUSP6. To evaluate the overall activity of the major canonical signaling pathways, we performed evaluation of phosphoproteome using a panel of validated antibody reagents.

We used Human Phosphokinase Proteome Profiler Array (R&D Systems) according to the manufacturer's protocol. A431 cells stably expressing control, DUSP6, or DUSP7 shRNA were grown in DMEM supplemented with 1% FBS for 48 hours (h), then treated with 1 μM erlotinib or vehicle alone for 3 h, and cell lysates were collected. Quantification was done with ImageJ software. For some proteins, their phosphorylation status was confirmed by Western blot. In our experiments, we established that shDUSP7 had no effect on drug sensitivity, thus serving as a negative control allowing us to discern DUSP6-specific targets from pathways regulated by both phosphatases regardless of the treatment with erlotinib (Fig. 2A).

Treatment with the EGFR inhibitor erlotinib for 3 h reduced phosphorylation of GSK3A/B and AKT in DUSP6-silenced cells and in a larger set of phosphorylated targets (P70S6K, PYK2, PLCγ, Paxillin, beta-catenin, ERK1/2, GSK3B and AKT) in DUSP7-silenced cells (Fig. 2B). Downregulation of pERK and pAKT in response to anti-EGFR in A431 cells has been

previously established in our laboratory. Since DUSP7-depleted cells showed little or no difference in sensitivity to anti-EGFR or DNA-cytotoxic agents (Fig 1B,D), we concluded that decreased phosphorylation of P70S6 kinase (T389), Pyk2 (Y402), PLCgamma-1 (Y783), β-catenin, and paxillin (Y118) (Fig. 2B) in these cells were not critical for drug sensitivity in our model. Changes in several phosphoepitopes of AMPKa1, MEK1/2, and MSK1/2 were similar in both sensitized shDUSP6 and non-sensitized shDUSP7 cells compared to control cells under basal conditions (Fig. 2A) or unchanged under erlotinib treatment and therefore were not pursued further.

Among several phosphoepitopes uniquely and consistently upregulated in DUSP6-silenced cells under basal or erlotinib-treated conditions were MAPK14 (p38 MAP kinase) and CHEK2, a checkpoint kinase. We further confirmed by Western blot that both p38 and CHEK2 had higher level of phosphorylation in DUSP6-silenced cells, and this status did not depend on the erlotinib treatment (Fig. 2C, D). Although MAPK14 is not considered a DUSP6 substrate , its close homolog, DUSP7, can bind and dephosphorylate p38 and JNK. Another possibility may be that increased p38 phosphorylation may be regulated indirectly. In this context, both MAPK14 and CHEK2 proteins are involved in the G2/M checkpoint regulation, thus suggesting a possibility that increased chemosensitivity upon DUSP6 silencing can be mediated via cell cycle regulation.

3. *Loss of DUSP6 is associated with slower growth and delayed G2/M cell cycle progression.*

To examine the possibility that DUSP6 might be involved in cell cycle regulation, we first assessed the growth rates of shRNA-modified cell lines in 10% FBS-DMEM. We noted that DUSP6-depleted cells grew much slower than control and DUSP7-silenced cells (Fig. 3A). Morphological evaluation of the DUSP6-deficient cells showed reduced cell size (not shown), confirmed also by flow cytometry.

To evaluate the effects of DUSP6 loss on cell cycle progression, we synchronized shRNA-modified A431 cells in the G1/S transition using double thymidine block. Cells plated in 6-well plates were synchronized with 2 mM thymidine for 16 h. Then the thymidine was washed out and cells were supplemented with fresh medium for 10 h before adding thymidine. After the second 16-h incubation with thymidine, cells were released from the block and collected at the indicated time points. Cells were fixed and stained with propidium iodide, and the DNA content was analyzed using the FACS flow cytometry (Fig. 3B). We found that DUSP6-depleted cells have significantly prolonged mitoses.

We used an alternative method to characterize the time required for completing mitosis using time-lapse video (Fig. 3C). The number of frames was used to calculate the length of mitosis for 50 individual cells. We further investigated if activation of the G2/M checkpoint and delayed mitosis are the result of persistent DNA open ends generated during the S-phase which may lead to G2/M checkpoint activation via ATM/CHEK2, ATR/CHEK1 or p38/MK2 pathways.

4. DUSP6 depletion activates ATM/CHEK2 and p38/MK2, but not ATR/CHEK1.

In our phosphoproteome survey, we found that DUSP6-silenced cells had higher level of phosphorylated CHEK2 (Fig. 2C), but not CHEK1. ATM is a sensor kinase upstream of CHEK2

and is activated in response to DNA damage. Similar to CHEK2, ATM was activated in DUSP6-silenced cells, and the level of its phosphorylation was significantly higher than in control or DUSP7-depleted cells (Fig. 4A). Histone H2AX is another target of ATM and is phosphorylated early after DNA damage. Unexpectedly, without any treatment with cytotoxic agents, DUSP6-depleted cells had higher level of phosphorylated histone H2AX (pH2AX) as shown by quantitative immunofluorescence analysis (Fig. 4B, C). These data suggest that CHEK2 activation occurred due to an upstream event, thus warranting further studies in the role of DUSP6 in regulation of cell cycle.

Figure 1. DUSP6 depletion sensitizes cells to multiple drugs. (A) Validation of knockdown by qPCR and WB. **(B)** Sensitivity of DUSP-depleted A431 cells to anti-EGFR antibody cetuximab. **(C)** DUSP6 depletion enhances apoptosis of A431 cells treated with 1 μM erlotinib as measured by AnnexinV surface labeling on Guava flow cytometer. Averages of three independent experiments are shown; asterisk indicates statistical differences with erlotinib-treated control cells, $P < 0.05$. **(D)** DUSP6 depletion sensitizes A431 cells to inhibitors of the EGFR (erlotinib), SRC/EGFR/ABL kinases (dasatinib), and a topoisomerase I inhibitor CPT11 over a broad range of concentrations. In **(B)** and **(D)**, cells were incubated with indicated concentrations of inhibitors for 72 h. The viability was measured by the Alamar blue method. Asterisks indicate significant differences with control cells, pound signs - between DUSP6- and DUSP7-depleted cells, $P < 0.05$.

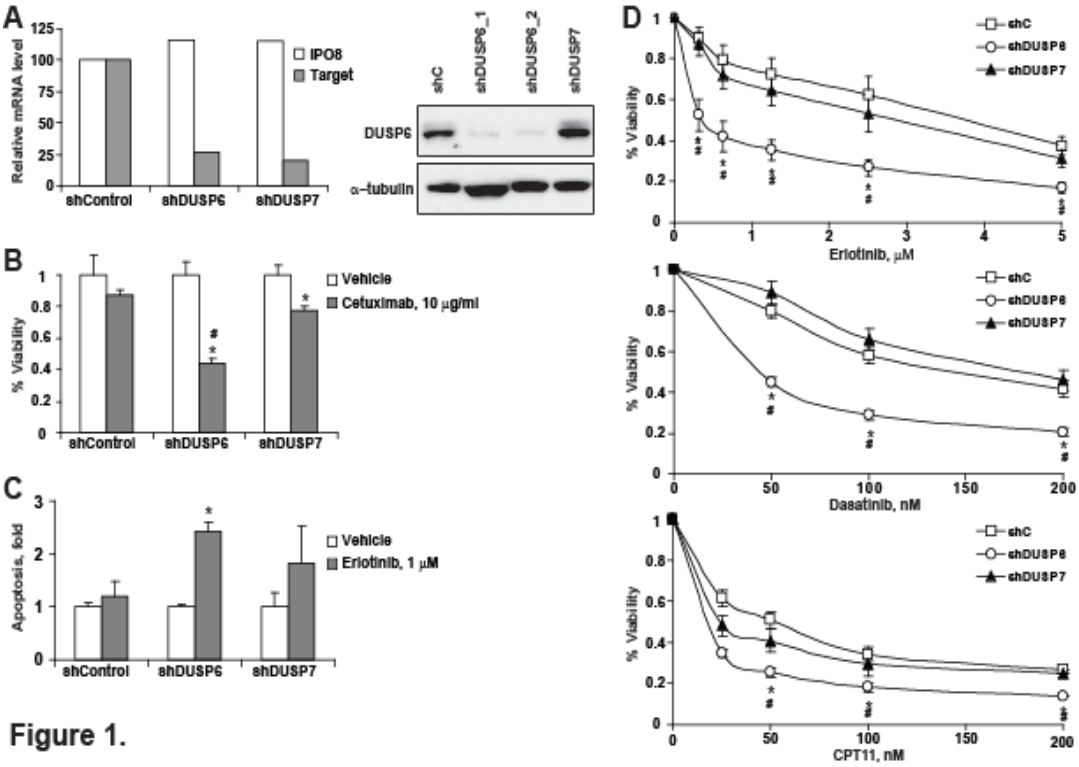


Figure 1.

