THE MOLECULAR MECHANISMS OF CANCER DEVELOPMENT AND BIOMARKER DISCOVERY

STUDY OBJECTIVES

The long-term research goal in my lab is to investigate the molecular mechanisms of cancer initiation, progression and metastasis. Regulation of gene expression at post-transcriptional level by signal transduction is the major target to study. The projects include, but not limited to, mRNA translation, miRNA, kinases in regulation and epigenetics. Novel pathways, translational factors and the specific phosphorylation states of proteins can be developed into the therapeutic targets to treat cancers. Biomarker discovery will focus on the identification of new proteins, peptides, circulating DNAs, RNAs by advanced mass spectrometry (MS), and exfoliated cells by flow cytometry and primary cell culture. Live cancerous tissues from biopsy and surgery, the Formalin-fixed, paraffin-embedded (FFPE) samples as well as patient urine and blood samples will be utilized to investigate the real carcinogenesis process in humans. The use of patient samples will advance the understanding of molecular mechanisms of cancer development, and facilitate the development of new therapeutic targets and diagnostic biomarkers.

Prostate and colorectal cancers will be specifically studied in this project. Prostate cancer is the most common malignancy in men and the third leading cause of cancer death; colorectal cancer is the third most common cancer in both men and women and the third leading cause of cancer-related mortality in the United States. The incidence and mortality rates of colorectal cancer in NEPA are higher than the average of the US, prostate cancer is also a common cancer affecting the population in NEPA. Therefore, it is very imperative to investigate these cancers with our local patients to advance the understanding of cancer etiology and mechanism; this will also help improve the healthcare of cancer patient in NEPA.

PARTICIPANTS

Patient Profiles Patients will be recruited through voluntary participation after signing informed consent. This translational research project aims to use patient samples for basic science investigation; the standard clinical treatments for each malignancy will be used and this study in no way affects the discretion of the treating physicians for each unique case. Dr. Christopher Peters in NROC (Northeast Radiation Oncology Center), a prostate cancer oncologist, and Dr. Bill Auriemma in CMC (Community Medical Center), a colorectal cancer surgeon, are the co-investigators on this project. They will be responsible for patient treatment, sample collection and provide all information associated with each case, including the general physical exam information, medical history, cancer stages and treatments, surgical procedure and pathological information etc.

Patient confidentiality will be maintained using an identification system containing a numbered code. The code will be entered into the database and used as the identification during the experiment in the research lab. Surgical Pathology information will be handled according to the protocol and policy in the corresponding hospital. Researchers will access the patient data with the coded number system without compromising patient confidentiality.

Inclusion/Exclusion Criteria Children under the age of 18 will be excluded from this research; prostate and colorectal cancers are also very rare in this age group. Gender for

colorectal cancer is not discriminated, but will be recorded to assist later data analysis. The ethnic background is not a standard to recruit patients.

RISKS

The risks to subjects of this study are minimal. Tissues obtained for this study are part of materials required by diagnosis and surgical treatment. By the nature of experiments in this research, only small amounts of samples are needed, so there is no additional pressure for the patient. Five to ten milliliter of additional urine and blood sample will be obtained with the consent of patient. The coding system will keep the confidentiality at high priority.

INFORMED CONSENT

Informed consent will be obtained using the consent form included in this protocol.

EXPERIMENTAL MATERIALS AND METHODS

Sample collection

1. **Tissue sample** Cancer tissues will be obtained mainly through biopsy and surgery. If possible the surrounding normal tissue will be taken as the best control from the same patient, which is practical for colon cancer. The polyps, tumor mass and nearby lymph node in colon cancer will be examined and distinguished by the doctor on site, except for the portion needed for pathology lab, 0.2-2 grams of fresh tissue (if possible up to 5g) will be dissected and briefly washed with DNase/RNase-free PBS containing protease and phosphatase inhibitors, stored on ice and immediately transferred to TCMC research lab for further processing (see below). For prostate cancer, biopsy sample is preferred as all patients diagnosed with prostate cancer will first have a biopsy regardless of the eventual treatment received (radiotherapy, surgery, active surveillance, etc). As limited by the size of biopsy, only small amount of tissue (~10mg) or the left-over from the diagnosis test will be obtained. Prostate gland tissue will be treated as above, and only used for protein and RNA extraction. If a whole gland is available from a surgery (i.e. radical prostatectomy), a large amount of tissue (up to 5g) will be acquired for more comprehensive experiments *(see below).*

2. **Urine sample** Urine is collected as a part of routine procedure of clinical test as instructed by the doctor. Briefly, urine after the first urination of the day is good for experiment. The midstream clean catch of urine (~10ml) will be collected in a sterile conical tube (e.g. Falcon, Cat#352098) under sterile conditions. Sample will be put on ice and transferred to TCMC research lab immediately for processing.

3. **Blood sample** Blood is also collected as a part of routine procedure required by clinical test. Five milliliter of whole blood is transferred immediately to a sterile tube containing citrate anti-coagulant (BD Vacutainer® CPT™ Cell Preparation Tube). The blood will be transferred at room temperature to TCMC research lab for immediate processing.

4. **Achieved FFPE sample** FFPE cancer tissues represent a vast and important resource for molecular analyses and retrospective clinical studies. Drs. Peters and Auriemma will coordinate with pathologists to access those samples for basic research. With archived information about the patient, cancer stage, treatment, pathological characteristics and

outcome, the analysis of FFPE sample will generate more integrated data and facilitate the interpretation of molecular mechanism. FFPE tissues are mainly used for gene expression and proteomics analysis *(see below).*

Sample processing

Samples are normally divided into five parts: a smaller amount will be snap frozen and stored at -80C for later use; one part each will be used for DNA, mRNA/miRNA and protein isolation; the other part will be used for cell isolation. If the sample amount is limited, protein, mRNA/miRNA and DNA analyses will be the top priority. All kinds of samples will be processed within maximum 2 hours after collection.

Protein extraction

- **Tissue sample** Fresh tissue sample will be used for total protein extraction with T-PER reagent (Pierce) to ensure the higher coverage of proteins from every cell compartments even including extracellular proteins. Tissue will be homogenized in 1:5 (w/v) of tissue to the reagent buffer containing the whole spectra of mammalian protease inhibitors (cocktail set III, EMD) and phosphatase inhibitors (Set I and III, EMD). Centrifuge at 13,000rpm for 10min at 4C to pellet cell/tissue debris. The supernatant is taken as the tissue lysate, the pellet is also saved to study some extremely insoluble proteins. The protein concentration will be determined by 660nm protein assay kit (Pierce) to eliminate the interference of various detergents and other reagents for better accuracy and sensitivity. Equal amount of protein will be adjusted for analysis immediately or aliqouted and stored at -80C for later use. The proteins remained in the insoluble pellet will be extracted with full strength RIPA buffer (25 mM Tris-HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% Sodium deoxycholate and 0.1 % SDS) or directly extracted with SDS-loading buffer plus heating for SDS-PAGE analysis.

- **Urine sample** Total urine proteins will be extracted by 90% ethanol precipitation. Nine fold of cold 90% ethanol is added to urine and centrifuged at 2500g for 25 min at 4C. The pellet is washed once by 90% ethanol and air-dried as the total protein. Dissolve the pellet with ½ strength of RIPA buffer for various protein analyses.

- **Blood sample** The whole blood collected in a Ficoll-based gradient centrifuge tube (BD Bioscience) as above will centrifuged at 1800g for 20min at RT. The separated plasma, mononuclear cells and platelets, granulocytes and red blood cells will be collected individually. This will increase protein coverage and the sensitivity of analysis. The plasma protein will be extracted by TRI REAGENT®BD according to manufacturer's instruction. Protein from blood cells will be extracted with T-PER as described as above.

- **FFPE sample** Proteins from slide mounted section (100mm2) or fixed tissue (2-3mm3) will be extracted with "Qproteome FFPE tissue kit" (Qiagen) according to manufacturer's protocol. Normally 20-100μg of protein will be obtained that is much enough for SDS-PAGE/western blotting and MS analysis.