Figure 2. DUSP6 depletion increases phosphorylation of p38 and CHEK2.
(A) Phosphoproteomic analysis of 46 signaling proteins in A431 cells stably expressing shDUSP6 or shDUSP7. Cells were grown in 1% DMEM and treated with vehicle (left panel) or 1 mM erlotinib (right panel). The data are relative increase in phosphorylation signal. Proteins shown in red have consistent factor of ≥2 increase in signal intensity in independent biological replicates, in blue ≤2 decrease. **(B)** Relative change in phosphorylation epitopes in A431 cells treated with erlotinib compared to vehicle. The proteins are listed in the same order in (A). **(C)** Phosphorylation of P38 and CHEK2 was confirmed by Western blot. **(D)** Averaged results of three independent quantitated Western blot measurements. Asterisks, $P < 0.05$ compared to shControl, pound signs -between DUSP6 and DUSP7-depleted cells.

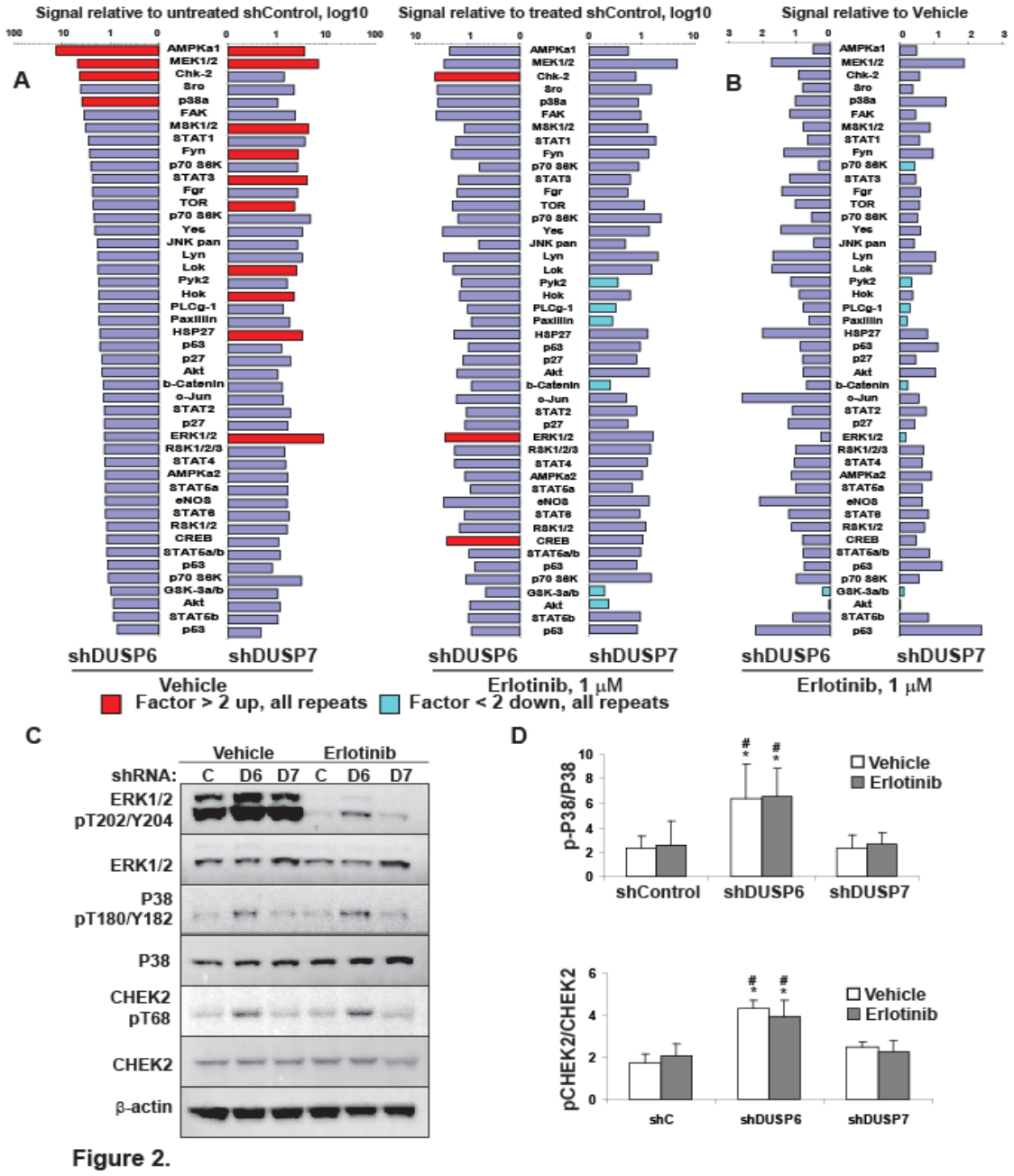
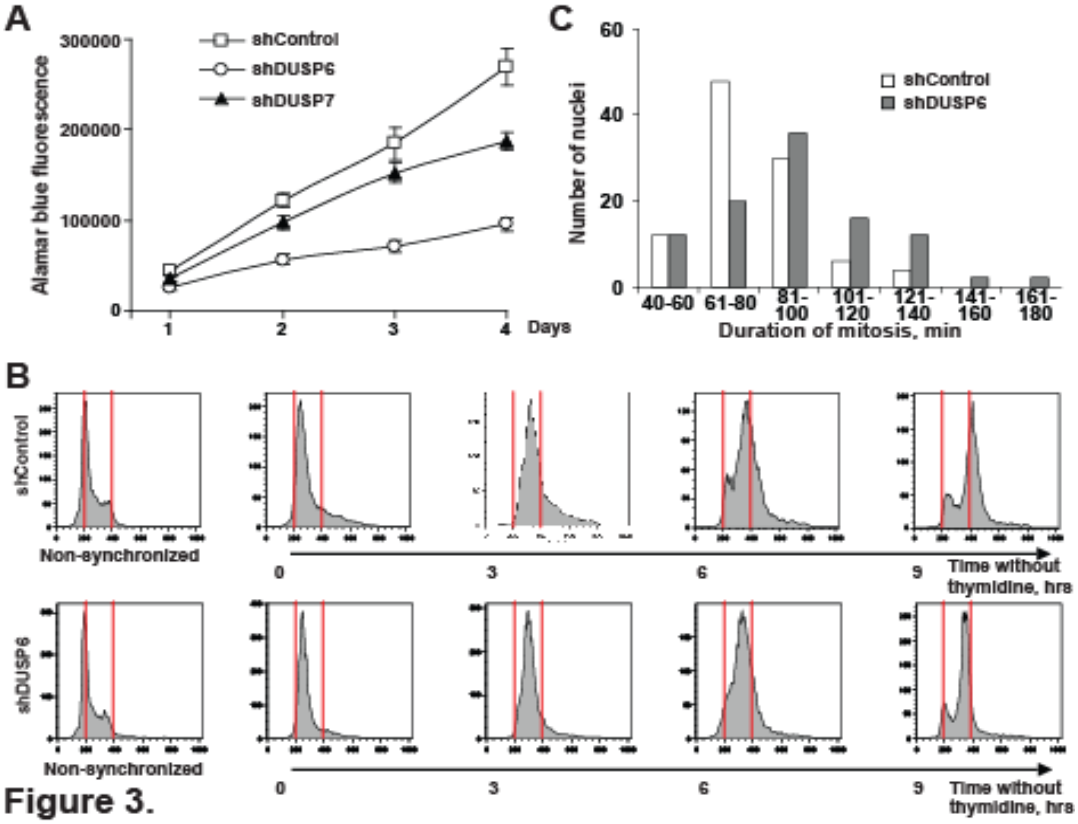


Figure 2.

Figure 3. Depletion of DUSP6 delays cell cycle progression.
(A) The growth rates of shRNA-modified assessed A431 cells was measured in full 10% FBS/L-Glutamine-supplemented DMEM.
(B) Synchronized shRNA-modified A431 cells were plated in 6-well plates and synchronized with 2 mM thymidine twice for 16 h. After the second 16-h incubation with thymidine, cells were collected at the indicated time points, fixed and stained with propidium iodide, and the DNA content was analyzed using the FACS flow cytometry.
(C) The time required for completing mitosis was estimated using time-lapse video. The number of frames was used to calculate the length of mitosis for 50 individual cells.



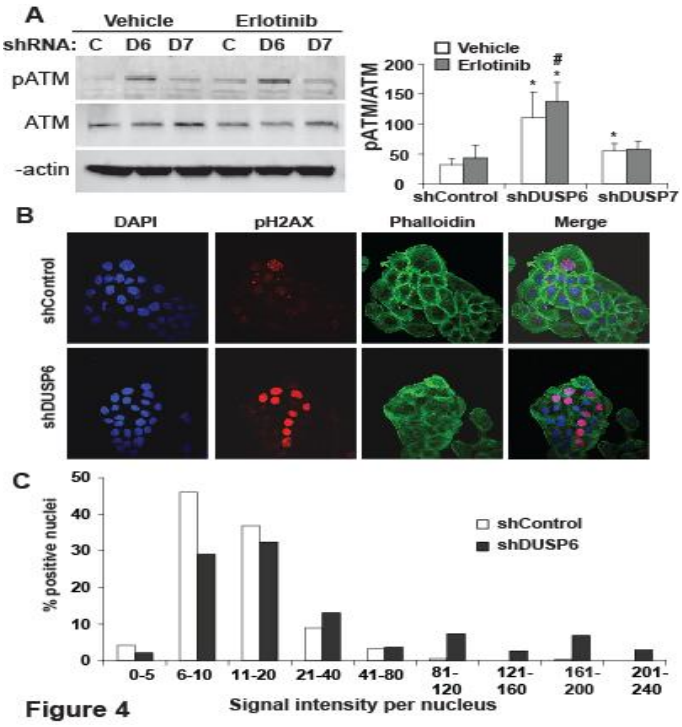


Figure 4. DUSP6 depletion activates ATM pathway.
(A) DUSP6 depletion increased phosphorylation of ATM.
*, P < 0.05 for shDUSP6 vs. shControl;
#, P<0.05 for DUSP6 vs. shDUSP7.
(B) Phospho-histone H2AX immunelabeling in control and DUSP6-depleted cells. Objective, 40x.
(C) Quantification was done using MetaMorph. Approximately 400 nuclei were analyzed per cell line.

Research Project 3: Project Title and Purpose

A Gene Methylation Progression Model of Bladder Cancer - The purpose of this exploratory project is to identify the difference in aberrant gene methylation between non-invasive, low grade and high grade stromal invasive, and muscle-invasive bladder tumors. The central question in the management of superficial bladder tumors, which comprise the majority of bladder cancer, is how to identify the superficial tumors that will progress from those that will not, and thereby avoid overtreatment or undertreatment. Knowledge of the molecular steps in progression of bladder cancer will likely provide insight into the biology of the disease as well as identify novel targets for therapy. It is likely that genes identified will also be involved in other types of cancer with similar management issues such as prostate cancer.

Anticipated Duration of Project

1/1/2011 - 12/31/2012

Project Overview

Bladder cancer is the sixth most common cancer in the western world. Currently, the clinician is unable to differentiate between more aggressive or indolent superficial bladder cancer. Despite the finding of pathologically confined disease at the time of surgery, 10-20% of patients will have progression to life-threatening muscle-invasive cancer. There is no molecular marker for this risk. Furthermore, the biological basis for invasion is poorly understood.

Our group has helped to pioneer the study of methylation in cancer. We have evidence that aberrant gene methylation in bladder cancer is frequent, early, neoplastic cell specific, and occurs in all pathways to tumorigenesis.

The goal of this project is to have our team use our expertise and state of the art genome-wide technology to rapidly screen for genes differentially methylated between pathological stages of bladder cancer. This will address the key clinical issue that the subset of individuals with superficial bladder cancer that will progress cannot be identified in advance.

The successful completion of this project could have future impact through improved care and outcome of bladder cancer. Our work could lead to an aggressiveness index to stratify aggressive superficial tumors for potentially curative cystectomy while still allowing patients with more indolent lesions to have bladder-sparing treatment. The genes identified may also have clinical utility in other tumor types with similar management issues, e.g., indolent versus aggressive prostate tumors. Lastly, we have expertise in translational application to bladder cancer and gene methylation tests have reached the clinic.

Our Specific Aims are
1. To profile genome-wide aberrant methylation of genes in pathological stages of bladder cancer
2. To validate aberrant methylation by direct bisulfite sequencing of the genes identified
3. To identify functional themes, pathways and networks in bladder tumor invasion and progression.

Principal Investigator

Paul Cairns, PhD
Assistant Professor
The Fox Chase Cancer Center
333 Cottman Avenue
Philadelphia PA 19111

Other Participating Researchers

Ilsiya Ibragimova, PhD - employed by The Fox Chase Cancer Center

Expected Research Outcomes and Benefits

We expect to identify genes with a statistically significant different frequency of aberrant methylation between pathological stages based on depth of invasion of bladder cancer.

We expect to verify 25 genes aberrantly methylated in bladder cancer but unmethylated in normal transitional cells.

We expect to identify functional themes, pathways or networks within the 25 genes.

The project will improve health status through:

- Further elucidation of the methylome of the human bladder cancer cell
- Candidate markers for risk stratification of superficial bladder tumors to be tested further
- Candidate markers for diagnosis, and novel targets for chemoprevention or therapy
- Insight into the biology of bladder cancer and, in particular, tumor invasion
- Insight into the biology of, and markers for, other cancers e.g. prostate cancer.

Summary of Research Completed

Fresh-frozen bladder tumor tissue was mounted in optimum cold temperature compound (OCT), a hematoxylin and eosin (H&E) stained slide made, and tumor cell-rich areas dissected with the assistance of a pathologist. DNA was extracted from primary tumor tissue using a standard technique of proteinase K digestion in the presence of sodium dodecyl sulfate at 37°C overnight followed by phenol/chloroform extraction. Small amounts of DNA were precipitated with one-tenth volume of 10M ammonium acetate and 2μl of glycogen with 2.5 volumes of 100% ethanol, incubation at –20°C and centrifugation. The yield and quality of DNA was examined by a nanodrop.