DNA extraction

- Samples will be processed similarly to those used for protein preparation. For tissues and cells and plasma, genomic DNA will be isolated by DNAzol reagent (invitrogen). This method is very efficient so that only a small amount of tissue (25-50mg) and cells ($1\text{-}3X10^7$) is required. 70-100% of DNA can be recovered. Genomic DNA and DNA

fragment down to 100bp can be all purified, the latter of which is beneficial to the identification of circulated DNA biomarkers. The high quality of DNA preparation by this method will allow all DNA analyses.

RNA and miRNA extraction

- Total RNA and miRNA can be co-purified with miRNeasy kit (Qiagen) to save sample and increase consistence and comparability. miRNA can be specifically enriched by this kit with slight modification. This kit is very efficient to purify RNAs from low amount of starting materials (2mg of tissue and $1X10^5$ cells), and the high quality meets the requirements of most downstream analysis, including qRT-PCR, microarray, northern blot and miRNA identification.

Electrophoretic analysis and Western blotting

- Proteins are normally separated by SDS-PAGE (polyacrylamide gel electrophoresis) by Bio-Rad Mini-Protean Tetra Cell system based on the improved Laemmli protocol. To increase the resolution for difficult proteins, gradient gel and tricine gel system will be employed. Native PAGE will be run occasionally to study protein complex and interaction. Identification and quantitation of targets in total proteins will be mainly achieved by western blotting (WB). The quantitative western blot is carried out using Odyssey infrared system (Li-Cor).Equal amount of protein is separated by SDS-PAGE and transferred onto nitrocellulose or PVDF membrane with fast semi-dry transfer procedure (Pierce). Membrane is blocked for 1 hr at room temperature in the blocking buffer, hybridization with primary antibody is carried out by incubation of primary antibody at 1:1000 dilution in TBST (50mM Tris/HCl, pH7.5, 150mM NaCl, 0.1% Tween-20) containing 1% BSA for 2 hr at room temperature or overnight at 4C. The dilution of antibody and the length of incubation are dependent on different antibody. The membrane is then washed with TBST for 3-5 times, followed by hybridization to infrared IRDye-labeled secondary antibody at 1:5000 dilution in TBST/1% BSA for 1 hr. After the wash for 5 times with TBST and twice with PBS, the image is acquired by Odyssey infrared scanner. The data are analyzed and quantitated by Odyssey software 3.0.

Protein identification by Mass spectrometry (MS)

- MS-based proteomics approach will be employed to identify novel proteins, post-translational modifications (PTM), particularly phosphorylation, and quantitation. The bottom-up (digesting protein first) and top-down (starting with full length protein) strategies will be used to increase the resolution and coverage. *To identify interesting proteins revealed by SDS-PAGE or 2-PAGE*, protein bands will be in-gel digested by trypsin, V8, Glu-C and Lys-C, respectively, to fully display peptides in different sequence contexts. The resultant peptides are then extracted, or purified by titanium dioxide (TiO_2) affinity column to enrich the phosphopeptides. Nano-UPLC-ESI-Q-TOF-MS/MS protocol will be run to obtain the MS data. MS/MS spectra are processed and searched against the NCBInr non-redundant protein database using the MASCOT algorithm (http://www.matrixscience.com) to identify protein. The phosphorylation site is identified by the typical neutral-loss dependent mass shifts observed from y- and/or b- ions. MALDI-TOF-MS/MS will be used as an alternative method of ionization to increase protein coverage. *To identify new proteins from total protein without pre-separation*, multidimensional liquid chromatography techniques coupled with tandem mass spectrometry will provide a robust method and increase the through-put.