One microgram of genomic DNA was bisulfite-modified using the Zymo Research EZ DNA Methylation Kit according to the manufacturer's instructions (noting the special requirements for use with the Infinium BeadChip). Bisulfite treated DNA was isothermally amplified overnight and enzymatically fragmented then hybridized to the Illumina Infinium humanmethylation27 beadchip. Four runs of 24 samples at a time were performed. We therefore have data from 75 (22 pTa, 21 low grade pT1, 7 high grade pT1, 25 pT2-4) transitional cell carcinoma (TCC) from patients with no prior chemotherapy as well as 4 normal transitional cell specimens obtained from age-matched individuals with no evidence of bladder cancer. To date, five duplicate samples have been run to test for batch effects.

The gene methylation data is provided in raw form after analysis by the Illumina BeadStudio software. This methylation data has undergone preliminary examination for quality control (QC) by the FCCC Biostatistics and Bioinformatics Facility using the R/Bioconductor platform. DNA methylation values, described as beta-values, vary between 0 (unmethylated) and 1 (fully methylated), representing the ratio of the intensity of the methylation bead type to the combined locus intensity. Beta values generated by Illumina BeadStudio were used to exclude unlikely candidate probes prior to analysis: 639 out of 27,578 probes with missing betas were removed. Log ratios of methylated signals to unmethylated signals (each offset by 100 to avoid large ratios between low-intensity signals) were compared between pathological stage subgroups using empirical Bayes moderated t-statistics implemented in the *limma* package. From QC analysis, there was evidence of variable beadchip quality and possible batch effects; we therefore incorporated empirical array quality measurements to downweight lower quality arrays, and included assay date as a blocking factor. P-values were adjusted for multiplicity using the Benjamini-Hochberg procedure to control the false discovery rate (FDR). Overall, 72 of the 75 (96%) tumor beadchips initially yielded data of sufficient quality to be used in the final analysis.

In the first 6 months of the project, we have therefore successfully completed the milestones as stated in the original application:

Months 1-3 Prepare DNA from tumor specimens
Months 4-6 Complete Infinium beadchip analysis of 75 tumor DNAs.

Research Project 4: Project Title and Purpose

Targeting STAT3 in Ovarian Cancer - The purpose of this project is to determine whether therapeutic targeting of signal transducer and activator of transcription 3 (STAT3) is effective for the treatment of ovarian cancer. Using an immunocompetent genetically engineered mouse model of epithelial ovarian cancer, we will determine the effect of STAT3 inhibition on tumor incidence latency and progression. Since activation of STAT3 in both tumor cells and immune/inflammatory cells in the tumor microenvironment is thought to contribute to cancer, this project is designed specifically to determine the effects of STAT3 inhibition on both cell populations.

Anticipated Duration of Project

1/1/2011 - 12/31/2012

Project Overview

Epithelial ovarian cancer (EOC) is the most fatal of all of the gynecological malignancies. Approximately 22,000 new cases are diagnosed and nearly 15,000 women die of ovarian cancer each year in the United Sates. The disproportionately high mortality rate of EOC is largely related to the fact that most cases are diagnosed at advanced stage, when disease has spread beyond the ovary. Patients generally respond well to aggressive surgical debulking and a standard combination chemotherapy regimen consisting of a platinum agent and taxane. However, the majority of patients eventually experience disease recurrence and, ultimately, drug resistant disease. There is a critical unmet need in identifying therapeutic agents that can improve patient outcomes. In order to discover alternative therapeutic agents, considerable effort has been devoted to defining the underlying molecular and genetic changes that contribute to the development and progression of EOC.

One target that has been identified is signal transducer and activator of transcription 3 (STAT3). STAT3 has been shown to be persistently activated in EOC and is associated with highly aggressive and invasive cases of EOC. Although STAT3 activation has been recognized in EOC for several years, until now, agents targeting STAT3 activation that are suitable for clinical use have been lacking. Recently, a small molecule Janus kinase 2 (JAK2) inhibitor, AZD1480, was identified that has direct effects on inhibition of STAT3 activity. The objective of our project is to test the therapeutic efficacy of this agent in an immunocompetent genetically engineered mouse model of EOC. In addition to evaluating the efficacy of AZD1480 in inhibiting ovarian cancer progression in mice, a second major objective is to assess the effects of AZD1480 treatment on the immune tumor microenvironment and ovarian cancer associated inflammation.

Our specific aims are to:
1) Evaluate the effects of a novel small molecule JAK2 inhibitor on epithelial ovarian cancer cells *in vitro*.

2) Determine the therapeutic efficacy of STAT3 inhibition on ovarian carcinoma development and dissemination *in vivo*.
3) Test the hypothesis that targeting the STAT3 pathway will inhibit *in vivo* tumor development by direct effects on both tumor cells and on inflammatory cells present in the tumor microenvironment.

Principal Investigator

Denise C. Connolly, PhD
Assistant Professor
The Fox Chase Cancer Center
333 Cottman Avenue
Philadelphia, PA 19111

Other Participating Researchers

Samuel Litwin, PhD, Richard R. Hardy, PhD, Glenn Rall, PhD, Russell Schilder, MD - employed by The Fox Chase Cancer Center

Expected Research Outcomes and Benefits

Based on extensive *in vitro* studies of STAT3 activation and signaling in our laboratory, we are convinced that it is a highly relevant target for therapeutic inhibition for the treatment of ovarian cancer. Activation of STAT3 is known to have pleiotropic effects supporting the growth and progression of many solid tumor types including epithelial ovarian cancer (EOC). Moreover, recent studies show that STAT3 activation in the tumor microenvironment also supports tumor initiation, growth and progression.

The development of a highly promising new Janus kinase 2 (JAK2) inhibitor, AZD1480, shows significant activity in preclinical cancer models. The goal of our project is to evaluate the therapeutic efficacy of AZD1480 for treatment of EOC using a genetically engineered mouse model of EOC developed in our lab. Importantly, EOC occurs spontaneously and these mice have a completely intact immune system. Therefore, using this model, we are uniquely capable of evaluating the effects of AZD1480 on not only ovarian tumor growth and progression, but on the immune tumor microenvironment as well. We predict that AZD1480 will show both tumor inhibitory activity as well as shifting the tumor microenvironment from tumor supporting inflammation to anti-tumor immunity. As AZD1480 is currently in Phase I development, demonstration of therapeutic efficacy in our mouse model of EOC would be expected to be rapidly translatable to a preclinical trial for ovarian cancer patients.

Summary of Research Completed

To initiate this project, it was necessary to establish a Material Transfer Agreement (MTA) with Astra Zeneca to obtain the small molecule JAK2 inhibitor AZD1480. With the assistance of the Fox Chase Cancer Center Office of Research and Development Alliances, an MTA was developed and finalized on February 11, 2011 and the AZD1480 compound was shipped to our

laboratory immediately thereafter.

After receiving the AZD1480, we began studies described in Specific Aim 1 to validate the activity of this drug in ovarian carcinoma cells. A previously published study showed that treatment of the ovarian carcinoma cell line MDA2774 with AZD1480 resulted in inhibition of STAT3 phosphorylation. We first evaluated the effects of AZD1480 drug treatment in ovarian carcinoma cell lines, including MDAH2774 as a benchmark. Briefly, cells were grown in monolayer culture in the presence of 0, 0.05, 0.1, 1.0, 5.0 and 10 μM AZD1480. Cells were lysed and protein concentrations determined by BCA protein assay. Equal amounts (10 μg) of protein were resolved on SDS-PAGE gels, transferred to PVDF membrane and probed with antibodies recognizing phosphorylated JAK2, total and phosphorylated STAT3 and Actin. Unexpectedly, the immunoblot analyses showed a dose dependent increase in phosphorylated JAK2 using commercially available antibodies from Cell Signaling and Invitrogen. However, in communication with our colleagues at Astra Zeneca, we have learned that there are issues related to the specificity of the commercially available anti-JAK2 antibodies and that they have abandoned this assay for AZD1480 activity. Instead, the favored approach is to determine the effects of AZD1480 on STAT3 activation since STAT3 is a direct downstream target of JAK2. Activation of STAT3 is commonly assessed by detection of the levels of STAT3 phosphorylated on tyrosine 705 (Y^{705}). Detection of total and Y^{705} phosphorylated STAT3 showed a dose dependent decrease in STAT3 activation in all five human ovarian carcinoma lines (MDAH 2774, SKOV-3, A1847, OVCAR-8 and PE-01) tested. Greater than 50% inhibition of phospho-$STAT3^{Y705}$ was evident at concentrations as low as 0.05 μM and nearly complete inhibition was observed at 0.1 μM of AZD1480. Similar results were observed in murine ovarian carcinoma (MOVCAR) cell lines MOVCAR 12-3, 5009, 5447 and 5612. The results indicate that concentrations as low as 0.1 μM of AZD1480 are effective at inhibiting the activation of the JAK2 target STAT3 and that AZD1480 mediated inhibition of STAT3 activation is similar in human and murine ovarian carcinoma cells.

The effects of AZD1480 treatment on cell growth/viability was measured using the Cell Titer Blue® Cell Viability Assay (Promega). The effects of AZD on several human (MDAH 2774, SKOV-3, A1847, OVCAR-8 and PE-01) and murine (MOVCAR 12-3, 5009, 5447 and 5612) ovarian carcinoma cell lines was tested. For each cell line and concentration of vehicle or AZD1480 (e.g., 0, 0.05, 0.1, 1.0, 5.0 and 10 μM), 3×10^3 cells/well were plated in triplicate in 96-well plates, incubated at 37°C and assayed 72 hours later. No significant change in cell viability was observed in the presence of 0.05 – 1.0 μM AZD1480. These results were similar to previous results obtained in our lab using the JAK inhibitor AG490. Decreased cell viability was only observed at the highest concentrations (5 and 10 μM) of AZD1480 suggesting that the effects on cell viability may be attributable to off-target effects. These data suggest that at lower more physiologically relevant concentrations that AZD1480 has little effect on cell/growth viability in cells grown in monolayer.

Expression and activation of STAT3 is known to result in the induction of anti-apoptotic genes. Therefore, inhibition of STAT3 activation may result in increased apoptosis. To determine whether AZD1480 treatment and the resulting inhibition of STAT3 activation results in increased apoptosis, we analyzed vehicle and drug treated cells by Guava-Nexin assay to assess levels of Annexin V and by western blot to assess induction of PARP and caspase cleavage.

Cells were prepared for Annexin V analysis by plating 6 x 10^5 cells/well in regular 6-well plates or 1.5 x 10^6 in Low Attachment 6-well plates in the absence or presence of 1 μM AZD1480. After 24 or 48 hours incubation at 37°C, cells were detached by trypsin digestion, counted and suspended at a concentration of 1x10^6 cells/ml. Aliquots of 100 μl (1x10^5 cells) were transferred to a tube and 100 μl of Guava Nexin reagent was added and incubated for 20 minutes. The percent of Annexin V positive cells were determined using the Guava Personal Cell Analysis machine exactly as described by the manufacturer (Guava Technologies). Treatment of MDAH 2774 and OVCAR 8 cells with AZD1480 resulted in increased levels of Annexin V positive cells compared to vehicle treated cells. Likewise, drug treatment resulted in increased levels PARP and caspase cleavage detected by western blot analysis. These results show that AZD1480-mediated inhibition of the JAK2/STAT3 pathway results in induction of apoptosis in ovarian carcinoma cells.

To assay the effects of AZD 1480 treatment on cellular adhesion capacity, 3 x 10^3 cells/well were plated in triplicate in 96-well plates coated with Fibronectin or Collagen I and incubated at 37°C for 1 hour. Plates were washed three times with phosphate buffered saline (PBS) and attached cells were fixed, stained and counted. Cells plated in the presence of 1 μM AZD1480 exhibited significantly diminished capacity of ovarian carcinoma cells to adhere to fibronectin or collagen I coated surfaces compared to untreated cells. Additional *in vitro* assays are underway to assess the effects of AZD1480 treatment on cellular migration and invasion.

For the *in vivo* studies described in Specific Aim 2, over the past three months, we have expanded our breeding of TgMISIIR-TAg transgenic mice. Briefly, since female TgMISIIR-TAg transgenic mice are infertile, male C57Bl/6 TgMISIIR-TAg mice are bred to wild-type female C57Bl/6 mice to generate female, ovarian cancer prone TgMISIIR-TAg transgenic mice. We have accrued sufficient wild type and transgenic mice to first conduct pilot pharmacokinetic (PK) and pharmacodynamic (PD) studies in AZD1480 treated mice to assess the effects of short-term drug treatment. These studies are currently being planned and coordinated with the assistance of our collaborators at Astra Zeneca, who will perform the PK analysis. In addition to the animals needed for the pilot studies, we have begun to accrue mice that will be placed on the drug treatment studies while breeding continues to generate additional mice.

Research Project 5: Project Title and Purpose

Classification and Prediction of Protein-Protein Interactions in Biology and Medicine - Protein interactions are central to all biological processes, including those leading to disease. Many proteins exist in complexes with other copies of the same protein and/or with other kinds of proteins critical to their function. Structural information on these interactions is essential for understanding basic biological processes and the development of therapeutics and diagnostics. However, the structures of these interactions are often unknown or ambiguous. We hypothesize that we can obtain far more information than currently available by exploiting evolutionary relationships among proteins and analysis of data from multiple biophysical experiments. To do this, we will classify domains in multi-domain proteins, cluster similar forms of interaction, and predict the structures of protein interactions that occur in signaling pathways associated with cancer.

Anticipated Duration of Project

1/1/2011 - 12/31/2012

Project Overview

In our research we apply protein structure prediction to develop hypotheses of how proteins carry out their biological functions and to design experiments in cellular biology and developmental therapeutics. For instance, for a protein that interacts with multiple partners, structure prediction of protein complexes may be used to design mutations to disrupt some interactions while maintaining others. The broad objective of this project is to develop methods for structure prediction of protein complexes through identification of biologically relevant interactions in protein crystal structures. A majority of X-ray crystal structures contain information on how proteins and protein domains interact with each other. However, in up to 20% of structures, the biological relevance of interactions in the crystal is uncertain, ambiguous, or unknown. In addition, the evolutionary significance of observed interactions cannot be identified from single structures, and must be inferred from multiple structures of homologous proteins. Our ability to extrapolate from known structures to predict the structures of other complexes requires the development of new tools and new forms of analysis.

We will pursue the following specific aims. First, we will use advanced bioinformatics methods to identify and classify protein domains within single and multi-domain proteins in the Protein Data Bank (PDB), the public repository of three-dimensional structures of proteins. Current sequence and structure-based classifications cover only 60-75% of the PDB at any one time, which is inadequate for exploiting structural information in understanding biological processes. The second aim will be to extend our recently introduced ProtCID database of common interfaces of proteins, which has enabled identification of important interfaces conserved through evolution among related proteins. We will compare interfaces of single domains contained within multi-domain proteins in protein crystals, including both intra- and interprotein interactions. This will greatly increase the inferences possible about the structures of protein-protein interactions. The third aim is to apply these new methods to develop hypotheses about how kinases active in cell signaling events related to cancer are regulated through protein-protein interactions. In particular, we will develop and test hypotheses on the structures of auto-phosphorylation complexes of tyrosine kinases.