Immunoprecipitation

- Immunoprecipitation (IP) is a very useful technique to study protein interaction and purify the target protein from mixture complex quickly for activity studies. Protein lysate prepared above will be adjusted to lower detergent concentrations for IP. Briefly, lysate will be pre-cleaned by normal IgG (same species and isotype as IP antibody) and protein A/G bead by incubation for 30min at 4C with gentle agitation. The bound bead will be removed by centrifugation to decrease the nonspecific binding. The IP antibody (1ug antibody to 250ug of protein, or tested empirically) is added to bind the target protein by incubation for 2hrs to overnight at 4C with gentle rotation. 50ul of net protein A/G bead equilibrated with IP buffer will be added to capture the antibody-antigen complex by incubation for another 2-4hrs at 4C. The bound bead will be spun down and washed 3 times with IP buffer, followed by the elution of protein by glycine buffer pH2.2-2.8 or directly denatured by SDS-loading buffer for SDS-PAGE analysis.

Kinase activity assay

- Kinase activation is one of major events in most cancer development; it is also an important therapeutic target and a diagnosis marker. Kinase activity will be measured by two different methods in our projects: a) WB with phosphospecific antibody agaist the activation sites of the kinase, for instance Ser473 of Akt1 and Thr202/Tyr204 for Erk1/2 MAPK. b) in vitro kinase activity assay with γ-^{32}P-ATP and purified substrate protein followed by autoradiography.

DNA methylation and SNP identification

- Global DNA hypomethylation has been observed to be one of the earliest molecular abnormalities described in human neoplasia. Technically it is also one of easiest epigenetic modifications to analyze. Global DNA methylation level will be measured by "Methylamp™ Global DNA Methylation Quantification Kit" (Epigentek) with an ELISA-like high throughput procedure. SNP (single nucleotide polymorphism) is another signature genetic mutation resulted from transition or transversion that is much more stable than microsatellites; it is strongly associated with cancer at some high-risk loci; therefore becoming a very useful biomarker for cancers. APC Icoi in colon cancer and androgen-responsive gene,TMPRSS2, in prostate cancer will be specifically focused. The genome-wide SNP will be assayed with "Cancer SNP Panel" based on the NCI Cancer Genome Anatomy Project SNP500 Cancer Database on "GoldGate" beadchip array system (Illumina). The SNP in the specific gene and loci will be identified by next-generation sequencing with "Genome analyzer" also from Illumia. These experiments will be collaborated with the "Institute for Integrative Genome Biology" (http://www.genomics.ucr.edu/) at University of California, Riverside, and will be eventually operated in our own core facility.

Gene expression profiling by Microarray

- The genome-wide gene expression will be analyzed by "Affimatrix microarray" and "Genome analyzer" systems. The efficiency and cost will be evaluated carefully for the choice of one system. Microarray will be performed through the collaboration with Dr. Yoon Gi Choi from the functional genomics core facility at UC, Berkeley (http://microarrays.berkeley.edu). Briefly, the mRNA purified above will be used. The quantity and quality are checked by RNA 6000 Nano Chip kit on 2100 Bioanalyzer (Agilent) prior to microarray analysis. cDNA will be synthesized by the Superscript cDNA synthesis kit (invitrogen) with a T7-(dT)24 primer, then cRNA will be *in vitro* transcribed and labeled by biotin using BioArray labeling kit (Enzo). cRNA is fragmented into 35-200 nucleotides in length and hybridized with Human Genome U133 Plus 2.0 Array (47,000

transcripts, Affymetrix) according to manufacturer's instruction. The array will be scanned by GeneChip scanner 3000. All scanned images are normalized and uploaded to R packages or Partek. All files are quality assessed with the appropriate packages for Bioconductor in R. The quality passed data are used for clustering and further analysis if the p value of a moderated F statistic for either model is ≤ 0.05, and the maximum fold change over time (between any two time points) is at least 1.5. A false discovery rate (FDR) test will also run, as well. To analyze effects of specific genes, microarray data are clustered using a clustering algorithm. Clusters are visualized in Java Treeview 1.0.4 after sorting by a clustering order. To annotate the biological functions of HOPACH clusters, we use Gene Map Annotator and Pathway Profiler (GenMAPP). Each cluster list will be used to create a gene expression database in GenMAPP, with first or second level HOPACH cluster numbers as filter criteria. MAPPFinder will be used to search for over-representation of Gene Ontology (GO) terms in each cluster. Significant biological associations will be indicated if three or more genes in a given GO term are changed, with a permuted p value ≤ 0.05. The data will be confirmed with quantitative PCR. If the genome analyzer system is chosen, the experiment will be done through the collaboration with the "Institute for Integrative Genome Biology" at University of California, Riverside, according to their protocols.