Principal Investigator

Roland L. Dunbrack, Jr., PhD
Associate Professor
The Fox Chase Cancer Center
333 Cottman Avenue
Philadelphia, PA 19111

Other Participating Researchers

Qiong Wei, PhD - employed by The Fox Chase Cancer Center

Expected Research Outcomes and Benefits

The expected outcome of the research is greater knowledge of the structures of protein-protein interactions and an expanded ability to predict these structures in support of biomedical research. We will develop a web-based resource for the classification of all proteins of known structure in the Protein Data Bank (PDB) into domain families and superfamilies. The classification of proteins will be more accurate and more complete than existing sequence-based and structure-based resources. For every protein domain family, the resource will identify all interactions with other domain families and with ligands with available structural information. These structural interactions will be compared and clustered so that interactions common among many homologous proteins and found in multiple crystallographic experiments can be readily identified, leading to greater confidence in the biological relevance of the observed interactions. The evolutionary relationships of these protein interactions will be analyzed to determine what biological inferences may be made for proteins of unknown structure.

We will use these methods to investigate the protein interactions associated with signaling events that occur in cancer. We have recently identified a homodimer observed in the crystals of two different tyrosine kinases, LCK and IGF1R that is likely to be the main autophorphorylation event leading to activation of all similar tyrosine kinases. We will carry out further studies of the interactions of kinases in protein crystals that may be associated with activation or inhibition events in normal and cancerous cells.

The resources developed in this project will be made publicly available for other researchers. The impact on health is through obtaining detailed structural information on how proteins interact with other molecules, since most drugs are designed to interfere or alter these interactions. The new resource will enable researchers to exploit the full compendium of structural knowledge to a much greater extent than is currently possible.

Summary of Research Completed

We have made substantial progress on Aim 1 of this project, developing robust methods for assigning PFAM domains to structures in the Protein Data Bank (PDB). We have approached this task in three steps. First, we have used consensus sequences derived from PSI-BLAST profiles for each PDB sequence to search against the hidden Markov models (HMMs) in PFAM-A. These consensus sequences often provide better E-values and longer alignments than the original sequences. Second, we have used structure alignments to compare weak hits or those with short alignments to structures with the highest statistical significance for each PFAM domain. These alignments are often of high statistical significance and are sometimes much longer than the sequence-based alignments. Third, we have used these data as input to a greedy algorithm that accurately handles domains split by large insertions or that includes separate chains in a PDB structure. This algorithm is much more realistic and sensible than the one that PFAM itself uses. We describe the results for each of these studies below.

<u>Consensus sequences</u>
We derived consensus sequences from PSI-BLAST profiles for every unique sequence in the PDB. One round of PSI-BLAST was used to search Uniprot with an E-value cutoff of 0.001 for

sequences used to produce the profile. The "log-odds consensus sequence" for each PDB sequence was derived by taking the highest scoring amino acid type at each position of the position-specific scoring matrix produced by PSI-BLAST. The "percentage consensus sequence" was derived by taking the most common amino acid at each position of the multiple sequence alignment produced by PSI-BLAST.

To investigate whether the consensus sequences improve the sensitivity of PFAM assignments by HMMER 3.0, we took the single best-scoring PFAM domain assignment for each PDB unique sequence from both consensus sequence types and the original sequences ("PDB-orig") with E-value < 0.001, then compared the distributions of E-values. Figure 1 shows the kernel density estimates and cumulative functions of $\log_{10}$(E-value). The consensus results (cyan, and green) are shifted to significantly better E-values (more negative $\log_{10}$(E-value)) than the original PDB sequences (red). The purple curve shows the result of choosing the best E-value of all three sequences ("PDB-best"). The mode of the density for the original sequences is at about 1.0E-20 while for the consensus sequences it is about 1.0E-25.

For each unique sequence, we selected the PFAM assignments which are in the same PFAM and cover approximately the same sequence regions with E-value < 0.001, then produced scatter plots of the PDB-orig alignment lengths against the consensus sequences and the PDB-best result (Figure 2). About 80% of the consensus sequence alignments are longer than the original sequence alignments. In some cases, PFAM assignments of consensus sequences are broken down into two or more fragments, producing separate alignments that are shorter than the single alignment produced by the original sequence, producing points far below the diagonal. This situation will be fixed in the algorithm that pieces together the separate PFAM assignments into an assignment for each whole sequence.

Table 1 shows the results of scoring each PFAM HMM against each unique PDB sequence and the consensus sequences. The number of PFAMs, the number of unique sequences with PFAM hits, etc. are shown in the table for those PFAMs with E-values better than 0.001. The consensus sequences improve both in the coverage of PDB sequences and the coverage of PFAM A models. The consensus sequences find 6.1% more PFAMs in the PDB than the original sequences, and cover 3.5% more residues of the unique PDB sequences.

Structure alignment

We used structure alignment to verify PFAM assignments with poor E-values and/or alignments that do not fully cover the HMM. To do this, we chose "exemplars" for each PFAM found in the PDB, if the best E-value over all PDB sequences or consensus sequences was $\leq 10^{-5}$ and the alignment covered at least 80% of the HMM positions with coordinates from a structure. Of the 7,785 PFAMs we find in the PDB (with any E-value), 63% of them (4,884) have an exemplar under these definitions.

For each PFAM assignment with E-value $> 10^{-5}$ or missing more than 10 residues of the HMM on either end of the alignment, we checked whether there was an exemplar for that PFAM or any PFAM in the same clan. PFAM defines clans for two or more PFAMs that are homologous to each other. This is similar to superfamily definitions in such databases as SCOP. These PDB sequences with poor hits fall into 5,258 PFAMs, and comprise 20,685 (14.8%) of assignments to

all PDB sequences. We used the FATCAT program from Adam Godzik's group to align the structures of proteins with poor PFAM assignments to the exemplars in the same clan as the original PFAM assignment. FATCAT was able to produce alignments with p-values better than 0.001 for 15,580 (75.3%) of these assignments. This is 11.1% of the total assignments to PDB sequences.

We determined new alignments of PDB sequences to PFAMs if the FATCAT p-value was ≤ 0.001 and the new alignment to the PFAM was longer than the original one. The new PFAM alignment is determined transitively. The procedure is as follows:

(A to B) + (B to C) = (A to C)

A to B is the HMM-exemplar alignment, B to C is the sequence alignment derived from the structure alignment of the exemplar structure to the poor PFAM domain structure, and A to C is the resulting HMM to poor PFAM domain alignment. We were able to improve the domain boundaries of 11,961 PFAM assignments. Table 2 gives the number of residues aligned by HMMER 3.0 and by FATCAT structure alignments and exemplars. The number of PDB residues aligned to PFAM HMMs is 33% higher, and the number of aligned HMM positions is 29% higher. An example of extending PFAM assignments using FATCAT structure alignment is shown in Figure 3.

<u>A greedy algorithm for determining PFAM architectures</u>
We developed a greedy algorithm for assigning multiple PFAMs to multi-domain protein structures in the PDB. Such an assignment is referred to as an "architecture" by PFAM itself. We used both the sequences in the PDB and the full Uniprot sequences from which these PDB sequences were derived. In some cases, the full-length sequences allow PFAM assignment when the truncated sequences used in structure determination do not (e.g. when peptides from large sequences are present in the structure).

Using the original and consensus sequence alignments to the PFAM HMMs as well as the structure alignments, we start with the highest scoring PFAM and assign that first. If a region of the PDB sequence of the first assignment longer than 30 residues is not aligned to HMM positions, then this region is labeled "unassigned" and the assignment is deemed a "split" assignment. That is, it has a long insertion in the PDB sequence that is not part of the HMM. We assign subsequent PFAMs if they do not overlap by more than 10 residues of any existing assignment. These assignments can follow into the regions relabeled "unassigned" in the previous step. If two (or more) assignments are to the same PFAM and they cover different regions of the PFAM (allowing short overlaps), these are also considered "split" assignments. A single PFAM HMM may cover more than one sequence in a PDB entry, if the regions covered in the HMM do not overlap by more than 10 residues). We also assign repeats at weaker E-values if the sequence contains the same repeat PFAM at significant E-values (≤10e-05).

This algorithm is simple but more robust than the one PFAM itself uses. In v.25 of PFAM, which covers PDB entries deposited through March 2011, PFAM assigns 85,446 domains to the PDB, including 3,216 repeats. Using our algorithm and the consensus and structure alignment data, we

assign 107,767 domains (+26%) including 6,267 repeats (+95%) to the same PDB sequences (i.e. those available in March 2011).

An example of poor domain assignments from the PFAM website is shown in Table 3. The IMPDH family appears in 28 structures in the PDB. Fifteen of these structures contain the alpha-beta barrel of IMPDH with two CBS domains inserted, a common feature of this family. For 8 of these 15, PFAM itself misses the CBS domains altogether. For the other 7, it has the two CBS domains after the IMPDH domain, since it has no annotation scheme for indicating that they are inserted between two portions of the IMPDH alpha-beta barrel domain. We annotated all 15 of these structures as (IMPDH[1-85])_(CBS)_(CBS)_(IMPDH[86-351]), indicating the insertion of the CBS domains after position 85 of the IMPDH HMM.

The coverage of the PDB sequences from the other well-known data resources are shown in Figure 4. PFAM v25 has 78% coverage of PDB sequences. SIFTS, which also uses PFAM HMMs, covers 88% of PDB sequences, and provides up-to-date PFAM assignments for the PDB by mapping PDB sequences to UniProt sequences and their PFAM assignments. The structure protein classification systems CATH and SCOP have a lower coverage because they involve manual assignments. They are often 1 to 2 years behind on analyzing the PDB with CATH covering 57% of PDB sequences and SCOP only 49%.

Our assignments used 367 additional PFAM A models than PFAM itself or SIFTS uses, and we were able to make assignments to 99% of PDB sequences. There are still 2,327 PDB polypeptide entity sequences which are not covered in our PFAM data set, of which 1,585 sequences are less than 20 amino acids and/or contain unknown or non-standard amino acids. A total of 96 of these 742 sequences with length ≥ 20 amino acids do not have Uniprot codes. The majority of these come from disordered regions and engineered proteins.

This work is being prepared for publication and the database of PFAM assignments to the PDB will be made publicly available and updated frequently.

Table 1. Summary of PFAM A assignments from the PDB and the consensus sequences with E-values ≤ 0.001.

	PDB-orig	PDB-consensus
#PFAM A	6,090	6,461 (↑6.1%)
#Entries	69,816	70,150 (↑0.5%)
#Unique sequences	45,805	46,516 (↑1.5%)
#Domains	60,547	63,869 (↑5.5%)
#Residues	8,740,432	9,047,704(↑3.5%)
#HMM positions	8,696,426	9,017,705 (↑3.7%)

Table 2. The number of residues before and after structure alignments for PDBs with weak PFAM assignments.

	#aligned Residues	#Envelope Residues	#HMM Positions
Before	1,227,544	1,428,456	1,246,328
After	1,636,550	1,636,550	1,605,329
Increased percentage ↑	33.3%	14.6%	28.8%

Table 3. Split domains in IMPDH family.

#PDB	**Our PFAM Architecture**	**http://pfam.sanger.ac.uk/ PFAM Architecture**
13	(IMPDH)	IMPDH
8	(IMPDH[1-85])_(CBS)_(CBS)_(IMPDH[86-351])	IMPDH
7	(IMPDH[1-85])_(CBS)_(CBS)_(IMPDH[86-351])	IMPDH_CBS_CBS

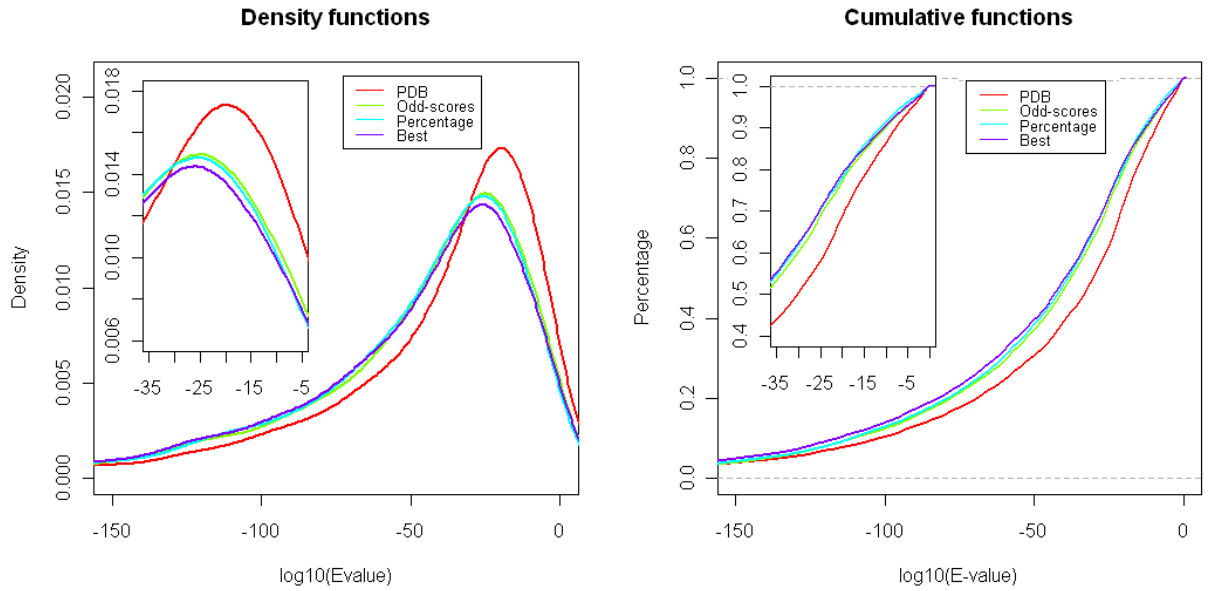


Figure 1. Comparison between PDB sequences and consensus sequences. Left: density functions. Right: cumulative density functions.

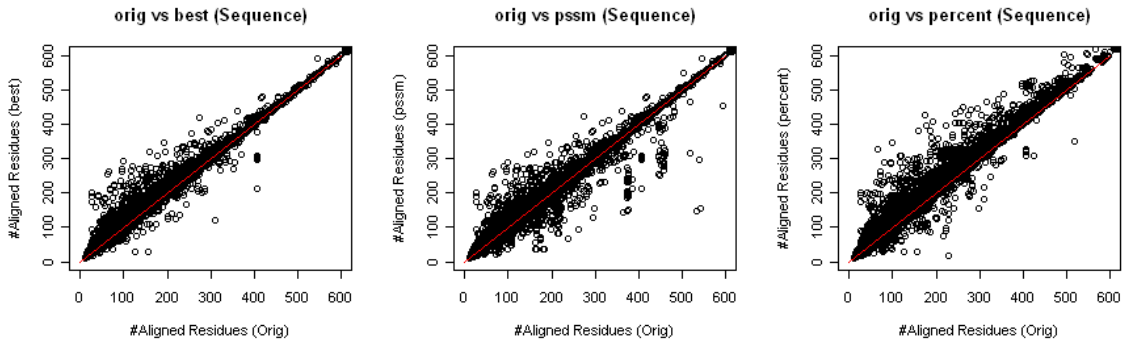


Figure 2. Scatter plots of the numbers of aligned residues of the original PDB sequences versus the numbers of aligned residues of the consensus sequences and PDB-best.