miRNA identification

- The miRNA profiling will be done with "RT2 miRNA PCR array" from SABiosciences. It is a high throughput method with low cost and easy data analysis, and can be operated at 96- or 384-well plate format on our own Applied Biosystems 7900HT Fast real time PCR instrument.

Cell isolation and establishment of primary cell culture

- Primary cells will be isolated from primary tumor and tumor in lymph nodes. Mechanical spill-out and limited enzymatic digestion methods will be used. For colon cancer, minced tumor cell aggregates will be first cultured in AR-5 initiation media until high density, then the cancer cells will be isolated from floating aggregates and adherent colonies and grown and maintained in ACL-4 medium supplemented with heat-inactivated 5% FBS. For prostate cancer, needle biopsy samples, tumor tissues obtained at radical prostatectomy and samples from different metastatic sites (e.g. bone and lymph node) will be used for cell culture. If the tissue is easily disaggregated into cell clumps microexplants by mince, then no enzymatic digestion is necessary. Otherwise, add collagenase (40 U/mL) to the growth medium to assist initial primary cell isolation. F12K media, with 5 ug/mL insulin, 30 n*M* sodium selenite, 250 n*M* hydrocortisone, 25 ng/mL cholera toxin, 5 ng/mL EGF, and 100 ug/mL BPE, will be used to establish primary prostate cancer cell culture. Primary cells will be applied to the detailed cell biological and biochemical studies to generate more confirmative data.

Cell analysis by flow cytometry

- Flow cytometry is a powerful technique to analyze cell surface antigens and intracellular protein expression with cell division and size distribution information. It is also a quantitative approach with minimum destructive invasion of cell. The cultured primary cells and the cells directly isolated from tumor tissue will be analyzed on our own Becton Dickinson FACSAria II flow cytometer. The ratio of S-phase DNA stained by propidium iodide will be used as a primary standard for identification of cancer cell. Other prostate and colon cancer specific markers will be also measured by this method with staining by

corresponding primary antibody and Qdot® fluorescence labeled secondary antibody (invitrogen).

Immunofluorescence staining

- Immunofluorescence staining (FISH) will be employed to measure protein localization and interaction within cell organelles. This information is also critical to reveal the molecular mechanism of cancer development. Cultured cell on chamber slide or fresh frozen tissue cryostat sections will be studied by FISH. Samples are fixed by 2-4% formaldehyde in PBS for 2-3 min for cells and 15 min for tissue sections, then blocked and incubated with primary antibodies at an appropriate dilution (1:100 to 1:500) for overnight at 4C, followed by hybridization with fluorescence labeled secondary antibody for 1-2 hrs at room temperature. The image will be acquired and analyzed by our Nikon A1 confocal fluorescence microscope.

Statistical analysis

- For quantitation and comparison of difference, data will be presented as mean $\pm$ standard deviation. Student's *t*-test will be applied to test hypotheses for paired samples or to compare 2 independent groups; ANOVA will be used to compare means in more than 2 independent groups. $P \leq 0.05$ is considered as statistically significant and p-value > 0.05 and ≤0.10 is regarded as border-line significant. All analyses are performed using GraphPad Prism 5 software.

LOCATION

All clinical treatment, sample collection, surgery and pathology procedures are performed at NROC, Dunmore, PA and Mercy Hospital NROC (Dr. Peters), Scranton, PA and CMC (Dr. Auriemma), Scranton, PA. All research procedures are carried out at The Commonwealth Medical College research facilities on the third floor of Lackawanna College, Scranton, PA.