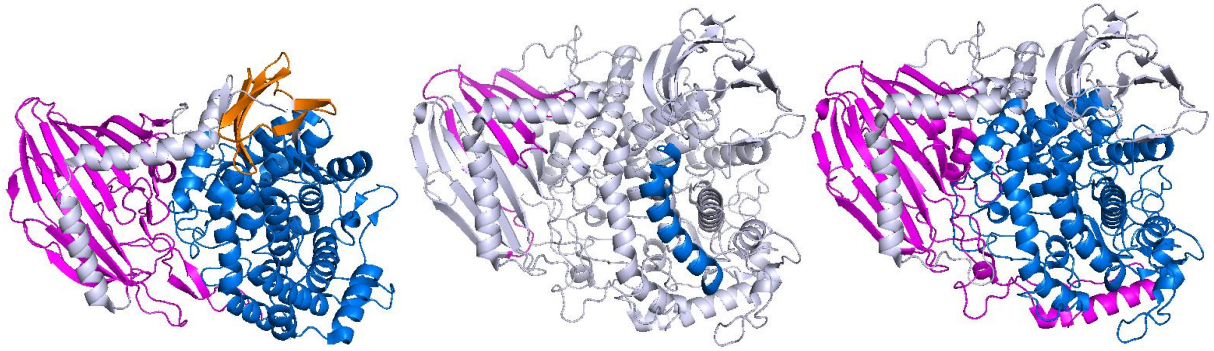


Figure 3. Structure alignments expand the PFAM assignments. Left: PDB entry 1H54 has three PFAM assignments in magenta, blue and orange. Center: PDB entry 2EAB has two PFAM assignments with weak E-values and short alignments, in magenta and blue. These assignments are in the same clans as the those in 1H54 of the same colors. Right: after structure alignment of 2EAB and 1H54, much longer PFAM assignments of these two domains can be made.

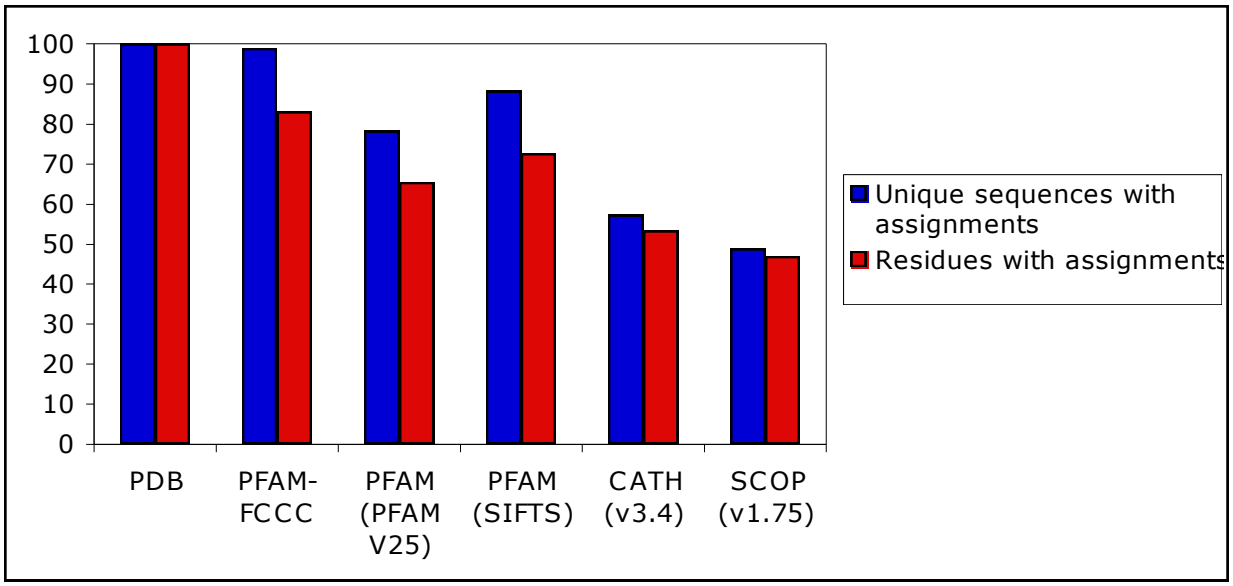


Figure 4. Comparison of the percentage of unique sequences and the percentage of residues with assignments in various domain assignment databases. The PDB columns simply represent the total proportions in the PDB (100%). PFAM-FCCC represents our assignments produced from the consensus alignments, structure alignments, and the greedy algorithm. PFAM v.25 covers the PDB only as of March 2011, and also assigns fewer domains in those sequences. PFAM assignments by the SIFTS database is more up-to-date than PFAM v.25, but uses the same assignment algorithm. CATH and SCOP are manual assignments of protein folds, superfamilies, and families, and are 1 to 2 years behind the PDB.

Research Project 6: Project Title and Purpose

Use of Somatic Cell Nuclear Transfer to Generate TCR Monoclonal Mice - A key approach in experimental immunology is to generate mice that express single T or B cell receptors (TCR, or BCR) in order to examine their specificity, function, and effect on development. Somatic cell nuclear transfer (SCNT) is a promising technique for the rapid generation of such TCR monoclonal mice, which involves replacing the oocyte's nucleus with a donor nucleus from a somatic cell. Here we propose experiments to improve the efficiency of SCNT for T lymphocytes and apply the approach to dissecting the mechanism of CD4/CD8 lineage commitment.

Anticipated Duration of Project

7/1/2011 - 06/30/2012

Project Overview

A key approach in experimental immunology is to generate mice that express single T or B cell receptors (TCR, or BCR) in order to examine their specificity, function, and effect on

development. Somatic cell nuclear transfer (SCNT) is a promising technique for the rapid generation of such TCR monoclonal mice, which involves replacing the oocyte's nucleus with a donor nucleus from a T cell. The nucleus of transferred T cells is reprogrammed, but nevertheless, retains the rearranged DNA encoding the TCR subunits and antigen specificity of the original T cell. The current study has two broad objectives: 1) To improve the efficiency of SCNT specifically for mouse T lymphocytes, and 2) To utilize SCNT to efficiently generate a number of TCR transgenic mouse lines from ThPOK deficient and ThPOK overexpressing mice, which will facilitate characterization of their developmental and functional capacities and elucidate the mechanism of CD4/CD8 lineage commitment.

This project has the following specific aims:

Aim 1: Improving SCNT efficiency for T lymphocytes. The efficiency of generation of cloned progeny by SCNT in mice has remained low for lymphocytes, but recent advances in our understanding of nuclear reprogramming suggest approaches for improvement, in particular exposure of donor cell nuclei to reprogramming factors and inhibitors of the p53 pathway. In this aim, we will compare procedures whereby T cells are reprogrammed in cell culture prior to SCNT, or in which reprogramming factors are injected directly into the oocyte prior to nuclear transfer.

Aim 2: Developmental and functional characterization of ThPOK-independent T cell subsets. We previously demonstrated that the transcription factor ThPOK plays a key role in deciding CD4/CD8 lineage choice for the majority of developing T cells. However, a distinct fraction of cells seem to make this decision independently of ThPOK. It is of considerable interest to elucidate the developmental origin of cells that utilize this alternate mechanism(s). In this aim, we will use SCNT to generate TCR monoclonal mice from such cells. Flow cytometry will be used to assess the developmental and functional attributes of T cells from these mice.

Principal Investigator

Dietmar Kappes, PhD
Professor
The Fox Chase Cancer Center
333 Cottman Avenue
Philadelphia, PA 19111

Other Participating Researchers

None

Expected Research Outcomes and Benefits

We anticipate 3 main outcomes from this project. First we expect to develop significant improvements to current procedures for SCNT using lymphocytes. The Transgenic Mouse Facility at Fox Chase is exceptionally qualified to carry out such studies given the extensive expertise of the personnel in microinjection and embryo manipulation, and its long-standing

record of success. Secondly, we expect to generate important information on the mechanism of CD4/CD8 lineage commitment, in particular whether alternate ThPOK-independent control mechanisms exist. Finally, establishing the SCNT approach at Fox Chase will facilitate future important studies of both normal development and aberrant development leading to cancer. For instance, certain mutant mouse lines which express autoreactive γδTCRs are prone to develop lymphomas and the SCNT approach can be utilized to probe the underlying mechanism.

Summary of Research Completed

The work on this project is scheduled to begin on July 1, 2011.